

50376
1990
45

N° d'ordre: 1442

68208

50376
1990
45

Année: 1990

THESE

présentée à

L'UNIVERSITE DES SCIENCES ET TECHNIQUES DE LILLE FLANDRES ARTOIS

en vue de l'obtention du titre de

DOCTEUR DE 3ème CYCLE

Spécialité : Automatique & Informatique Industrielle

par

Abdelhamid HAOUAT



**ANALYSE DES SYSTEMES COMPLEXES:
APPROCHES INFORMATIONNELLES DE DECOMPOSITION,
D'EVALUATION DE LA QUALITE D'UN MODELE ET DE
DETERMINATION DE L'ORDRE D'UN SYSTEME.**

Thèse soutenue le 09 Février 1990 devant la commission d'examen

Membres du Jury : MM.

P. VIDAL

Président

M. STAROSWIECKI

Rapporteur

J.G POSTAIRE

Examinateur



AVANT-PROPOS

Le travail présenté dans ce mémoire a été effectué dans le centre d'Automatique et de Productique de l'Université des Sciences et Techniques de Lille Flandre Artois.

Je tiens à remercier 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 tiens à exprimer mes vifs remerciements à Monsieur Marcel STAROSWIECKI, professeur à l'Université de Lille I, qui a dirigé l'ensemble de mes travaux. Ses conseils ainsi que sa disponibilité, ont été pour moi, une source permanente dans la poursuite de mes recherches.

Je suis très sensible à l'honneur que me fait Monsieur Jaques-Gérard POSTAIRE, professeur à l'Université de Lille I, de participer au jury. Je le remercie très sincèrement.

Je tiens à remercier vivement celles et ceux qui, de loin ou de près, dont l'apport tant moral que matériel, a contribué à la réalisation de cette thèse.

ANALYSE DES SYSTEMES COMPLEXES :
APPROCHES INFORMATIONNELLES DE DECOMPOSITION,
D'EVALUATION DE LA QUALITE D'UN MODELE ET DE
DETERMINATION DE L'ORDRE D'UN SYSTEME.

Ames parents

"Toute pensée exige que l'on prenne une chose pour une autre."

Paul Valéry

Introduction à la méthode de Léonard de Vinci. Gallimard

**Today information theory provides the only existing narrow
bridge between the two different worlds of chaotic and organised systems.**

Prof.L. Levitin

*IEEE. Info Theory Group Newsletter
September 1988.*

SOMMAIRE

**Chapitre I : DESCRIPTION D'UN SYSTEME COMPLEXE . ELEMENTS EN
THEORIE DE L'INFORMATION .**

I. Introduction.	5
II. Description d'un système complexe.	7
II.1. Système-Source.	7
II.2. Génération des données .	8
II.3 Estimation des probabilités.	10
III. Théorie de l'Information.	13
III.1 Approche de Shannon.	13
III.1.1 Définitions.	13
III.1.2 Propriétés.	16
III.1.2.1 Cas discret.	16
III.1.2.2 Cas continu.	19
III.1.2.3. Représentation graphique.	20
III.2. Approche de Kullback.	22
III.2.1 Définitions.	22
III.2.2. Propriétés.	24
III.3 Autres approches.	31
III.3.1 Mesures d'entropie paramétriques.	31
III.3.2 Mesures de divergence paramétriques.	33
III.3.3 Entropie du Max.	36
III.3.4 Mesures d'inexactitude.	37
IV. Conclusion.	39

Chapitre II. ANALYSE DES SYSTEMES STATIQUES PAR LA T I.

I. Introduction.	40
II. Analyse de couplage statique par la T I.	42
II.1. Indices de couplage statique.	42
II.2. Relations fondamentales.	45
III. Décomposition d'un système statique.	47
III.1. Approches binaires.	47
III.2. Approches multidimensionnelles.	49
III.2.1. Méthodes hiérarchiques.	50
III.2.1.1. Algorithme hiérarchique ascendant.	50
III.2.1.2. Algorithme hiérarchique descendant.	51
III.2.1.3. Algorithme des approximations successives.	51
III.2.2. Méthode de programmation dynamique.	53
III.2.3. Optimalité des méthodes.	54
IV. Approche d'explication.	56
IV.1. Le concept de modélisabilité.	56
IV.1.1. Définitions.	56
IV.1.2. Remarques.	57
IV.2. Formulation de l'approche d'explication.	58
IV.3. Cas déterministe.	59
IV.4. Cas non déterministe.	62
IV.4.1. Séparation des modalités mal expliquées.	63
IV.4.2. Recherche de variables explicatives supplémentaires.	71
V. Conclusion.	73

Chapitre IV : EVALUATION DE LA QUALITE D'UN MODELE PAR LA T I.

