

N° d'ordre : 1087

50376  
1983  
249

50376  
1983  
249

# THÈSE

présentée à

L'UNIVERSITE DES SCIENCES ET TECHNIQUES DE LILLE

pour obtenir le titre de

**DOCTEUR DE 3<sup>ème</sup> CYCLE**

Spécialité : Automatique

par

Mouloud SBAI

Maître ès-Sciences



## **ANALYSE STRUCTURALE DES SYSTEMES COMPLEXES : METHODES D'EXPLICATION ET DE PARTITION**



Thèse soutenue le 29 Septembre 1983 devant la Commission d'Examen

Membres du Jury :	MM.	P.	VIDAL	Président
		M.	STAROSWIECKI	Rapporteur
		C.	LANGRAND	Examineurs
		J.C.	POSTAIRE	
		C.	VASSEUR	

A mes parents

A mes frères et soeurs

## A V A N T - P R O P O S

---

Le travail présenté dans ce mémoire a été effectué au Centre d'Automatique de L'Université des Sciences et Techniques de LILLE I.

Je tiens à exprimer ma grande reconnaissance à Monsieur Le Professeur Pierre VIDAL, pour l'accueil qu'il m'a réservé dans son laboratoire et pour l'honneur qu'il me fait de présider le jury de cette thèse.

Je remercie vivement Monsieur Marcel STAROSWIECKI, Professeur à l'Université de LILLE I pour m'avoir initié à la recherche dans le domaine de l'analyse structurale. Ses conseils, ses encouragements, et son temps généreusement consacré à mes recherches ont été pour l'essentiel dans les résultats de mes travaux. Qu'il trouve ici l'expression de ma profonde gratitude.

A Monsieur LANGRAND, Professeur à L'Université de LILLE I, je tiens à dire combien j'ai apprécié ses conseils et les fructueux échanges d'idées sur ces travaux. Je lui adresse mes remerciements avec toute ma reconnaissance, pour sa participation au jury.

Je tiens à adresser mes vifs remerciements à Messieurs J.G POSTAIRE, Professeur à l'Université de LILLE I et C.VASSEUR Maître Assistant à l'Université de LILLE I pour l'honneur qu'ils me font en acceptant de participer à mon jury de thèse.

Mes remerciements vont aussi à toutes les personnes qui ont d'une façon ou d'une autre participé à la réalisation matérielle de ce mémoire.

Enfin, je tiens à remercier Monsieur B.BARFETY, Directeur du C.R.O.U.S., Monsieur C.VANHAECKE et toutes les personnes du service d'Accueil des Etudiants Etrangers, où chacun s'ingénie à vous rendre le séjour plus agréable en FRANCE et je leur en sais gré.



## TABLE DES MATIERES

## INTRODUCTION GENERALE

CHAPITRE I : L'ANALYSE STRUCTURALE, LA DECOMPOSITION DES SYSTEMES COMPLEXES  
ET LA THEORIE DE L'INFORMATION.

I.1	Introduction .....	2
I.2	Nature de la donnée .....	6
I.3	Estimation des probabilités .....	8
I.4	L'apport de l'analyse des données .....	8
I.41	Analyse en composantes principales .....	9
I.42	Analyse factorielle des correspondances .....	10
I.43	Régression linéaire .....	11
I.44	Analyse canonique .....	11
I.45	La décomposition des systèmes Complexes .....	12
I.46	Conclusion	
I.5	La théorie de l'information .....	13
I.52	Eléments de la théorie de l'information .....	14
I.52a	Définition de l'entropie .....	14
I.52b	Entropie conjointe et entropie conditionnelle .....	16
I.52c	Entropie d'une partie $S$ de $\Sigma$ .....	17
I.52d	Entropie d'une partition de $\Omega$ .....	17
I.52e	Entropie conditionnelle de $S_1$ par rapport à $S_2$ .....	18
I.52f	Entropie conditionnelle d'une partition R par rapport à une partition Q .....	18
I.52g	Distance entre partitions .....	18
I.52h	Définition de la transinformation .....	19
I.52i	Transinformation interne .....	19
I.52j	Transinformation externe .....	20
I.52k	Propriétés .....	20
I.52l	Cas des variables continues .....	21
I.52m	Analogie entre la théorie de l'information et la théorie de la mesure .....	22
I.6	Conclusion .....	23

## CHAPITRE II : METHODES D'EXPLICATION

II.1	Introduction .....	26
II.2	Graphes d'interconnexion .....	26
II.2.1	Analyse par paires de variables .....	27
II.2.1a	Cas des systèmes statiques .....	27
II.2.1b	Cas des systèmes dynamiques .....	30
II.2.1c	Insuffisance de l'approche binaire .....	31
II.2.2	Analyse multidimensionnelle .....	31
II.2.2a	Indices de liaison .....	34
II.2.2b	Algorithme de détermination de $\gamma_i^k$ .....	35
II.2.3	Graphe d'interconnexion .....	36
II.3	Influence du bruit .....	38
II.3.1	Relations fonctionnelles bruitées .....	38
II.3.1a	Entropie conditionnelle nulle .....	39
II.3.1b	Entropie conditionnelle maximale .....	39
II.3.1c	Relation fonctionnelle bruitée .....	40
II.3.2	Détermination de l'entropie conditionnelle en fonction du bruit .....	40
II.3.2a	Bruit à densité uniforme .....	41
II.3.2b	Bruit à densité de probabilité exponentielle .....	42
II.3.2c	Bruit à densité de probabilité normale .....	43
II.4	Influence du nombre de modalités du vecteur explicatif .....	49
II.4.1	Algorithme hiérarchique ascendant .....	49
II.4.2	Cas d'une relation fonctionnelle bruitée lorsque les lignes n'ont pas même probabilité ..	54
II.5	Conclusion .....	58

<b>CHAPITRE III : METHODES DE DECOMPOSITION</b>	<b>p.59</b>
III.1 Introduction	p.60
III.2 Décomposition à partir des Graphes	p.61
III.2.1 Concepts sur la théorie des Graphes	p.61
III.2.2 Décomposition des Graphes orientés	p.64
III.2.3 Décomposition des Graphes symétriques ou non orientés	p.66
III.2.4 Décomposition des Graphes valués	p.69
III.3 Mesures de similarité	p.71
III.3.1 Position du problème	p.71
III.3.2 Quelques définitions	p.73
III.3.3 Calcul des indices	p.74
III.4 Méthodes hiérarchiques	p.78
III.4.1 Hiérarchie de parties	p.78
III.4.2 Algorithmes ascendants	p.78
III.4.3 Algorithmes descendants	p.79
III.4.4 Algorithme de transfert généralisé	p.80
III.5 Méthodes à centres variables	p.82
III.6 Conclusion	p.83

<b>CHAPITRE IV : METHODE DES NUÉES DYNAMIQUES</b>	p.84
IV.1 Introduction	p.85
IV.2 Principe de l'algorithme	p.86
IV.3 Choix du triplet (f, g, w)	p.86
IV.4 Variantes	p.87
IV.41 $\mathcal{L} = \mathcal{P}(\Sigma)$ , $R(\mathcal{E}_j, \lambda_j) = \sum_{x \in \mathcal{E}_j} D(x, \lambda_j)$	p.87
IV.42 $\mathcal{L} = \mathcal{P}(\mathcal{L})$ , $R(\mathcal{E}_j, \lambda_j) = D(P_x, P_{\lambda_j})$	p.89
IV.43 Convergence de ces algorithmes	p.91
IV.44 Détermination des noyaux	p.91
IV.5 Exemple d'application	p.100
IV.6 Conclusion	p.101
CONCLUSION GÉNÉRALE	p.102
ANNEXE	p.104
BIBLIOGRAPHIE	p.109

INTRODUCTION GÉNÉRALE

---



- INTRODUCTION GENERALE -

Un système complexe est le plus souvent constitué d'un nombre élevé d'éléments qui interagissent de façon non simple (SIM, 62). Il pose des problèmes importants et nouveaux d'analyse et de commande dûs essentiellement au nombre élevé de variables nécessaires pour sa description. On peut réduire la dimension du problème en définissant des variables explicatives, soit parmi l'ensemble des variables du système, soit par une procédure d'agrégation de certaines d'entre elles. Cette approche peut faire appel aux méthodes descriptives de l'analyse des données, bien connues actuellement : Analyse en composantes principales, analyse canonique, etc....

La plupart de ces méthodes utilisent explicitement des hypothèses de linéarité des relations liant les variables ou des fonctions de ces variables.

Pour notre part, nous nous intéressons à une autre approche de l'analyse des données utilisant les concepts de la théorie de l'information. Cette approche présente certains avantages par rapport aux méthodes classiques. Elle permet de mesurer des couplages dûs à des mécanismes non linéaires et de mettre en évidence des relations globales dans un groupe de variables ; alors que par les autres méthodes, la relation entre groupes de variables est estimée à partir d'une moyenne de relations entre les variables de ces groupes.

Le but de l'analyse est de trouver pour chaque variable, un groupe de variables qui l'explique le mieux possible ; la relation de dépendance étant la principale raison qui fait qu'une variable explique une autre.

Une autre méthode consiste à décomposer le système global en sous-systèmes faiblement couplés pour mettre en oeuvre des méthodes de modélisation par modèles partiels, de commande décentralisée, d'optimisation hiérarchisée, dont la convergence est d'autant plus rapide que les sous-systèmes sont indépendantes (RICH 75). Elle débouche sur les méthodes de classification automatique.

### Présentation des chapitres suivants

Dans le premier chapitre, nous allons décrire succinctement quelques méthodes de visualisation (analyse factorielle), d'explication (analyse canonique, régression linéaire) et de structuration (décomposition). On montre aussi leur apport au domaine qui nous intéresse. On donne ensuite les définitions de base et quelques propriétés de la théorie de l'information, outil essentiel de notre travail.

Le deuxième chapitre développe les méthodes d'explication d'une variable à partir d'un groupe de variables, ce qui débouche sur la construction d'un graphe d'interconnexions prenant en compte l'aspect multidimensionnel des relations. Des algorithmes sont développés pour réduire le nombre de modalités du vecteur explicatif. L'influence du bruit sur les relations fonctionnelles sera étudiée.

Dans le troisième chapitre, on examine quelques méthodes de décomposition des systèmes. On utilise la théorie des graphes, connaissant les couplages entre toutes les paires de variables ou groupes de variables. Après avoir défini quelques indices permettant la mesure de ces couplages, on présente certaines méthodes de décomposition hiérarchique. Le chapitre se termine par l'introduction des méthodes à centres variables.

Le quatrième chapitre est consacré à la méthode des nuées dynamiques. Plusieurs variantes sont étudiées et comparées. Les résultats sont interprétés pour chacune d'elles.

CHAPITRE I

---

L'ANALYSE STRUCTURALE, LA DÉCOMPOSITION DES SYSTÈMES COMPLEXES ET

LA THÉORIE DE L'INFORMATION

## I.1. INTRODUCTION

L'analyse d'un système conduit généralement à l'élaboration d'un modèle dont les paramètres sont identifiés à partir d'un certain nombre d'observations. Les difficultés de modélisation croissent directement avec le nombre de variables du système et le nombre de leurs interconnexions. La complexité doit être réduite par structuration de l'information par laquelle le système est appréhendé.

Si l'on connaît le modèle du système (relations fonctionnelles entre les variables du système, relation binaire liant les variables....), on utilise un graphe d'interconnexions qui permettra de mettre en évidence les différents sous-systèmes.

Mais le modèle d'un système complexe est en général difficile à obtenir. En fait, la présence des couplages est toujours décelable dans les données produites par le système et se concrétise par des évolutions non indépendantes de certains groupes de variables. Dans ce sens, l'analyste dispose généralement de trois types de méthodes :

- Les méthodes de visualisation telles que l'analyse en composante principale et l'analyse factorielle des correspondances.
- Les méthodes d'explication telles que l'analyse canonique et la régression linéaire.

Ces deux méthodes permettent de traiter des grands tableaux de données sans trop les appauvrir. Elles aboutissent à la mise en évidence des liaisons linéaires entre les variables du système ; le plus souvent, elles en offrent une représentation graphique.

- Les méthodes de structuration qui consistent à définir la structure du système en terme de sous-systèmes.

Le principal inconvénient de ces méthodes tient au fait que les résultats de l'analyse sont fondés sur les hypothèses assez restrictives et souvent difficilement vérifiables. Or la confrontation avec la complexité des systèmes étudiés, montre que les variables sont quelconques (quantitatives, qualitatives...) et les relations entre elles sont souvent non linéaires.

La théorie de l'information se prête bien à l'analyse des systèmes car elle est sensible à toutes les relations, qu'elles soient linéaires, non linéaires, logiques... Nous précisons les avantages et les inconvénients de l'approche "informationnelle" et le chapitre se termine par quelques définitions et propriétés essentielles de la théorie de l'information.

## I.2. NATURE DE LA DONNÉE

Soit  $\Sigma = \{X_1, X_2, \dots, X_i, \dots, X_N\}$ , l'ensemble des N variables nécessaires pour la description du système étudié. Soit  $\Omega = \{\omega_1, \omega_2, \dots, \omega_j, \dots, \omega_L\}$ , un ensemble d'échantillonnage. L'ensemble des modalités prises par une variable X sera désigné par  $M_i$ . L'ensemble des variables se présente sous la forme des différentes composantes de l'application suivante :

a) Cas des systèmes statiques (s)

$$X_i^{(s)} : \Omega^s \longrightarrow M_i^s = M_i = \{a_1^i, a_2^i, \dots, a_{m_i}^i\} \quad i = 1, 2, \dots, N$$

$$\omega_j \longrightarrow X_i^s(\omega_j) = X_i(\omega_j)$$

b) Cas des systèmes dynamiques (D)

$$X_i^D : \Omega^D \longrightarrow M_i^D = M_i^P = M_i \times M_i \times \dots \times M_i \quad (p \text{ fois})$$

$$\omega_j \longrightarrow X_i^D(\omega_j) = (X_i(\omega_j), X_i(\omega_{j-1}), \dots, X_i(\omega_{j-(p-1)}))$$

Chaque vecteur  $X_i^D$  est composé des p dernières valeurs prises par la variable  $X_i$  à des instants donnés. La période d'échantillonnage est fixée par le théorème de Shannon (LARM 75 ; ASH 65).

En fait, dans la pratique, on conserve le même tableau initial de données (Fig. I.1.) et la prise en compte de ces p valeurs se fait au niveau des algorithmes.

$\Omega \backslash \Sigma$	$X_1$	$X_2$	$\dots$	$X_i$	$\dots$	$X_N$
$\omega_1$						
$\omega_2$						
$\vdots$						
$\omega_j$				$X_i(\omega_j)$		
$\vdots$						
$\omega_L$						

Figure I.1

Ce tableau initial de données constitue la seule connaissance de départ pour toute notre analyse. Il s'agit aussi bien des cas statiques que dynamiques.

Chaque variable  $X_i$  (respectivement un sous ensemble de variables  $S$ ) définit par elle-même une partition  $P_{X_i}$  (resp  $P_S$ ) de l'ensemble  $\Omega$  où chaque classe est formée des  $\omega$  identiques du point de vue  $X_i$  (resp  $S$ ). En notant  $\mathcal{P}(\Omega)$ , l'ensemble des partitions de  $\Omega$ , et  $\mathcal{P}(\Sigma)$ , l'ensemble de toutes les parties non vides de  $\Sigma$ , on peut dire que  $\Sigma$  (resp  $\mathcal{P}(\Sigma)$ )  $\subseteq \mathcal{P}(\Omega)$ .

Soit  $\ll$  : "être plus fine que", la relation d'ordre définie par :

$$\forall R = \{R_1, R_2, \dots, R_r\} \in \mathcal{P}(\Omega), \forall Q = \{Q_1, Q_2, \dots, Q_q\} \in \mathcal{P}(\Omega)$$

$$R \ll Q \iff \forall R_i \in R, \exists Q_j \in Q \text{ tel que } R_i \subseteq Q_j$$

$\mathcal{P}(\Omega)$  muni de la relation  $\ll$  constitue un treillis.

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

Considérons deux variables quelconques X et Y définies sur  $\Omega$ , on dit que :

- X est équivalent à Y  $\iff P_X = P_Y$
- X est plus fine que Y  $\iff P_X \ll P_Y$

Nous verrons au paragraphe I.5. que la théorie de l'information ne distingue pas entre variables équivalentes.

### I.3. ESTIMATION DES PROBABILITES

Le caractère aléatoire des variables observées est dû soit à la nature du système à analyser, soit à la façon dont il a été excité .

Soit  $S \in \mathcal{P}(\Sigma)$ ,  $J_S$  l'ensemble des indices des variables  $X_i$  appartenant à S et  $\vec{S}$  le vecteur aléatoire constitué par ces variables. On note  $p(\vec{S} = \alpha)$  la probabilité pour que ce vecteur prenne une valeur  $\alpha \in M_S = \prod_{i \in J_S} M_i$

Sous certaines hypothèses concernant le système (ergodicité, stationnarité), le nombre d'échantillons et la période d'échantillonnage, ces probabilités peuvent être estimées au moyen des fréquences relatives. Du tableau initial de données (Fig I.1), on peut extraire le tableau des probabilités dont l'élément (i, j) est la probabilité d'occurrence  $P_i^j$  de la j<sup>ième</sup> valeur  $\beta$  de la i<sup>ème</sup> variable  $X_i$

$$P_i^j = P_r(X_i = \beta) = \frac{\text{nombre de } \omega \text{ pour lesquels } X_i = \beta}{L} = \frac{|\{\omega : X_i = \beta\}|}{|\Omega|}$$

### I.4. L'APPORT DE L'ANALYSE DES DONNEES

Il existe plusieurs méthodes de traitement des données actuellement disponibles pour l'analyse structurale des systèmes. Ces méthodes se distinguent les unes des autres par le type de variables qu'elles traitent et par le choix de la métrique qui sert à calculer les couplages et les inerties. Dans chaque cas, ces choix se justifient par les buts que l'on se donne et par les données dont on part.



On présente dans ce paragraphe quelques méthodes pouvant être appliquées au domaine qui nous intéresse. L'analyse en composantes principales, l'analyse des correspondances, qui sont les principales méthodes d'analyse factorielle, la régression linéaire, l'analyse canonique et quelques techniques de décomposition.

Ces méthodes ont fait l'objet d'un grand nombre d'études dans les domaines les plus divers (BENZ 80) (VOL 82) (CAIL 76) (BERT 75).

#### I.4.1. Analyse en composantes principales

Cette méthode peut être appliquée à tout tableau fournissant sur chacun des individus d'une population quelconque, les mesures d'une série de variables quantitatives. On peut l'utiliser pour étudier une matrice de corrélations.

Le tableau initial de données (Fig. I.1.) peut-être représenté par un nuage de L points dans l'espace  $\mathbb{R}^N$  des variables dont le grand nombre de dimensions interdit la visualisation du nuage. Il s'agit de trouver les axes d'inertie du nuage et d'obtenir des visualisations dans un espace de faible dimension.

- On cherche une droite telle que l'inertie du nuage par rapport à cette droite soit minimum. C'est le premier axe principal, le nuage sera alors représenté par la projection des points sur cet axe.

-L'opération est répétée dans  $\mathbb{R}^{L-1}$ , on obtient alors un deuxième axe et un nouveau nuage projetée dans  $\mathbb{R}^{L-2}$ .

#### Facteurs principaux

Ce sont des nouvelles variables obtenues à partir des variables initiales de la manière suivante :

$$y_1 = \sum_{i=1}^N y_i^1 X_i$$

telle que la variance de  $y_1$  soit maximum

$$y_2 = \sum_{i=1}^N y_i^2 X_i$$

$y_2$  orthonormale et non corrélée à  $y_1$  et telle que la variance de  $y_2$  soit maximum et ainsi de suite.

On peut vérifier facilement que les facteurs principaux sont des vecteurs propres de la matrice  $V$  de variance - covariance des variables de  $\Sigma$ .

Facteurs principaux et axes principaux sont liés. Les facteurs principaux sont les variables suivant lesquelles la dispersion de l'ensemble d'échantillonnage est maximale.

#### 1.4.2. Analyse factorielle des correspondances

Cette méthode est essentiellement une analyse métrique qui s'applique à des données de contingence croisant deux ensembles de modalités. Elle permet de décrire la dépendance entre deux variables qualitatives.

Soit  $P_{ik}$  ( $i=1$  à  $N$ ,  $k=1$  à  $q$ ) la probabilité d'occurrence conjointe des modalités  $i$  et  $k$ .

$$\text{On note } P_i = \sum_{k=1}^q P_{ik} \quad P_k = \sum_{i=1}^N P_{ik} \quad P = \sum_{k=1}^q P_k = \sum_{i=1}^N P_i$$

Le tableau des données (Fig. I.1) est remplacé par un tableau d'éléments  $P_{ik}$  ( $i = 1$  à  $N$ ,  $k = 1$  à  $q$ ).

La distance entre deux éléments  $i$  et  $i'$  :

$$d^2(i, i') = \sum_{k=1}^q \frac{1}{P_k} \left[ \frac{P_{ik}}{P_i} - \frac{P_{i'k}}{P_{i'}} \right]$$

est la distance du  $\chi^2$  associée à la loi de probabilité  $P_k$  entre deux distributions  $\frac{P_{ik}}{P_i}$  et  $\frac{P_{i'k}}{P_{i'}}$

L'étude des relations entre variables revient à l'étude des proximités entre les lignes de ce tableau.

Le choix de la distance de  $\chi^2$  est lié essentiellement au principe d'équivalence distributionnelle. C'est-à-dire qu'on ne change pas les résultats de l'analyse en remplaçant les modalités  $a_i^1$  et  $a_i^2$  de profils égaux par une seule,  $\alpha_{1,2}$  représentant la réunion de  $a_i^1$  et  $a_i^2$ .

Une analyse en composante principale peut être faite en utilisant comme métrique la distance de  $\chi^2$  au lieu de la distance euclidienne dans l'étude précédente.

#### I.4.3. Régression linéaire

Du tableau initial de données on extrait des variables à expliquer notées  $Y_i, i = 1, \dots, N$  et des variables explicatives  $Z_i^j, i = 1, \dots, N, j = 1, \dots, p$ . On suppose que ces variables sont quantitatives et qu'il existe des liaisons linéaires entre elles.

On recherche  $N+1$  scalaires  $b_i, i = 0 \dots N$  tels que

$$Y_i - \{b_0 + b_1 Z_i^1 + b_2 Z_i^2 + \dots + b_N Z_i^p\} \text{ soit minimum } \forall i = 1 \text{ à } N.$$

Si l'on pose

$$b = \begin{bmatrix} b_0 \\ \vdots \\ b_N \end{bmatrix} \quad Z = \begin{bmatrix} 1 & Z_1^1 & \dots & Z_1^p \\ \vdots & \vdots & \dots & \vdots \\ 1 & Z_N^1 & \dots & Z_N^p \end{bmatrix} \quad Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix}$$

Le problème se pose d'une manière plus précise :

trouver  $b \in \mathbb{R}^{N+1}$  tel que  $\|Y - Zb\|$  soit minimum.

La qualité de la représentation de  $Y$  par  $Zb$  peut être mesurée à l'aide de l'indice  $\frac{\|Zb\|}{\|Y\|}$ , coefficient de corrélation multiple entre  $Y$  et l'ensemble des variables explicatives.

Avec cette méthode, on peut mesurer l'intensité de liaison entre une variable et un groupe de variables (par exemple : relation entrée-sortie).

#### I.4.4. Analyse canonique

Il s'agit dans cette méthode de trouver des relations entre deux groupes de variables quantitatives extraits du tableau initial de données, on note  $Y$  et  $Z$  les vecteurs représentant ces groupes de variables.

$$Y^T = (Y_1, Y_2, \dots, Y_q) \quad Z^T = (Z_1, Z_2, \dots, Z_p)$$

on cherche une forme linéaire des  $Y_j, a_i = \sum_{j=1}^p a_i^j \in \mathbb{R}^p$

espace des formes linéaires de dimension  $p$  et une forme linéaire des  $Z_k,$

$b_i = \sum_{k=1}^q b_i^k \quad Z_k \in \mathbb{R}^q$  espace des formes linéaires de dimension  $q$ , tels que

$a_i$  et  $b_i$  ont une corrélation maximum.

Le problème de la recherche de  $a_i$  et  $b_i$  peut-être formulé ainsi :

- chercher  $a_1 \in \mathbb{R}^p$  et  $b_1 \in \mathbb{R}^q$  telles que

La corrélation entre  $a_1$  et  $b_1$  soit maximum avec la contrainte

$$\text{var}(a_1) = 1$$

$$\text{var}(b_1) = 1$$

- chercher  $a_2 \in \mathbb{R}^p$  et  $b_2 \in \mathbb{R}^q$  vérifiant les mêmes conditions que  $a_1$  et  $b_1$  et les conditions de non corrélation.

$$\text{cor}(a_1, a_2) = 0$$

$$\text{cor}(b_1, b_2) = 0$$

Cette procédure peut-être répétée au plus  $p$  fois ( $p \leq q$ ) les  $a_i$  et  $b_i$  sont appelées variables canoniques associées aux variables  $Y_j$  (resp. aux variables  $Z_k$ ). En comparant les couples  $(a_i, b_i)$ , on peut chercher à mettre en évidence des groupes de variables liés entre eux.

#### I.4.5. La décomposition des systèmes complexes

Il est naturel de dire qu'un système complexe doit être décomposé en sous-systèmes dans le but de réduire sa complexité apparente. Il s'agit de décomposer le système en sous-systèmes faiblement couplés et le problème global est ainsi décomposé en sous-problèmes ayant peu de relations entre eux. Ceci accélère très notablement la vitesse de résolution probable.

Dans leur ensemble, les méthodes de décomposition sont très variées. On peut les regrouper en trois catégories, les méthodes dites "ascendantes" qui sont les plus utilisées, les méthodes "descendantes" et enfin des méthodes de recherche d'une partition.

##### I.4.5.a) Les méthodes ascendantes

On définit tout d'abord un indice de distance  $d(X,Y)$  pour chaque couple de variables et un indice de distance  $D$  entre sous-ensembles de  $\Sigma$ , compatible avec  $d$ . On commence par chercher le couple de variables  $(X,Y)$  tel que  $d(X,Y)$  soit minimum. On agrège alors ces deux variables. On calcule à l'aide de  $D$ , l'intensité de la liaison entre l'agrégat ainsi formé et les autres variables, puis on considère cet agrégat comme un élément parmi les autres variables. On recherche ensuite les deux éléments les plus proches au sens de  $D$  et ainsi de suite.

#### I.4.5.b) Les méthodes descendantes

Elles parcourent le treillis des partitions dans le sens opposé : on part du système  $\Sigma$  et on le décompose en deux parties, puis chacune d'elles est décomposée à son tour en deux etc...

#### I.4.5.c) Les méthodes de recherche d'une partition

Il s'agit non plus de donner une hiérarchie de toutes les parties de  $\Sigma$ , mais simplement une partition. On peut citer l'algorithme des transferts et l'algorithme des nuées dynamiques. Ce dernier sera étudié au chapitre IV.

#### . Algorithme des transferts

L'algorithme des transferts est proposé par REGNIER (REGN 65). Il consiste à améliorer une partition estimée a priori en déplaçant chaque élément d'une classe à une autre. La nouvelle partition doit être meilleure au sens d'un critère d'optimisation que la partition obtenue à l'étape précédente. Le processus s'arrête lorsqu'aucun transfert n'améliore plus la partition.

#### I.4.6. Conclusion

La plupart des méthodes précédemment décrites utilisent des hypothèses de linéarité des relations liant les variables du système. Chacune d'elles convient particulièrement bien à un certain type de variables (quantitatives, qualitatives ...).

Chaque méthode ayant sa propre spécificité, nous sommes amenés à pratiquer différentes analyses du même tableau, car dans chaque cas, on peut obtenir des résultats qui s'enrichissent mutuellement.

### I.5 - LA THEORIE DE L'INFORMATION

#### I.5.1 Introduction

La théorie de l'information est née vers 1920 avec les travaux de NYQUIST et HARTLEY sur l'aptitude des divers systèmes à transmettre l'intelligence d'un message (NYQ.24, HART.28).

En fait, la théorie de l'information a été réellement fondée en 1948 par Claude E. SHANNON pour étudier la transmission des messages par une ligne téléphonique ou un canal hertzien brouillé (SHAN.48,49). A peu près en même temps que SHANNON, WIENER a publié deux ouvrages très connus (WIEN 48,49).

Ce sont les travaux de WIENER qui ont donné naissance aux études, dans le domaine militaire, sur la prévision automatique de trajectoire et sur la commande automatique de tir.

Conçus à l'origine pour la résolution des problèmes pratiques des télécommunications, la théorie de l'information devient actuellement l'outil mathématique indispensable pour l'étude des processus de commande les plus divers. Mc.GILL, W.R. ASHBY, R.C. CONANT, M.RICHETIN, J.DUFOUR parmi d'autres utilisent l'information pour mesurer la connexion et l'interdépendance entre les sous-systèmes d'un système donné ( GIL.54, ASHB.65, CON.68, RICH.75, DUF.79). Après les travaux de B.FORTE et J.KAMPE de FERIET sur l'information généralisée (KAMP.67), C.LANGRAND, J.LOSFELD, A.DUSSAUCHOY, J.P BARTHELEMY obtiennent des résultats importants pour d'autres types d'information que celle de SHANNON (LANG.23, LOSF.74, DUSS.80, BART.79). R.LOPEZ DE MANTARA, S.GUIASU et C. REISCHER définissent une distance entropique. (LOP.77, GUIA.79).

W.R. ASHBY, R.C CONANT, HEIDEMANN, B.PORTER, W.J.H.KICKERT et al étudient les relations entre la transmission de l'information et la capacité de régulation d'un système (ASH.58, CONA.69, WEID.69, PORT.76, KICK.78).

On utilise la théorie de l'information pour l'analyse structurale des systèmes complexes pour plusieurs raisons:

- Elle permet de mesurer l'intensité de liaison pour des relations quelconques (linéaires, non linéaires ...) entre variables ou groupes de variables quelconques (quantitatives, qualitatives...).
- Elle met en évidence des relations globales dans un groupe de variables, alors que par les autres méthodes, la relation entre groupes de variables est estimée à partir d'une moyenne des relations entre les variables de ces groupes.
- Elle permet aussi la construction d'indices de similarité présentant des propriétés intéressantes (cf §.III.3).