I. Introduction.	110
II. Variables et partitions.	112
II.1. Partition associée à une variable.	112
II.1.1. Définition.	112
II.1.2. Quelques propriétés.	113
II.2. Partition associée à une relation: $f(X)=Y$.	114
II.2.1 Partition de l'ensemble d'observation .	115
II.2.2. Partition de l'ensemble des modalités de Y.	116
II.2.2. Partition de l'ensemble des modalités de X.	119
II.3. Pourquoi s'intéresse-t-on aux partitions induites?	120
III. Qualité d'un modèle	121
III.1 Indice informationnel de Shannon.	121
III.2 Indice informationnel de Kullback.	123
III.3. Autres indices.	126
III.4. Cas de variables continues.	127
IV. Application au modèle linéaire.	128
IV.1. Détermination des partitions induites.	132
IV.2. Performance d'un modèle linéaire.	133
IV.3. Exemple d'application.	135
V. Conclusion.	140

NOTATIONS

$\Sigma = \{ X_1 , X_2 , \dots , X_N \}$, l'ensemble des variables du système.

M_X , l'ensemble des modalités d'une variable X de Σ .

$\Omega = \{ w_p , w_{p+1} , \dots , w_L \}$, l'ensemble d'observations.

$\mathcal{P}(\Omega)$, l'ensemble des parties de Ω .

$P(\Omega)$, une partition de Ω .

$\mathbb{P}(\Omega)$, l'ensemble des partitions de Ω .

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

f relation définie de M_X vers M_Y telle que $f(X)=Y$

$P_f(\Omega)$, partition de Ω induite par la relation f .

p_i , probabilité marginale.

p_{ij} , probabilité conjointe.

$p_{j/i}$, probabilité conditionnelle.

$H(X)$, entropie de X .

$H(Y/X)$, entropie résiduelle de Y sachant X .

$I(X:Y)$, traninformation entre X et Y .

y^p , variable Y considérée sur l' horizon d'observation :
[k, k-1, ..., k-p].

\bar{y}^p , variable Y considérée sur l' horizon d'observation :
[k-1, ..., k-p].

\bar{c}^p , indice informationnel de couplage dynamique.

INTRODUCTION GENERALE

Les travaux présentés dans cette thèse se situent dans le cadre de l'analyse des systèmes complexes au moyen de la Théorie de l'Information, particulièrement autour des trois thèmes suivants :

L'analyse structurale et la décomposition des systèmes complexes.

L'évaluation de la qualité d'un modèle.

La détermination de l'ordre d'un système.

Au premier chapitre, nous nous intéressons à la théorie de l'information (T I): Nous présentons d'abord les principales définitions et propriétés de l'entropie et de la transinformation de Shannon. De telles mesures, par leurs propriétés et leurs interprétations, permettent de disposer d'indices intéressants pour évaluer les couplages dans un système.

Nous développons ensuite les définitions et propriétés des mesures de divergence de Kullback. Nous montrerons que celles-ci présentent des propriétés similaires à celles de l'information de Shannon, mais ne vérifient pas la propriété de sous-additivité.

Nous présentons enfin des mesures d'information du type paramétrique telles l'entropie d'ordre α , l'entropie du type β , la divergence d'ordre α et la divergence du type β . Nous montrerons que celles-ci ne possèdent pas toutes la propriété de sous-additivité et ne permettent pas en général d'obtenir des conditions nécessaires et suffisantes pour évaluer les couplages entre les variables d'un système.

Le deuxième chapitre est consacré à l'analyse structurale et la décomposition des systèmes statiques, par le biais de l'information de Shannon. Nous étudierons des indices issus de la T I, et nous soulignerons leurs limites respectives. Ces indices, appliqués aux variables prises deux à deux, conduisent à une matrice de couplage dont l'analyse permet une décomposition du système en sous-systèmes faiblement couplés. Une telle approche peut s'avérer non satisfaisante pour les systèmes où une grande part de l'information est échangée entre groupes de plus de deux variables.

Nous présenterons ensuite des approches multidimensionnelles. Celles-ci examinent l'information échangée entre groupes de variables et permettent d'aboutir à des partitions du système à travers des algorithmes du type hiérarchique ou de programmation dynamique. Cependant, l'application de ces approches peut conduire à des résultats insuffisants dans le cas de systèmes très complexes, puisque leur mise en oeuvre nécessite un volume de calcul important.

Nous étudierons enfin une approche s'appuyant sur le concept de modélisabilité. Cette approche suppose que l'ensemble des variables du système est constitué de deux groupes de variables : un groupe de variables à expliquer et un groupe de variables explicatives. Elle consiste à déterminer pour chaque variable à expliquer, un sous-ensemble explicatif optimal dans le sens où l'information est maximale. Une telle approche se trouve limitée dans le cas de systèmes soumis à des perturbations d'origines diverses. Dans ces conditions, on considère une procédure qui consiste à séparer pour chaque variable à expliquer, les modalités bien expliquées des modalités mal expliquées à cause du bruit, ou dont l'explication nécessite l'introduction de variables supplémentaires. Les sous-

ensembles explicatifs optimaux sont alors recherchés à partir des variables initialement retenues dans le système et les variables supplémentaires.

Le troisième chapitre sera consacré à l'analyse structurale des systèmes dynamiques et à la détermination de l'ordre d'un système. Dans un premier temps, nous étudierons à partir de la T I, des coefficients de couplage dynamique qui prennent en compte l'information véhiculée entre les variables au cours de l'évolution du système.