### I.5.2 - Eléments de la théorie de l'information

#### 1.5.2.a) Définition de l'entropie

Convenons que l'information transmise concerne une variable  $X$  pouvant prendre les modalités  $a_1, a_2, \dots, a_n$  avec les probabilités respectives  $p_1, p_2, \dots, p_n$  c'est-à-dire une variable caractérisée par un certain degré d'incertitude. Il est évident que les renseignements que l'on peut obtenir sur  $X$  seront

d'autant plus importants que son incertitude à priori est grande. Que signifie alors un degré d'incertitude plus ou moins grand et comment peut-il être mesuré ?

Comparons tout d'abord entre-elles deux variables X et Y, telles que : X prend ses valeurs dans l'ensemble  $\{a_1, a_2\}$  avec la même probabilité

Y prend ses valeurs dans l'ensemble  $\{b_1, b_2, b_3, b_4\}$  avec la même probabilité.

Il est évident que Y a une incertitude plus grande que X, car le nombre de ses modalités est plus important.

Soit maintenant une variable Z qui prend ses valeurs dans l'ensemble  $\{c_1, c_2\}$  avec  $P_r(Z=c_1) = 0,99$  et  $P_r(Z=c_2) = 0,01$ . Le degré d'incertitude de z est très faible. On peut dire d'une manière presque certaine que Z prend la valeur  $c_1$ . Lorsque l'on considère la variable X, il y a également deux modalités; mais le degré d'incertitude est bien plus important.

On voit ainsi que le degré d'incertitude d'une variable est déterminé par le nombre de ses modalités et leurs probabilités, peu important leurs valeurs.

L'incertitude à priori d'une variable est appelée entropie. Pour une variable discrète, elle est définie par :

$$H(X) = - \sum_{i=1}^n p_i \log p_i \quad [I.1]$$

Comme nous allons le voir, l'entropie vérifie certaines propriétés intéressantes, justifiant son utilisation en qualité de degré d'incertitude.

$H(X)$  prend ses valeurs dans l'intervalle  $[0, \log n]$ . Tout d'abord elle s'annule dans le cas d'un évènement sûr. Puis pour un nombre donné de modalités, elle est maximale et égale à  $\log n$  si la probabilité de chaque évènement est la même. Enfin, l'entropie est additive, c'est-à-dire que lorsque plusieurs variables indépendantes forment un groupe de variables, leurs entropies s'ajoutent.

L'unité de l'entropie dépend de la base choisie pour le logarithme. Si l'on choisit pour base, le nombre 10, l'unité est le Hartley (dit), si c'est le nombre 2 l'unité est le bit. Pour le logarithme népérien l'unité est le nat (natural unit). Ultérieurement, sauf mention spéciale, par le symbole  $\log$  on entendra des logarithmes de base 2 ; il est facile de voir que dans ce cas l'unité de mesure est l'entropie d'une variable ayant deux modalités, chacune avec la probabilité  $\frac{1}{2}$ .



I.5.2.b) Entropie conjointe et entropie conditionnelle

Pour deux variables  $\{X, Y\}$  de probabilités définies par  $P_{ij} = P_r(X=a_i, Y=b_j)$ ,  $1 \leq i \leq n$ ,  $1 \leq j \leq m$ , l'entropie (conjointe) est définie par :

$$H(X, Y) = - \sum_{i=1}^n \sum_{j=1}^m P_{ij} \log P_{ij} \quad [I.2.]$$

Supposons maintenant que  $X = a_i$ . Désignons par  $P(b_j/a_i)$  la probabilité conditionnelle pour la variable  $Y$  d'avoir la modalité  $b_j$  lorsque la variable  $X$  a la modalité  $a_i$ . L'entropie conditionnelle de  $Y$  sachant que  $X$  a la modalité  $a_i$  est par définition:

$$H(Y|a_i) = - \sum_{j=1}^m P(b_j/a_i) \log P(b_j/a_i) \quad [I.3]$$

L'entropie moyenne de  $Y$  lorsque  $X$  peut prendre des modalités différentes, appelée entropie conditionnelle totale est :

$$H(Y|X) = \sum_{i=1}^n P_i H(Y/a_i)$$

ou compte tenu de [I.3]

$$\begin{aligned} H(Y|X) &= - \sum_{i=1}^n P_i \sum_{j=1}^m P_r(b_j/a_i) \log Pr(b_j/a_i) \\ &= - \sum_{i=1}^n \sum_{j=1}^m P_i P_r(b_j/a_i) \log Pr(b_j/a_i) \end{aligned}$$

or  $P_i P_r(b_j/a_i) = P_{ij}$  (théorème des probabilités composées)

$$H(Y|X) = - \sum_{i=1}^n \sum_{j=1}^m P_{ij} \log \frac{P_{ij}}{P_i} \quad [I.4]$$

La généralisation à une partie  $S$  de  $\Sigma$  (ensemble des variables étudiées) est immédiate,  $S \in \mathcal{P}(\Sigma)$ , ensemble des parties de  $\Sigma$ . Etant donné que chaque élément de  $\mathcal{P}(\Sigma)$  donne lieu à une partition de  $\Omega$ , l'entropie peut être définie de deux manières équivalentes.

I.5.2.c. Entropie d'une partie  $S$  de  $\Sigma$  :

C'est la quantité d'information que fournit la connaissance de la valeur prise par le vecteur  $\vec{S}$

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

$$S = \{X_i, i \in J_S\} \rightsquigarrow H(S) = - \sum_{\alpha \in M_S} P(\vec{S}=\alpha) \log P(\vec{S}=\alpha)$$

I.5.2.d. Entropie d'une partition de  $\Omega$  :

L'entropie d'une variable dépend des probabilités avec lesquelles elle prend une certaine modalité  $\alpha$ , c'est-à-dire du nombre de  $\omega \in \Omega$  pour lesquels  $X(\omega) = \alpha$  pour chaque  $\alpha \in M_X$ , c'est justement la partition  $P_X$  de  $\Omega$ . On peut donc considérer  $\Omega$  muni d'une mesure de probabilité  $P$  et définir alors :

$$H : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$$

$$R = \{R_1, R_2, \dots, R_s\} \rightarrow H(R) = - \sum_{i=1}^s P(R_i) \log P(R_i)$$

- Relation entre les deux définitions :

Soit  $X : \Omega \rightarrow M_X$  une variable quelconque et sa partition

$$P_X = \{ \{\omega : X(\omega) = \alpha\} \mid \alpha \in M_X \} \setminus \{\emptyset\} \text{ alors}$$

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

L'entropie d'une variable est égale à l'entropie de sa partition associée. Deux variables équivalentes ont donc la même entropie.

I.5.2.e. Entropie conditionnelle de  $s_1$  par rapport à  $s_2$  :

C'est la quantité d'information que fournit encore  $s_1$  lorsqu'on aurait rendu constante la partie du système concernée par  $s_2$  (ASHB, 65).

$$S_1 = \{ X_i, i \in J_{S_1} \} \quad S_2 = \{ X_i, i \in J_{S_2} \} \quad S_1 \text{ et } S_2 \in \mathcal{P}(\Sigma)$$

$$H(./. ) \quad \mathcal{P}(\Sigma) \rightarrow \mathbb{R}$$

$$H(S_1/S_2) = - \sum_{\alpha \in M_{S_1}} \sum_{\beta \in M_{S_2}} P(S_1 = \alpha, S_2 = \beta) \log \left( \frac{P(S_1 = \alpha, S_2 = \beta)}{P(S_1 = \alpha)} \right)$$

I.5.2.f. Entropie conditionnelle d'une partition R par rapport à une partition Q

$$R = \{R_1, R_2, \dots, R_i, \dots, R_r\} \quad Q = \{Q_1, Q_2, \dots, Q_j, Q_q\} \quad R \text{ et } Q \in \mathcal{P}(\Omega)$$

$$H(./. ) : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$$

$$H(R/Q) = - \sum_{i=1}^r \sum_{j=1}^q P(R_i \cap Q_j) \log \left( \frac{P(R_i \cap Q_j)}{P(R_i)} \right)$$

Bien sûr, si R et Q sont des partitions associées respectivement à  $S_1$  et  $S_2$  alors

$$H(S_1|S_2) = H(R|Q).$$

I.5.2.g. Distance entre partitions

Dans sa thèse, Lopez de Mantaras propose une distance entre partitions basée sur la théorie de l'information (Lop. 77) :

$$D(R, Q) = H(R|Q) + H(Q|R)$$

Il est évident que cette notion de distance peut-être utilisée pour deux variables quelconques X et Y définies sur  $\Omega$

$$D(X, Y) = H(X|Y) + H(Y|X)$$

La propriété  $D(X, Y) = 0 \iff X = Y$  indique alors que X et Y sont équivalentes.

I.5.2.h. Définition de la transinformation

L'entropie étant la mesure d'incertitude d'une variable définie sur  $\Omega$ . Il est évident que si l'on obtient certaines informations sur la variable, cette incertitude diminue. Il semble donc tout naturel de mesurer la quantité d'information obtenue sur une variable par la diminution de son entropie.

Soit une variable  $X$  définie sur  $\Omega$  ; estimons l'information accumulée au fur et à mesure que  $X$  devient connu. Avant que l'on ait ces données, l'entropie (à priori) de  $X$  était  $H(X)$  ; après l'obtention des données, la variable  $X$  est entièrement déterminée, c'est-à-dire que l'entropie est devenue égale à zéro. L'information obtenue avec la détermination de  $X$  notée par convention  $\overleftarrow{I}(X : X)$  est évidemment égale à la diminution de l'entropie.

$$\overleftarrow{I}(X : X) = H(X) - 0 = H(X)$$

Considérons maintenant deux variables  $X$  et  $Y$  définies sur  $\Omega$  ; la quantité  $H(X) - H(X|Y)$  correspond à une diminution de l'entropie de  $X$  avec l'obtention des renseignements concernant  $Y$ . Cette quantité d'information contenue dans  $Y$  au sujet de  $X$  est appelée transinformation externe entre  $X$  et  $Y$ , noté  $\overleftarrow{I}(X : Y)^*$  ou transinformation interne de l'ensemble  $\{X, Y\}$  noté  $\overleftarrow{I}(X : Y) = H(X) - H(X|Y)$

$I(X, Y)$ , sachant que  $H(X|Y) = H(X, Y) - H(Y)$  on a

$$\overleftarrow{I}(X : Y) = H(X) + H(Y) - H(X, Y)$$

La généralisation est immédiate pour plusieurs variables ou groupes de variables.

I.5.2.i. Transinformation interne

C'est la quantité d'information échangée à l'intérieur d'un groupe de variables :

$$I : \mathcal{P}(X) \rightarrow \mathbb{R}$$

$$S \rightarrow I(S) = \sum_{i \in J_S} H(\{X_i\}) - H(S)$$

\* Remarque: on convient de poser  $\overleftarrow{I}(\{X, X\}) = \overleftarrow{I}(X) = 0$

I.5.2.j. Transinformation externe

C'est la quantité d'information échangée entre groupes de variables :

$$\begin{aligned} \overleftarrow{I} : \mathcal{P}(\mathcal{P}(\Sigma)) &\rightarrow \mathbb{R} \\ R = \{ R_1, \dots, R_n \} &\rightarrow \overleftarrow{I}(R) = \sum_{i=1}^K H(R_i) - H(\bigcup_{i=1}^k R_i) \end{aligned}$$

I.5.2.k. Propriétés

L'entropie et la transinformation vérifient certaines propriétés intéressantes ( SHAN 49, ASH 65, MILL 63, TORO 82)

$$\forall S_1, S_2, S_3 \in \mathcal{P}(\Sigma) \quad \forall R, Q \in \mathcal{P}(\Sigma)$$

(1)  $H(S_1) \geq 0 \quad \dot{I}(S_1) \geq 0, H(S_1/S_2) \geq 0, \overleftarrow{I}(R) \geq 0$

(2)  $H(S_1 \cap S_2) + H(S_1 \cup S_2) \leq H(S_1) + H(S_2)$ , l'égalité est vérifiée si et seulement si  $S_1 \cap S_2 = \emptyset$

(3)  $S_1 \subseteq S_2 \implies H(S_1) \leq H(S_2) \quad \dot{I}(S_1) \leq \dot{I}(S_2), H(S_3/S_1) \geq H(S_3/S_2)$

(4)  $R \subseteq Q \implies \overleftarrow{I}(R) \geq \overleftarrow{I}(Q)$

(5)  $\forall R = \{ R_1, \dots, R_r \} \in \mathcal{P}(\Sigma)$

$\overleftarrow{I}(R) = 0 \iff R_1, R_2, \dots, R_r$  sont des sous-ensembles de variables indépendants entre eux

(6)  $\overleftarrow{I}(R) + \sum_{i=1}^r \dot{I}(R_i) = \text{constante} = \overleftarrow{I}(\{X_i\}, i=1, \dots, N)$

$$\forall R = \{ R_1, R_2, \dots, R_i, \dots, R_j, \dots, R_r \} \in \mathcal{P}(\Sigma)$$

(7) soit  $R' = \{ R_1, R_2, \dots, R_i \cup R_j, \dots, R_r \}$  alors

$$\overleftarrow{I}(R') = \overleftarrow{I}(R) - \overleftarrow{I}(R_i, R_j)$$

(8) Soit  $R' = \{ R_1, R_2, \dots, R_i \setminus S, S, \dots, R_r \}$  avec  $S \subseteq R_i, S \neq R_i, S \neq \emptyset, |R_i| \geq 2$

$$\text{on a } \overleftarrow{I}(R') = \overleftarrow{I}(R) + \overleftarrow{I}(R_i \setminus S, S)$$

(9) Principe de conditionnement uniforme (GARN 56)

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

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

On a d'après ce principe :

$$\forall C \in \mathcal{P}(\Sigma) \quad H(A \cup B / C) = H(A/C) + H(B/C) - \overleftrightarrow{I}_C(A:B)$$

Ces propriétés seront utilisées tout au long des chapitres ultérieurs.

I.5.2.b. Cas des variables continues

Dans le cas continu, une variable prend ses valeurs dans un intervalle  $[a, b]$  et qui a une densité de probabilité  $\rho$  telle que

$$\begin{aligned} \rho(X) &\geq 0 \\ \int \rho(X) \, dX &= 1 \end{aligned}$$

a pour entropie

$$H_c(X) = - \int \rho(X) \log \rho(X) \, dX$$

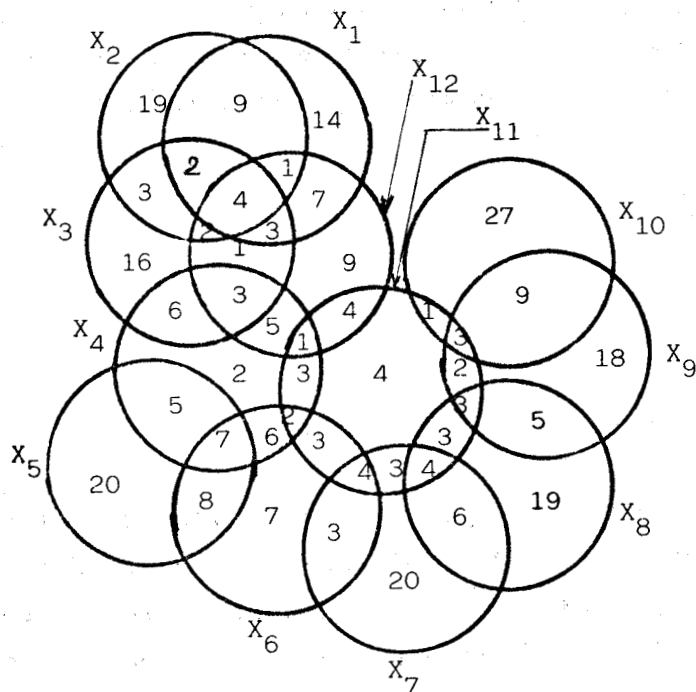
L'entropie continue n'a pas les mêmes propriétés que dans le cas discret et elle peut prendre des valeurs négatives. Le calcul de l'entropie d'une variable continue par discrétisation n'est pas sans inconvénient et impose des choix judicieux pour que cette discrétisation n'affecte pas les résultats de l'analyse (DUF. 75).

Notre étude sera consacrée par la suite aux variables discrètes.

I.5.2.m Analogie entre les théorie de l'information et la théorie de la mesure

Miller a établi une analogie entre la théorie de l'information et la théorie de la mesure en comparant leurs propriétés respectives (MILL. 63). Cette analogie permet de représenter chaque variable par un cercle de surface proportionnelle à son entropie. La surface de l'intersection entre deux cercles représente la transinformation entre les deux variables correspondantes.

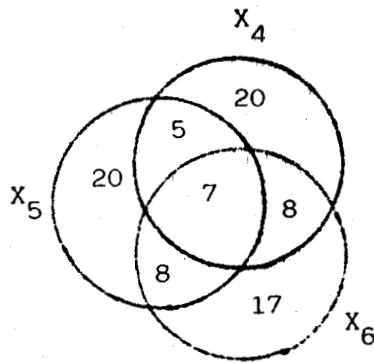
Cette analogie est illustrée par un exemple (Fig. I.4.)



(Fig. I.4.) : Représentation plane du système.



Prenons par exemple les variables  $X_4, X_5, X_6$ .



On a  $H(X_4) = H(X_5) = H(X_6) = 40$

$$H(X_4|X_5) = 28$$

$$\overleftrightarrow{I}(X_4 : X_5) = 12$$

$$\overleftrightarrow{I}(X_4 : X_5 : X_6) = 7 \times 2 + 8 + 8 + 5 = 35$$

le chiffre 7 compte deux fois car la surface correspondante est à l'intersection de 3 cercles. De façon générale, lorsqu'on a l'intersection de  $n$  variables, on multiplie le chiffre correspondant à cette intersection par  $n-1$ , d'où la transinformation entre ces variables.

Cette représentation graphique n'est pas réciproque, en effet, TORO a montré qu'à toute représentation graphique correspond un système "réel", mais tout système réel n'a pas forcément de représentation graphique (TORO 82).

### I.6, CONCLUSION

Ce chapitre introductif avait pour but de faire une présentation générale des données et d'exposer un ensemble de méthodes d'analyse et de décomposition des systèmes complexes.

Chaque méthode possède son domaine privilégié d'application qui est fonction de la nature particulière du système considéré et des buts précis d'analyse. Ainsi, l'analyse en composantes principales présente un grand intérêt lorsque les données sont quantitatives. L'analyse canonique et la régression linéaire permettent de mesurer l'intensité de liaison entre variables ou groupes de variables, les relatins étant linéaires.

Nous avons décrit une autre approche qui repose sur la théorie de l'information. La transinformation permet de mesurer des couplages quelconques dans des systèmes discrets, mais nécessite pour les systèmes continus une discrétisation des intervalles des valeurs des variables qui peut entraîner une perte d'information sur ces couplages. Nous utilisons cet outil pour l'analyse des systèmes complexes. Nous aborderons dans un premier temps les méthodes d'explication.

CHAPITRE II

---

MÉTHODES D'EXPLICATION

## II.1 - INTRODUCTION

Dans le tableau initial des données (fig.I.1), on considère chaque variable  $X_i$  comme une variable aléatoire, dont les données constituent des réalisations (au sens des probabilités). Cependant, la nature du système, déterministe ou stochastique, n'intervient pas dans l'analyse puisqu'il n'est pas nécessaire de spécifier comment les données ont été obtenues.

Le but de l'analyse est de déterminer les intensités de couplage entre les variables. C'est en ce sens que les concepts probabilistes sont les outils de l'analyse des relations causales dans un système complexe.

Nous exposons dans ce chapitre quelques méthodes d'explications d'une variable par un groupe de variables. La détermination des couplages entre les variables du système prises deux par deux, permet de construire une matrice de couplage. La structure du système peut alors être extraite de cette matrice. On montre que l'analyse par paires de variables est insuffisante, et on fait une analyse multidimensionnelle des relations qui débouche sur la construction d'un graphe d'interconnexion.

L'influence du bruit sur les relations fonctionnelles est étudiée. Un algorithme permettant de réduire le nombre de modalités du vecteur explicatif est exposé.

## II.2 - GRAPHES D'INTERCONNEXION

L'analyse des couplages entre variables ou groupes de variables permettra de construire le graphe d'interconnexion à partir duquel on dégagera la structure du système étudié.

La théorie de l'information permet de calculer des indices de couplages. Ces indices nous donnent une valuation des arêtes d'un graphe, les variables étant représentées par les sommets.

On montre qu'une analyse par paires de variables est parfois très approximative. On étudie alors une approche permettant la prise en compte des relations globales par groupes de variables.

II.2.1. - Analyse par paires de variables  
=====

Le but de ce paragraphe est de déterminer les intensités de couplage entre les variables prises deux par deux, au moyen des indices de similarité. Ces indices permettent de construire une matrice de couplage ou un graphe d'interconnexion.

Nous étudierons ces indices dans le cas statique et dans le cas dynamique et nous soulignons leurs limites respectives.

II.2.1.a - Cas des systèmes statiques

La transinformation  $\overleftrightarrow{I}(X_i, X_j)$  est une mesure du couplage statique entre  $X_i$  et  $X_j$ . Dans le but d'éliminer l'influence du nombre de valeurs possibles pour chaque variable, RICHTIN propose le coefficient de couplage  $r_{ij}$  :

$$r_{ij} = \frac{\overleftrightarrow{I}(X_i : X_j)}{\sqrt{H(X_i)H(X_j)}} \quad (\text{RICH.75})$$

CONANT utilise un coefficient plus général  $C_{ij} = \frac{\overleftrightarrow{I}(X_i, X_j)}{H(X_j)}$  (CONA.72)

Une fois calculées les intensités de couplages entre les variables, les auteurs donnent la structure du système en reconnaissant les sous-systèmes faiblement couplés. Le grand risque des indices cités ci-dessus est qu'ils détectent des dépendances apparentes. La dépendance statistique n'implique pas nécessairement l'effet causal direct entre  $X_i$  et  $X_j$ .

Une autre approche consiste à déterminer les variables explicatives de variables dépendantes. On utilise l'indice  $\overleftrightarrow{I}_{\Sigma} - (X_i, X_j)(X_i : X_j)$  qui est une mesure du couplage direct entre  $X_i$  et  $X_j$ .

Nous donnons ici le tableau de cet indice, calculé sur l'exemple de la figure I.4 :

	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>	X <sub>9</sub>	X <sub>10</sub>	X <sub>11</sub>	X <sub>12</sub>
X <sub>1</sub>	40	9	0	0	0	0	0	0	0	0	0	7
X <sub>2</sub>		40	3	0	0	0	0	0	0	0	0	0
X <sub>3</sub>			40	6	0	0	0	0	0	0	0	1
X <sub>4</sub>				40	5	6	0	0	0	0	3	5
X <sub>5</sub>					40	8	0	0	0	0	0	0
X <sub>6</sub>						40	3	0	0	0	3	0
X <sub>7</sub>							40	6	0	0	3	0
X <sub>8</sub>								40	5	0	3	0
X <sub>9</sub>									40	9	2	0
X <sub>10</sub>										40	1	0
X <sub>11</sub>											40	4
X <sub>12</sub>												40

Figure II.1

Cette matrice de couplage est schématisée par la figure II.2, sur laquelle les intensités de liaison sont proportionnelles à l'épaisseur de la flèche,



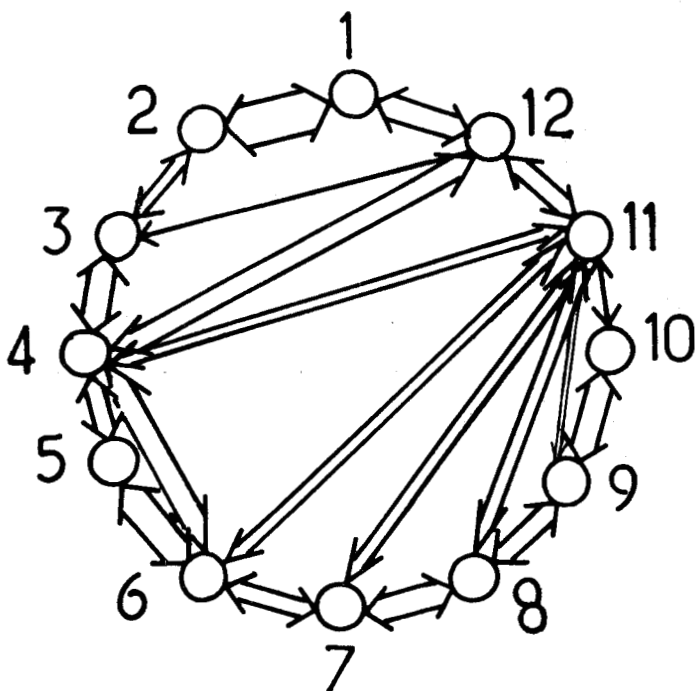


Figure II.2

En examinant ce graphe, on peut avoir le modèle suivant :

$$X_1 = f_1 (X_2, X_{12})$$

$$X_7 = f_7 (X_6, X_8, X_{11})$$

$$X_2 = f_2 (X_1, X_3)$$

$$X_8 = f_8 (X_7, X_9, X_{11})$$

$$X_3 = f_3 (X_2, X_4)$$

$$X_9 = f_9 (X_8, X_{10}, X_{11})$$

$$X_4 = f_4 (X_3, X_5, X_6, X_{12})$$

$$X_{10} = f_{10} (X_9, X_{11})$$

$$X_5 = f_5 (X_4, X_6)$$

$$X_{11} = f_{11} (X_4, X_6, X_7, X_8, X_9, X_{10}, X_{12})$$

$$X_6 = f_6 (X_4, X_5, X_7, X_{11})$$

$$X_{12} = f_{12} (X_1, X_3, X_4, X_{11})$$

Le coefficient de couplage de CONANT  $C_{ij} = \frac{\overleftrightarrow{I}(X_i : X_j)}{H(X_j)}$  est plus réaliste

que celui de RICHTERIN  $r_{ij} = \frac{\overleftrightarrow{I}(X_i : X_j)}{\sqrt{H(X_i)H(X_j)}}$ , en effet :  $0 \leq C_{ij} \leq 1$   $C_{ij} = 1$  ssi

il existe une application de  $M_i$  dans  $M_j$ .  $M_i$  (respectivement  $M_j$ ) est l'ensemble des valeurs de  $X_i$  (respectivement  $X_j$ )  $0 \leq r_{ij} \leq 1$   $r_{ij} = 1$  ssi il existe une application biunivoque de  $M_i$  dans  $M_j$ .



Le coefficient de couplage direct  $\overleftrightarrow{I}_{\Sigma} - (X_i, X_j) (X_i, X_j)$  est certes le plus efficace pour la recherche des variables explicatives de variables dépendantes. Cependant, certains couplages entre  $X_i$  et  $X_j$  ne sont pas mesurés dans le cas par exemple où  $\overleftrightarrow{I}(X_i : (X_j, X_k)) \neq 0$  et  $\overleftrightarrow{I}_{\Sigma} - (X_i, X_j) (X_i : X_j) = 0$  (cf proposition II.2 p.32)

### II.2.1.b - Cas des systèmes dynamiques

La recherche d'un modèle dans un système dynamique concerne pour chaque variable, sa valeur à un instant quelconque et certaines de ses valeurs prises à des instants antérieurs. On note  $X'_j$  la valeur de  $X_j$  à un instant décalé d'une période d'échantillonnage  $T$  par rapport à l'instant de la mesure de  $X_i$ .

Comme précédemment, la matrice de couplage dynamique peut être obtenue à partir des coefficients :

$$r'_{ij} = \frac{\overleftrightarrow{I}(X_i : X'_j)}{\sqrt{H(X_i) \cdot H(X'_j)}}$$

$0 \leq r'_{ij} \leq 1$   $r'_{ij} = 1$  ssi il existe une application biunivoque de  $M_i$  dans  $M'_j$  ( $M'_j =$  ensemble des valeurs de  $X'_j = M_j$ ).

$$\text{ou } C_{ij} = \frac{\overleftrightarrow{I}(X_i : X'_j)}{H(X'_j)}$$

$0 \leq C'_{ij} \leq 1$   $C'_{ij} = 1$  ssi il existe une application de  $M_i$  dans  $M'_j$ . Ce coefficient est donc plus général que le précédent.

RINK a proposé le coefficient suivant :

$$m_{ij} = \frac{\overleftrightarrow{I}_{\Sigma} - X_i (X_i : X'_j)}{H_{\Sigma} - X_i (X'_j)}$$

qui est une mesure du couplage dynamique direct entre  $X_i$  et  $X_j$ .

Ce coefficient élimine la mesure des couplages dynamiques apparents dûs à des couplages statiques (RINK.73). Cependant, il nécessite beaucoup d'échantillons pour que les résultats soient significatifs.

II.2.1.c - Insuffisance de l'approche binaire

Si la somme des interactions est importante, les couplages par paires de variables sont faibles et ne permettent pas une analyse correcte du système. Les propositions ci-dessous montrent l'insuffisance de l'analyse par paires.

Proposition II.1 : Décomposition de l'information interne dans un système

Soient  $X_1, X_2 \dots X_N$  des variables définies sur  $\Omega$ , alors :

$$\begin{aligned} \overline{I}(X_1 : X_2 : X_3 \dots : X_N) &= \sum_{\substack{i,j \\ i \neq j}} \overleftarrow{I}(X_i : X_j) + \sum_{\substack{i,j,k \\ i \neq j \neq k}} Q(X_i, X_j, X_k) \\ &+ \sum_{\substack{i,j,k,l \\ i \neq j \neq k \neq l}} Q(X_i, X_j, X_k, X_l) + \dots Q(X_1, \dots, X_N). \end{aligned}$$

Q étant une mesure de l'interaction à l'intérieur de groupes de variables.

$$Q(X_i, X_j, \dots, X_m) = Q_{X_i}(X_j, \dots, X_m) - Q(X_j, \dots, X_m)$$

$$\text{avec } Q(X_i, X_j, X_k) = \overleftarrow{I}_{X_i}(X_j : X_k) - \overleftarrow{I}(X_j : X_k)$$

Cette formule a été démontrée par Mc-GILL (GILL.54), elle montre bien que l'information interne d'un système ne dépend pas seulement des informations entre variables prises deux par deux.

Proposition II.2 : Explication d'une variable par d'autres variables du système.

Soient  $X_1, X_2, \dots, X_N$ , un ensemble de variables définies sur  $\Omega$ , alors

$$\overleftrightarrow{I}(X_1 : X_2, X_3, \dots, X_N) = \sum_{k=2}^{N-1} \overleftrightarrow{I}_{X_{k+1}, \dots, X_N}(X_1 : X_k) + \overleftrightarrow{I}(X_1 : X_N)$$

Démonstration

On démontre tout d'abord la relation dans le cas de trois variables  $X_1, X_2, Z$  définies sur  $\Omega$

$$\begin{aligned} \overleftrightarrow{I}(X_1 : X_2, Z) - \overleftrightarrow{I}_Z(X_1 : X_2) &= H(X_1) + H(X_2, Z) - H(X_1, X_2, Z) - H(X_1/Z) - H(X_1/Z) + \\ &\quad H(X_1, X_2/Z) \\ &= H(X_1) + \cancel{H(X_2, Z)} - \cancel{H(X_1, X_2, Z)} - H(X_1, Z) + H(Z) - \cancel{H(X_2, Z)} + \cancel{H(Z)} + \cancel{H(X_1, X_2, Z)} - \cancel{H(Z)} \\ &= H(X_1) + H(Z) - H(X_1, Z) = \overleftrightarrow{I}(X_1 : Z) \end{aligned}$$

$$\overleftrightarrow{I}(X_1 : X_2, Z) = \overleftrightarrow{I}_Z(X_1 : X_2) + \overleftrightarrow{I}(X_1 : Z)$$

On remplace  $Z$  par le vecteur  $(X_3, X_4, \dots, X_N)$

$$\overleftrightarrow{I}(X_1 : X_2, X_3, \dots, X_N) = \overleftrightarrow{I}_{X_3, \dots, X_N}(X_1 : X_2) + \overleftrightarrow{I}(X_1 : X_3, X_4, \dots, X_N)$$

En développant de la même façon le deuxième terme du deuxième membre, il vient :

$$\overleftrightarrow{I}(X_1 : X_2, X_3, \dots, X_N) = \overleftrightarrow{I}_{X_3, \dots, X_N}(X_1 : X_2) + \overleftrightarrow{I}_{X_4, \dots, X_N}(X_1 : X_3) + \overleftrightarrow{I}(X_1 : X_4, X_5, \dots, X_N)$$

De proche en proche, on aboutit au résultat

$$\begin{aligned} \overleftrightarrow{I}(X_1 : X_2, X_3, \dots, X_N) &= \overleftrightarrow{I}_{X_3, \dots, X_N}(X_1 : X_2) + \overleftrightarrow{I}_{X_4, \dots, X_N}(X_1 : X_3) + \dots + \overleftrightarrow{I}_{X_N}(X_1 : X_{N-1}) + \overleftrightarrow{I}(X_1 : X_N) \\ &= \sum_{k=2}^{N-1} \overleftrightarrow{I}_{X_{k+1}, \dots, X_N}(X_1 : X_k) + \overleftrightarrow{I}(X_1 : X_N) \end{aligned}$$

Pour illustrer ces résultats, nous considérons un exemple de trois variables binaires :

$\Sigma$ $\Omega$	$X_1$	$X_2$	$X_3$
$\omega_1$	0	0	1
$\omega_2$	0	1	0
$\omega_3$	1	0	0
$\omega_4$	1	1	1

Figure II.3

Le calcul des probabilités conjointes et de l'entropie pour tous les sous-ensembles de  $\Sigma = \{ X_1, X_2, X_3 \}$  est présenté sous forme de tableaux.

	0	1	H
$X_1$	$\frac{1}{2}$	$\frac{1}{2}$	1
$X_2$	$\frac{1}{2}$	$\frac{1}{2}$	1
$X_3$	$\frac{1}{2}$	$\frac{1}{2}$	1

	00	01	11	10	H
$X_1 X_2$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	2
$X_2 X_3$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	2
$X_3 X_1$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	2

	000	001	011	010	110	111	101	100	H
$X_1 X_2 X_3$	0	$\frac{1}{4}$	0	$\frac{1}{4}$	0	$\frac{1}{4}$	0	$\frac{1}{4}$	2

D'après ces entropies, on obtient :

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



Une analyse utilisant cet indice de couplage va conclure qu'il n'y a aucun lien entre les variables, alors qu'on a en réalité :

$$\forall i \neq j \neq k \quad \overleftarrow{I}_{X_k}(X_i : X_j) = 1 \quad \text{et} \quad \overleftarrow{I}(X_i : X_j : X_k) = 1$$

Cela implique qu'il y a bien un lien entre les variables. Une étude multidimensionnelle est donc nécessaire, ainsi dans cet exemple, on vérifie que  $H(X_i / X_j, X_k) = 0 \forall i \neq j \neq k$ , c'est-à-dire qu'une variable est complètement déterminée lorsqu'on connaît les deux autres.

En conclusion, si les interactions sont relativement élevées, l'analyse à partir des couplages entre paires de variables est imprécise. Des dépendances apparentes peuvent être mesurées. Pour pallier cette lacune, on peut utiliser la transinformation conditionnelle  $\overleftarrow{I}_{\Sigma - \{X_i, X_j\}}(X_i : X_j)$  qui correspond au couplage direct entre  $X_i$  et  $X_j$ ; mais certains couplages entre  $X_i$  et  $X_j$  ne sont pas mesurés dans le cas par exemple où  $\overleftarrow{I}(X_i : (X_j, X_k)) \neq 0$  et  $\overleftarrow{I}_{\Sigma - \{X_i, X_j\}}(X_i : X_j) = 0$ .

L'étude globale des relations qui existent entre toutes les variables d'un système nécessite une analyse multidimensionnelle.

## II.2.2, - Analyse multidimensionnelle

=====

Il s'agit de déterminer la structure de liaisons entre les variables du système. Ceci aboutit à la construction d'un graphe de dépendance. Notre objectif est alors de déterminer pour chaque variable de  $\Sigma$ , le groupe de variables minimum qui l'explique le mieux.

### II.2.2, a - Indices de liaison

Pour mesurer l'intensité des relations entre sous-systèmes, on se propose d'expliquer chaque variable  $X_i$  de  $\Sigma$  à partir d'autres. On convient d'appeler les premières, variables à expliquer  $\{ X_i \} i = 1, \dots, N$ , et les secondes, variables explicatives  $\Theta_i \subseteq \Sigma - \{ X_i \}$ .

Soit  $S$ , un ensemble de variables, telles que  $X_i \notin S$ . L'intensité de liaison entre  $X_i$  et  $S$  peut être mesurée par l'entropie conditionnelle  $H(X_i/S)$ . En fait, on a :

. Entropie de  $X_i$  expliquée par  $S$  :

$$I(X_i : S) = H(X_i) - H(X_i/S)$$

. Entropie de  $X_i$  non expliquée par  $S$  :

$$H(X_i/S) = H(X_i, S) - H(S)$$

Notre but est de chercher le meilleur ensemble explicatif  $\xi_i$ , tel que :

$$(P) H(X_i/\xi_i) \leq H(X_i/S) \quad \forall S \in \mathcal{P}(\theta_i)$$

Soit  $\mathcal{P}_K(\theta_i)$ , l'ensemble des parties de  $\theta_i$  dont le cardinal est  $K$ .

Soit  $\xi_i^k$  le meilleur ensemble explicatif de la variable  $X_i$ ,  $\xi_i^k \in \mathcal{P}_K(\theta_i)$ .

On a d'après la relation (3.§.I.5.2.k)

$$H(X_i/\xi_i^K) \leq H(X_i/\xi_i^{K-1})$$

de sorte que le problème (P) conduit à la solution triviale  $\xi_i = \theta_i$ . Nous allons alors chercher des optimums locaux. Le problème P devient, pour  $K$  donné trouver  $\xi_i^K$  tel que

$$H(X_i/\xi_i^K) \leq H(X_i/S) \quad \forall S \in \mathcal{P}_K(\theta_i).$$

Algorithmes de détermination de  $\xi_i^K$

---

### Algorithme agrégatif AA

Cet algorithme construit une chaîne de sous-ensembles explicatifs  $S_1, S_2, \dots, S_{/ \theta_i /}$  avec  $S_{K+1} = S_K \cup \{ X_\alpha \}$  la variable  $X_\alpha$  étant choisie de façon à minimiser  $H(X_i/S_K \cup \{ X_\alpha \})$  pour chaque variable  $X_i \in \Sigma$ . Les sous-ensembles explicatifs optimaux  $\xi_i^1, \xi_i^2, \dots, \xi_i^{/ \theta_i /}$  ne constituent pas forcément une chaîne ordonnée par la relation  $\underline{C}$ . En conséquence, tout algorithme dont le résultat est une chaîne peut ne pas trouver tous les optimums locaux.

Algorithme désagrégatif AD

Cet algorithme construit une chaîne  $S_{/0_i/}, S_{/0_i/-1}, \dots, S_2, S_1$  avec  $S_{K-1} = S_K \setminus \{X_\alpha\}$ , la variable  $x_\alpha$  étant choisie de façon à minimiser  $H(X_i/S_{K-1})$  pour chaque variable  $X_i \in \Sigma$ . Pas plus que l'algorithme AA, il n'est pas sûr que l'algorithme AD trouve tous les optimums locaux.

Algorithme d'approximation successives AAS

Cet algorithme consiste à appliquer successivement les deux précédents jusqu'à stabilisation. Il cherche pour chaque variable  $X_i \in \Sigma$  et pour chaque  $K$   $1 \leq K \leq /0_i/ - 1$ , un sous-ensemble de  $0_i$  vérifiant une condition nécessaire d'optimalité (TORO.82).

II.2.3. - Graphe d'interconnexion

=====

On peut construire les graphes d'interconnexion ainsi :

$$G_K = (\Sigma, U_K) \quad K = 1, 2, \dots, N$$

avec  $\Sigma = \{X_1, X_2, \dots, X_N\}$  l'ensemble de sommets,

$$U_K = \{(x_i, x_j) : x_i \in \xi_j^K\} \text{ l'ensemble d'arcs}$$

Un arc  $(X_i, X_j)$  peut être interprété comme : "la variable  $x_i$  contribue avec d'autres à déterminer ou à expliquer la variable  $X_j$ ". La caractéristique la plus importante du graphe  $G_K$  est son caractère multidimensionnel, qui prend en compte des aspects échappant aux graphes classiques construits sur des critères binaires.

Les différents résultats concernant l'exemple de la figure I.4 sont représentés dans le tableau ci-dessous. Ceux-ci ont été obtenus à partir de l'algorithme d'approximation successives.

	k	1	2	3	4	5	6	7...
X <sub>1</sub>	$\xi^k$	X <sub>2</sub>	X <sub>2</sub> X <sub>12</sub>	"	"	"	"	"
	H(X <sub>1</sub> / $\xi^k$ )	24	14	"	"	"	"	"
X <sub>2</sub>	$\xi^k$	X <sub>1</sub>	X <sub>1</sub> X <sub>3</sub>	"	"	"	"	"
	H(X <sub>2</sub> / $\xi^k$ )	24	19	"	"	"	"	"
X <sub>3</sub>	$\xi^k$	X <sub>12</sub>	X <sub>4</sub> X <sub>12</sub>	X <sub>2</sub> X <sub>4</sub> X <sub>12</sub>	"	"	"	"
	H(X <sub>3</sub> / $\xi^k$ )	27	21	16	"	"	"	"
X <sub>4</sub>	$\xi^k$	X <sub>6</sub>	X <sub>3</sub> X <sub>6</sub>	X <sub>3</sub> X <sub>6</sub> X <sub>12</sub>	X <sub>3</sub> X <sub>5</sub> X <sub>6</sub> X <sub>12</sub>	X <sub>3</sub> X <sub>5</sub> X <sub>6</sub> X <sub>11</sub> X <sub>12</sub>	"	"
	H(X <sub>4</sub> / $\xi^k$ )	25	16	10	5	2	"	"
X <sub>5</sub>	$\xi^k$	X <sub>6</sub>	X <sub>4</sub> X <sub>6</sub>	"	"	"	"	"
	H(X <sub>5</sub> / $\xi^k$ )	25	20	"	"	"	"	"
X <sub>6</sub>	$\xi^k$	X <sub>5</sub>	X <sub>5</sub> X <sub>11</sub>	X <sub>4</sub> X <sub>5</sub> X <sub>11</sub>	X <sub>4</sub> X <sub>5</sub> X <sub>7</sub> X <sub>11</sub>	"	"	"
	H(X <sub>6</sub> / $\xi^k$ )	26	16	10	7	"	"	"
X <sub>7</sub>	$\xi^k$	X <sub>8</sub>	X <sub>6</sub> X <sub>8</sub>	X <sub>6</sub> X <sub>8</sub> X <sub>11</sub>	"	"	"	"
	H(X <sub>7</sub> / $\xi^k$ )	30	23	20	"	"	"	"
X <sub>8</sub>	$\xi^k$	X <sub>7</sub>	X <sub>7</sub> X <sub>9</sub>	X <sub>7</sub> X <sub>9</sub> X <sub>11</sub>	"	"	"	"
	H(X <sub>8</sub> / $\xi^k$ )	30	22	19	"	"	"	"
X <sub>9</sub>	$\xi^k$	X <sub>10</sub>	X <sub>8</sub> X <sub>10</sub>	X <sub>8</sub> X <sub>10</sub> X <sub>11</sub>	"	"	"	"
	H(X <sub>9</sub> / $\xi^k$ )	28	20	18	"	"	"	"
X <sub>10</sub>	$\xi^k$	X <sub>9</sub>	X <sub>9</sub> X <sub>11</sub>	"	"	"	"	"
	H(X <sub>10</sub> / $\xi^k$ )	29	27	"	"	"	"	"
X <sub>11</sub>	$\xi^k$	X <sub>7</sub>	X <sub>7</sub> X <sub>9</sub>	X <sub>4</sub> X <sub>7</sub> X <sub>9</sub>	X <sub>4</sub> X <sub>7</sub> X <sub>9</sub> X <sub>12</sub>	X <sub>4</sub> X <sub>6</sub> X <sub>7</sub> X <sub>9</sub> X <sub>12</sub>	X <sub>4</sub> X <sub>6</sub> X <sub>7</sub> X <sub>9</sub> X <sub>10</sub> X <sub>12</sub>	"
	H(X <sub>11</sub> / $\xi^k$ )	29	21	15	11	5	4	"
X <sub>12</sub>	$\xi^k$	X <sub>1</sub>	X <sub>1</sub> X <sub>4</sub>	X <sub>1</sub> X <sub>4</sub> X <sub>11</sub>	X <sub>1</sub> X <sub>3</sub> X <sub>4</sub> X <sub>11</sub>	"	"	"
	H(X <sub>12</sub> / $\xi^k$ )	25	16	12	9	"	"	"

BUS LILLE

On constate qu'à partir d'un certain k, les résultats ne peuvent plus être améliorés.



II.3 - INFLUENCE DU BRUIT

Quelle que soit la variable  $Y \in \Sigma$ , la quantité  $H(Y/\lambda)$  indique la possibilité d'obtenir une relation de la forme

$$Y = f(\lambda) \text{ avec } \lambda \in \Sigma - Y$$

$H(Y/\lambda)$  a une valeur proche de zéro lorsque la relation de causalité entre  $Y$  et  $\lambda$  est forte. Par contre, une valeur élevée (proche de  $H(Y)$ ) de  $H(Y/\lambda)$  indique qu'il n'est pas possible de construire un modèle déterministe, le bruit étant prépondérant ou la relation causale n'existant pas. Nous allons étudier dans ce paragraphe, la perte d'information sur  $Y$  en fonction du bruit, dans une relation fonctionnelle (ROMBAUT.82).

II.3.1. - Relations fonctionnelles bruitées

Soient  $Y$  une variable à expliquer, de modalités  $\{ C_1, \dots, C_n \}$  et  $\lambda$  un vecteur explicatif défini sur  $\Omega$  de modalités  $\{ g_1, g_2, \dots, g_N \}$ . Considérons deux partitions de l'ensemble  $\Omega$ ,  $P_Y$  et  $P_\lambda$  associées respectivement à  $Y$  et  $\lambda$ .

$$P_Y = \{ \mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_n \}$$

$$P_\lambda = \{ G_1, G_2, \dots, G_N \}$$

Formons la matrice de contingence notée  $M(Y/\lambda)$  de la façon suivante :

$$P_{ij} = \frac{\text{card}(G_i \cap \mathcal{C}_j)}{\text{card}(\Omega)}$$

$P_{ij}$  est une estimation de la probabilité  $P(G_i \cap \mathcal{C}_j)$

Y λ	$\mathcal{C}_1$	$\mathcal{C}_2$	...	$\mathcal{C}_j$	...	$\mathcal{C}_n$
$G_1$						$P_1$
$G_2$						
...						
$G_i$				$P_{ij}$		$P_i$
...						
$G_N$						$P_N$

Figure II.1

On peut étudier l'existence du couplage entre  $Y$  et  $\lambda$ , se traduisant par une relation analytique  $Y = f(\lambda, \epsilon)$ .  $\epsilon$  est une perturbation due aux effets d'autres variables du système étudié ou du bruit. Pour estimer la part qui peut être accordée à l'ensemble des diverses perturbations, on va étudier la perte d'information en fonction du bruit, dans une relation fonctionnelle.

II.3.1.a - Entropie conditionnelle nulle

S'il existe une relation univoque entre  $Y$  et  $\lambda$ , la matrice de contingence prend la forme suivante :

$\lambda \backslash Y$	$\mathcal{C}_1$	$\mathcal{C}_2 \dots \mathcal{C}_j \dots \mathcal{C}_n$	
$G_1$	$P_1$	0 ... 0	$P_1$
$G_2$	$P_2$	0 ... 0	$P_2$
$\vdots$			$\vdots$
$G_i$	0	0 ... $P_i$ ... 0	$\vdots$
$\vdots$			$\vdots$
$G_N$	0	0 ... 0	$P_N$

Figure II.2

Dans ce cas, la matrice de contingence a un seul élément différent de zéro par ligne et l'entropie conditionnelle de  $Y$  par rapport à  $\lambda$  est nulle.

Réciproquement, si  $H(Y/\lambda) = 0$ , alors dans la matrice de contingence, une seule probabilité par ligne est non nulle, et à chaque  $g_i$  correspond un  $\mathcal{C}_j$  et un seul.

II.3.1.b - Entropie conditionnelle maximale

Si les variables  $Y$  et  $\lambda$  sont indépendantes, les probabilités sur chaque ligne sont égales et l'entropie conditionnelle de  $Y$  par rapport à  $\lambda$  est maximale.

$$H(Y/\lambda) = H(Y)$$

La matrice de contingence deviendra alors :

$\lambda \backslash Y$	$C_1$	$C_2$	$\dots$	$C_j$	$\dots$	$C_n$	
$G_1$	$\frac{P_1}{n}$	$\frac{P_1}{n}$	$\dots$			$\frac{P_1}{n}$	$P_1$
$G_2$	$\frac{P_2}{n}$	$\frac{P_2}{n}$	$\dots$			$\frac{P_2}{n}$	$P_2$
$\vdots$							$\vdots$
$G_i$	$\frac{P_i}{n}$	$\frac{P_i}{n}$	$\dots$			$\frac{P_i}{n}$	$P_i$
$\vdots$							$\vdots$
$G_N$	$\frac{P_N}{n}$	$\frac{P_N}{n}$	$\dots$			$\frac{P_N}{n}$	$P_N$

Figure III.3

Réciproquement, si  $H(Y/\lambda) = H(Y)$  alors, dans la matrice de contingence les probabilités sur chaque ligne sont égales.

II.3.1.c - Relation fonctionnelle bruitée

Lorsque la connaissance de  $\lambda$  ne permet plus de déterminer avec certitude la valeur de  $Y$ , on aura une relation fonctionnelle bruitée :

$$0 < H(Y/\lambda) < H(Y)$$

II.3.2. - Détermination de l'entropie conditionnelle en fonction du bruit

Supposons que  $\lambda$  a le même nombre de modalités que  $Y$ ,  $N=n$ , et que toutes les modalités de  $\lambda$  sont équiprobables,  $P_i = \frac{1}{n}$   $i=1, \dots, N$ . S'il y a une relation fonctionnelle entre  $\lambda$  et  $Y$ , la matrice de contingence est carrée diagonale dont les éléments sont égaux à  $\frac{1}{n}$ . Si cette relation fonctionnelle est bruitée, alors les éléments hors diagonaux ne sont plus nuls. Plus le bruit sera important, plus les éléments  $P_{ij}$  vont s'uniformiser et plus l'incertitude de  $Y$  connaissant  $\lambda$  va augmenter. Afin d'étudier l'évolution de  $H(Y/\lambda)$  en fonction du bruit. Nous avons choisi trois densités de bruits différentes.

II.3.2.a - Bruit à densité uniforme

Le profil des lignes correspondant au bruit à densité de probabilité uniforme est représenté par la figure II.4 et la matrice de probabilités correspondante par la figure II.5.

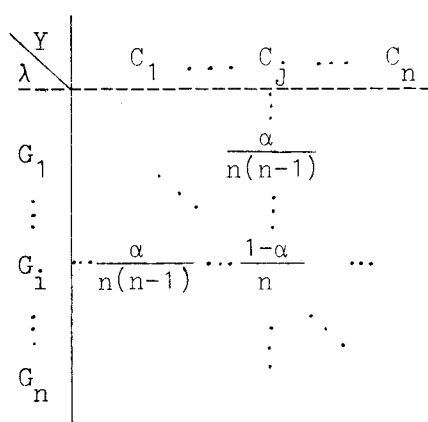
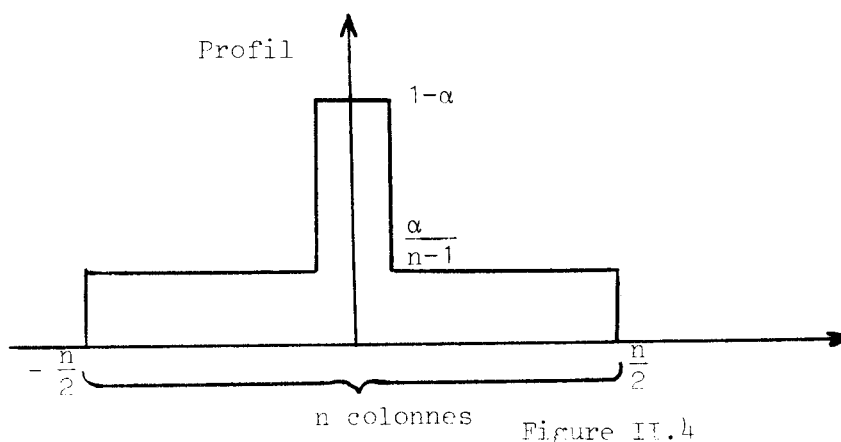


Figure II.5

On vérifie bien que la somme des probabilités sur chaque ligne est égale à  $\frac{1}{n}$ . Pour  $\alpha = 0$ , la relation fonctionnelle entre  $Y$  et  $\lambda$  n'est pas bruitée. Pour  $\alpha = \frac{n-1}{n}$ , le bruit est maximum, l'entropie conditionnelle  $H(Y/\lambda)$  aussi.

On définit un indice de bruit  $\beta$  permettant de normaliser la représentation graphique

$$\beta = \frac{\alpha n}{n-1} \quad 0 \leq \beta \leq 1$$

$\beta = 0 \rightarrow$  bruit nul

$\beta = 1 \rightarrow$  bruit maximum

Calcul de l'entropie conditionnelle : on remplace  $P_i, P_j, P_{ij}$  par leur valeur, il vient :

$$H(Y/\lambda) = -\log n - (1-\alpha) \log \frac{(1-\alpha)}{n} - \alpha \log \frac{\alpha}{n(n-1)}$$

Des courbes donnant l'entropie conditionnelle normalisée  $\frac{H(Y/\lambda)}{H(Y)}$  en fonction de  $\beta$  ont été tracées pour différentes valeurs de  $n$ .

Lorsque  $n$  tend vers l'infini, ces courbes tendent vers une courbe limite d'équation  $L = \alpha$ .

En effet, en développant l'expression de  $H(Y/\lambda)$  ci-dessus, il vient :

$$H(Y/\lambda) = -\alpha \log \alpha - (1-\alpha) \log (1-\alpha) + \alpha \log (n-1)$$

L'entropie conditionnelle normalisée est égale à

$$\frac{H(Y/\lambda)}{H(Y)} = \frac{-\alpha \log \alpha - (1-\alpha) \log (1-\alpha) + \alpha \log (n-1)}{\log n}$$

$$L = \lim_{n \rightarrow \infty} \frac{H(Y/\lambda)}{H(Y)} = \lim_{n \rightarrow \infty} \frac{\alpha \log (n-1)}{\log n} = + \alpha$$

$$\text{or } \lim_{n \rightarrow \infty} \alpha = \lim_{n \rightarrow \infty} \beta \text{ donc } L = + \alpha$$

### II.3.2.b - Bruit à densité de probabilité exponentielle

Le profil des lignes correspondant au bruit à densité de probabilité exponentielle est représenté par la figure ci-dessous :

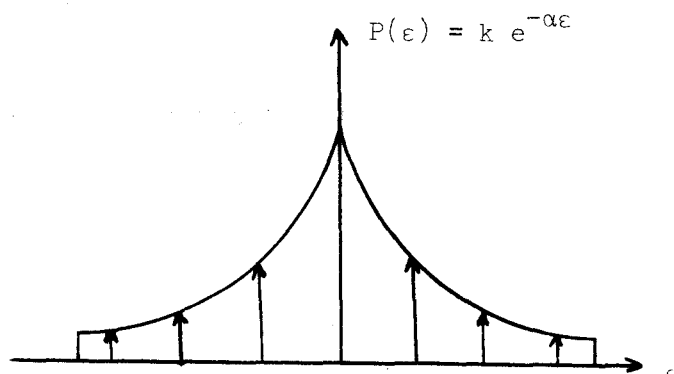


Figure III.6

avec  $\epsilon$  : distance entre le point et l'axe en nombre de modalités.

Le tableau de probabilités dans ce cas possède une relation d'ordre sur les colonnes. Les probabilités sur chaque colonne sont calculées en fonction d'une distance en nombre de modalités par rapport à l'élément principal, c'est-à-dire par rapport à la case de probabilité maximale. Dans ce cas, l'élément principal se trouve sur la diagonale.

$\alpha$  représente l'importance du bruit  $0 \leq \alpha \leq \infty$

$$\alpha = 0 \Rightarrow H(Y/\lambda) = H(Y)$$

$$\alpha = \beta \Rightarrow H(Y/\lambda) = 0$$

$k$  peut être calculé en utilisant le fait que la somme des probabilités sur chaque ligne est égale à  $\frac{1}{n}$ .

$$P_i = \frac{1}{n} \sum_{j=1}^n k(i) e^{-\alpha|i-j|} = \frac{1}{n} \Rightarrow k(i) = \frac{1}{\sum_{j=1}^n e^{-\alpha|i-j|}}$$

$$\text{et } P_{ij} = \frac{e^{-\alpha|i-j|}}{n \sum_{j=1}^n e^{-\alpha|i-j|}}$$

On souhaiterait que le bruit varie de 0 à 1. Pour cela, on fait un changement de variable  $\beta = \frac{1}{1 + \alpha}$  ou  $\alpha = \frac{1 - \beta}{\beta}$

On calcule l'entropie conditionnelle normalisée en fonction de  $\beta$

$$\frac{H(Y/\lambda)}{H(Y)} = \frac{\sum_{j=1}^n p_j \log p_j - \sum_{i=1}^N \sum_{j=1}^n p_{ij} \log p_{ij}}{\log n}$$

Les courbes correspondantes à cette fonction sont tracées pour différentes valeurs de  $N$ . (cf. figure III.9)

II.3.2.c - Bruit à densité de probabilité normale

Le profil des lignes correspondant au bruit à densité de probabilité normale est représenté par la figure suivante :

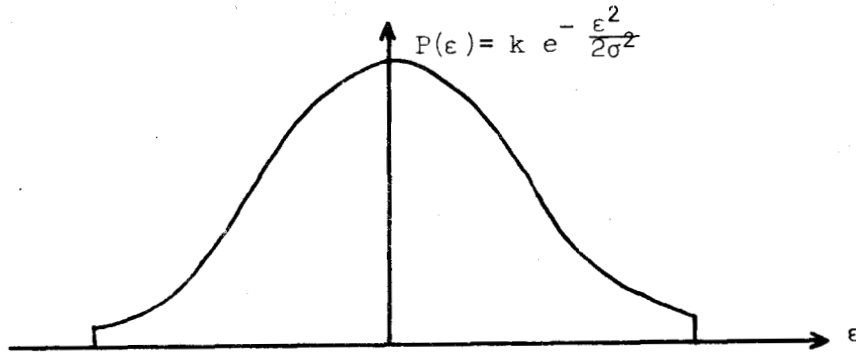


Figure III.7

avec  $\epsilon$  : distance entre le point et l'axe en nombre de modalités

Comme dans le cas du bruit exponentiel, le tableau de probabilités dans ce cas possède une relation d'ordre sur les colonnes. Les probabilités sur chaque colonne sont calculées en fonction d'une distance en nombre de modalités par rapport à l'élément principal.