Nous considérerons ces coefficients dynamiques dans le cadre de la décomposition des systèmes dynamiques. Par ailleurs, nous verrons que de tels coefficients permettent d'indiquer dans quelle mesure un système dynamique est modélisable suivant un modèle d'ordre fixé au préalable.

Dans la dernière partie de ce chapitre, nous présenterons quelques approches informationnelles classiques pour la détermination de l'ordre d'un système. Ces approches s'appuient sur la donnée d'un modèle dont il faut identifier les paramètres. Nous proposerons ensuite une approche basée sur l'utilisation d'un coefficient informationnel dynamique. Celle-ci conduit à la détermination de l'ordre d'un système .

Au chapitre IV, nous nous intéresserons à l'évaluation de la qualité d'un modèle statique du type entrée-sortie au moyen de la T.I. Pour cela, on définit ce qu'est une partition induite par un tel modèle et on étudie ses propriétés. Une telle partition peut être considérée similaire au modèle dans le sens où l'information extraite du modèle est celle contenue dans cette partition. Nous proposerons alors deux approches pour apprécier la qualité d'un modèle :

--> Par l'évaluation de l'information extraite de la partition-modèle au sujet du système.

--> Par la mesure de la divergence entre les données fournies par le modèle et celles du système.

A partir de ces approches, on peut également évaluer la performance d'un modèle par rapport à un autre.

Enfin , nous étudierons la qualité d'un modèle linéaire à travers un exemple numérique.

CHAPITRE I :

DESCRIPTION D'UN SYSTEME COMPLEXE.

ELEMENTS EN THEORIE DE L'INFORMATION.

I. INTRODUCTION.

Les domaines d'application de "l'Approche Système" ne cessent de s'élargir pour inclure aujourd'hui des systèmes divers, technologiques, écologiques, biologiques, socio-économiques... Parallèlement, la complexité des systèmes pris en compte s'accroît et pose de sérieux problèmes d'analyse dont la résolution nécessite la mise en oeuvre d'outils et de méthodes de plus en plus fiables et performantes.

Ainsi toute approche d'analyse vise en priorité à rendre le système moins complexe et à mettre à jour son organisation interne.

Parmi ces approches, nous nous intéressons aux méthodes d'analyse structurale; celles-ci font appel à l'analyse des données, la théorie de l'information, la théorie des graphes, etc... De telles méthodes apportent une contribution intéressante pour l'appréhension des systèmes complexes.

L'application de ces approches suggère une méthodologie qui, à partir d'un niveau de connaissance du système, conduit à un niveau de connaissance où le système est mieux appréhendé. Ces niveaux s'organisent d'après G.Klir [Klir 69, 75, 76, 77] comme suit:

Niveau 0 ou Système-Source:

Ce niveau constitue une phase initiale dans laquelle le système est décrit par un ensemble de variables jugées

pertinentes par l'analyste. Chaque variable est connue par ses modalités ou états.

Niveau 1 ou Système-Données:

Le système est connu par une succession de modalités prises par les variables sur un ensemble d'observations. A ce niveau, le système est décrit sous forme d'un tableau de données.

Niveau 2 ou Système-Générateur:

Ce niveau constitue la phase où le système est défini par un ensemble de relations génératrices qui peuvent être sous forme de lois de probabilités et/ou de relations fonctionnelles de nature diverses(algébriques, analytiques, différentielles, etc...).

Niveau 3 ou Système-Structure:

A ce niveau, le système est composé d'un ensemble de sous-systèmes qui interagissent entre eux et susceptibles de véhiculer un flux important d'information.

Compte tenu de ces différents niveaux, toute approche d'analyse peut être considérée comme une suite de déplacements d'un niveau vers un autre. Dans ce travail, les approches que nous développerons ultérieurement se situent aux niveaux 2 et 3 et supposent que le système est connu aux niveaux 0 et 1.

II. DESCRIPTION D'UN SYSTEME COMPLEXE.

II.1 Système source.

Un système est initialement décrit par un ensemble de variables et/ou paramètres jugés pertinents. Il peut s'agir de variables d'entrée, de sortie, de commande, d'état, ou de paramètres tels la constante de temps, le coefficient d'amortissement, l'ordre présumé du système etc... Les variables retenues pour la description du système doivent mettre en évidence son fonctionnement et caractériser au mieux son comportement.

Selon la nature du système, certaines variables sont accessibles, mesurées ou non, d'autres inaccessibles, donc non mesurables.

Aucune restriction n'est considérée au sujet de la nature des variables. Elles peuvent être dépendantes ou non du temps, aléatoires ou non, qualitatives ou quantitatives, logiques ou floues, à ensemble de modalités structuré ou non.

On note par: $\Sigma = \{ X_1 , X_2 , \dots , X_N \}$,

l'ensemble des variables du système (supposées ordonnées suivant une numérotation fixe);

et par: $M_X = \{ x_1 , x_2 , \dots , x_n \}$

un ensemble fini de modalités ou états de chaque variable X de Σ .

II.2 Génération des données.

Les modalités des variables sont relevées à partir d'un ensemble d'observations (population d'individus, horizon de temps, ...). Chaque variable est alors connue par la succession des modalités qu'elle prend pour différentes observations.