$\sigma$  est l'écart type qui correspond à l'importance du bruit  $0 \leq \sigma \leq \infty$

$$\sigma = 0 \Rightarrow H(Y/\lambda) = 0$$

$$\sigma = \infty \Rightarrow H(Y/\lambda) = H(Y)$$

$k$  peut être calculée en utilisant le fait que la somme des probabilités sur chaque ligne est égale à  $\frac{1}{n}$

$$P_i = \frac{1}{n} k(i) \sum_{j=1}^n e^{-\frac{(i-j)^2}{2\sigma^2}} = \frac{1}{n} \Rightarrow k(i) = \frac{1}{\sum_{j=1}^n e^{-\frac{(i-j)^2}{2\sigma^2}}}$$

$$\text{et } F_{ij} = \frac{1}{n} \frac{e^{-\frac{(i-j)^2}{2\sigma^2}}}{\sum_{j=1}^n e^{-\frac{(i-j)^2}{2\sigma^2}}}$$

Pour la normalisation de  $\sigma$ , on fait le changement de variable suivant

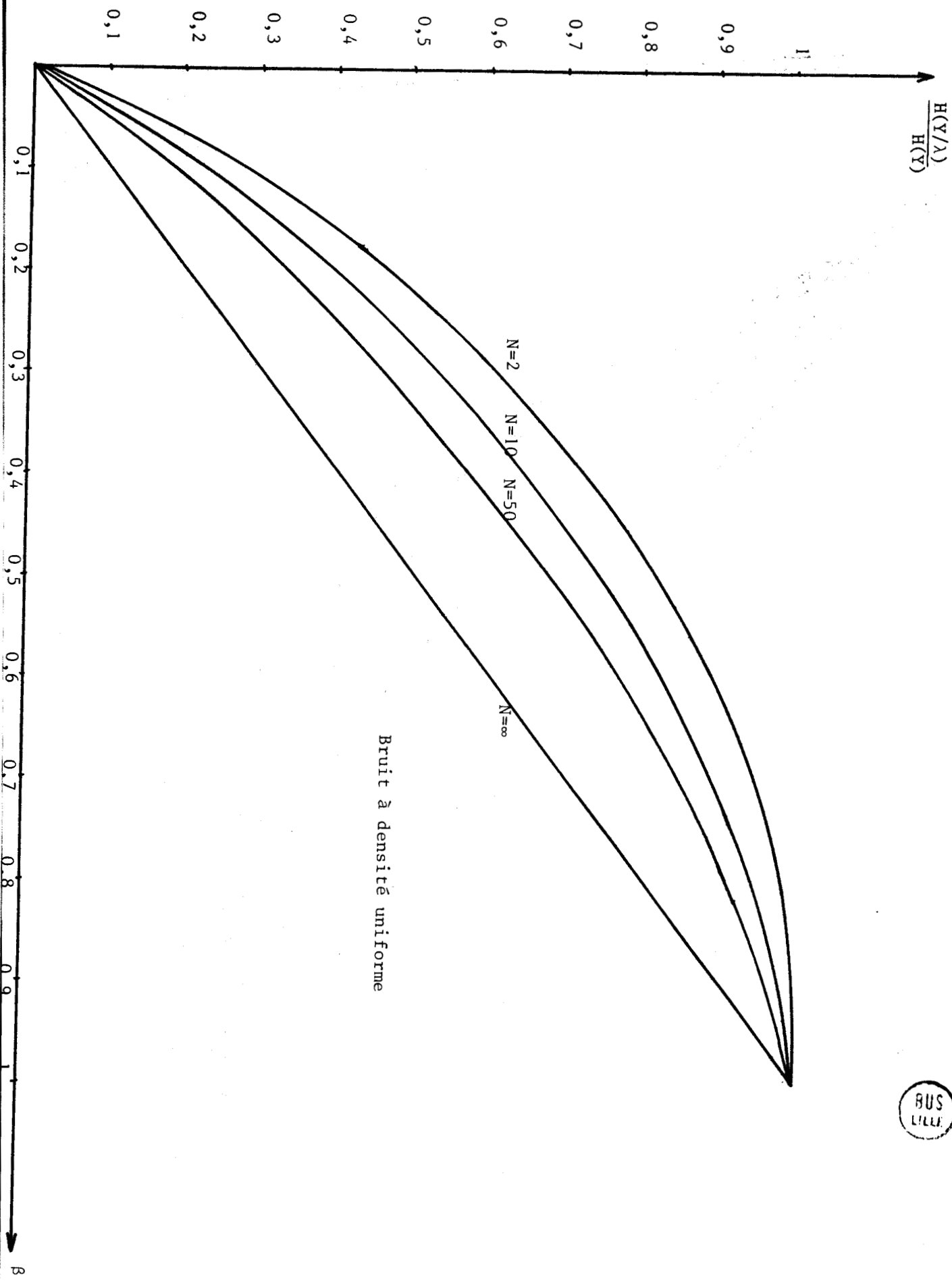
$$\beta = \frac{\sigma}{\sigma + 1} \text{ ou } \sigma = \frac{\beta}{1 - \beta}$$

On calcule l'entropie conditionnelle normalisée en fonction de  $\beta$

$$\frac{H(Y/\lambda)}{H(Y)} = \frac{\sum_{j=1}^n p_j \log p_j - \sum_{i=1}^N \sum_{j=1}^n p_{ij} \log p_{ij}}{\log n}$$

Les courbes correspondantes à cette fonction sont tracées pour différentes valeurs de  $N$ . (cf. Fig. III. 10)





Bruit à densité uniforme

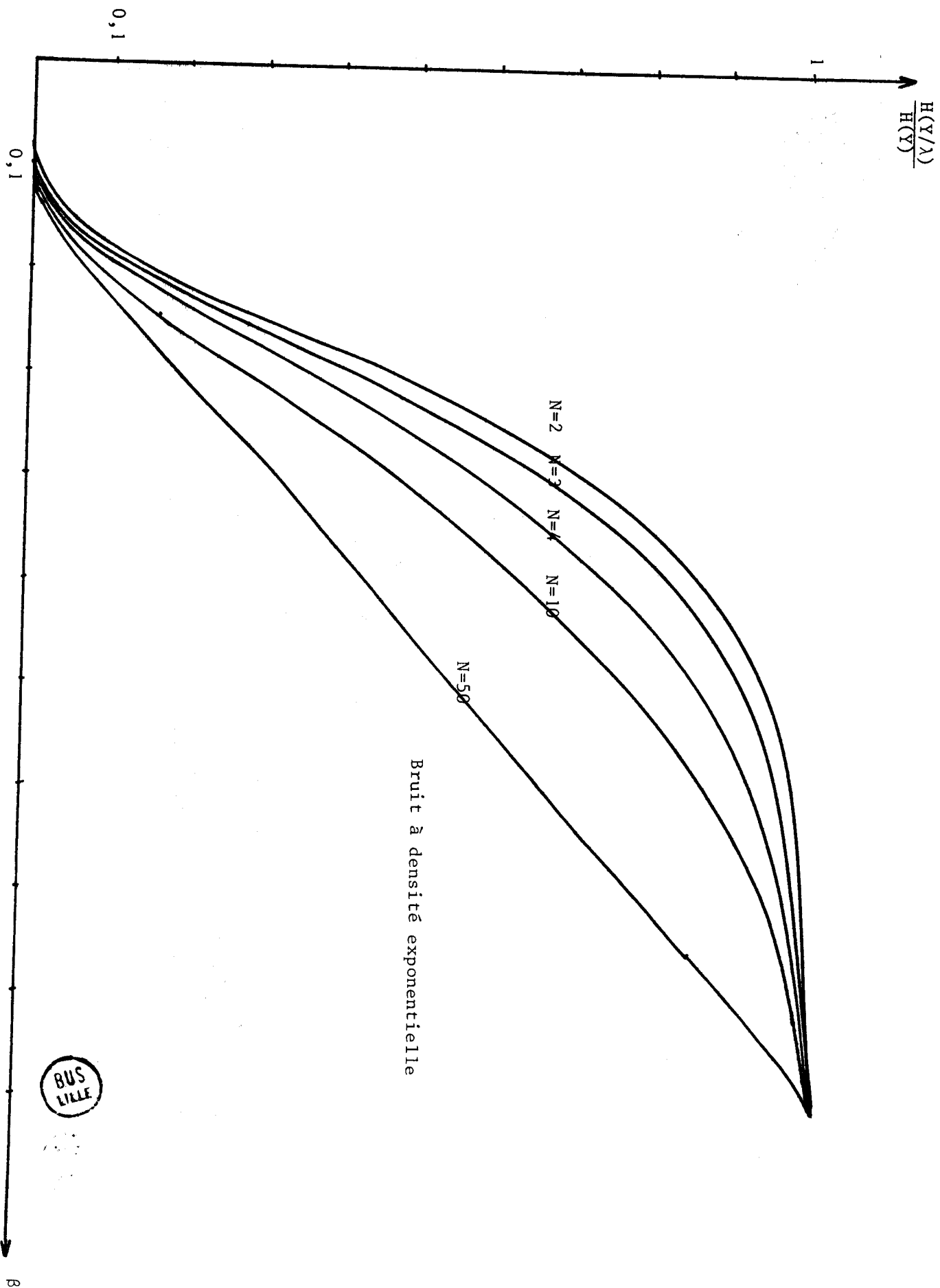


Figure III.9

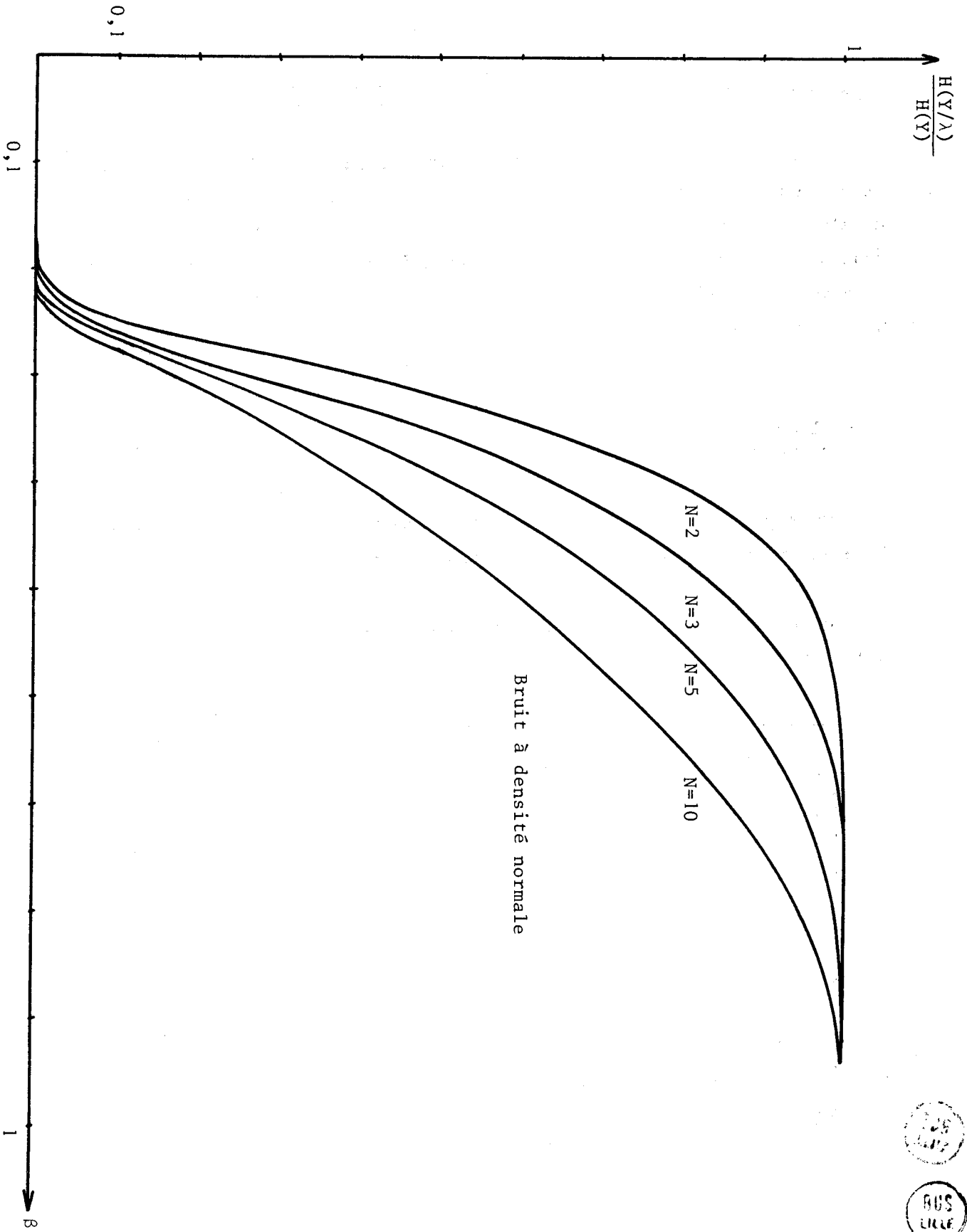


Figure III.10

II.4 - INFLUENCE DU NOMBRE DE MODALITES DU VECTEUR EXPLICATIF

Le vecteur explicatif  $\lambda$  peut posséder un grand nombre de modalités, souvent beaucoup plus grand que le nombre de modalités de la variable à expliquer. On peut alors réduire le nombre de modalités de  $\lambda$  en les regroupant, c'est-à-dire, en créant une partition de l'ensemble des modalités de  $Y$ . Pour cela, nous appliquons l'algorithme hiérarchique ascendant et nous étudions l'évolution de l'information lors de son application. (ROMB.82).

II.4.1. - Algorithme hiérarchique ascendant

On regroupe à chaque itération les deux modalités de  $\lambda$  dont le regroupement occasionnera le moins de perte d'information sur  $Y$ .

Soit la matrice de contingence  $M(Y/\lambda)$  telle que  $N \gg n$  avec  $N = \text{card } \lambda$  et  $n = \text{card } Y$

$\lambda \backslash Y$	$\epsilon_1$	.....	$\epsilon_j$	.....	$\epsilon_n$
$G_1$					
.....					
$G_i$			$P_{ij}$		
.....					
$G_k$			$P_{kj}$		
.....					
$G_N$					

$P_i$

$P_k$

Le regroupement de deux lignes  $i$  et  $k$  correspond à additionner, pour tout  $j = 1$  à  $n$   $P_{ij}$  et  $P_{kj}$ .

L'entropie conditionnelle avant le regroupement est :

$$H_0(Y/\lambda) = - \sum_{l=1}^N \sum_{j=1}^n P_{lj} \log P_{lj} + \sum_{l=1}^N P_l \log P_l$$

L'entropie conditionnelle après le regroupement est :

$$H_1(Y/\lambda) = - \sum_{l=1, l \neq i, k}^N \sum_{j=1}^n P_{lj} \log P_{lj} - \sum_{j=1}^n (P_{ij} + P_{kj}) \log (P_{ij} + P_{kj})$$

$$+ \sum_{l=1, l \neq i, k}^N P_l \log P_l + (P_i + P_k) \log (P_i + P_k)$$

Le critère à minimiser est la différence entre l'entropie conditionnelle après et avant le regroupement c'est-à-dire la perte d'information due au regroupement.

$$\Delta H = H_1(Y/\lambda) - H_0(Y/\lambda) = - P_i \log P_i - P_k \log P_k + (P_i + P_k) \log (P_i + P_k)$$

$$- \sum_{j=1}^n [- P_{ij} \log P_{ij} - P_{kj} \log P_{kj} + (P_{ij} + P_{kj}) \log (P_{ij} + P_{kj})]$$

$$= \sum_{j=1}^n \left( P_{ij} \log \frac{P_{ij}(P_i + P_k)}{P_i(P_{ij} + P_{kj})} + P_{kj} \log \frac{P_{kj}(P_i + P_k)}{P_k(P_{ij} + P_{kj})} \right)$$

Il s'agit de minimiser  $\Delta H$  par rapport aux indices  $i$  et  $k$ , c'est-à-dire par rapport aux lois  $P_{ij}$  et  $P_{ik}$ . On démontre que  $\Delta H$  est minimale et égale à zéro lorsque  $P_{ij}$  et  $P_{kj}$  sont proportionnelles, quelque soit  $j = 1$  à  $n$ . C'est le principe d'équivalence distributionnelle.

Démonstration

En utilisant la méthode des multiplicateurs indéterminés de Lagrange, cherchons l'extrémum de la fonction

$$L = - P_i \log P_i - P_k \log P_k + (P_i + P_k) \log (P_i + P_k) \\ - \sum_{j=1}^n (-P_{ij} \log P_{ij} - P_{kj} \log P_{kj}) + (P_{ij} + P_{kj}) \log (P_{ij} + P_{kj}) \\ + \lambda \left( \sum_{j=1}^n P_{ij} - P_i \right) + \mu \left( \sum_{j=1}^n P_{kj} - P_k \right)$$

En décrivant L par rapport à  $P_i, P_k, P_{ij}, P_{kj}$  et en annulant les dérivées; on obtient le système d'équations suivant :

$$\left\{ \begin{array}{l} \frac{\partial L}{\partial P_i} = - \log P_i - 1 + \log (P_i + P_k) + 1 - \lambda = 0 \\ \frac{\partial L}{\partial P_k} = - \log P_k - 1 + \log (P_i + P_k) + 1 - \mu = 0 \\ \frac{\partial L}{\partial P_{ij}} = \log P_{ij} + 1 - \log (P_{ij} + P_{kj}) - 1 + \lambda = 0 \quad j=1, \dots, n \\ \frac{\partial L}{\partial P_{kj}} = \log P_{kj} + 1 - \log (P_{ij} + P_{kj}) - 1 + \mu = 0 \quad j=1, \dots, n \end{array} \right.$$

$$\left\{ \begin{array}{l} \lambda = \log \frac{P_i + P_k}{P_i} = \log \frac{P_{ij} + P_{kj}}{P_{ij}} \quad j=1, \dots, n \\ \mu = \log \frac{P_i + P_k}{P_k} = \log \frac{P_{ij} + P_{kj}}{P_{kj}} \quad j=1, \dots, n \end{array} \right.$$

$$\left\{ \begin{array}{l} \frac{P_i + P_k}{P_i} = \frac{P_{ij} + P_{kj}}{P_{ij}} \\ \frac{P_i + P_k}{P_k} = \frac{P_{ij} + P_{kj}}{P_{kj}} \end{array} \right. \quad j=1, \dots, n$$

$$\frac{P_i + P_k}{P_{ij} + P_{kj}} = \frac{P_i}{P_{ij}} = \frac{P_k}{P_{kj}} \quad j=1, \dots, n$$

donc  $\frac{P_i}{P_k} = \frac{P_{ij}}{P_{kj}}$

Ce qui montre que l'extrémum (ici le minimum) correspond à des lignes de probabilités proportionnelles entre elles :

$$P_{ij} = \alpha P_{kj} \quad \forall j=1, \dots, n$$

$$P_i = \alpha P_k$$

En remplaçant  $P_i$  et  $P_{ij}$  par leurs valeurs, on obtient :

$$\Delta H = \sum_{j=1}^n \alpha P_{kj} \log \frac{\alpha P_{kj} (\alpha P_k + P_k)}{\alpha P_k (\alpha P_{kj} + P_{kj})} + P_{kj} \log \frac{P_{kj} (\alpha P_k + P_k)}{P_k (\alpha P_{kj} + P_{kj})}$$

$$= \sum_{j=1}^n \alpha P_{kj} \log \frac{P_{kj} P_k (\alpha+1)\alpha}{\alpha P_k (\alpha+1)P_{kj}} + P_{kj} \log \frac{P_{kj} (\alpha+1)P_k}{P_k (\alpha+1)P_{kj}} = 0$$

De même, le regroupement de deux colonnes peut être fait sans perte d'information si leurs profils sont proportionnels. La démonstration est semblable à la précédente.

Des courbes  $\Delta H = f(N)$  sont tracées pour des relations fonctionnelles avec ou sans bruit,  $n$  étant égal à 10. (Figure III.11)

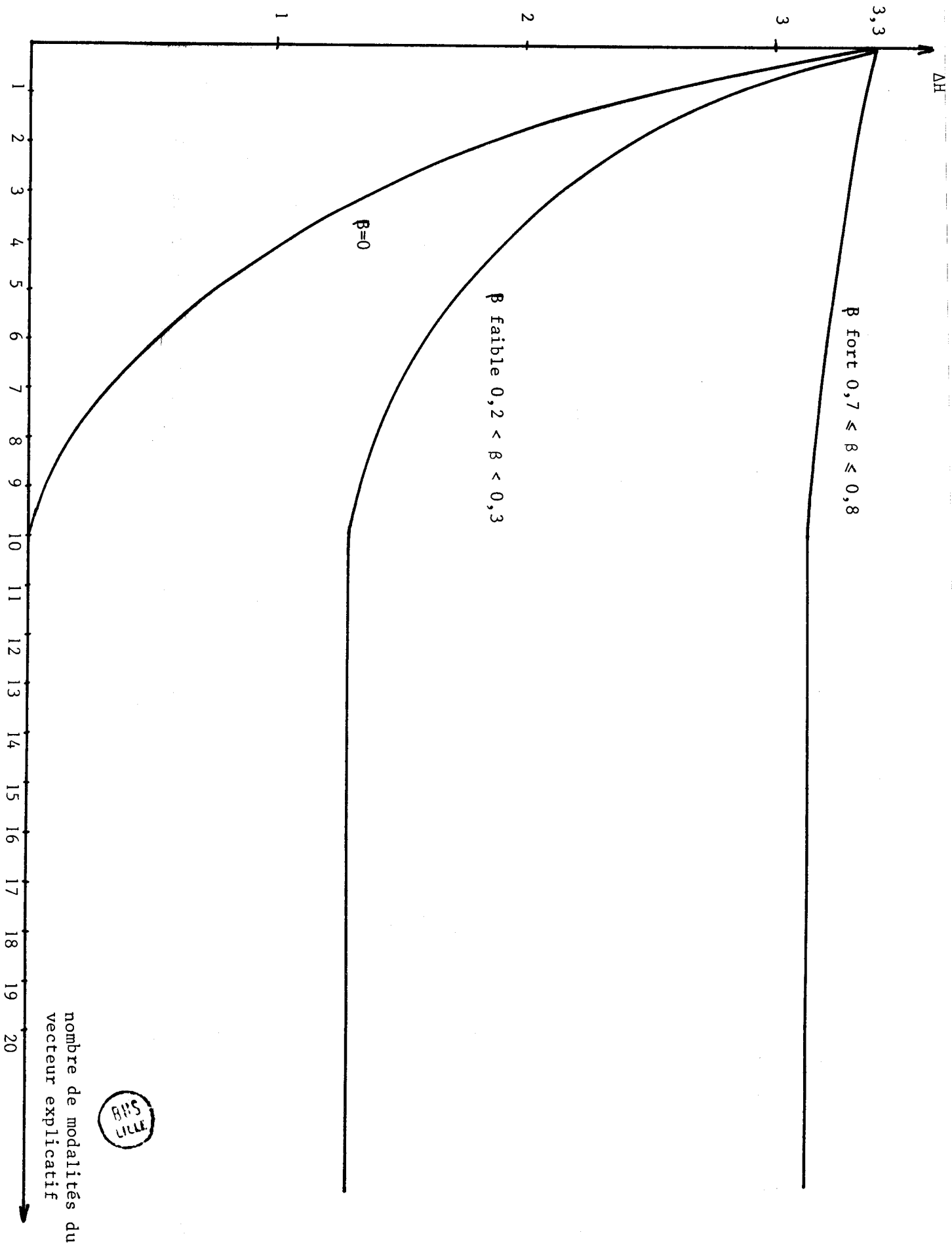


Figure III.1



II.4.2. - Cas d'une relation fonctionnelle bruitée, lorsque les lignes n'ont pas même probabilité

Pour chaque ligne, l'élément principal est la case  $j_m$ , telle que  $P_{ij}$  est maximal sur la ligne de la matrice  $M(Y/\lambda)$ .

On étudie la perte d'information associée au regroupement des lignes  $i$  et  $k$  dont l'élément principal ne se trouve pas sur la même colonne.

Proposition

A poids égal, la perte d'information  $\Delta H$ , obtenue en regroupant deux lignes  $i$  et  $k$  bruitées est plus petite que celle  $\Delta H'$  obtenue en regroupant deux mêmes lignes non bruitées. En plus pour que  $\Delta H$  soit inférieure à un certain seuil  $\gamma$ , il suffit que  $P_i$  et  $P_k$  soient inférieurs à  $\epsilon$  tel que

$$\epsilon = \frac{\gamma}{2 \log 2}$$

Démonstration

Considérons la matrice  $M(Y/\lambda)$  de la figure II.1 . Il s'agit par exemple de regrouper les lignes  $i$  et  $k$ .

$$\begin{array}{l} P_{i1}, P_{i2}, \dots, P_{in} \\ P_{k1}, P_{k2}, \dots, P_{kn} \end{array}$$

avec

$$\begin{array}{l} P_{i1} + P_{i2} + \dots + P_{in} = P_i \\ P_{k1} + P_{k2} + \dots + P_{kn} = P_k \end{array}$$

supposons par exemple que  $P_{i1} > P_{ij} \quad j=2 \dots n$   
 $P_{k2} > P_{kj} \quad j=1 \dots n \quad j \neq 2$

En regroupant ces deux lignes, il vient :

$$\begin{aligned} \Delta H &= \sum_{j=1}^n \left[ P_{ij} \log \frac{P_{ij} (P_i + P_k)}{P_i (P_{ij} + P_{kj})} + P_{kj} \log \frac{P_{kj} (P_i + P_k)}{P_k (P_{ij} + P_{kj})} \right] \\ &= \left( \sum_{j=1}^n P_{ij} \right) \log \frac{P_i + P_k}{P_i} + \left( \sum_{j=1}^n P_{kj} \right) \log \frac{P_i + P_k}{P_k} + \sum_{j=1}^n P_{ij} \log \frac{P_{ij}}{P_{ij} + P_{kj}} \\ &= P_i \log \frac{P_i + P_k}{P_i} + P_k \log \frac{P_i + P_k}{P_k} + \sum_{j=1}^n P_{ij} \log \frac{P_{ij}}{P_{ij} + P_{kj}} \end{aligned}$$

$P_{ij} > 0$  et  $P_{kj} > 0 \forall j=1, \dots, n$  on a

$$\frac{P_{ij}}{P_{ij} + P_{kj}} < 1 \implies P_{ij} \log \frac{P_{ij}}{P_{ij} + P_{kj}} < 0$$

de même  $P_{kj} \log \frac{P_{kj}}{P_{ij} + P_{kj}} < 0$

Il en résulte

$$\Delta H < P_i \log \frac{P_i + P_k}{P_i} + P_k \log \frac{P_i + P_k}{P_k}$$

Considérons maintenant deux lignes où le bruit est nul :

$$P_i \ 0 \ \dots \ 0$$

$$0 \ P_k \ \dots \ 0$$

En regroupant ces deux lignes, on a :

$$\Delta H' = P_i \log \frac{P_i + P_k}{P_i} + P_k \log \frac{P_i + P_k}{P_k}$$

Il vient

$$\Delta H < \Delta H'$$

C'est-à-dire, la perte d'information dûe au regroupement de deux lignes bruitées, peut être majorée par celle dûe au regroupement des deux mêmes lignes non bruitées.

Si  $P_i$  et  $P_k$  sont inférieures à un certain seuil  $\epsilon$ , alors

$$\Delta H < \Delta H' < \epsilon \log \frac{2\epsilon}{\epsilon} + \epsilon \log \frac{2\epsilon}{\epsilon} = 2 \epsilon \log 2 = \gamma$$

plus précisément, pour que le regroupement des lignes  $i$  et  $k$  occasionne une perte d'information inférieure à  $\gamma$ , il suffit que  $P_i$  et  $P_k$  soient inférieures à  $\epsilon$ , tel que  $\epsilon = \frac{\gamma}{2 \log 2}$ .

c.f.d

Pour  $m$  regroupements, on peut se fixer une perte d'information maximale  $\gamma' = m \gamma$ , sachant que les probabilités des lignes à regrouper ne doivent pas dépasser  $\epsilon$ . Il faut bien sûr qu'il y ait au moins  $m$  valeurs de probabilités inférieures à  $\epsilon$ .

#### Exemple de courbe dans le cas général

Dans les exemples précédents, à chaque colonne  $j$  correspondaient deux lignes dont l'élément principal était situé sur la colonne  $j$ .

Dans cet exemple, le nombre de lignes correspondant à chaque colonne est tiré au sort entre 1 et 5. Aussi, les probabilités de chaque ligne sont tirées au sort, et l'intensité du bruit  $\beta$  est tirée au sort entre deux valeurs fixées. De la même façon que dans les cas précédents, le bruit est choisi. L'algorithme est appliqué de la même manière que précédemment. La courbe que nous avons tracée (figure III. 12) correspond à un bruit gaussien. Le bruit  $\beta$  prend des valeurs aléatoires entre 0,25 et 0,35. Le nombre total de modalités de  $\lambda$  est de 30.

On peut observer, comme sur les autres courbes, que la cassure se situe vers 10 et que l'information perdue au cours des premiers regroupements est très faible.

$\Delta H$

cas général  $0,25 \leq \beta \leq 0,35$   $N=30$

Nombre de modalités du vecteur explicatif



1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27

Figure III. 12

## CONCLUSION

---

Les résultats présentés dans la première partie de ce chapitre permettent de construire un graphe d'interconnexion. Afin de reconnaître des sous-systèmes faiblement couplés, nous devons dans le chapitre suivant, étudier des algorithmes de décomposition de ces graphes.

Dans la deuxième partie, les différents résultats nous permettent de simplifier le vecteur explicatif en regroupant certaines de ses modalités. Dans la plupart des cas, les premiers regroupements correspondent à ceux des lignes ayant même élément principal. Des exceptions peuvent se produire quand le bruit est très important sur deux lignes n'ayant par même élément principal.

CHAPITRE III

METHODES DE DECOMPOSITION

### III - I INTRODUCTION

La décomposition d'un système complexe en sous-systèmes faiblement couplés permet de réduire sa complexité apparente. Pour cela, on peut utiliser la théorie des graphes quand on connaît le modèle du système (système d'équations, matrices .....). Mais celui-ci est en général difficile à obtenir. En fait, la présence de couplages est toujours décelable dans les données produites par le système et se concrétise par des évolutions non indépendantes de certains groupes de variables. Cette dépendance se présente sous la forme de relations quelconques (linéaires, non linéaires, ...) entre des variables aussi bien quantitatives que qualitatives. On propose alors quelques indices de similarité, issus de la théorie de l'information pour mesurer l'intensité de ces relations. Plusieurs méthodes de décomposition sont présentées dans ce chapitre.

La première partie concerne l'utilisation des graphes.

Dans la deuxième partie sont exposées différentes méthodes de décomposition hiérarchique et à centres variables.

### III - 2 DECOMPOSITION A PARTIR DES GRAPHES

---

La théorie des graphes est souvent utilisée dans la modélisation structurale; dans le graphe obtenu, chaque variable est alors présentée par un noeud et chaque liaison par un arc.

Les méthodes de décomposition des graphes orientés ou non orientés sont basées sur les définitions de simple et forte connexité.

L'utilisation des graphes valués est complémentaire d'une étude purement qualitative des graphes décrivant un système donné.

#### III - 2 - 1 Concepts sur la théorie des graphes (ROY 70 . PICAR 72)

Nous rappelons quelques définitions utilisées pour la mise en oeuvre des algorithmes de décomposition à partir des graphes.

##### - Graphe orienté , graphe non orienté

Un graphe est un couple  $(E, U)$  constitué par :

1°) Un ensemble  $E = (X_1, X_2, \dots, X_n)$  d'éléments appelés sommets du graphe.

2°) Un sous-ensemble  $U$  du produit cartésien  $E \times E$ . Les éléments de  $U$  sont appelés arcs s'ils sont orientés et arêtes s'ils ne le sont pas.

##### - Graphe valué

Un graphe valué  $G$  est un quadruplet  $(E, U, f, g)$  formé d'un ensemble  $E$  (l'ensemble des sommets de  $G$ ), d'un ensemble  $U \subseteq E \times E$  (l'ensemble des arcs de  $G$ ); d'une application  $f : E \rightarrow \mathbb{R}^+$  (pondération des sommets) et d'une application  $g : E \rightarrow \mathbb{R}$  (pondération des arcs).

##### - Chemin

Un chemin est une suite d'arcs  $\mu = (u_1, u_2, \dots, u_i, \dots, u_p)$  telle que l'extrémité terminale de l'arc  $u_i$  coïncide avec l'extrémité initiale de l'arc  $u_{i+1}$  pour  $i = 1, 2, \dots, p-1$ .