Pour les systèmes dynamiques, la détermination d'une variable à un instant donné requiert en général son observation à un certain nombre d'instantes antérieurs. Une variable X_i se présente alors comme application d'un ensemble d'observation vers un ensemble de vecteurs de ses modalités :

Soit $\Omega = \{ w_p, w_{p+1}, \dots, w_L \}$ où p est le nombre d'observation nécessaire pour la description de l'évolution future du système.

On a :

$$X_i : \Omega \longrightarrow M_{X_i} * \dots * M_{X_i} \quad (p \text{ fois}).$$

$$w_j \longrightarrow X_i^*(w_j) = (X_i(w_j), X_i(w_{j-1}), \dots, X_i(w_{j-(p-1)})).$$

Compte tenu de cette formulation, un système dynamique peut être décrit par le tableau des données suivant :

$\Omega \setminus \Sigma$	X_1	\dots	X_i	\dots	X_N
w_p			.		
.			.		
.			.		
w_j	\dots		$X_i^*(w_j)$	\dots	
.			.		
.			.		
.			.		
w_L					

Fig.I.1 Tableau des données du système.

Pour les systèmes statiques, chaque variable est connue à un instant donné par la modalité qu'elle prend à cet instant. Elle se présente comme application d'un ensemble d'observations vers l'ensemble de ses modalités.

Comme pour le cas dynamique, un système statique peut être décrit par un tableau similaire où l'élément $X_i^*(w_j)$ est la modalité prise par X_i à l'observation w_j et p égal à 1.

REMARQUES :

La description des données du système dépend d'un certain nombre de paramètres. Le choix de ces paramètres est d'une grande importance :

- Le nombre n_i des modalités de X_i , le nombre d'observations, la période d'échantillonnage T de l'horizon de temps, sont choisis de façon à estimer au mieux les variables sous forme discrète (cette hypothèse obéit à des raisons pratiques de traitement des données) et ce, sans perte excessive d'information. A titre indicatif, on peut choisir une période d'échantillonnage appropriée à partir du théorème de Shannon, compte-tenu d'hypothèses de stationnarité et d'ergodicité [Ash 65].

Signalons qu'en général, il n'existe pas de règles systématiques efficaces permettant un choix optimal de ces paramètres. Souvent le praticien opère pour cela de façon empirique.

II.3 Estimation des probabilités.

Une mesure de probabilité sur un ensemble fini se présente sous sa forme générale [Parzen 60] comme une fonction p telle que:

$$p : \begin{array}{l} \mathcal{P}(\Omega) \text{ ----> } \mathbb{R}. \\ A \text{ ----> } p(A) \end{array}$$

vérifiant:

- pour tout $A \in \mathcal{P}(\Omega)$, $p(A) \geq 0$. $\mathcal{P}(\Omega)$ étant l'ensemble des parties de Ω .

- pour tout $A_i \in \mathcal{P}(\Omega)$, $A_j \in \mathcal{P}(\Omega)$,

tels que $A_i \cap A_j = \emptyset$ pour $i \neq j$; $i, j = 1, 2, \dots, n$;

$$\text{On a : } p\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n p(A_i).$$

$$- p(\Omega) = 1.$$

Du fait de la nature du système lui-même, ou de la façon dont il a été excité, les variables du système se présentent sous forme aléatoire.

Ainsi, pour $X: \Omega \rightarrow M_X$, variable quelconque de Σ (vectorielle ou non), on définit la probabilité d'occurrence de l'événement $X = x$, notée $p(X = x)$. Une telle probabilité peut être estimée au moyen des fréquences relatives évaluées sur Ω . On considérera:

$$p(X = x) = \frac{|\{ w \in \Omega / X(w) = x \}|}{|\Omega|}$$

où $|\Omega| = L$ si X est une variable statique,
et $|\Omega| = L - p + 1$ si X est une variable dynamique.

Notons que le choix du cardinal de $\Omega, |\Omega|$, joue un rôle important dans l'estimation de $p(X=x)$ par les fréquences relatives; en effet, un échantillon de faible taille peut conduire à des résultats peu significatifs. C'est pourquoi celui-ci doit être de taille suffisamment grande. Conant [Conant 72] propose une limite inférieure:

$$|\Omega| \gg \sum_{i=1}^N \pi (|M_{X_i}| - 1).$$

III. THEORIE DE L'INFORMATION.

La théorie de l'information, par ses diverses approches ne cesse d'investir des domaines liés à l'analyse des systèmes, tels l'analyse structurale, la modélisation, l'identification, le diagnostic etc... Différents outils informationnels ont été développés principalement à partir des approches de Shannon, de Kullback et de leurs généralisations.

Conçue à l'origine pour la résolution de problèmes pratiques de télécommunication, la théorie de l'information de Shannon [Shannon 48] fournit des outils intéressants pour l'étude des processus les plus divers, en particulier pour l'analyse structurale des systèmes complexes. En effet, elle permet de mesurer l'intensité des liaisons (linéaires ou non) entre les variables ou groupes de variables de nature quelconque.