##### - Circuit

C'est un chemin  $\mu$ , tel que l'extrémité initiale du premier arc coïncide avec l'extrémité terminale du dernier arc.

##### - Chaîne

Une chaîne est une suite d'arcs  $\mu' = (u_1, u_2, \dots, u_i, \dots, u_q)$  telle que chaque arc  $u_j$  de la suite est attaché à  $u_{j-1}$  par une de ses extrémités et à  $u_{j+1}$  par l'autre de ses extrémités.

##### - Cycle

C'est une chaîne  $\mu'$  telle que

1) tous les arcs de la suite  $\mu'$  sont différents

2) le sommet initial et le sommet terminal de la chaîne coïncident.



Sous graphe de  $G = (E, U)$  engendré par un ensemble  $A \subset E$  : c'est le graphe  $G_A = (A, U_A)$ , où

$$U_A = \{ u_i = (x, y), x \in A, y \in A \}$$

Graphe fortement connexe

C'est un graphe  $G = (E, U)$  tel que, pour tout  $x, y \in E$  avec  $x \neq y$ , il existe un chemin de  $x$  à  $y$  et un chemin de  $y$  à  $x$ .

Composante fortement connexe

C'est un sous-graphe  $G_A$  de  $G$  qui est fortement connexe, et tel que pour tout  $x \in E - A$ ,  $G_A \cup x$  n'est pas fortement connexe.

Proposition III - 1

Si  $G = (E, U)$  est un graphe, la relation binaire sur  $E$  défini par "x = y ou il existe un chemin de  $x$  à  $y$  et un chemin de  $y$  à  $x$ " est une relation d'équivalence. De ce fait, les différentes composantes fortement connexes définissent une partition de  $E$ . (La démonstration est évidente).

Graphe simplement connexe (ou graphe connexe)

C'est un graphe  $G = (E, U)$  tel que pour tout  $x, y \in E$ , avec  $x \neq y$ , il existe une chaîne dont les deux extrémités sont  $x$  et  $y$ .

Composante simplement connexe (ou composante connexe)

C'est un sous-graphe  $G_A$  de  $G$  qui est connexe, et tel que pour tout  $x \in E - A$ , le sous-graphe  $G_A \cup x$  n'est pas connexe.

Proposition III - 2

Si  $G = (E, U)$  est un graphe, la relation binaire sur  $E$  définie par "x = y ou il existe une chaîne dont les deux extrémités sont  $x$  et  $y$ ", est une relation d'équivalence, et, de ce fait, les différentes composantes connexes définissent une partition de  $E$ .

( La démonstration est immédiate. )

Graphe réduit de  $G = (E, U)$

C'est un graphe  $G_r = (C_f, U_r)$  tel que :

$C_f =$  ensemble des composantes fortement connexes de  $G$  (partition de  $E$ )

$\{ (C_{fi}, C_{fj}) \in U_r \mid C_{fi}, C_{fj} \in C_f \}$  ssi  $\exists x_k \in C_{fi}$  et  $x_f \in C_{fj}$  tels que

$(x_k, x_f) \in U$ .

Le graphe  $G_r$  est un graphe sans circuit. Il est s-connexe si  $G$  l'est.

Rang d'un sommet dans un graphe sans circuit

Dans un graphe sans circuit, un sommet  $x$  est de rang  $r$  si le nombre maximal d'arcs de tout chemin admettant  $x$  comme extrémité terminale est précisément  $r$ .

Exemple: III - 1 :

Soit le graphe G ( Fig III - 1) et sa matrice d'adjacence associée M ( Fig III - 2)

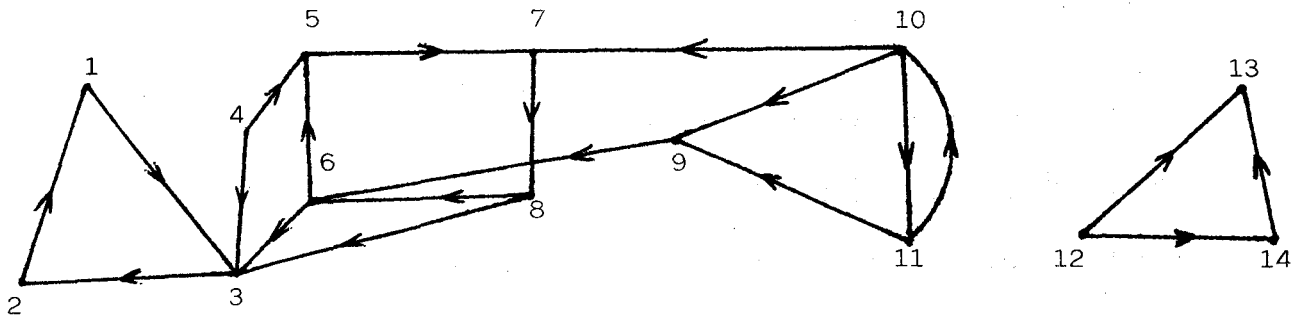


FIG III - 1 Graphe G

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	1	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	1	0	1	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	1	0	0	0	0	0	0	0
6	0	0	1	0	1	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	1	0	0	0	0	0	0
8	0	0	1	0	0	1	0	0	0	0	0	0	0	0
9	0	0	0	0	0	1	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	1	0	1	0	1	0	0	0
11	0	0	0	0	0	0	0	0	1	1	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	1	1
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	1	0

FIG III - 2 Matrice M



Les sous-graphes  $G_1$  et  $G_2$  sont aussi les composantes s - connexes de G.

Les composantes f - connexes de G sont :

(  $\{1,2,3\}$  ,  $\{5,6,7,8\}$  ,  $\{9\}$  ,  $\{10,11\}$  ,  $\{4\}$  )

III - 2 - 2 Décomposition des graphes orientés

La décomposition d'un graphe orienté peut être obtenue en faisant des permutations des lignes et colonnes de sa matrice associée, en vue d'obtenir une matrice bloc triangulaire. Nous allons citer quelques méthodes de décomposition de ce type de graphes en s'appuyant sur les différentes composantes f-connexes.

Algorithme de STEWARD (STEW 65)

On effectue des permutations des lignes et des colonnes de la matrice associée au graphe G, de manière à la rendre bloc - triangulaire, en s'appuyant sur les différentes composantes fortement connexes.

Algorithme du graphe réduit

Cet algorithme comporte trois étapes :

- 1) La recherche des composantes f- connexes  
( ROY 70 ) ( RAM 65 ) ( TARJ 75 )
- 2) La construction du graphe réduit et la détermination des rangs  
(ROY 70 )
- 3) La construction du graphe structuré par rang et d'une matrice bloc - diagonale.

- Algorithme de WARFIELD

Il permet de déterminer simultanément les composantes f - connexes et leur rang.

Une composante f - connexe est telle que :

$$\forall x \in cf_i \quad S(x) \cap P(x) = cf_i$$

avec  $S(x)$  = ensemble des suivants de  $x$

$P(x)$  = ensemble des précédents de  $x$

Le rang au niveau de chaque composante est déterminé ainsi :

$$\begin{array}{l} \text{Rang } 0 \quad x \in cf_0, \quad \{P(x) - cf_0\} = \emptyset \\ \text{Rang } 1 \quad x \in cf_1, \quad \{P(x) - cf_1\} = cf_0 \\ \vdots \\ \text{Rang } n \quad x \in cf_n, \quad \{P(x) - cf_n\} = \bigcup_{j=1}^{n-1} cf_j \end{array}$$

Pour le sous-graphe  $G_1$  de la figure III<sub>1</sub>, on a :

Sommet	S (X')	P (X')	S (X) ∩ P (X)
1	1 2 3	1 2 3 4 5 6 7 8 9 10 11	1 2 3
2	1 2 3	1 2 3 4 5 6 7 8 9 10 11	1 2 3
3	1 2 3	1 2 3 4 5 6 7 8 9 10 11	1 2 3
4	1 2 3 4 5 6 7	4	4
5	1 2 3 5 6 7 8	4 5 6 7 8 9 10 11	5 6 7 8
6	1 2 3 5 6 7 8	4 5 6 7 8 9 10 11	5 6 7 8
7	1 2 3 5 6 7 8	4 5 6 7 8 9 10 11	5 6 7 8
8	1 2 3 5 6 7 8	4 5 6 7 8 9 10 11	5 6 7 8
9	1 2 3 5 6 7 8 9	9 10 11	9
10	1 2 3 5 6 7 8 9 10 11	10 11	10 11
11	1 2 3 5 6 7 8 9 10 11	10 11	10 11

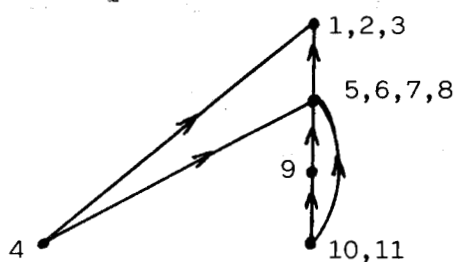
Graphe réduit du sous-graphe  $G_1$

Rang 3

Rang 2

Rang 1

Rang 0



Pour trouver une matrice bloc - triangulaire inférieure (resp. supérieure) il suffit d'écrire les composantes f - connexes par ordre décroissant (resp. croissant).

La matrice bloc triangulaire inférieure du sous-graphe  $G_1$  est donnée par la figure III -3



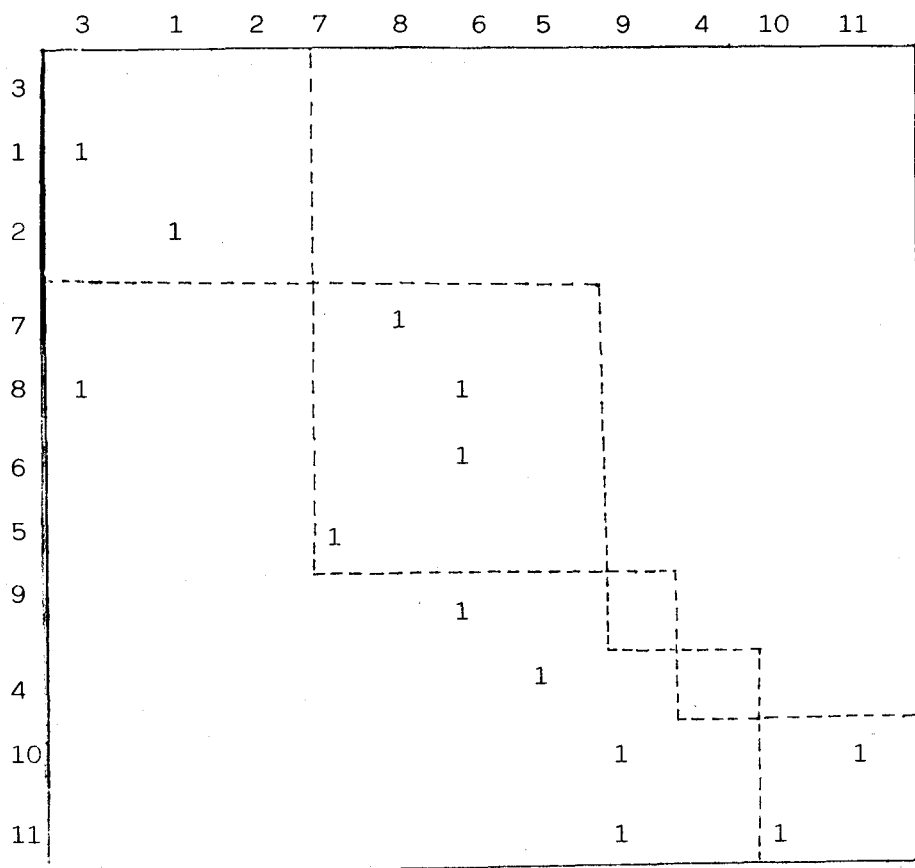


Fig III - 3

Les algorithmes que nous venons de citer donnent des structures hiérarchisées induites par la connexité forte d'un graphe orienté. Nous aborderons dans le paragraphe suivant la décomposition des graphes symétriques ou non orientés.

III - 2 - 3 Décomposition des graphes symétriques ou non orientés

Si un graphe n'est pas s-connexe, on effectue des permutations symétriques sur les lignes et colonnes de la matrice associée, pour aboutir à une matrice bloc - diagonale. chaque bloc correspond alors à la notion de sous-système.

Lorsque le graphe est s-connexe, on utilise le concept d'ensembles minimaux d'articulation, à partir desquels on obtient des sous-composantes connexes. On détermine aussi pour celles-ci des sous-ensembles d'articulations et on continue la décomposition. On définit ainsi des partitions de plus en plus fines de E.

Ensemble d'articulation

Dans un graphe s-connexe  $G = (E, U)$ , tout sous-ensemble de sommets (respectivement sommet)  $C$  dont la suppression engendre un sous-graphe  $G_A = (E - C, U_A)$  qui est non connexe est appelé ensemble (respectivement point) d'articulation.

Ensemble minimal d'articulation

Soit un graphe s-connexe  $G = (E, \Gamma)$  et soient  $k$  sous-ensembles disjoints  $A_1, \dots, A_k$  de  $E$ , tels que :

$$\bigcap A_i \cap A_j = \emptyset, i \neq j, i, j = 1, \dots, k$$

Soit une  $k + 1$  partition de  $E : (E_1, \dots, E_k, C)$  telle que

1)  $A_i \subseteq E_i \quad i = 1, \dots, k$

2)  $\bigcap E_i \cap E_j = \emptyset \quad i \neq j \quad i, j = 1, \dots, k$

3)  $C$  de cardinal minimum.

$C$  est alors un ensemble minimal d'articulation, il n'est pas unique en général.

Un ensemble d'articulation  $C$  d'un graphe s-connexe définit une partition de  $E = \{E_1, \dots, E_k, C\}$  telle que les sous-graphes  $G_{E_i} = (E_i, U_i)$  sont s-connexes.

De même un ensemble d'articulation  $C_i$  d'un sous-graphe définit une partition de  $E_i = \{E_{i1}, \dots, E_{ij}, \dots, E_{ik}, C_i\}$  telle que les sous-graphes  $G_{E_{ij}} = (E_{ij}, U_{ij})$  sont s-connexes.

Pour chaque ensemble d'articulation  $C$  d'un graphe s-connexe, il existe une matrice de permutation  $P$  telle que  $M^* = P^{-1} M P$  est une matrice de la forme suivante :

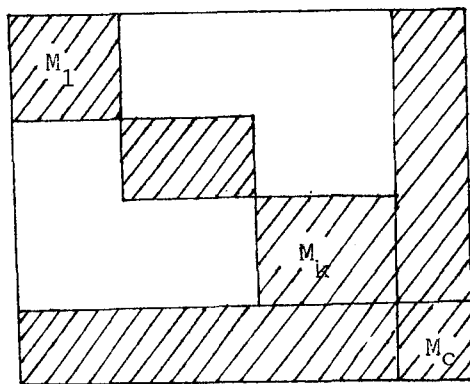


Fig.III.4.

Les matrices carrées  $M_i$  représentent les sous-systèmes.  $M_c$  étant la matrice associée à l'ensemble d'articulation.

Nous allons citer quelques algorithmes de détermination des ensembles minimaux d'articulation

a) Algorithmes de détermination des ensembles minimaux d'articulation

Algorithme de MALGRANGE (KAUF 68)

Soit  $k = 2$ , alors, la partition cherchée est  $E = (E_1, E_2, C)$  avec  $A_1 \subseteq E_1$ .

$A_2 \subseteq E_2$

Soit  $M_c$  la matrice du graphe complémentaire de  $G$ . La sous-matrice de  $M_c$  définie par les lignes correspondant à  $E_1$  et par les colonnes correspondant à  $E_2$  a tous ses éléments égaux à 1. On démontre que cette sous-matrice complète est première et que toutes les sous-matrices premières de  $M_c$  définissent des ensembles d'articulation minimaux  $C$ . A partir d'une couverture de sous-matrices complètes de  $M_c$ , on détermine par des transformations logiques successives, toutes les sous-matrices premières de  $M_c$ .

Cet algorithme donne toutes les solutions, cependant lorsque  $n$  est grand, le nombre de solutions devient important et le programme de calcul nécessite beaucoup de temps d'exécution et de place mémoire.

Méthode du flot maximal (RICH 75)

A partir d'une transformation du graphe initial en un réseau de transport, on recherche le flot maximal dans le réseau. La coupe minimale obtenue permet de définir un ensemble d'articulation minimal.

Exemple III - 2: Soit un graphe  $G'$  déduit obtenu en supprimant les orientations du graphe  $G$  ( Fig III -1)

La permutation  $p = \{ 10, 11, 8, 6, 5, 4, 9, 7, 1, 2, 3 \}$  permet de faire apparaître les deux structures liées à la connexité simple Fig III -5

	10	11	8	6	5	4	9	7	1	2	3	12	13	14
10	0	1						1	1		0			
11	1	0						1	0		0			
8			0	1	0	0	0	1				1		
6			1	0	1	0	1	0				1		
5			0	1	0	1	0	1				1		
4			0	0	1	0	0	0				1		
9			0	1	0	0	0	0				0		
7			1	0	1	0	0	0				0		
1									0	1	1			
2									1	0	1			
3	0	0	1	1	1	1	0	0	1	1	0			
12												0	1	1
13												1	0	1
14												1	1	0

Fig. III.5.

La représentation sous forme d'un graphe non valué d'un système permet une analyse qualitative de sa structure en vue de sa décomposition. Lorsque le système peut être représenté par un graphe valué, son analyse sera plus fine et sa décomposition plus compliquée.

### III - 2 - 4 Décomposition des graphes valués

Le paragraphe précédent étudie les graphes, traduisant l'existence ou non d'une liaison entre éléments. Pour tenir compte de l'intensité de cette liaison qui est nécessaire par exemple pour la coordination de sous-systèmes, on utilise les graphes valués. La valuation porte sur les sommets et / ou les arcs du graphe. Il existe plusieurs méthodes de décomposition des graphes valués. Nous en citerons une parmi les méthodes proposées par MILLGRAM (MIL 75)

#### Méthodes des nuées dynamiques

La technique fondamentale des nuées dynamiques est rappelée au chapitre IV .

Soit  $G (E, U, f, g)$  un graphe valué.

Soit  $a : E \times E \rightarrow R^+$  une fonction de similarité définie sur  $E$

$$\text{On pose } a(x, y) = \frac{1}{g(x, y)} \quad \{x, y\} \in U$$

$$a(x, y) = 0 \quad \{x, y\} \notin U$$

On construit pour chaque noyau  $A_i$  une suite

$$A_i = E_i \subseteq E_i^1 \subseteq \dots \subseteq E_i^{n(i)} = E_i$$

$$\text{avec } E_i^{p+1} = E_i^p \cup \{x_i^p\}$$

où le sommet  $x_i^p$  est celui qui rend maximum une fonction d'attraction.

$$D_i(x, E_i^p) = \sum_{y \in E_i^p} a(x, y) W_i(y)$$

Les  $W_i(y)$  sont des pondérations qui font jouer des rôles plus importants aux éléments  $y \in E_i^p$  avec  $r$  assez grand.

Pour déterminer les noyaux à partir de la partition, on définit pour chaque classe  $E_i$  une frontière, c'est à dire l'ensemble des sommets de  $E_i$ ,

voisins des sommets de  $E - E_i$ , et on retire cette frontière. Cette procédure est répétée, les derniers  $n_i$  sommets constituent le noyau numéro  $i$ .



## Conclusion

Nous avons exposé dans ce paragraphe plusieurs méthodes de décomposition des graphes . Le choix de l'une d'elles dépend du but de l'analyse et de la nature du système étudié. Ces méthodes supposent en général, la connaissance du modèle du système ( système d'équation, matrices..... ) . L'absence des modèles des systèmes justifie la recherche d'autres méthodes basées uniquement sur l'analyse des données. C'est ce que nous allons étudier dans les paragraphes suivants .

III - 3 mesures de similarité

III - 3 - I Position du problème

En général, un algorithme de décomposition hiérarchique utilise un tableau de similarités ( ou de dissimilarité ); celui-ci est calculé à partir d'un tableau de description du système étudié. Dans un premier temps, nous allons considérer un tableau de données logiques. C'est un tableau contenant uniquement des zéros et des uns matérialisant ainsi en général la présence d'une modalité (1) ou son absence (0). Nous rappellerons plusieurs indices de similarités ou de dissimilarité entre individus (observations). Une étude semblable peut être faite pour les indices de similarité (ou dissimilarité) entre variables. Nous retrouvons par analogie plusieurs indices de similarité ou distances (certains sont nouveaux) issus de tableaux plus généraux que le tableau de données logiques.

Soient  $\omega_1$  et  $\omega_2$  deux individus de  $\Omega$  qui peuvent être représentés par deux points  $a(\omega_1)$  et  $a(\omega_2)$  de  $\{0,1\}^T$ .

$$a(\omega_1) = (\omega_1^1 \dots \omega_1^T)$$

$$a(\omega_2) = (\omega_2^1 \dots \omega_2^T)$$

$\omega_1^1 \dots \omega_1^T$  (resp  $\omega_2^1 \dots \omega_2^T$ ) représente une suite de 0 ou de 1 qui figurent dans un tableau, pour décrire l'individu  $\omega_1$  (respectivement  $\omega_2$ )

On associe au couple  $(\omega_1, \omega_2)$  les cardinaux suivants :

$$s = \sum_i \omega_1^i \omega_2^i$$

$$t = \sum_i (1 - \omega_1^i) (1 - \omega_2^i)$$

$$u = \sum_i (1 - \omega_1^i) \omega_2^i$$

$$v = \sum_i \omega_1^i (1 - \omega_2^i)$$

s (respectivement t) = nombre de variables pour lesquelles  $\omega_1$  et  $\omega_2$  ont simultanément la modalité 1 ( resp la modalité 0)

u (resp v) = nombre de variables pour lesquelles  $\omega_1$  (resp  $\omega_2$ ) prend la modalité 1 et  $\omega_2$  (resp  $\omega_1$ ) la modalité 0 ( Fig III -1)

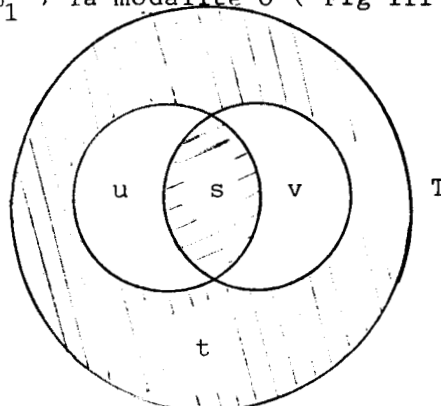


Fig. III.1.

Dualement, sur les colonnes du tableau, on peut définir les quatre nombres ( s', u', v', t' ), où s' est le nombre d'éléments de  $\Omega$  qui possèdent à la fois la modalité j et la modalité j', etc .....

Dans la pratique, on dispose d'un tableau de données plus général (par exemple le tableau initial des données Fig I - 1). On utilise alors la notion de semi-valuation et plus particulièrement celle de l'entropie.

En fait, l'application "card" n'est qu'un exemple de semi-valuation (ABID 79 ).

Soient deux variables  $x, y \in \Sigma$ , on peut avoir la représentation plane analogue à celle de la figure III - 1

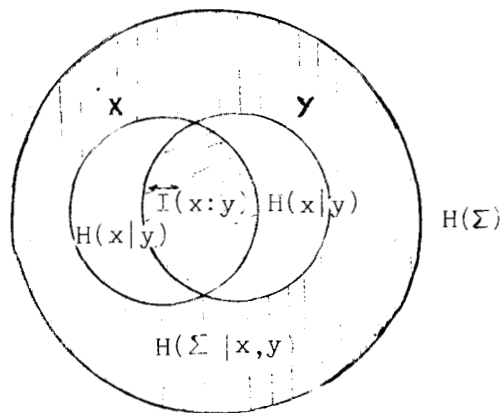


Fig. III.2.

Par analogie avec la figure III - 1, la figure III -2 fait apparaître la représentation de  $\Sigma$  par son entropie. Si on considère les surfaces représentant deux variables  $x$  et  $y$  proportionnelles à leurs entropies respectives, l'analogie peut être faite de la façon suivante :

- S  $\longrightarrow$   $I(x : y)$
- u  $\longrightarrow$   $H(x/y)$
- v  $\longrightarrow$   $H(y/x)$
- t  $\longrightarrow$   $H(\Sigma | x, y)$
- T  $\longrightarrow$   $H(\Sigma)$

Cette analogie nous permettra de définir quelques indices de similarités ou de dissimilarité qui peuvent être utilisés pour mettre en oeuvre des algorithmes de décomposition.

III - 3 - 2 Quelques définitions

Les définitions ci-dessous sont tirées de la référence (CAILL 76 )

Indice de similarité sur un ensemble E : C'est une application S de  $E \times E$  dans l'ensemble  $\mathbb{R}^+$  des nombres positifs ou nuls.

$$E \times E \xrightarrow{S} \mathbb{R}^+$$

$$(i, j) \longrightarrow S(i, j)$$

telle que  $S(i, j) = S(j, i)$  pour tout  $(i, j) \in E \times E$

$$S(i, i) = S(j, j) = S_{\max} > S(i, j) \text{ pour tout } (i, j) \in E \times E \text{ avec } i \neq j$$

Indice de dissimilarité sur un ensemble E : C'est une application d de  $E \times E$  dans  $\mathbb{R}^+$

$$E \times E \xrightarrow{d} \mathbb{R}^+$$

$$(i, j) \longrightarrow d(i, j)$$

telle que  $d(i, j) = d(j, i)$  pour tout  $(i, j) \in E \times E$

$$d(i, i) = 0 \text{ pour tout } i \in E$$

Les définitions ci-dessus sont peu contraignantes, mais elles ont l'inconvénient de permettre des incohérences (CAILL 76 )

On peut avoir par exemple simultanément :

$$\begin{cases} i \neq j \\ d(i, j) = 0 \\ \text{il existe } k \text{ tel que } d(i, k) \neq d(j, k) \end{cases}$$

Cette incohérence disparaît si l'indice de dissimilarité possède l'une au moins des deux propriétés suivantes :

$$(1) \quad d(i, j) = 0 \implies i = j$$

$$(2) \quad d(i, j) \leq d(i, k) + d(j, k) \quad \forall (i, j, k) \in E^3$$

Les indices de dissimilarité vérifiant la propriété (1) sont appelés des indices de distance . Ceux qui vérifient la propriété (2) sont appelés des écarts; ceux, enfin, qui vérifient ces deux propriétés sont appelés des distances.

En décomposition hiérarchique, on utilisera aussi des distances ultramétriques, qui vérifient la propriété.

$$d(i, j) \leq \sup \{ d(i, k), d(k, j) \} \quad \forall (i, j, k) \in E^3$$

III - 3 - 3 Calcul des indices

Après un rappel de quelques indices de similarité, nous allons déduire par analogie leurs équivalents utilisant le concept de l'entropie. Cette procédure nous a permis d'obtenir quelques indices utilisés en analyse structurale, ainsi que d'autres qui sont à notre connaissance nouveaux.

(1) Indice de JACCARD ( I908 )

$$I_1 = \frac{S}{S + u + v}$$

$$I_1' = \frac{\overleftrightarrow{I}(x:y)}{\overleftrightarrow{I}(x:y) + H(x(y)) + H(y(x))} = \frac{\overleftrightarrow{I}(x:y)}{H(x) + H(y) - H(x,y) + H(x,y) - H(y) + H(x,y) - H(x)}$$

$$= \frac{\overleftrightarrow{I}(x:y)}{H(x,y)} \quad 0 \leq I_1' \leq 1$$

DUCHAUSOY démontre que ce coefficient de couplage est une relation de similarité floue transitive. (DUSS 80, 82)

(2) Indice de CZEKANOWSKI ( I913 )

DICE (I945)

SORENSEN ( I948 )

$$I_2 = \frac{S}{S + \frac{1}{2}(u+v)}$$

$$I_2' = \frac{\overleftrightarrow{I}(x:y)}{\overleftrightarrow{I}(x:y) + \frac{1}{2}[H(x/y) + H(y/x)]} = \frac{\overleftrightarrow{I}(x:y)}{H(x) + H(y) - H(x,y) + \frac{1}{2}[2H(x,y) - H(x) - H(y)]}$$

$$= \frac{2\overleftrightarrow{I}(x:y)}{H(x) + H(y)}$$

L'indice  $I_2'$  est un indice de distance, en effet:

Soient  $x, y, z \in E$  : on a

$$2 \frac{\overleftrightarrow{I}(x:z)}{H(x) + H(z)} + 2 \frac{\overleftrightarrow{I}(y:z)}{H(y) + H(z)} \geq 2 \frac{\overleftrightarrow{I}(x:y)}{H(x) + H(y)}$$

démonstration :

$$\frac{\overleftrightarrow{I}(x:y)}{H(x) + H(y)} = \frac{\overleftrightarrow{I}(x:y)}{H(x) + H(y) - \overleftrightarrow{I}(x:y) + \overleftrightarrow{I}(x:y)} = \frac{\overleftrightarrow{I}(x:y)}{H(x,y) + \overleftrightarrow{I}(x:y)} = 1 + \frac{H(x,y)}{\overleftrightarrow{I}(x:y)}$$

on pose  $d'(x,y) = \frac{\overleftarrow{I}(x;y)}{H(x,y)}$

il vient

$$d(x,y) = \frac{2}{1 + \frac{1}{d'(x,y)}} = 2 \frac{d'(x,y)}{1 + d'(x,y)}$$

$$\frac{d'(x,y)}{1+d'(x,z)} + \frac{d'(y,z)}{1+d'(y,z)} \geq \frac{d'(x,z)}{1+d'(x,z)+d'(y,z)} + \frac{d'(y,z)}{1 + d'(y,z) + d'(x,z)} =$$

$$\frac{d'(x,z) + d'(y,z)}{1 + d'(x,z)+d'(y,z)} = \frac{1}{1 + \frac{1}{d'(x,y)+d'(y,z)}} \geq \frac{1}{1 + \frac{1}{d'(x,y)}}$$

car  $d'(x,y) \leq d'(x,z) + d'(y,z)$  (indice de DUSSAUCHOY. (DUSS.80)

$$\begin{aligned} \frac{1}{1 + \frac{1}{d'(x,y)}} &= \frac{d'(x,y)}{1 + d'(x,y)} = \frac{\frac{\overleftarrow{I}(x;y)}{H(x,y)}}{1 + \frac{\overleftarrow{I}(x;y)}{H(x,y)}} = \frac{\overleftarrow{I}(x;y)}{H(x,y) + \overleftarrow{I}(x;y)} \\ &= \frac{\overleftarrow{I}(x;y)}{H(x) + H(y)} \quad \text{c.f.d} \end{aligned}$$

3 - Indice de RUSSEL et RAO (1940)

$$I_3 = \frac{S}{\overline{I}}$$

$$I'_3 = \frac{\overline{I}(x;y)}{H(\Sigma)}$$

On démontre que  $1 - I'_3$  est une distance, en effet :

Soit  $x_1, x_2, \dots, x_n \in \Sigma$ , la reflexivité et la symétrie sont évidentes. l'inégalité triangulaire est vérifiée :

$$1 - \frac{\overline{I}(x_1 : x_2)}{H(\Sigma)} = \frac{H(\Sigma) - \overline{I}(x_1 : x_2)}{H(\Sigma)} \leq \frac{H(\Sigma) - \overline{I}(x_1 : x_3)}{H(\Sigma)} + \frac{H(\Sigma) - \overline{I}(x_2 : x_3)}{H(\Sigma)}$$

$$H(\Sigma) - \overline{I}(x_1 : x_2) \leq 2H(\Sigma) - \overline{I}(x_1 : x_3) - \overline{I}(x_2 : x_3)$$

Démonstration :

$$H(\{\Sigma - x_1, x_2\} / x_1, x_2) \geq 0 \quad (\text{cf propriété (I) § I.52.k})$$

$$H(\{\Sigma - x_1, x_2\}, x_1, x_2) - H(x_1, x_2) \geq 0$$

$$H(\{\Sigma - x_1, x_2\} + H(x_1) + H(x_2) - H(\Sigma - x_1, x_2) - H(x_1) - H(x_2) + H(\{\Sigma - x_1, x_2\}; x_1, x_2) - H(x_1, x_2)) \geq 0$$

$$H(\Sigma - x_1, x_2) + \overline{I}(x_1 : x_2) - \overline{I}(\Sigma - x_1, x_2 : x_1 : x_2) \geq 0$$

En ajoutant les nombres positifs à cette expression; l'inégalité reste valable.

$$H(\Sigma -x_1, x_2) + \overleftrightarrow{I}(x_1 : x_2) - \overleftrightarrow{I}(\Sigma -x_1, x_2 : x_1 : x_2) + H(x_1/x_3) + H(x_2/x_3) \geq 0$$

$$H(\Sigma -x_1, x_2) + \overleftrightarrow{I}(x_1 : x_2) - \overleftrightarrow{I}(\Sigma -x_1, x_2 : x_1 : x_2) + H(x_1) - \overleftrightarrow{I}(x_1 : x_3) + H(x_2) - \overleftrightarrow{I}(x_2 : x_3) \geq 0$$

$$H(\Sigma) + \overleftrightarrow{I}(x_1 : x_2) - \overleftrightarrow{I}(x_1 : x_3) - \overleftrightarrow{I}(x_2 : x_3) \geq 0$$

$$2H(\Sigma) - \overleftrightarrow{I}(x_1 : x_3) - \overleftrightarrow{I}(x_2 : x_3) \geq H(\Sigma) - \overleftrightarrow{I}(x_1 : x_2)$$

(4) Indice de KULEZYNSKI (1927)

$$I_4 = \frac{1}{2} \left[ \frac{S}{S+u} + \frac{S}{S+v} \right]$$

$$I'_4 = \frac{1}{2} \left[ \frac{\overleftrightarrow{I}(x:y)}{H(x)} + \frac{\overleftrightarrow{I}(y:x)}{H(y)} \right]$$

$I'_4$  permet de rendre symétrique l'indice de CONANT  $\frac{\overleftrightarrow{I}(x:y)}{H(x)}$   
(CON 72)

(5) Indice de OCHIAI (1957)

$$I_5 = \frac{S}{(S+u)(S+v)}$$

$$I'_5 = \frac{\overleftrightarrow{I}(x:y)}{\sqrt{H(x)H(y)}}$$

$I'_5$  est l'indice de Richetin (RICH 75)

(6) Distance du cardinal de la différence symétrique

$d_1 = u + v$  est une distance (CAILL 76)

Soient trois éléments de l'ensemble des parties de  $\Omega$ , A, B et C, en normalisant la distance  $d_1$  CAILLEZ et PAGES ont proposé deux autres distances,  $d_2$  et  $d_3$ .

$$d_2(A, B) = \begin{cases} = 0 & \text{si } A = B = \emptyset \\ = \frac{u+v}{S+u+v} & \text{si non.} \end{cases}$$

$$d_3(A, B) = \begin{cases} 0 & \text{si } A = B = \Omega \\ \frac{u+v}{|\Omega| - S} & \text{si non} \end{cases}$$

L'utilisation des distnaces  $d_1$  et  $d_2$  permettent de retrouver la distance de LOPEZ DEMANTARA (LOP 75)

Soient  $x,y$  deux variables appartenant à un ensemble  $\Sigma$

$$d'_1(x,y) = H(x/y) + H(y/x)$$

ou  $d'_2(x,y) = \frac{H(x/y) + H(y/x)}{H(x;y)}$  qui n'est que  $d'_1$  normalisé.

A partir de la distance  $d_3$ , on aboutit à

$$d'_3(x,y) = \frac{H(x/y) + H(y/x)}{H(\Sigma) - I(x;y)}$$

$d'_3(x,y)$  est une distance, en effet :

$$\forall x,y \in \Sigma$$

$$d'_3(x,y) = \frac{2 d'_1(x,y)}{d'_1(x,y) + d'_1(x, \Sigma) + d'_1(y, \Sigma)}$$

est forcément une distaance puisque  $d'_1(x,y)$  l'est aussi (CAILL 76)

Les indices précédents peuvent être calculés de la même façon pour des sous-ensembles de variables. Il suffit de remplacer les variables  $x$  et  $y$  par des parties de  $\Sigma$ .

Remarque

Nous remarquons que l'indice  $I'_3$  est fonction de  $\overleftrightarrow{I}(x;y)$  alors que les indices  $I'_1, I'_2, I'_4$  et  $d'_2$  font apparaître la quantité  $\frac{\overleftrightarrow{I}(x;y)}{H(x/y) + H(y/x)}$ .

Le dénominateur  $H(x/y) + H(y/x)$  sert à normaliser  $\overleftrightarrow{I}(x;y)$  et renforcer la mesure de liaison entre  $x$  et  $y$ .

Le choix de l'un de ces indices dépend du but de l'analyse et de la méthode de décomposition utilisée. Nous examinons dans le paragraphe suivant quelques méthodes hiérarchiques.



III - 4 Méthodes hiérarchiques

Après avoir rappelé les définitions fondamentales utilisées dans la décomposition hiérarchique, on présentera des méthodes diverses sous trois titres: Algorithmes ascendants (ou agglomératifs); algorithmes descendants (ou divisifs) et algorithmes basés sur l'optimisation d'un critère.

III - 4 - 1 Hiérarchie de parties

Définition :

Soit  $\Sigma$  un ensemble fini de variables et H une famille de parties de  $\Sigma$ . H est une hiérarchie de parties si :

- 1)  $\forall h, h' \in H : h \cap h' \in \{h, h', \emptyset\}$
- 2)  $\forall h \in H \quad \cup \{h' \in H ; h' \neq h \text{ et } h' \subset h\} \in \{h, \emptyset\}$

Une hiérarchie de parties sur  $\Sigma$  est dite totale si

- 1)  $\forall h, h' \in H ; h \cap h' \in \{h, h', \emptyset\}$
- 2)  $\exists h \in H ; \forall x \in \Sigma , \{x\} \in h$

Une hiérarchie est dite indicée s'il existe une application  $v$  de H dans  $\mathbb{R}^+$  telle que :

$$\forall h, h' \in H ; \begin{matrix} h \subset h' \\ \text{et } h \neq h' \end{matrix} \rightarrow \{v(h) < v(h')\} \text{ et } \{v(h) \neq v(h')\}$$

Pour permettre les comparaisons des hiérarchies indicées on ajoute les conditions suivantes :

$$v\{x\} = 0 \quad \forall x \in \Sigma$$

$$v\{\Sigma\} \leq 1$$

Il y a équivalence entre ultramétrie et hiérarchie indicée (BERT 75). Chercher une partition de  $\Sigma$  revient donc à définir une ultramétrie sur  $\Sigma \times \Sigma$ . On distingue deux façons de construire une décomposition hiérarchique, soit par dichotomies successives à partir de l'ensemble  $\Sigma$  soit par agrégations successives à partir des éléments de  $\Sigma$ .

III - 4 - 2 Algorithmes ascendants

Soit  $\Sigma$  un ensemble de N éléments muni d'un indice de dissimilarité S (cf § IV -2). Pour construire une hiérarchie indicée, il suffit de "transformer" l'indice  $\delta$  en distance ultramétrique. Comme il existe plusieurs distances ultramétriques sur  $\Sigma$ , il convient de rechercher la plus proche de  $\delta$  parmi l'ensemble des ultramétriques inférieures à  $\delta$ . Il est facile de montrer qu'il en existe une, supérieure à toutes les autres, appelée "ultramétrie sous-dominante de  $\delta$ " (CAILL 76). Il existe plusieurs algorithmes permettant d'obtenir l'ultramétrie sous-dominante, on décrit succinctement quelques unes dans ce paragraphe.

a/ L'algorithme de Johnson

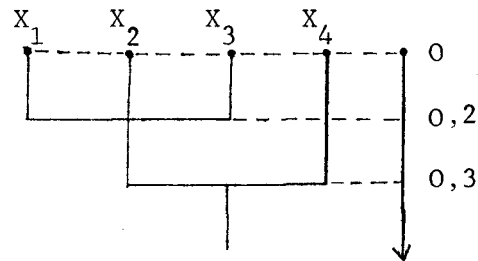
L'algorithme de Johnson procède par "regroupements successifs" de deux classes les plus "proches" (JOHN 67)

Partant d'un indice de dissimilarité de départ  $\delta_{ij}$  entre deux variables  $X_i$  et  $X_j$ , on choisit comme distance entre classes :

$$d(A, B) = \inf \{ \delta_{ij} / X_i \in A, X_j \in B \}$$

exemple :

	$X_1$	$X_2$	$X_3$	$X_4$
$X_1$	0	0,2	0,4	0,5
$X_2$		0	0,2	0,6
$X_3$			0	0,3
$X_4$				0



b/ Algorithme de Roux

L'Algorithme de ROUX examine successivement tous les triangles de  $\Sigma (X_i, X_j, X_k)$  pour rendre isocèle chacun d'eux en réduisant le plus grand côté au côté immédiatement inférieur. (ROU 68)

c/ Algorithme de LERMAN (LER 70)

L'Algorithme de LERMAN consiste à ranger les dissimilarités par ordre croissant (ordonnance) puis les examiner à partir de la plus petite. Dès qu'un segment a une extrémité commune avec l'un de ceux examinés, on rend isocèle le triangle que forment ces deux segments en diminuant le plus grand côté. L'ordonnance est par conséquent modifiée et l'on poursuit l'examen des dissimilarités réordonnées. Chaque modification nouvelle apportée n'influe pas sur les segments déjà examinés. Les modifications s'effectuent donc en un seul passage. C'est là un avantage de cet algorithme. L'inconvénient est que l'on doit établir l'ordonnance.

Nous signalons l'existence d'autres algorithmes ascendants. L'algorithme de Prim (PRIM 57), l'algorithme de Sibson (SIB 72), l'algorithme de SOKAL et MICHENER (SOK 58)

III - 4 - 3 Algorithmes descendants

Au contraire des algorithmes ascendants, il existe des algorithmes descendants dans lesquels on part de l'ensemble  $\Sigma$  (partition en une classe) sur laquelle on opère des partitions de plus en plus fines.

III - 4 - 4 Algorithme de transfert généralisé

Le principal inconvénient des algorithmes présentés jusqu'ici tient au fait que rien n'est dit sur leur optimalité. Nous en décrivons quelques uns basés sur la théorie de l'information et nous renvoyons à la référence (TORO 81) pour plus de détail.

A partir du tableau initial des données ( Fig I-1), on cherche une partition  $R = \{R_1, R_2, \dots, R_k\} \in \mathbb{P}(\Sigma)$  telle que l'interaction dans chaque sous-système soit aussi grande que possible et que l'interaction entre sous-systèmes soit minimum.

La première approche dans ce sens est due à CONANT (CONA 72) qui a proposé un critère basé sur la théorie de l'information. Plusieurs recherches sont développées sur la même voie. (RICH 75) (DUF 79) (TORO 82). Il convient alors de minimiser  $\overleftrightarrow{I}(R)$ . Mais le problème posé de cette façon conduit à la solution triviale en un seul sous-système  $R = \Sigma$ . Pour cela, R doit appartenir à  $\mathbb{P}_r(\Sigma)$ , ensemble des partitions de  $\Sigma$  à r classes, r étant fixé.

Les propriétés (7) et (8) (cf § I.5.2j) permettent de calculer la transformation externe lorsqu'on se déplace sur le treillis des partitions, d'un niveau à l'autre, par regroupement ou éclatement de classes. Ces deux opérations seront utilisées pour définir deux algorithmes heuristiques explorant le treillis  $\mathbb{P}(\Sigma)$  en sens inverse, l'algorithme ascendant noté (AA) et l'algorithme descendant (AD)

a/ Algorithme ascendant (AA)

Cet algorithme peut se résumer à l'application répétée de l'opérateur (AA)

$$AA : \mathbb{P}_r(\Sigma) \longrightarrow \mathbb{P}_{r-1}(\Sigma) \quad 2 \leq r \leq n$$

$$R = \{R_1, \dots, R_r\} \longrightarrow AA(R) = \{R_1, \dots, R_i \cup R_j, \dots, R_r\}$$

tel que

$$\overleftrightarrow{I}(R_i : R_j) = \max \{ \overleftrightarrow{I}(R_k, R_l) ; K, l = 1, 2, \dots, r \}$$

en utilisant la formule (7 § I52 k) on a :

$$\overleftrightarrow{I}(AA(R)) = \overleftrightarrow{I}(R) - \overleftrightarrow{I}(R_i, R_j)$$

b/ Algorithme descendant (AD)

Il peut se résumer à l'application répétée de l'opérateur (AD):

$$AD : \mathbb{P}_r(\Sigma) \longrightarrow \mathbb{P}_{r+1}(\Sigma) \quad 1 \leq r \leq n-1$$

$$Q = \{Q_1, \dots, Q_r\} \longrightarrow DA(Q) = \{Q_1, \dots, Q_i \setminus S^*, S^*, \dots, Q_r\}$$

tel que

$$\overleftrightarrow{I}(Q_i \setminus S^* ; S^*) = \min \{ \overleftrightarrow{I}(Q_k \setminus S; S) ; K = 1, \dots, r, Q_k \in Q \}$$

$$S \subset Q_k, \emptyset \neq S \neq Q_k \}$$

En utilisant la formule (8, § I52 k) on a :

$$\overleftrightarrow{I} (AD(Q)) = \overleftrightarrow{I}(Q) + \overleftrightarrow{I}(Q_i \setminus S^*; S^*)$$

c/ Transfert généralisé (TG)

L'algorithme TG utilise deux opérateurs qui sont obtenus par la composition de AA et AD et notés AA o AD et AD o AA. TORO a démontré dans sa thèse (TOR 82) les propriétés suivantes :

Soit  $Q \in \mathbb{P}_r(\Sigma)$ ,  $2 \leq r \leq n-1$  on a :

$$\overleftrightarrow{I}(AD \circ AA(Q)) \leq \overleftrightarrow{I}(Q) \text{ et}$$

(10)

$$\overleftrightarrow{I}(AA \circ AD) \leq \overleftrightarrow{I}(Q)$$

A partir d'une partition initiale donnée, l'algorithme TG applique les opérateurs AA o AD et AD o AA autant de fois que cela est possible, jusqu'à la stabilisation qui est assurée par (10). La partition finale satisfait la condition nécessaire d'optimalité.

$$\overleftrightarrow{I}(AD \circ AA(Q)) = \overleftrightarrow{I}(AA \circ AD(Q)) = \overleftrightarrow{I}(Q)$$

La convergence est généralement obtenue en quelques itérations: plusieurs partitions initiales peuvent alors être essayées.

Exemple d'application

Considérons l'exemple de la figure I.1 pour lequel on se propose de chercher une partition en 3 classes.

Soit la partition initiale

$$P_0 = (\{x_1, x_2, x_3, x_4, x_5, x_6, x_{11}, x_{12}, x_7, x_8, x_9, x_{10}\})$$

$$\text{dont la transinformation externe } \overleftrightarrow{I}(P_0) = 53$$

L'algorithme TG fournit deux solutions.

$$P = (\{x_1, x_2, x_3, x_{12}\} \{x_4, x_5, x_6, x_7, x_8, x_9, x_{11}\} \{x_{10}\}) \overleftrightarrow{I}(P) = 32$$

$$P' = (\{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_{11}, x_{12}\} \{x_9\} \{x_{10}\}) \overleftrightarrow{I}(P') = 29$$

A partir de ces solutions, il est possible d'avoir les relations suivantes :

$$\left. \begin{aligned} f(x_1, x_2, x_3, x_{12}, \alpha) &= 0 \\ g(x_4, x_5, x_6, x_7, x_8, x_9, x_{11}, \beta) &= 0 \end{aligned} \right\} \text{ Pour la première solution}$$

$$h(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_{11}, x_{12}, \gamma) = 0 \text{ pour la deuxième solution}$$

$\alpha, \beta, \gamma$  représentent les effets du bruit ou d'autres variables.

### III - 5 Méthodes à centres variables

---

Les méthodes de décomposition hiérarchique fournissent une description minutieuse de l'ensemble  $\Sigma$ . En fait, dans la pratique, il est parfois inutile d'avoir une hiérarchie des parties de  $\Sigma$ , de plus ces méthodes nécessitent beaucoup de place mémoire et deviennent prohibitives dès que les données dépassent 350 variables à classer. Nous allons décrire quelques algorithmes de partitionnement assez simples, qui ont l'avantage de permettre la décomposition d'un grand nombre de variables qui serait difficile selon d'autres méthodes.

#### a/ Algorithme de FORGY (FOR 65)

L'ensemble  $\Sigma$  est muni d'une distance euclidienne, on se donne une partition de départ. Le 1er pas de l'algorithme consiste à déterminer les centres de gravité des classes de cette partition. Le 2ème pas de l'algorithme permet de construire une nouvelle partition en agglomérant les éléments autour de ces centres de gravité. La classe  $C_j$  contient les éléments plus proches du  $j^{\text{ème}}$  centre de gravité que des autres. Les éléments qui sont à une même distance de plusieurs centres de gravité sont affectés arbitrairement. La qualité d'une partition est mesurée par la somme des inerties des classes par rapport à leurs centres de gravité.

#### b/ Algorithme de Mac Queen (MAC 67)

Soit  $k$  le nombre de classes souhaité. On choisit arbitrairement  $k$  éléments de  $\Sigma$  comme centres d'agrégation et on constitue une classe avec chacun d'eux. On prend parmi les  $N - k$  autres éléments le 1er qui se présente et on le réunit au centre le plus proche. On remplace ce dernier par le centre de gravité des deux points réunis. On prend parmi les  $n - k - 1$  autres éléments le 1er qui se présente et on l'affecte au centre le plus proche. On remplace ce dernier par le centre de gravité de la classe formée..... et ainsi de suite jusqu'à ce que tous les éléments soient affectés. On obtient ainsi une partition  $P$  et on agglomère les points autour des centres de gravité des classes de  $P$ , ce qui fournit directement la partition finale  $P'$ .

#### c / Algorithme de DIDAY ( DID 72)

Sous le nom de méthode des nuées dynamiques, E. DIDAY a mis au point un algorithme de partition qui peut être considéré comme une généralisation des algorithmes de FORGY et de MAC QUEEN. En effet, l'agglomération dans cette méthode se fait autour d'ensembles d'éléments appartenant à  $\Sigma$ , appelés noyaux. Une classe d'une partition est alors représentée par plusieurs de ses éléments. Les noyaux forment un résumé des classes plus riches que ne peut l'être un centre de gravité.

Les algorithmes de transfert ainsi que ceux à "centres variables" semblent les mieux adaptés pour la recherche d'une partition. Ils sont d'autant plus rapides que la partition de départ est judicieusement choisie. Ils possèdent aussi l'avantage d'avoir un critère d'optimisation permettant de comparer la qualité des partitions obtenues.

C O N C L U S I O N

=====

Plusieurs méthodes de décomposition sont exposées dans ce chapitre. Le principal inconvénient des méthodes de décomposition par les graphes non valués réside dans le fait que l'intensité de liaison entre variables n'est pas prise en compte, ce qui pose problème par exemple pour la coordination des sous-systèmes. On utilise alors les graphes valués qui donnent une décomposition plus fine, mais ils sont difficiles à mettre en oeuvre.

L'analogie entre deux semi-valuations, l'entropie et l'application 'card' nous a permis de retrouver les principaux indices utilisés en analyse structurale et d'en proposer quelques uns qui sont à notre connaissance nouveaux.

Deux méthodes de décomposition directe sont ensuite présentées. La première est basée sur la classification hiérarchique. Plusieurs algorithmes sont rappelés. Ces méthodes nécessitent beaucoup de place mémoire. Les méthodes d'agrégation autour des centres variables ne nécessitent pas la mémorisation des  $\frac{N(N-1)}{2}$

"distances" entre éléments de  $\Sigma$ , car il suffit de connaître les distances des variables au centre d'agrégation. En plus, l'expérience montre que la partition finale est obtenue rapidement. Le principal inconvénient de ces méthodes tient au fait que la partition finale dépend de la partition initiale. La méthode des nuées dynamiques s'adapte bien à l'analyse structurale des systèmes complexes. Elle sera étudiée en détail au chapitre suivant.

CHAPITRE IV

---

METHODE DES NUÉES DYNAMIQUES (M.N.D.)

---

#### IV 1 - Introduction

Les algorithmes AA et AD explorent tous les niveaux du treillis des partitions, alors que les algorithmes de transfert par groupe n'explorent, pour un niveau  $r$ , que les niveaux adjacents  $r-1$  et  $r+1$ . L'algorithme des nuées dynamiques travaille sur un seul niveau et permet de plus, une description des classes au moyen de représentants caractéristiques, les noyaux. S'il est bien choisi, le noyau formera un résumé de sa classe, plus riche que ne peut l'être un centre de gravité, dans l'algorithme de FORGY par exemple.

Après avoir défini une fonction d'affectation  $f$ , une fonction de représentation  $g$ , et un critère d'optimisation  $w$ , l'algorithme des nuées dynamiques nous permet de construire par itération, à partir d'une famille quelconque de noyaux ou d'une partition initiale, une partition en  $k$  classes.

Plusieurs variantes sont étudiées et comparées en faisant simplement varier le choix du triplet  $f, g, w$  : celles basées sur une étude par paires de variables, et celles permettant la prise en compte des relations globales par groupes de variables. Les résultats sont interprétés et permettent d'établir un modèle possible du système étudié.



#### IV 2 - Principe de l'algorithme

En suivant les principales étapes de cette méthode telles qu'elles ont été décrites par DIDAY (DID 72) , on définit :

$\mathbb{L}_k$  est une partie de l'ensemble  $\{ L/L = (\lambda_1, \lambda_2, \dots, \lambda_j, \dots, \lambda_k) \mid \lambda_j \in \mathcal{L}, |\lambda_j| = n_j \text{ fixé} \}$ .  
ou selon les cas,  $\mathcal{L}$  pourra représenter  $\mathcal{P}(\Sigma)$  ou  $\mathcal{P}(\Omega)$  . Un élément de  $\mathbb{L}_k$  est appelé un système de noyaux et les parties  $\lambda_j$  , des noyaux.

$\mathbb{P}_k(\Sigma)$  désigne l'ensemble des partitions de  $\Sigma$  en k classes.

$\mathbb{P}_k(\Omega)$  désigne l'ensemble des partitions de  $\Omega$  en k classes.

Soit  $L_0 = (\lambda_1^0, \lambda_2^0, \dots, \lambda_k^0)$  la suite initiale de noyaux estimée ou prise au hasard dans  $\mathcal{L}$ . Après ce choix, le premier pas de l'algorithme consiste à associer chacune des variables du système au noyau le plus proche (au sens de D défini ci-dessous) partageant ainsi  $\Sigma$  en k classes  $\{ \mathcal{C}_1^{(1)}, \mathcal{C}_2^{(1)}, \dots, \mathcal{C}_k^{(1)} \} = P_1$ .  
A chaque classe, on associe le noyau le plus représentatif, formant de ce fait une nouvelle suite  $:(\lambda_1, \lambda_2, \dots, \lambda_j, \dots, \lambda_k)$  et ainsi de suite.

A chaque itération, on mesure la qualité du couple,  $(L_i, P_i)$  par un critère adéquat  $w$  que nous définissons ci-dessous. DIDAY étudie la convergence de ce type d'algorithme (DID 72) . Son formalisme fait appel à deux applications  $f$  et  $g$ .  $f(L) = P$ ,  $f$  est l'application qui à un ensemble de k noyaux associe une partition à au plus k classes.  $g(P) = L$ ,  $g$  est l'application qui, à une partition  $P$  à k classes associe un ensemble  $L$  de k noyaux . On passe d'une partition  $P_i$  à une autre  $P_{i+1}$  par

$$P_{i+1} = (f \circ g)(P_i) = (f \circ g)^{i+1}(P_0)$$

Nous utilisons les concepts d'entropie pour le choix de  $f$  et  $g$  ainsi que  $w$ . On peut aussi partir de  $P_0 = \{ \mathcal{C}_1^{(0)}, \mathcal{C}_2^{(0)}, \dots, \mathcal{C}_k^{(0)} \}$ , on en déduit  $L_1 = g(P_0)$  et  $P_1 = f(L_1)$  etc ...

#### IV 3 - Choix du triplet (f, g, w)

- l'application f : On choisit un indice  $D$  permettant de mesurer le degré de liaison entre une variable et un élément de  $\mathcal{L}$  . La jème classe  $\mathcal{C}_j$  de la partition  $P$  associée au noyau  $\lambda_j$  est alors définie par :

$$\mathcal{C}_j = \{ X \in \Sigma \mid D(X, \lambda_j) \leq D(X, \lambda_i), \forall j \neq i \}$$

en cas d'égalité, on affecte  $X$  à la partie du plus petit indice.

- l'application g : pour juger dans quelle mesure un élément de  $\Lambda$  peut-être considéré comme représentatif de la classe  $\mathcal{C}_j$  de la partition P, on choisit un indice  $R(\mathcal{C}_j, \lambda_j)$  faisant apparaître la nature du couplage entre  $\mathcal{C}_j$  et  $\lambda_j$ . Le jème noyau est celui qui minimise  $R(\mathcal{C}_j, \lambda_j)$

- critère w :

La qualité du couple  $V = (L, P)$  est mesurée par le critère suivant :

$$w(v) = \sum_{j=1}^k R(\mathcal{C}_j, \lambda_j)$$

$$V_0 = (L_0, P_0)$$

⋮

$$V_i = (L_i, P_i)$$

⋮

$$V_n = (L_n, P_n)$$

à la suite  $V_0, V_1, \dots, V_i, \dots, V_n$  on associe la suite  $u_0, u_1, \dots, u_i, \dots, u_n$  avec  $u_i = W(V_i)$

#### IV 4 - Variantes

Le choix des fonctions f et g et de l'ensemble des noyaux  $\Lambda$  dépend du but de l'analyse qu'on se fixe. Ce choix permet de définir différentes variantes que nous allons examiner.

$$\text{IV 4.1} \quad \Lambda = \mathcal{P}(\Sigma) \quad R(\mathcal{C}_j, \lambda_j) = \sum_{X \in \mathcal{C}_j} D(X, \lambda_j)$$

Dans ce cas, les noyaux sont choisis parmi l'ensemble des variables du système étudié. L'indice D mesure l'intensité de couplage entre une variable et une partie de  $\Sigma$ . C'est le choix de D qui définit les propriétés des partitions obtenues. On va alors faire plusieurs choix. Si l'on cherche à construire des classes bien compactes, on choisira l'indice  $D(X, \lambda_j) = \max_{X_i^j \in \lambda_j} H(X/X_i^j)$  noté  $\Delta$  ; si l'on cherche à bien reconstituer les classes filiformes, on prendra  $S(X, \lambda_j) = \min_{X_i^j \in \lambda_j} H(X/X_i^j)$  noté  $\delta$ .

Ces deux indices sont illustrés sur la figure suivante :

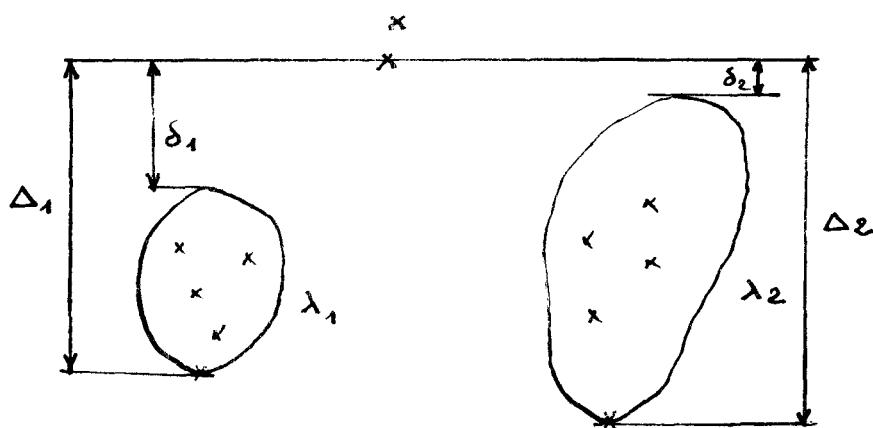


FIG. IV.1

Ces deux indices conduisent à construire des partitions très différentes comme on peut le concevoir en regardant le schéma ci-dessus.

Avec l'indice  $\delta$ ,  $X$  est dans la classe de noyau  $\lambda_2$ , et avec l'indice  $\Delta$ ,  $X$  est dans la classe de noyau  $\lambda_1$ . D'autres distances peuvent être imaginées, dont la plus simple est la moyenne de l'entropie de  $X$  non expliquée par les  $X_i^j$

$$D(X, \lambda_j) = \frac{1}{n_j} \sum_{X_i^j \in \lambda_j} H(X/X_i^j)$$

On examinera aussi des indices tenant compte de l'aspect multidimensionnel des variables tel que par exemple  $D(X, \lambda_j) = H(X/\lambda_j)$ . Dans tous les cas, l'entropie conditionnelle minimum suppose l'existence d'une relation fonctionnelle, éventuellement bruitée. (cf §II.3.1). Un seuil de l'entropie conditionnelle peut être fixé à partir duquel l'hypothèse de l'existence de ces relations sera rejetée.