L'approche informationnelle de Kullback [Kullback 59] fut, à l'origine, le résultat de développements en Statistique, qui ont permis de définir des critères probabilistes. Ces critères s'appliquent particulièrement à des problèmes de test d'hypothèses, de séparabilité, de reconnaissance des formes, de validation de modèles, etc...

Dans ce paragraphe, nous décrivons les principales définitions et propriétés de ces deux approches, et nous donnerons un aperçu sur certaines de leurs généralisations.

III.1. Approche de Shannon.

Dans la suite, nous considérons les variables du système munies de distributions de probabilités. Celles ci peuvent être extraites d'un tableau des données du système.

III.1.1. Définitions.

Entropie d'une variable discrète.

Soient X et Y deux variables discrètes à modalités respectivement dans :

$$M_X = \{x_1, x_2, \dots, x_n\} \text{ et } M_Y = \{Y_1, Y_2, \dots, Y_m\}.$$

On suppose que X et Y sont munies de distributions de probabilité : $p_i = \text{pr}(X = x_i)$, $q_j = \text{pr}(Y = Y_j)$,
 $p_{ij} = \text{pr}(X=x_i, Y=Y_j)$ pour $i = 1, 2, \dots, n$; $j = 1, 2, \dots, m$.

On définit l'entropie de X par:

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

L'entropie d'une variable est une mesure de son indétermination à priori. C'est une quantité qui ne dépend pas des modalités de la variable; en conséquence, elle s'applique aux variables quantitatives, qualitatives, à ensemble de modalités structuré ou non.

Pour deux variables X et Y , on définit l'entropie conjointe de X , Y et l'entropie conditionnelle de Y sachant X , respectivement par:

$$H(X,Y) = - \sum_{i=1}^n \sum_{j=1}^m p_{ij} \log p_{ij} .$$

$$H(Y/X) = \sum_{i=1}^n p_i H(Y/x_i); \text{ où } H(Y/x_i) = - \sum_{j=1}^m p_{j/i} \log p_{j/i} .$$

où $p_{j/i} = \text{pr}(X=x_i / Y=y_j)$

Entropie d'une partition.

On considère $P(\Omega)$, l'ensemble des parties de l'espace d'observation du système, muni d'une mesure de probabilité, et $P(\Omega)$ une partition de Ω :

$$P(\Omega) = \{P_1, P_2, \dots, P_n\} .$$

On définit l'entropie de $P(\Omega)$ par:

$$H(P) = - \sum_{i=1}^n p_r(P(\Omega)=P_i) \log p_r(P(\Omega)=P_i) .$$

Pour deux partitions $P(\Omega)$ et $Q(\Omega)$, l'entropie conjointe et l'entropie conditionnelle sont définies de façon similaire au cas de deux variables.

Entropie d'une variable continue.

Soit X une variable aléatoire continue qui prend ses valeurs dans M_X (sous-ensemble de R^n) suivant une densité de probabilité $p(x)$.

On définit l'entropie de X par:

$$H(X) = - \int_{M_X} p(x) \log p(x) dx.$$

Pour deux variables continues X et Y , on définit l'entropie conjointe de X , Y et l'entropie conditionnelle de Y sachant X (dite entropie résiduelle), respectivement par:

$$H(X, Y) = - \int_{M_{XY}} p(x, y) \log p(x, y) dx dy$$

et

$$H(Y/X) = \int_{M_X} p(x) H(Y/x) dx$$

où

$$H(Y/x) = - \int_{M_Y} p(y/x) \log p(y/x) dy.$$

Transinformation entre deux variables.

Soient X et Y deux variables aléatoires continues ou discrètes.

On définit la transinformation entre X et Y par :

$$I(X:Y) = H(X) + H(Y) - H(X, Y).$$

$I(X:Y)$ mesure l'information transmise entre les variables X et Y .

De même, si $P(\Omega)$ et $Q(\Omega)$ désignent deux partitions de Ω , la transinformation entre $P(\Omega)$ et $Q(\Omega)$ est définie par:

$$I(P:Q) = H(P) + H(Q) - H(P, Q).$$

Transinformation conditionnelle.

Soient X , Y et Z trois variables aléatoires continues ou discrètes. On définit la transinformation conditionnelle entre X et Y sachant Z par :

$$I_Z(X:Y) = H(X/Z) + H(Y/Z) - H(X,Y/Z).$$

$I_Z(X:Y)$ est une mesure de l'information transmise entre X et Y connaissant les valeurs prises par Z .

III.1.2. Propriétés : [Dufour 79] [Toro 82].

III.1.2.1. Cas discret.

En discret, on montre que:

P1: $0 \leq H(X) \leq \log n$. De plus, on a:

$H(X) = 0$ si et seulement si X est constante.

$H(X) = \log n$ si et seulement si $p(X=x) = 1/n$ pour tout $x \in M_X$.

P2: $H(X,Y) \leq H(X) + H(Y)$ (Sous-additivité).

De plus, il y a égalité si et seulement si X et Y sont statistiquement indépendantes.

P3: $H(X,Y) = H(Y) + H(X/Y) = H(X) + H(Y/X)$.