IV 4.1.a 
$$D(X, \lambda_j) = \min_{X_i^j \in \lambda_j} H(X/X_i^j)$$

L'appartenance d'une variable  $X$  à une classe de noyau  $\lambda_j$  implique la possibilité de construire un modèle de la forme suivante :

$$\exists X_i^j \in \lambda_j \quad \text{telle que} \quad X = f(X_i^j)$$

c'est-à-dire aussi 
$$X = f'(X_j)$$

$$\text{IV 4.1.b} \quad D(X, \lambda_j) = \max_{X_i^j \in \lambda_j} H(X/X_i^j)$$

Dans cette variante, le modèle sera de la forme suivante  
 $\forall X_i^j \in \lambda_j$  on a  $X = F(X_i^j)$  et par conséquent  $X = F'(\lambda_j)$   
 $X$  étant une variable appartenant à la classe de noyau  $\lambda_j$ .

$$\text{IV 4.1.c} \quad D(X, \lambda_j) = \frac{1}{n_j} \sum_{X_i^j \in \lambda_j} H(X/X_i^j)$$

Cette variante se situe en quelque sorte entre les deux précédentes et le modèle qu'on peut en déduire est :  $X = h(\lambda_j)$ , si  $X$  est une variable de la classe de noyau  $\lambda_j$ .

$$\text{IV 4.1.d} \quad D(X, \lambda_j) = H(X/\lambda_j)$$

L'avantage de cette variante sur les précédentes est son caractère multidimensionnel, dans le sens où elle tient compte de l'information que toutes les variables de  $\lambda_j$  prises dans leur ensemble fournissent sur  $X$ . Ce qui peut être très différent d'une addition de l'information que chacune des variables de  $\lambda_j$  fournit séparément sur  $X$ .

Dans les variantes IV 4.1.c et IV.4.1.d,  $0 \leq D(X, \lambda_j) \leq H(X)$   
 $D(X, \lambda_j) = 0$  si et seulement si il existe une application  $g$  entre  $X$  et  $\lambda_j$  telle que  $X = g(\lambda_j)$

- variantes utilisant une distance :

$$\text{IV 4.1.e} \quad D(X, \lambda_j) = \frac{1}{n_j} \sum_{X_i^j \in \lambda_j} [H(X/X_i^j) + H(X_i^j/X)]$$

$$0 \leq D(X, \lambda_j) \leq H(X) + \frac{1}{n_j} \sum_{X_i^j \in \lambda_j} H(X_i^j)$$

$$\text{IV 4.1.f} \quad D(X, \lambda_j) = H(X/\lambda_j) + H(\lambda_j/X)$$

$$0 \leq D(X, \lambda_j) \leq H(X) + H(\lambda_j)$$

Dans les variantes IV 4.1.e et IV 4.1.f,  $D(X, \lambda_j) = 0$  si et seulement si il existe une bijection  $F$  telle que  $X = F(\lambda_j)$ . Cet indice est donc moins général que celui dans les variantes IV 4.1.c et IV 4.1.d

$$\text{IV 4.2.} \quad \Lambda = \mathbb{P}(\Omega) \quad R(\mathcal{C}_j, \lambda_j) = \sum_{X \in \mathcal{C}_j} D(P_X, P_{\lambda_j})$$

Les noyaux peuvent être constitués par un vecteur  $\lambda_j$  défini sur  $\mathbb{P}(\Omega)$ .  
 Ce vecteur n'appartient donc pas forcément à  $\Sigma$ .

Ceci nous permet de raisonner sur les partitions de  $\Omega$  sachant que l'entropie d'une variable et celle de sa partition associée sont égales

$$\text{IV 4.2.a } D(X, \lambda_j) = H(P_X / P_{\lambda_j})$$

$D(X, \lambda_j) = 0 \iff X = g'(\lambda_j)$   $g'$  est une application entre  $X$  et  $\lambda_j$

$$\text{IV 4.2.b } D(X, \lambda_j) = H(P_X / P_{\lambda_j}) + H(P_{\lambda_j} / P_X)$$

$D(X, \lambda_j) = 0 \iff X = G'(\lambda_j)$   $G'$  est une bijection.

On a pour toutes ces variantes le résultat suivant :

Si les  $m_j$  variables  $X_1, X_2, \dots, X_{m_j}$  se trouvent dans la même classe  $\mathcal{C}_j$  de noyaux  $\lambda_j$ , cela implique la possibilité d'établir  $m_j$  équations dépendant de la valeur de  $D(X_i, \lambda_j)$  :

Une valeur élevée de  $D(X_i, \lambda_j)$  nous indiquerait que  $\lambda_j$  n'est pas suffisamment représentatif de  $X_i$ , ou qu'il n'est pas possible

de construire un modèle déterministe et qu'il faut introduire un nombre plus ou moins important de paramètres  $b_i$  tel que

$$X_1 = f_1(\lambda_j, b_1)$$

$$X_2 = f_2(\lambda_j, b_2)$$

⋮

$$X_{m_j} = f_{m_j}(\lambda_j, b_{m_j})$$

Lorsque  $D(X_i, \lambda_j)$  tend vers 0, le modèle devient

$$X_1 = f_1(\lambda_j)$$

$$X_2 = f_2(\lambda_j)$$

$$X_{m_j} = f_{m_j}(\lambda_j)$$

Si  $X_1$  prend  $n_{X_1}$  modalités

$X_2$  "  $n_{X_2}$  " "

$X_{m_j}$  "  $n_{X_{m_j}}$  " "

On pose  $m = \max(n_{X_1}, n_{X_2}, \dots, n_{X_{m_j}})$ . On a vu au paragraphe (II.4) qu'on n'a pas besoin de découper  $\lambda_j$  en plus de  $m$  modalités. On prend alors comme noyau de la classe  $\mathcal{C}_j$ , le vecteur  $\lambda_j$  obtenu en regroupant les modalités de  $\lambda_j$  de façon à en avoir  $m$  sans perdre beaucoup d'informations.

IV.4.3. Convergence de ces algorithmes

définition 1 : R est semi-carré si

$$\forall L \in \mathbb{L}_k \quad W(g \circ f(L), f \circ g(L)) \leq W(g \circ f(L), f(L))$$

définition 2 : un élément  $v = (L, P) \in \mathbb{L}_k \times \mathbb{P}_k$  est dit sans biais pour les fonctions f et g si  $P = f \circ g(P)$  et  $L = g \circ f(L)$ .

définition 3 : une suite  $(L^{(n)}, P^{(n)})$  de  $\mathbb{L}_k, \mathbb{P}_k$  est dite convergente si elle est stationnaire, c'est-à-dire s'il existe un entier M tel que

$$\forall n \geq M \quad (L^{(n)}, P^{(n)}) = (L^*, P^*)$$

On trouve dans (DIDAY.1972) la preuve des théorèmes et propriétés suivantes :

Théorème 1: Si R est semi-carré, alors le triplet (f,g,w) définit une suite  $V_n$  telle que la suite de réels  $W(V_n)$  soit décroissante.

Propriété 1:  $\forall \lambda_j \in L \quad \forall P \in \mathbb{P}_k \quad P = \{\varrho_1, \varrho_2, \dots, \varrho_j, \varrho_k\}$

$$R(\varrho_j, \lambda_j) = \sum_{X \in \varrho_j} D(X, \lambda_j)$$

alors R est semi-carré.

Théorème 2: Si R a les deux propriétés suivantes

1) semi-carré

2)  $\forall P \in f(\mathbb{L}_k), \forall j = 1, \dots, k, R(\varrho_j, \lambda_j)$  est minimum pour un unique  $\lambda$

Alors,  $(V_n)$  est convergente et sa limite est sans biais.

Les variantes précédemment décrites vérifient la condition 1 du théorème 2. Lorsque le noyau n'est pas unique, c'est-à-dire lorsque plusieurs noyaux minimisent la fonction R, le noyau retenu sera par exemple : celui qui contient une variable ayant l'indice i minimum. La condition 2 étant alors à son tour vérifiée, il s'ensuit la convergence des algorithmes précédents.

IV.4.4. Détermination des noyaux

Pour certaines variantes, le problème de recherche du noyau  $\lambda_j$  d'une classe

$\varrho_j$  conduit à la solution triviale  $\lambda_j = \varrho_j$ . Nous allons examiner deux exemples :

variante IV.4.1.d

Le noyau  $\lambda_j$  est formé de  $n_j$  éléments de  $\Sigma$  qui minimisent  $R(\varrho_j, \lambda_j)$ , c'est-à-dire

$$R(\varrho_j, \lambda_j) \leq R(\varrho_j, S) \quad \forall S \in \mathcal{P}(\Sigma)$$

avec  $R(\varrho_j, \lambda_j) = \sum_{X \in \varrho_j} H(X/\lambda_j)$

Le problème posé de cette façon conduit, dans le cas où le cardinal de  $\lambda_j$  n'est pas fixé à la solution triviale  $\lambda_j = \mathcal{C}_j$ ,  $j = 1, 2, \dots, k$ . Or, cette solution triviale ne présente pas d'intérêt pour la M N D, en effet l'algorithme converge en une itération puisque les noyaux forment une partition de  $\Sigma$ , et la partition cherchée sera la même.

Nous allons donc chercher des optimums locaux, et poser le problème suivant :

$$\forall \mathcal{C}_j \in \Sigma \text{ et pour } l = 1, 2, \dots, |\mathcal{C}_j| \text{ trouver } \lambda_j^l \subset \mathcal{P}_1(\Sigma)$$

$$\text{avec } |\lambda_j^l| = l \text{ tel que } \forall S \in \mathcal{P}_1(\Sigma), |S| = l$$

$$R(\mathcal{C}_j, \lambda_j^l) \leq R(\mathcal{C}_j, S)$$

La formulation de ce problème résulte des problèmes d'explication. On va pouvoir utiliser le même type d'algorithmes.

ALGORITHME AGREGATIF ( AA )

Pour chaque classe  $\mathcal{C}_j \in \mathcal{P}(\Sigma)$  cet algorithme construit une chaîne de noyaux  $S_1, S_2, \dots, S_{|\mathcal{C}_j|-1}$  avec  $S_{l+1} = S_l \cup \{x_\alpha\}$ , la variable  $x_\alpha$  étant choisie de façon à rendre  $R(\mathcal{C}_j, S_{l+1})$  minimum.

ALGORITHME DESAGREGATIF ( AD )

Pour chaque classe  $\mathcal{C}_j \in \mathcal{P}(\Sigma)$ , cet algorithme construit une chaîne de noyaux  $S_{|\mathcal{C}_j|-1}, \dots, S_2, S_1$  avec  $S_{l-1} = S_l \setminus \{x_\alpha\}$  la variable  $x_\alpha$  étant choisie de façon à rendre  $R(\mathcal{C}_j, S_{l-1})$  minimum.

ALGORITHME D'APPROXIMATION SUCCESSIVE. ( AAS )

Cet algorithme consiste à appliquer successivement les deux précédents jusqu'à stabilisation. Il cherche pour chaque  $\mathcal{C}_j \in \mathcal{P}(\Sigma)$  et pour chaque  $1 \leq l \leq |\mathcal{C}_j| - 1$  un noyau vérifiant la condition nécessaire d'optimalité.

Les algorithmes AA et AD se résument à une application répétée pour chaque  $\mathcal{C}_j \in \mathcal{P}(\Sigma)$  des opérateurs suivants :

$$(ajouter) A : \mathcal{P}_1(\Sigma) \rightarrow \mathcal{P}_{l+1}(\Sigma), 0 \leq l \leq |\mathcal{C}_j| - 1$$

$$S \rightsquigarrow A(S) = S \cup \{x_\alpha\}$$

$$\text{où } x_\alpha \in \Sigma \setminus S, \text{ et } \forall x \in \Sigma \setminus S : R(\mathcal{C}_j, S \cup \{x_\alpha\}) \leq R(\mathcal{C}_j, S \cup \{x\})$$

(retirer)  $R : \mathcal{P}_\ell(\Sigma) \rightarrow \mathcal{P}_{\ell-1}(\Sigma), 1 \leq \ell \leq |\mathcal{E}_j|$   
 $S \rightsquigarrow R(S) = S \setminus \{x_\alpha\}$

où  $x_\alpha \in S$  et  $\forall x \in S, R(\mathcal{E}_j, S \setminus \{x_\alpha\}) \leq R(\mathcal{E}_j, S \setminus \{x\})$

A l'aide des opérateurs A et R nous pouvons définir les deux opérateurs suivants, pour  $\ell = 1, 2, \dots, |\mathcal{E}_j| - 1$

AOR :  $\mathcal{P}_\ell(\Sigma) \rightarrow \mathcal{P}_\ell(\Sigma)$   
 $S \longmapsto AOR(S) = A(R(S))$

ROA :  $\mathcal{P}_\ell(\Sigma) \rightarrow \mathcal{P}_\ell(\Sigma)$   
 $S \longmapsto ROA(S) = R(A(S))$

Proposition IV.1

Quel que soit  $\mathcal{E}_j \in \mathcal{P}(\Sigma)$ , et quel que soit  $\ell, 1 \leq \ell \leq |\mathcal{E}_j| - 1$   
 $\forall S \in \mathcal{P}_\ell(\Sigma)$  :

$$R(\mathcal{E}_j, AOR(S)) \leq R(\mathcal{E}_j, S)$$

$$R(\mathcal{E}_j, ROA(S)) \leq R(\mathcal{E}_j, S)$$

Démonstration

Soit  $\mathcal{E}_j \in \mathcal{P}(\Sigma)$  et  $S \in \mathcal{P}_\ell(\Sigma)$ , pour  $\ell : 1 \leq \ell \leq |\mathcal{E}_j| - 1$   
 $R(\mathcal{E}_j, R(S)) = \min \{ R(\mathcal{E}_j, S \setminus \{x\}) : x \in S \} = R(\mathcal{E}_j, S \setminus \{x_\beta\})$   
 pour un certain  $x_\beta \in S$ . D'autre part, nous avons :  
 $R(\mathcal{E}_j, AOR(S)) = R(\mathcal{E}_j, A(R(S))) = R(\mathcal{E}_j, A(S \setminus \{x_\beta\}))$   
 $= \min \{ R(\mathcal{E}_j, S \setminus \{x_\beta\} \cup \{x\}) : x \in \Sigma \setminus (S \setminus \{x_\beta\}) \}$   
 or  $x_\beta \in \Sigma \setminus (S \setminus \{x_\beta\})$  donc  
 $R(\mathcal{E}_j, AOR(S)) \leq R(\mathcal{E}_j, S \setminus \{x_\beta\} \cup \{x_\beta\}) = R(\mathcal{E}_j, S)$

La démonstration pour  $ROA(S)$  est analogue à celle-ci.

(c.f.d.)

Proposition IV.2

Tous les noyaux optimaux  $\lambda_j^\ell$  sont des points fixes des opérateurs AOR et ROA, c'est-à-dire,

$$\forall \mathcal{E}_j \in \Sigma, \forall \ell : 1 \leq \ell \leq |\mathcal{E}_j| - 1$$

$$R(\mathcal{E}_j, \lambda_j^\ell) = R(\mathcal{E}_j, AOR(\lambda_j^\ell)) = R(\mathcal{E}_j, ROA(\lambda_j^\ell)).$$



Démonstration

Etant donné que AOR  $(\lambda_j^1)$  et ROA  $(\lambda_j^1)$  appartiennent à  $\mathcal{P}_1(\Sigma)$  par définition de  $\lambda_j^1$  on a :

$$R(\mathcal{C}_j, \text{AOR}(\lambda_j^1)) \geq R(\mathcal{C}_j, \lambda_j^1) \leq R(\mathcal{C}_j, \text{ROA}(\lambda_j^1)).$$

On a aussi d'après la proposition IV.1

$$R(\mathcal{C}_j, \text{AOR}(\lambda_j^1)) \leq R(\mathcal{C}_j, \lambda_j^1) \geq R(\mathcal{C}_j, \text{ROA}(\lambda_j^1)).$$

Condition d'optimalité

On dira qu'une partie de  $\Sigma$ ,  $S \subset \mathcal{P}(\Sigma)$  est stable si c'est un point fixe aussi bien de AOR que de ROA. On vient de voir par ailleurs que tout noyau optimal  $\lambda_j^1$  est stable. Cependant, il ne s'agit que d'une condition nécessaire d'optimalité, car il peut y avoir des parties de  $\Sigma$  stables qui ne soient pas optimales.

L'algorithme d'approximations successives que nous proposons cherche pour chaque classe  $\mathcal{C}_j$  et pour  $l$  fixé  $1 \leq l \leq |\mathcal{C}_j| - 1$  une partie de  $\Sigma$  vérifiant la condition nécessaire d'optimalité. Pour cela et pour  $l$  fixé, on se donne un sous ensemble à  $l$  éléments de  $\Sigma$ , et on lui applique les opérateurs AOR et ROA autant de fois que cela est possible jusqu'à stabilisation. La convergence est assurée par le caractère contractant des opérateurs.

Convergence de l'algorithme principal

Les noyaux ne vérifient pas une condition d'optimalité globale, c'est-à-dire que l'algorithme trouve le meilleur noyau parmi ceux qui lui sont accessibles. Pour que l'inégalité suivante

$$w(P_{i+1}, L_{i+1}) \leq w(P_{i+1}, L_i)$$

nécessaire à la convergence de l'algorithme principal soit vérifiée, et pour calculer  $L_{i+1}$ , on part de  $L_i$  et on applique l'algorithme AOR ou ROA. Puisque  $R(\mathcal{E}_{j+1}, \text{AOR}(\lambda_j)) \leq R(\mathcal{E}_{j+1}, \lambda_j)$  l'inégalité ci-dessus est vérifiée.

Variante IV.4.2.a

Le noyau  $\lambda_j$  est constitué par un vecteur défini sur  $\mathbb{P}(\Omega)$ . Ce vecteur n'appartient donc pas forcément à  $\Sigma$ . Ceci nous permet de raisonner sur les partitions de  $\Omega$  sachant que l'entropie d'une variable et celle de sa partition associée sont égales.

On cherche donc une partition de  $\Omega$ ,  $P_{\lambda_j}$  telle que

$$\sum_{X \in \mathcal{E}_j} D(P_X, P_{\lambda_j}) \leq \sum_{X \in \mathcal{E}_j} D(P_X, P) \quad \forall P \in \mathbb{P}(\Omega)$$

Comme pour la variante **a**, nous allons chercher des optimums locaux de la façon suivante.

$\forall \mathcal{E}_j \in \Sigma$  et pour  $l = 1, 2, \dots, |\mathcal{E}_j|$  trouver  $P_{\lambda_j}^l \in \mathbb{P}_1(\Omega)$ , tel que

$$\forall P \in \mathbb{P}_1(\Omega)$$

$$\sum_{X \in \mathcal{E}_j} D(P_X, P_{\lambda_j}^l) \leq \sum_{X \in \mathcal{E}_j} D(P_X, P)$$

Algorithme ascendant AH

Pour chaque classe  $\mathcal{C}_j \in \mathcal{P}(\Sigma)$  à laquelle est associée la partition  $P_{\mathcal{C}_j}$  et à partir d'une partition du niveau  $l$ , on recherche parmi toutes celles du niveau  $l-1$  moins fines celle qui minimise le critère d'optimalité. On passe d'une partition  $R = \{R_1, \dots, R_l\}$  du niveau  $l$  à une partition du niveau  $l-1$  par regroupement en une seule de deux de ses classes,  $R_\alpha, R_\beta$ . Cet algorithme part de la partition triviale  $\{\omega_1, \omega_2, \dots, \omega_L\}$ .

Algorithme descendant DR

Cet algorithme est en quelque sorte l'inverse du précédent. On passe d'une partition  $R = \{R_1, \dots, R_l\}$  du niveau  $l$  à une partition de niveau  $l+1$  par l'éclatement d'une de ses classes en deux. Pour chaque classe  $\mathcal{C}_j \in \mathcal{P}(\Sigma)$ , on part de la partition triviale  $P_{\mathcal{C}_j}$ , et on traverse les différents niveaux du treillis en minimisant à chaque étape le critère d'optimalité.

Les algorithmes (AH) et (DR) peuvent se résumer à une application répétée des deux opérateurs suivants.

Opérateur AH

$$AH : \mathcal{P}_l(\Omega) \longrightarrow \mathcal{P}_{l-1}(\Omega) \quad 2 \leq l \leq |\mathcal{C}_j|$$

$$R = \{R_1, \dots, R_l\} \longrightarrow AH(R) = \{R_1, \dots, R_\alpha \cup R_\beta, \dots, R_l\}$$

avec  $R_\alpha, R_\beta \in R$  tel que

$$\forall R_i, R_j \in R \quad \sum_{X \in \mathcal{C}_j} D(P_X, AH(R)) \leq \sum_{X \in \mathcal{C}_j} D(P_X, \{R_1, \dots, R_i \cup R_j, \dots, R_l\})$$

Opérateur DR

$$DR : \mathcal{P}_l(\Omega) \longrightarrow \mathcal{P}_{l+1}(\Omega) \quad 1 \leq l \leq |\mathcal{C}_j| - 1$$

$$R = \{R_1, \dots, R_l\} \longrightarrow DR(R) = \{R_1, \dots, R_\alpha \setminus S^*, S^*, \dots, R_l\}$$

avec  $R_\alpha \in R, S^* \subset R_\alpha, \emptyset \neq S^* \neq R_\alpha$ , et tel que

$$\sum_{X \in \mathcal{E}_j} D(P_X, DR(R)) \leq \sum_{X \in \mathcal{E}_j} D(P_X, \{R_1, \dots, R_i \setminus S, S, \dots, R_1\})$$

$$\forall R_i \in R, \forall S \subseteq R_i, \emptyset \neq S \neq R_i.$$

La composition des deux opérateurs AH et DR permet alors de définir les deux nouveaux que nous utilisons dans la M.N.D.

Opérateurs contractants

$$AH \text{ ODR} : \mathbb{P}_1(\Omega) \rightarrow \mathbb{P}_1(\Omega) \quad 2 \leq l \leq |\mathcal{E}_j| - 1$$

$$R \longrightarrow AH \text{ ODR}(R) = AH(DR(R))$$

$$DROAH : \mathbb{P}_1(\Omega) \longrightarrow \mathbb{P}_1(\Omega)$$

$$R \longrightarrow DROAH(R) = DR(AH(R))$$

Proposition IV.3

Soit Q une partition quelconque du niveau l,  $2 \leq l \leq |\mathcal{E}_j| - 1$

alors :

$$1) \sum_{X \in \mathcal{E}_j} D(P_X, AHODR(R)) \leq \sum_{X \in \mathcal{E}_j} D(P_X, R)$$

$$2) \sum_{X \in \mathcal{E}_j} D(P_X, DROAH(R)) \leq \sum_{X \in \mathcal{E}_j} D(P_X, R)$$

Démonstration

Soit  $R = \{R_1, R_2, \dots, R_l\}$ . Moyennant une rénumérotation éventuelle, on peut écrire  $AH(R) = \{R_1 \cup R_2, R_3, \dots, R_l\}$

$$\text{avec } \sum_{X \in \mathcal{E}_j} D(P_X, AH(R)) \leq \sum_{X \in \mathcal{E}_j} D(P_X, \{R_1 \cup R_2, R_3, \dots, R_l\})$$

Pour le calcul de  $D(Pe_j, DR(AH(R)))$ , deux cas peuvent se présenter

1°) L'application de DR divise la classe  $R_1 \cup R_2$

$$\sum_{X \in \mathcal{E}_j} D(P_X, AHODRO(R)) \leq \sum_{X \in \mathcal{E}_j} D(P_X, \{R_1 \cup R_2 \setminus S, S, \dots, R_1\}) \quad \forall S \subset R_1 \cup R_2 \\ \emptyset \neq S \neq R_1 \cup R_2$$

en particulier, on a, pour  $S = R_2$

$$\sum_{X \in \mathcal{E}_j} D(P_X, AHODRO(R)) \leq \sum_{X \in \mathcal{E}_j} D(P_X, R).$$

2°) L'application de DR divise une autre classe,  $R_3$  par exemple

$$\sum_{X \in \mathcal{E}_j} D(P_X, AHODRO(R)) = \sum_{X \in \mathcal{E}_j} D(P_X, \{R_1 \cup R_2, R_3 \setminus S^*, S^*, \dots, R_1\}) \\ \forall S \subset R_3 \quad \emptyset \neq S \neq R_3$$

en particulier, on a :

$$\sum_{X \in \mathcal{E}_j} D(P_X, AHODRO(R)) \leq \sum_{X \in \mathcal{E}_j} D(P_X, \{R_1 \cup R_2, R_1 \cup R_2 \setminus R_2, R_2, \dots, R_1\}) \\ \leq \sum_{X \in \mathcal{E}_j} D(P_X, \{R_3, R_1, R_2, \dots, R_1\}) \\ \leq \sum_{X \in \mathcal{E}_j} D(P_X, R)$$

La démonstration est semblable pour le calcul de  $D(Pe_j, AH(DR(R)))$

(c.f.d)

Proposition IV.4

Toute partition optimale  $M_1^*$  du niveau 1, est un point fixe des applications DROAH et AHODR, c'est-à-dire, qu'on a :

$$\sum_{X \in \mathcal{E}_j} D(P_X, AHODR(M_1^*)) = \sum_{X \in \mathcal{E}_j} D(P_X, DROAH(M_1^*)) = \sum_{X \in \mathcal{E}_j} D(P_X, M_1^*)$$

$$\forall 2 \leq l \leq |\mathcal{E}_j| - 1$$

Démonstration

AHODR ( $M_1^*$ ) et DROAH( $M_1^*$ ) sont deux partitions appartenant à  $\mathbb{P}_1(\Omega)$  cela implique les inégalités suivantes :  $D(P_{\mathcal{E}_j}, AHODR(M_1^*)) \geq D(P_{\mathcal{E}_j}, M_1^*)$

et  $\sum_{X \in \mathcal{E}_j} D(P_X, DROAH(M_1^*)) \geq \sum_{X \in \mathcal{E}_j} D(P_X, M_1^*)$ . Par ailleurs, l'utilisation du théorème

précédent fournit l'inégalité dans l'autre sens, ce qui prouve l'égalité annoncée.

Condition d'optimalité



On dira qu'une partition  $M_1 \in \mathbb{P}_1(\Omega)$  est une partition stable ssi c'est un point fixe à la fois pour les opérateurs DROAH et AHODR. On vient de voir que toute partition optimale  $M_1^*$  ( $2 \leq l \leq |\mathcal{E}_j| - 1$ ) est stable. Il s'agit d'une condition nécessaire d'optimalité, car il se peut qu'il y ait des décompositions stables qui ne soient pas des optimums locaux.

L'algorithme d'approximations successives que nous proposons cherche pour chaque classe  $\mathcal{E}_j$  et pour l fixé  $1 < l \leq |\mathcal{E}_j| - 2$ , une partition de  $\Omega$  vérifiant la condition nécessaire d'optimalité. Pour cela, et pour l fixé, on se donne une partition de  $\Omega$  à l classes, et on lui applique les opérateurs AHODR et DROAH autant de fois que cela est possible, jusqu'à stabilisation. La convergence est assurée par le caractère contractant des opérateurs.

Convergence de l'algorithme principal

Les noyaux  $P_{\lambda_j}$  ne vérifient pas une condition d'optimalité globale, c'est-à-dire que l'algorithme trouve le meilleur noyau parmi ceux qui lui sont accessibles. De la même façon que pour la variante  $d$  on part à chaque itération de  $P_{\lambda_j}$  et on lui applique l'algorithme AHODR ou DROAH pour être sur que l'algorithme converge.

Exemple d'application

Nous reprenons le système représenté par la figure I.1 . Cet exemple permet de comparer les résultats d'une variante, en faisant varier le choix des noyaux initiaux.

$$\mathbf{A} = \mathcal{P}(\Sigma) \quad D(X, \lambda_j) = H(X/\lambda_j) \quad R(\mathbf{e}_j, \lambda_j) = \sum_{X \in \mathbf{e}_j} D(X, \lambda_j)$$

On choisit trois vecteurs à deux variables chacun, comme noyaux initiaux.

$L_0 = (\{X_1 X_2\}, \{X_5 X_6\}, \{X_9 X_{10}\})$ . On applique l'algorithme pour la variante  $d$ .

il vient  $P_0 = (\{X_1 X_2 X_3 X_{12}\}, \{X_4 X_5 X_6 X_7\}, \{X_8 X_9 X_{10} X_{11}\}) \quad W = 167$

$L_1 = (\{X_2 X_{12}\}, \{X_4 X_7\}, \{X_8 X_{10}\}) \quad W = 128$

$$P_1 = (\{X_1 X_2 X_3 X_{12}\}, \{X_4 X_5 X_6 X_7\}, \{X_8 X_9 X_{10} X_{11}\}) = P_0 \quad W = 128$$

Cette solution indique la possibilité d'avoir les relations suivantes :

$$X_1 = f_1(X_2 X_{12}, \varepsilon_1) \quad X_5 = f_5(X_4 X_7, \varepsilon_5) \quad X_9 = f_9(X_8 X_{10}, \varepsilon_9)$$

$$X_3 = f_3(X_2 X_{12}, \varepsilon_3) \quad X_6 = f_6(X_4 X_7, \varepsilon_6) \quad X_{11} = f_{11}(X_8 X_{10}, \varepsilon_{11})$$

$\varepsilon_1, \varepsilon_3, \varepsilon_5, \varepsilon_6, \varepsilon_9, \varepsilon_{11}$  représentent les effets du bruit ou d'autres variables.

On choisit maintenant trois vecteurs à trois variables chacun comme noyaux initiaux.

$$L_0 = (\{X_1 X_3 X_{12}\}, \{X_4 X_5 X_7\}, \{X_8 X_{10} X_{11}\})$$

$$\text{il vient } P_0 = (\{X_1 X_2 X_3 X_{12}\}, \{X_4 X_5 X_6 X_7\}, \{X_8 X_9 X_{10} X_{11}\})$$

$$W = 65$$

$$L_1 = (\{X_2 X_3 X_{12}\}, \{X_4 X_5 X_7\}, \{X_4 X_{10} X_{11}\})$$

$$W = 60$$

$$P_1 = (\{X_1 X_2 X_3 X_{12}\}, \{X_4 X_5 X_6 X_7\}, \{X_8 X_9 X_{10} X_{11}\})$$

$$W = 60$$

d'où les relations possibles

$$X_1 = g_1(X_2 X_3 X_{12}, \alpha_1) \quad X_6 = g_6(X_4 X_5 X_7, \alpha_6) \quad X_9 = g_9(X_8 X_{10} X_{11}, \alpha_9)$$

$\alpha_1, \alpha_6$  et  $\alpha_9$  représentent les effets du bruit ou d'autres variables ( cf. §.II.3)

## VII - CONCLUSION

Si les différentes variantes ne sont pas fondamentalement différentes, il faut cependant souligner qu'elles ne traduisent pas de la même manière les résultats des décompositions qu'elles opèrent. Le choix d'une variante dépend de la nature des données à traiter et des buts précis de l'analyse. Ainsi lorsqu'on ne dispose que d'un tableau de "distances" entre chaque couple de variables, les stratégies "min" et "max" sont souvent utilisées suivant qu'on cherche à construire des agrégats filiformes ou compacts.