De façon générale, on a:

$$H(X_1, X_2, \dots, X_N) = H(X_1) + H(X_2/X_1) + H(X_3/X_1, X_2) + \\ \dots + H(X_N/X_1, X_2, \dots, X_{N-1}).$$

P4: $H(Y/X) \leq H(Y).$

De plus, il y a égalité si et seulement si X et Y sont statistiquement indépendantes.

P5: $H(Y/X) = 0$ ssi il existe une application f de M_X vers M_Y telle que : $f(X) = Y$.

P6: $I(X:Y) = H(X) - H(X/Y) = H(Y) - H(Y/X).$

P7: $I(X:Y) \geq 0.$

De plus, il y a égalité si et seulement si X et Y sont statistiquement indépendantes.

P8: $I(X:Y) \leq \text{Max} (H(X), H(Y)).$

De plus, $I(X:Y) = \text{Max} (H(X), H(Y))$ si et seulement si:

il existe une relation univoque f soit de M_X vers M_Y telle que: $f(X) = Y$, soit de M_Y vers M_X telle que: $f(Y)=X$.

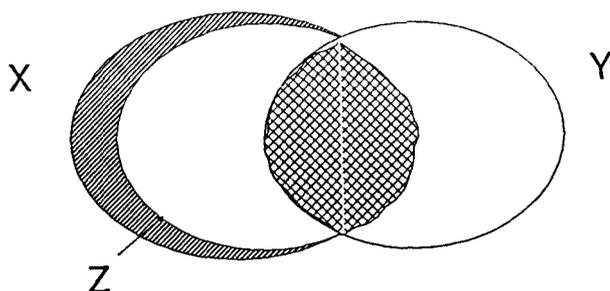
En particulier, si $H(X) = H(Y)$, cette relation est biunivoque.

P9: $I(X:Y) \geq I(f(X):Y).$

De plus, il y a égalité si f est biunivoque.

La réciproque n'est pas vraie comme le montre le contre-exemple suivant:

Soient X, Y, Z telles que : $H(X/Z) > 0$, $H(Z/X) = 0$ et $I(X:Y) = I(Z:Y)$.



puisque $H(X/Z) > 0$ et $H(Z/X) = 0$, il existe une relation non biunivoque f telle que: $Z = f(X)$.

$$\text{P10: } I(X_1: X_2: \dots : X_N) = \sum_{i=1}^N H(X_i) - H(X_1, X_2, \dots, X_N).$$

De façon générale, si $R = \{R_1, R_2, \dots, R_S\}$ est une partition d'un ensemble de variables, on a:

$$I(R_1: R_2: \dots : R_S) = \sum_{i=1}^S H(R_i) - H(R_1, R_2, \dots, R_S).$$

III.1.2.2. Cas continu [Dubuisson 78].

En général, les propriétés que vérifient l'entropie et la transinformation en discret ne sont pas vérifiées en continu.

On montre que :

P1: L'entropie peut être négative, voire infinie.

P2: L'entropie d'une variable continue est généralement sensible aux transformations:

$$\text{Si } T \text{ de } X = Z, \text{ alors } H(Z) = H(X) - \int_{M_X} \log |J(x)| dx.$$

Où $J(x)$ est le jacobien de T .

En particulier, si T est une rotation, alors $H(Z) = H(X)$.

P3: Si X est une variable continue et X_D une variable obtenue par discrétisation de X de pas δX , alors:

$$H(X) = \lim_{\delta X \rightarrow 0} (H(X_D) + \log(\delta X)).$$

P4: Si X est une variable vectorielle gaussienne sur \mathbb{R}^n , alors, on a:

$$H(X) = 1/2 \log(2\pi e)^n |\det V|$$

où V est la matrice de variance-covariance de X .

P5: Si X et Y sont deux variables gaussiennes, alors:

$$I(X:Y) = -\log \sqrt{1 - r^2(X,Y)}. \quad [\text{Dufour 79}]$$

$$H(X/Y) = 1/2 * \text{Log} [2\pi e \text{Var}(X) (1-r^2(X,Y))]$$

où $r(X,Y) = \text{Cov}(X,Y) / \sqrt{\text{Var}(X) \cdot \text{Var}(Y)}$ est le coefficient de corrélation linéaire.

Notons que la plupart des propriétés qui découlent de l'entropie et de la transinformation, permettent de disposer de coefficients et de critères intéressants relatifs aux lois de transmission et de l'évolution de l'information dans un système. (Chap II, §2. ChapIII, §2, §3. ChapIV, §3).

III.1.2.3. Représentation graphique.

L'existence d'une analogie entre la théorie de la mesure et l'entropie [Miller 63] permet d'avoir - sous quelques réserves - une représentation graphique d'un système. Cette analogie permet de représenter chaque variable par une surface proportionnelle à son entropie. Pour deux ou plusieurs variables, elle conduit à une représentation des différentes entropies et transinformations mises en jeu entre les variables, comme l'indique la figure suivante :

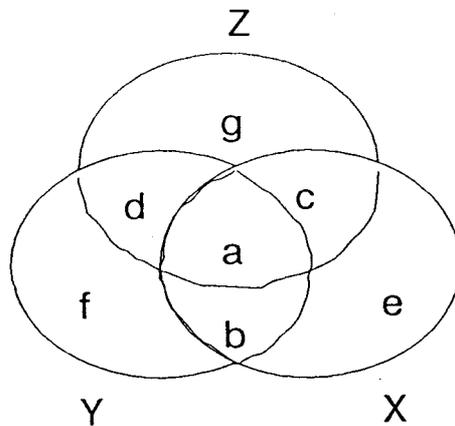


Fig1.2
Représentation
graphique
d'un système
à 3 variables.

avec:

$$H(X) = a + b + c + e .$$

$$H(X,Y) = a + b + c + e + f + d .$$

$$H(X,Y,Z) = a + b + c + e + f + g + d .$$

$$H(X/Y) = e + c .$$

$$I(X:Y) = a + b .$$

$$I_Z(X:Y) = b .$$

$$I(X:Y:Z) = 2a + b + c + d .$$

Pour $I(X:Y:Z)$, la surface a est comptée deux fois; pour N variables a est comptée $N-1$ fois

Cependant comme le montre [Toro 82], cette représentation doit être utilisée avec réserve; en effet: A toute représentation graphique correspond un système, mais tout système n'a pas de représentation graphique. C'est le cas par exemple du système: $\Sigma = \{X_1, X_2, X_3, X_4\}$ dont le tableau des données est défini par:

$\Omega \setminus \Sigma$	X_1	X_2	X_3	X_4
w_1	0	0	0	1
w_2	0	0	1	0
w_3	0	1	0	0
w_4	0	1	1	0
w_5	1	1	1	0
w_6	1	1	0	0
w_7	1	0	1	1
w_8	1	0	0	0

On obtient $I(X_i : X_j) = 0$ pour tout $i \neq j$, alors que $I(X_1 : X_2 : X_3 : X_4) = 1$. Ce qui conduit à représenter le système par quatre cercles disjoints deux à deux et dont l'intersection est non vide. Ceci est géométriquement impossible.

III.2. Approche de Kullback.

Dans ce paragraphe, nous allons étudier les propriétés des mesures de divergence de Kullback. Nous verrons plus loin (Chap IV) que de telles mesures permettent de disposer de critères applicables à l'évaluation de la qualité d'un modèle.

III.2.1. Définitions formelles.

Divergence relative à une variable.

Soit X une variable à modalités dans:

$$M_X = \{ x_1, x_2, \dots, x_n \},$$

munie de deux distributions notées $p=(p_i)$ et $q=(q_i)$, $p_i > 0$
 $q_i > 0$ $i=1$ à n .

On définit la divergence de p par rapport à q par :

$$J_X(p, q) = \sum_{i=1}^m p_i \log \frac{p_i}{q_i}$$

Et la divergence symétrique entre p et q par :

$$S_X(p, q) = J_X(p, q) + J_X(q, p).$$

Ces deux définitions ne dépendent pas des modalités de la variable X .

Pour deux variables X et Y munies de deux distributions conjointes $P=(p_{ij})$ et $Q=(q_{ij})$, on définit la divergence de P par rapport à Q par :

$$J_{XY}(P, Q) = \sum_{i=1}^n \sum_{j=1}^m p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

Et la divergence relative à X sachant Y par :

$$J_{X/Y}(P, Q) = \sum_{i=1}^n \sum_{j=1}^m p_{i/j} \log \frac{p_{i/j}}{q_{i/j}}$$

De façon similaire, on définit également les divergences symétriques conjointe et conditionnelle.

Divergence entre deux partitions.

Soient deux partitions $P(\Omega)$ et $Q(\Omega)$ de l'espace d'observation, muni d'une mesure de probabilité. On suppose que $P(\Omega)$ et $Q(\Omega)$ ont le même cardinal n .

On définit la divergence de $P(\Omega)$ par rapport à $Q(\Omega)$ par :

$$J(P(\Omega), Q(\Omega)) = \sum_{i=1}^n p_r(P_i) \log \frac{p_r(P_i)}{p_r(Q_i)}$$

III.2.2. Propriétés.

On montre que :

P1 : $J_X(p, q) \geq 0$;

de plus il y a égalité $\iff p_i = q_i$ pour $i=1 \dots n$.

Démonstration :

Sachant que pour $\alpha > 0$ l'expression $\log \alpha \leq \alpha - 1$ est vérifiée, on a :

$$\log \frac{q_i}{p_i} \leq \frac{q_i}{p_i} - 1, \text{ pour } i=1, \dots, n.$$

En multipliant par p_i et en sommant ensuite, on obtient :

$$\sum_{i=1}^n p_i \log \frac{q_i}{p_i} \leq \sum_{i=1}^n q_i - \sum_{i=1}^n p_i = 0 .$$

Il en résulte que : $J_X(p, q) \geq 0$.