Pour tenir compte de l'aspect multidimensionnel des noyaux, nous avons proposé plusieurs variantes, en fonction du modèle souhaité. Les résultats obtenus sont meilleurs que dans l'approche "par paires", ce qui constitue la part la plus importante et la plus originale dans ce chapitre.

La M.N.D. a l'inconvénient de fixer à priori le cardinal des noyaux initiaux. Ce qui signifie, imposer un modèle dont le praticien n'est pas en mesure de vérifier la validité. Par ailleurs, cette méthode rapide et efficace présente vis-à-vis des techniques existantes, des facilités au niveau de l'interprétation des résultats.



## CONCLUSION GÉNÉRALE

---

La modélisation et la commande des systèmes complexes peuvent en général être largement facilitées par une analyse structurale. Celle-ci conduit souvent à une décomposition (si elle existe) en sous-systèmes "faiblement couplés" du système global.

Dans la première partie, nous avons présenté un ensemble de méthodes d'analyse et de décomposition des systèmes complexes. Chaque méthode possède son domaine privilégié d'application qui est fonction de la nature des données à traiter et des buts précis de l'analyse. Ainsi, une analyse en composantes principales présente un grand intérêt lorsque les variables sont quantitatives, le modèle étant linéaire. En revanche, l'approche informationnelle présente l'avantage d'être indépendante de la nature des variables et de n'exiger aucune hypothèse restrictive.

Dans la deuxième partie, nous avons présenté quelques algorithmes permettant d'expliquer " une variable à partir d'autres ". On aboutit alors à la définition d'un graphe d'interconnexion qui permet d'exhiber la structure des liaisons entre variables. Ce graphe présente un caractère "multidimensionnel" qui prend en compte des aspects échappant aux graphes classiques construits sur des critères linéaires.

L'influence du bruit sur les relations fonctionnelles est étudiée. Un algorithme permettant de réduire le nombre de modalités du vecteur explicatif est exposé.

La troisième partie contient un ensemble de méthodes de décomposition. La théorie des graphes en est un outil privilégié. Lorsque le modèle du système est connu, deux méthodes de décomposition directe sont ensuite présentées. La première, basée sur la classification hiérarchique, définit plusieurs algorithmes. Nous avons proposé quelques indices de similarité permettant leur réalisation. La deuxième, dite à centres variables, permet de traiter rapidement de grands tableaux de données, mais elle suppose une connaissance minimale sur la structure du système, puisqu'il faut fixer à priori le nombre de centres initiaux.

La quatrième partie est consacrée à la méthode des nuées dynamiques dont les principaux avantages résident dans la facilité au niveau de l'interprétation des résultats, et la possibilité de choisir des noyaux parmi des éléments particuliers du système étudié.

Lors de la décomposition d'un système en sous-systèmes faiblement couplés, certains sous-systèmes peuvent être largement plus complexes que d'autres. Il convient d'exiger en plus qu'ils soient d'une complexité uniforme. Dans l'annexe, nous avons proposé plusieurs critères dans ce sens.

Outre la nécessité de l'expérimentation de ces méthodes, on peut discerner dès à présent quelques directions de recherche intéressantes :

Etudier des possibilités de création d'indices de similarité nouveaux et préciser les conditions de leur utilisation.

Les méthodes d'explication et de décomposition permettent de mettre en évidence les relations fonctionnelles entre les variables du système étudié. Celles-ci peuvent être redondantes; une méthode permettant l'identification de ces relations facilitera aisément la phase de la modélisation.

Enfin, l'approche multicritère que nous avons abordée en annexe devrait faire l'objet d'une étude plus approfondie : Recherche d'autres critères ainsi que des algorithmes permettant leur mise en oeuvre.

A N N E X E

A N N E X E

On a constaté que le critère de couplage minimal conduit quelquefois à des partitions dont les classes sont de complexités très différentes. Pour pallier cet inconvénient, on peut envisager une approche multicritère faisant intervenir les critères suivants :

1° )  $\bar{I}(R)$  minimum, ou ce qui revient au même (cf § I.52.j)

$$\sum_{i=1}^r \dot{I}(R_i) \text{ maximum}$$

2° ) Les  $\dot{I}(R_i)$  équivalents

On note  $\dot{I}_i$  au lieu de  $\dot{I}(R_i)$

Portons les valeurs des transinformatives internes  $I_i$  sous la forme de coordonnées cartésiennes dans un espace que nous appellerons espace des partitions. Considérons le cas des partitions en deux classes. A la partition  $P_1$  correspond le point  $x_1$  de coordonnées  $I_1$  et  $I_2$  dans le plan des partitions. (Fig 1)

La droite  $I_1 + I_2 = p = \text{constante}$  est une droite de pente  $-1$  dans le plan des partitions. Admettons que c'est la droite AB pour la partition  $P_1$  sur la figure 1. On voit sans peine que la valeur de  $p$  pour les points de cette droite est proportionnelle à la longueur de la perpendiculaire OE abaissée de l'origine des coordonnées sur la droite AB, et que OE est en même temps bissectrice de l'angle compris entre les axes de coordonnées.

La figure 1 montre qu'on peut obtenir une partition meilleure au sens du critère 1 que  $P_1$  pour laquelle  $p$  sera plus grand. A cette fin il faut déplacer la droite AB parallèlement à elle-même en s'éloignant de l'origine des coordonnées.

A la position extrême A'B' de cette droite correspond la plus grande valeur de  $p$  qu'on note  $p_0$ .

Au sens du critère 2, une partition est d'autant meilleure qu'une autre qu'elle est plus proche de la bissectrice OE.

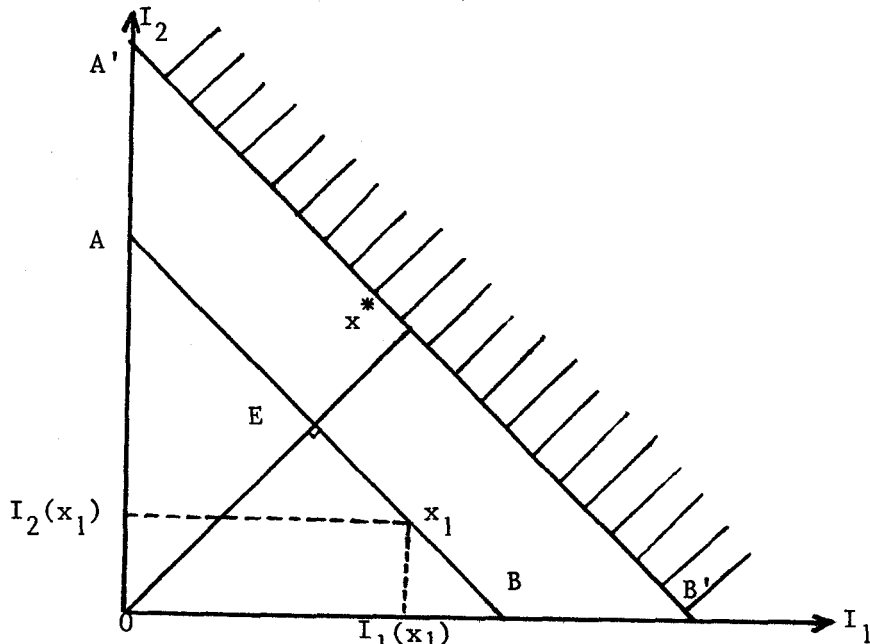


Figure 1

Comme il n'existe en général, aucun point  $x^*$  qui vérifie à la fois les deux critères, nous sommes obligés d'adopter l'un des points de vue ci-dessous :

1) Remplacer les deux critères par un seul en utilisant les normes (Fig 2)

Il est souvent commode de minimiser la distance des points  $x, x^*$  notée  $d(x, x^*)$ . Cette distance associée à une norme  $p$  sur l'espace des partitions est définie par  $d(x, x^*) = p(x - x^*)$

- Distances classiques dans  $R^n$

Désignons par  $I_i$  les coordonnées d'une partition  $x$  de  $X$  et  $\frac{\rho}{K}$  celles de  $x^*$

$$d(x, x^*) = \left[ \sum_{i=1}^k (I_i - \frac{\rho}{k})^2 \right]^{\frac{1}{2}}$$

$$d'(x, x^*) = \sup_{i=1 \dots K} \left| I_i - \frac{\rho}{k} \right|$$

$$d''(x, x^*) = \sum_{i=1}^k \left| I_i - \frac{\rho}{k} \right|$$

Interprétation de ces normes dans le plan des partitions

- Norme euclidienne

Pour chaque partition à laquelle on fait correspondre un point  $x_i$  ( $i=1, \dots, n$ ) on trace la boule qui est dans ce cas un cercle de centre  $x^*$  et de rayon  $R_i$  (FIG 2). La meilleure partition est donc celle qui minimise  $R_i$ . On voit immédiatement que tous les points  $x_{ik}$  ( $i \subset k$ ) qui se trouvent sur un même cercle de centre  $x^*$  sont équivalents au sens de cette norme, c'est à dire que toutes les partitions correspondantes à  $x_{ik}$  sont aussi bonnes les unes que les autres.

- Norme somme des valeurs absolues

Il en est de même pour cette norme, la boule étant le carré  $I$  (Fig 2) de centre  $x^*$ .

- Norme du max

Cette norme ne peut pas être utilisée. On va le montrer sur un exemple. La boule étant le carré  $II$  (Fig 2) prenons deux partitions dont les points correspondants se trouvent sur le même carré :  $x_1$  de coordonnées  $I_1$  et  $I_2$  tel que

$$I_1 = I_2 \text{ et } I_1 + I_2 = \rho_1 ; x_2 \text{ de coordonnées } I'_1 \text{ et } I'_2 ; I'_2 > I'_1, I'_1 + I'_2 = \rho_2 = \rho_1 :$$

donc  $x_1$  correspond à la meilleure partition, puisque pour la même somme que  $x_2$

les transformations sont égales.

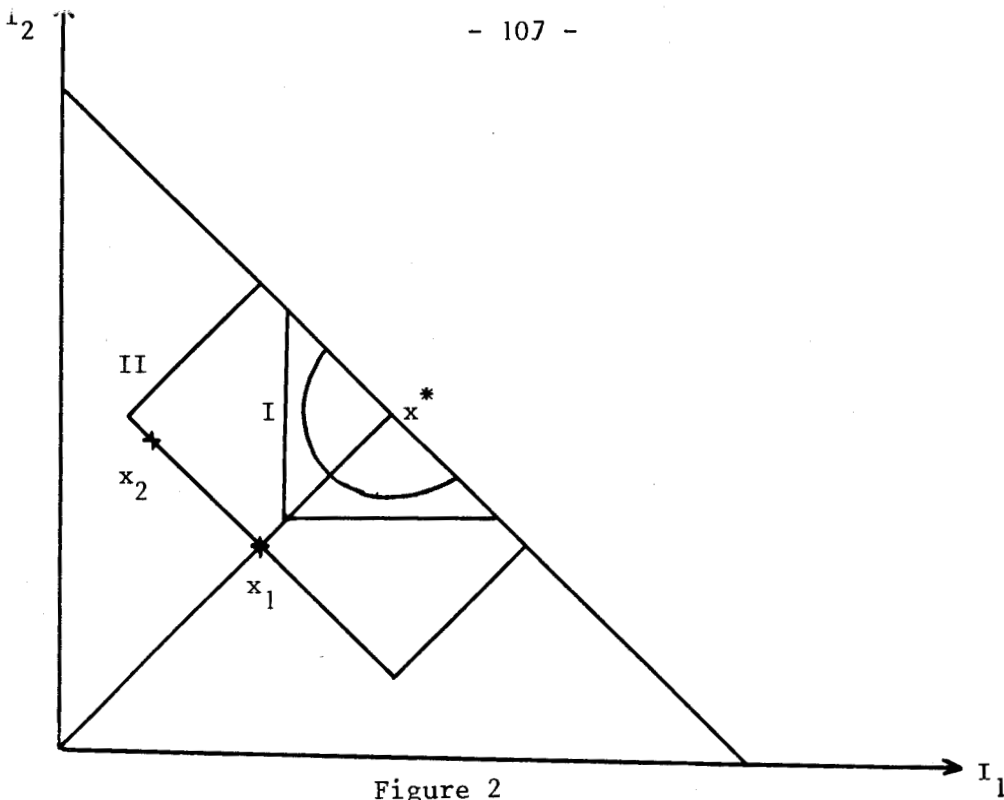


Figure 2

2) Critère multiplicatif

S'approcher le plus de la bissectrice OE (critère 2) peut se traduire par : trouver un coefficient d'atténuation multiplicatif  $\alpha(I_i)$  dont la valeur est maximum sur la bissectrice et s'atténue de part et d'autre de celle-ci (Fig 3b)

Considérons par exemple le critère

$$C = \alpha(I_i) \cdot \sum_{i=1}^k I_i$$

qu'il faut maximiser pour obtenir la meilleure partition.

- Calcul d'un coefficient d'atténuation

$$\cos \varepsilon = \frac{I_1 + I_2}{\sqrt{2} \sqrt{I_1^2 + I_2^2}} \quad \cos^2 \varepsilon = \frac{I_1^2 + I_2^2 + 2 I_1 I_2}{2 (I_1^2 + I_2^2)}$$

$$2 \cos^2 \varepsilon = 1 + 2 \frac{I_1 I_2}{I_1^2 + I_2^2}$$

$$\alpha(I_i) = 2 \cos^2 \varepsilon - 1 \Rightarrow \cos 2 \varepsilon = 2 \frac{I_1 I_2}{I_1^2 + I_2^2} = \frac{2}{\frac{I_1}{I_2} + \frac{I_2}{I_1}}$$

Dans l'espace des partitions en K classes , on a

$$\cos \varepsilon = \frac{I_1 + I_2 + \dots + I_k}{\sqrt{K} \sqrt{I_1^2 + I_2^2 + \dots + I_k^2}}$$

$$\cos^2 \epsilon = \frac{\sum_{i=1}^k I_i^2 + 2 \sum_{\substack{i=1 \\ i \neq j}}^k I_i I_j}{K \sum_{i=1}^K I_i^2} =$$

$$= \frac{1}{K} + 2 \frac{\sum_{\substack{i=1 \\ i \neq j}}^K I_i I_j}{K \sum_{i=1}^K I_i^2} \Rightarrow 2 \cos^2 \epsilon - 1 = \frac{2}{K} - 1 + \frac{4}{K} \frac{\sum_{\substack{i=1 \\ i \neq j}}^K I_i I_j}{\sum_{i=1}^K I_i^2}$$

Le critère à maximiser dans ce cas est donc :

$$\alpha(I_i) = \frac{2 - K}{K} + \frac{4}{K} \frac{\sum_{\substack{i=1 \\ i \neq j}}^k I_i I_j}{\sum_{i=1}^K I_i^2}$$

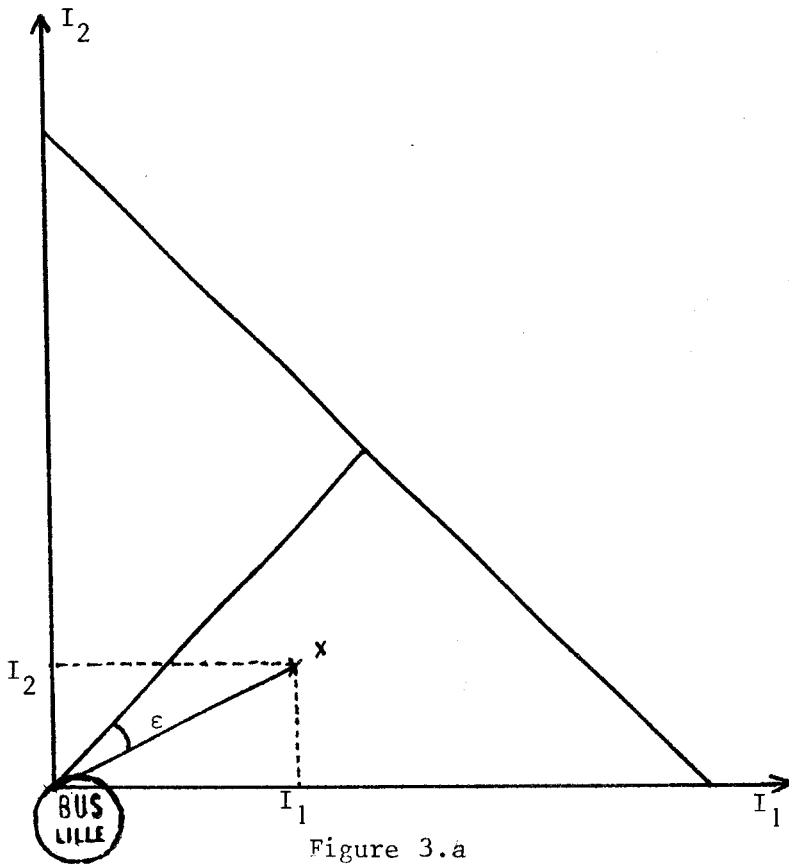


Figure 3.a

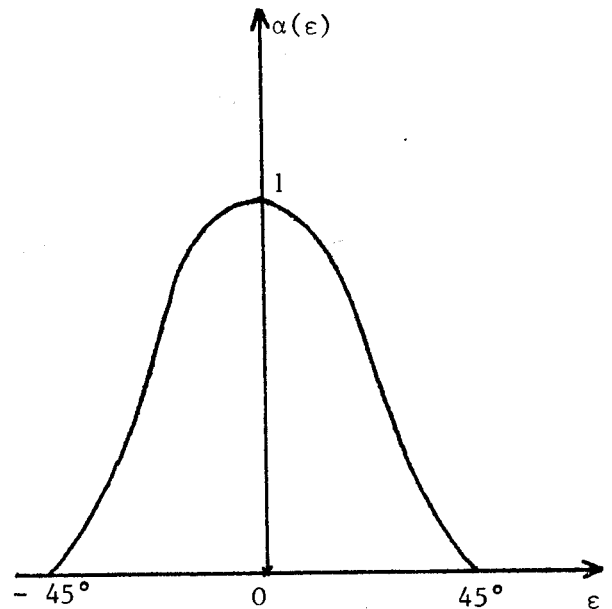


Figure 3.b

Remarque : la solution n'est pas unique et dépend du choix du coefficient  $\alpha(I_i)$ . Néanmoins, cette méthode a l'avantage sur la précédente de ne pas avoir à calculer

$$\rho = \max_{i=1} \sum I_i, \text{ et d'avoir } x \text{ unique pour } \alpha \text{ donné.}$$

B I B L I O G R A P H I E

---

( A S H 65 ) : R.ASH

" Information Theory"

John Wiley and Sons , 1965

( A S H B 65 ) : W.R ASHBY

"Measuring the internal Informational Exchange  
in a system"

Cybernetica (Namur - Belgique ), Vol 8 n° I pp 5 - 22 . 1965

( BART 79 ) = J.P BARTHELEMY

Propriétés métriques des ensembles ordonnés.

Thèse de doctorat d'état .Besançon 1979

( BENZ 80 ) J.P BENZECRI

"L'analyse des données "

(2 tomes ) Paris Dunod 1980

(BERT 75 ) P.BERTIER - J.M BOUROCHE

"L'analyse des données multidimensionnelles"

Presses universitaires de France - 1975

( CAI 76 ) : F CAILLIEZ - J.P PAGES

"Introduction à l'analyse des données "

SMASH Paris 1976

( CAV 65 ) A.W.F EDWARDS - L.I CAVALLI - SFORZA

"A Method for cluster analysis"

Biométrie , Vol 21 n° 2 - 1965

( CONA68 ) R.C CONANT

" Information transfer in complex systems with application to regulation"

PHD . Thesis, University of Illinois, U.S.A, Janvier 1968.

( CONA69 ) : R.C CONANT

"The information transfer required in regulatory process"

I.E.E.E Trans on systems, science and Cybernetics.

Vol SSC -5, n° 4, PP, 334 - 338 Octobre 1969



( CONA72 ) : R.C CONANT

" Detecting Subsystems of a complex system"

I.E.E.E trans on systems, Man and Cybernetics .Vol SM C-2, n°4 , PP  
550 - 553 , Septembre 1972.

( DID 72 ) : E.DIDAY

"Nouvelles méthodes et nouveaux concepts en classification automatique  
et reconnaissance des formes "

Thèse de doctorat es-sciences mathématiques, Univ Paris VI 4 Décembre 1972.

( DUSS 80 ) : A . DUCHAUSSOY

" Generalized information theory on some ordered sets and structure of  
systems"

14 th annual Conf on information sciences en systems,Princeton University  
Mars.

( DUF 75 ) : J.DUFOUR - G.GILLES - C.FOULARS

"analyse structurale at partition de s systèmes dynamiques complexes à  
l'aide de la théorie de l'information "

" Etude dell'influence du nombre des classes"

CRAS. Tome 282 Mars 1975

(FOR 65 ) E.W FORGY

"Cluster analysis of multivariate data:

Efficiency versus interpretability of classifications"

Biométric Soc - Meetings, Riverside, California.

Résumé dans Biométries Vol 2I n° 3 Septembre 1965

(GIL 54 ) W.J. MC GILL

'Multivariate information transmission"

Psychométrieka , Vol 19 ;n°2 ;PP 97 - 116 - June 1954

( GUIA 79 ) S.GUIASU and GREISCHER

"Some remarks on entropic distance, entropique measure of connexion and  
hamming distance"

RAIRO informatique théorique Vol 13 n° 4 P 395 à 407,1979

( HART 28 ) R.V.L HARTLEY

" Transmission of information "

Bell system technical Journal 1928,7 , P 535

(HIM 73 ) D.HIMMELBLAU

Morphology of decomposition in "Decomposition of large scale problems"

North Holland 1973

( HUB 74 ) L.HUBERT

"Some application of graph theory to clustering"

Psychometrika VOL 39 - N° 3 - Sept 1974

( JAC 08 ) P.JACCARD

"Nouvelles recherches sur la distribution florale"

Bull.soc Vand . Sci Nat,T 44,PP 223 à 270 (1908)

(JAN 80 ) JANICKE BIESS

" The effectivity of algorithms for the decomposition of large scale systems"

Hungarian Journal of Ind Chem ,Vol 8 ,1980,PP 45 - 58

(JOHN 67 ) : S.C JOHNSON

" Hierarchical clustering schemas"

Psychometrika - Vol 32 - n° 3 ; Sept 1967

(KAM 67 ) : J.KAMPE DE FERIET et B.FORTE

" Information et probabilité"

Compte rendu Acad Sciences Paris 269, Série A . 1967

P 110 - 114 ; P 142 - 146 ; P 350 - 353

( KAUF 68 ) : A.KAUFMANN

"Introduction à la combinatoire en vue des applications"

DUNOD PARIS 1968

( KICK 78 ) W.J.M KICKERT , J.W.M BERTRAN , J.PRAAGMAN

" Some comments on cybernetical and control "

I.E.E.E. trans on systems , Man and Cybernetics .

Vol SMC - 8 n° II ;PP 805 - 809 Novembre 78

( LANG 73 ) C.LANGRAND

"Information généralisée .Estiamtion et sélection"

Thèse d'état es-sciences Mathématiques.

Université de Lille I n° D'Ordre 271,7 MARS 1973

( LARM 75 \ P.LARMINANT - Y THOMAS

" Automatique des systèmes linéaires"

Tome I : Signaux et systèmes Flammarion sciences 1975

( LED \ LEDET - HIMMELBLAU

"Advances in chemical engineering 1979"

(LER 70 \ ) : I.C LERMAN

" Les bases de la classification automatique"

Paris - Gauthier - Villars 1970

( LOSF 74 \ ) J. LOSFELD

"Information fournie par un ensemble d'informateurs et application aux questionnaires et à l'analyse des données "

Thèse d'état en sciences Mathématiques, Université de Lille I -N° D'Ordre 304  
Juin 1974

(LOP 77 \ ) 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  
N° D'Ordre 1998 - Juin 1977

(MAC 67 \ ) J.B MAC QUEEN

"Some methods for classification and analysis of multivariate observations"

5 th Berkeley Symposium on mathematical Statistics and probability.  
Vol I - N° I Berkeley University of California Press - 1967

(MILL 63 \ ) G.A MILLER

" What is information Measurement"

American Psychologist Vol 8 n°2 - PP 50 - 51,1963

( NYQU 24 \ ) H.NYQUIST

"Certain Factors affecting telegraph Speed"

Bell System Technical Journal Vol 3 - P 324 - 1924

( OCHIA 57 \ ) A.OCHIAI

" Zoographic Studies on the solevid fishes found in Japan and its neighbouring regions"

Bull - Jap - Soc - Sci - Fish - T . 22,1957

( PICAR 72 ) C.F PICARD

"Graphes et questionnaires"

GAUTHIER - VILLARS - EDITEUR collection "programmation" 1972

(PORT 76 ) B.PORTER

" Requisite variety in the system end control science"

International Journal of General Systems Vol 2 - PP 225 - 229 - 1976

(PRIM 57 ) R.C PRIM

" Shortest connection networks and some generalizations"

The Bell system Technical Journal Nov 1957

( RAM 40 ) C.RAMAMOORTY

"Connectivity consideration of graph representing discret sequential systems"

I.E.E.E. Trans On Elect comp October ,1965 - PP 724 - 727

( RAN 71 ) : W.M RAND

" Objective criteria for the evaluation of clustering methods"

J .Am Stat Assoc Vol 66 - n° 336 - Dec 1971

( REGN 65 ) S REGNIER

"Sur quelques aspects mathématiques des problèmes de la classification automatique"

I.C.C BU 11 Vol 4 - P 175 - 191 - 1965

( RLCH 75 ) M.RICHETIN

"Analyse structurale des systèmes complexes en vue d'une commande hiérarchisée"

Thèse d'état Université Paul Sabatier Toulouse n° d'ordre 674 - Juillet 75

( RINK 73 ) R.E RINK

" Information theoretic methods for modelling and analysing large systems"

5th IFIP Conference on optimisation techniques : Rome 1973

Publié par lecture Notes on computer science N° 3

Allemagne 1973

( ROM 82 ) H.ROMBAUT

"determination des relations fonctionnelles dans un système de variables par l'analyse informationnelle des données"

Mémoire de D.E.A. - Université Lille 1 - Juin 1982.

(ROUX 68 ) M.ROUX

"Un algorithme pour construire une hiérarchie particulière"

Thèse de doctorat 3ème cycle - Université Paris VI - 1968

(ROY 70 ) B.ROY

"Algebre moderne et théorie des graphes "

DUNOD - PARIS 1970

(SHAN 48 ) C.E SHANNON

"A mathematical theory of communication "

Bell system Technical Journal, Vol 27 - PP 379 à 423 et 623 à 656 - 1948

( SHAN 49 ) C.E SHANNON

"Communication in the presence of noise "

Proc I.R.E - Vol 37 - P 10 - 1949

(SIM 62 ) H.A SIMON

" The Architecture of complexity"

Proc of the American Philosophical Society -Vol 106 - n° 6 -DECEMBRE 1962

( SIB 72 ) R.SIBSON

"Slink : an optimally efficient algorithm for the simple link cluster method"

The computer Journal - Volume 16 n° 1 - 1972

(SOK 58 ) R.R.SOKAL - C.D MICHENER

"A statistical method for evaluating systematic relation ships "

U niv Kansas SCI - Bull ;38 - 1958

(STAR 81 ) M.STAROSWIECKI - V.TORO

"Structural analysis of complex systems by means of canonical analysis"

VII International Conference on systems science  
Wroclaw ;Pologne ;15 - 18 Septembre 1981

(STEW 65 ) D.V STEWARD

" Partitionning and tearing systems of equation"

JournalSIAM of Numerical Analysis Série B - Vol 2 - N° 2 - 1965

(TARJ 75 ) R.TARJAN

"EFFICIENCY of good but not linear set union algorithm"

Journ ACM ;Vol 22 - N° 2 - 1975 - PP 215 - 225

( TORO 81 ) V.TORO - M.STAROSWIECKI

"Méthodes heuristiques et décomposition optimale des systèmes complexes"

Congrès AFCET Automatique I98I  
Nantes , 27 - 29 Octobre I98I

(TORO 82 ) V.TORO

"Contribution à l'analyse structurale de systèmes complexes à l'aide de l'entropie et ses généralisations"

Thèse de 3ème cycle Université de LILLE I - N° D'ordre 955 - 8 MARS I982

(VOL 82 ) M. VOLLE

"Analyse des données "  
2ème édition economics I982

(WAR 63 ) J.H WARD

"Hierarchical grouping to optimize an objective function"

J.Am Stat.Assoc . Vol 58 - MARS I963

( WIEN 48 ) N.WIENER

"Cybernetics"

The Technology Press and John Wiley and Sons I948

( WIEN 49 ) N.WIENER

" Extrapolation, Interpolation and Smoothing of stationary time series"

The Technology Press and John Wiley and Sons -I949





### RESUME

Le travail présenté dans cette thèse a pour objet, l'analyse structurale des systèmes complexes.

Dans un premier temps, nous avons rappelé quelques méthodes de l'analyse des données et montré comment elles peuvent être appliquées au domaine qui nous intéresse. On donne ensuite quelques définitions et propriétés de la théorie de l'information, outil essentiel de notre travail.

La deuxième partie de ce travail développe quelques méthodes d'explication d'une variable à partir d'un groupe de variables. On montre que l'analyse des couplages par paires de variables est insuffisante et on fait une analyse multidimensionnelle des relations qui débouche sur la construction d'un graphe d'interconnexion (les indices de couplages étant issus de la théorie de l'information). On étudie l'influence des perturbations sur la mise en évidence des relations fonctionnelles, ainsi que celle du nombre de modalités du vecteur explicatif.

Dans la troisième partie, on examine quelques méthodes de décomposition de systèmes. Plusieurs algorithmes de décomposition à partir des graphes sont étudiés. On expose aussi différents algorithmes de décomposition hiérarchique et à centre variables et on propose plusieurs indices de similarité permettant leur réalisation.

Dans la quatrième partie, on propose une approche très simple basée sur la méthode des nuées dynamiques. Plusieurs variantes sont étudiées et comparées. Les résultats sont interprétés pour chacune d'elles.

Mots clés : Systèmes complexes / Analyse structurale / Théorie de l'information /  
Décomposition / Nuées dynamiques.

12/10/77