Par ailleurs, $p_i = q_i$ si et seulement si $\log p_i/q_i = 0$

c'est à dire $J_X(p, q) = 0$.

$$P2 : J_{XY}(P, Q) = J_X(p, q) + J_{Y/X}(P, Q) .$$

Démonstration :

$$\begin{aligned} J_{XY}(P, Q) &= \sum_{i=1}^n \sum_{j=1}^m p_{ij} \log \frac{p_{ij}}{q_{ij}} \\ &= \sum_{i=1}^n \sum_{j=1}^m p_{ij} \log \frac{p_{j/i} \cdot p_i}{q_{j/i} \cdot q_i} \\ &= \sum_{i=1}^n \sum_{j=1}^m p_{ij} \log \frac{p_i}{q_i} + \sum_{i=1}^n \sum_{j=1}^m p_{ij} \log \frac{p_{j/i}}{q_{j/i}} \\ &= J_X(p, q) + J_{Y/X}(P, Q) . \end{aligned}$$

P3 : $J_{XY}(P, Q) = J_X(p, q) + J_Y(p, q)$ si : X et Y sont statistiquement indépendantes suivant les deux distributions P et Q.

Démonstration :

X et Y indépendantes, alors: $p_{i,j} = p_i * p_j$ et $q_{i,j} = q_i * q_j$
 $i = 1, 2, \dots, n.$ $j = 1, 2, \dots, n.$

$$\begin{aligned}
 J_{XY}(P, Q) &= \sum_{i=1}^n \sum_{j=1}^m p_{ij} \log \frac{p_{ij}}{q_{ij}} \\
 &= \sum_{i=1}^n \sum_{j=1}^m p_i * p_j \log [(p_i/q_i) * (p_j/q_j)] \\
 &= \sum_{i=1}^n \sum_{j=1}^m p_i * p_j (\log p_i/q_i + \log p_j/q_j) \\
 &= \sum_{i=1}^n p_i \log p_i/q_i + \sum_{j=1}^m p_j \log p_j/q_j \\
 &= J_X(p, q) + J_Y(p, q).
 \end{aligned}$$

Notons que la réciproque n'est pas vraie comme on peut le vérifier à partir du contre-exemple suivant:

Soient les distributions:

$$P = \begin{bmatrix} a/3 & 2a/3 \\ (1-a)/3 & (2-2a)/3 \end{bmatrix} \quad \text{et} \quad Q = \begin{bmatrix} 1/6 & 1/3 \\ 1/3 & 1/6 \end{bmatrix}$$

telles que $a = 3 \log 3 - 4 \approx 0.75$ ($\log 3 \approx 1.58$).

On vérifie que :

$$J_{XY}(P, Q) = J_X(p, q) + J_Y(p, q)$$

et que X et Y ne sont pas statistiquement indépendantes suivant Q.

P4 : Soient X_1, X_2, \dots, X_N N variables indépendantes munies des mêmes distributions p et q, alors :

$$J_{X_1 X_2, \dots, X_N}(p, q) = N \cdot J_{X_1}(p, q)$$

Cette propriété est une conséquence de la précédente.

P5 : X munie de deux distributions:

$$p = (p_1, p_2, \dots, p_n) \text{ et } q = (q_1, q_2, \dots, q_n) .$$

Et notons par :

X' la variable obtenue à partir de X en regroupant les deux premières modalités.

X'' la variable définie à partir des deux premières modalités de X, munie des deux distributions :

$$p'' = \left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2} \right) \text{ et } q'' = \left(\frac{q_1}{q_1+q_2}, \frac{q_2}{q_1+q_2} \right)$$

Alors, on a :

$$J_{X'}(p', q') = J_X(p, q) - (p_1+p_2) J_{X''}(p'', q'') .$$

$$\text{Où } p' = (p_1+p_2, p_3, \dots, p_n) \text{ et } q' = (q_1+q_2, q_3, \dots, q_n) .$$

Cette propriété montre que si l'on regroupe deux modalités de la variable X, la divergence relative à la variable obtenue X', diminue d'une quantité égale à $(p_1+p_2) J_{X''}(p'', q'')$.

Démonstration :

$$\begin{aligned}
 J_{X'}(p', q') &= (p_1+p_2) \log \frac{p_1+p_2}{q_1+q_2} + \sum_{j=3}^n p_j \log p_j/q_j \\
 &= J_X(p, q) - (p_1+p_2) \left[\frac{p_1}{p_1+p_2} \log p_1/q_1 + \frac{p_2}{p_1+p_2} \log p_2/q_2 \right. \\
 &\quad \left. - \log \frac{p_1+p_2}{q_1+q_2} \right]. \\
 &= J_X(p, q) - (p_1+p_2) \left[\frac{p_1}{p_1+p_2} \log p_1/q_1 + \frac{p_2}{p_1+p_2} \log p_2/q_2 \right. \\
 &\quad \left. - \left(\frac{p_1}{p_1+p_2} \log \frac{p_1+p_2}{q_1+q_2} + \frac{p_2}{p_1+p_2} \log \frac{p_1+p_2}{q_1+q_2} \right) \right]. \\
 &= J_X(p, q) - (p_1+p_2) \left[\frac{p_1}{p_1+p_2} \log \frac{p_1}{q_1} + \frac{p_2}{p_1+p_2} \log \frac{p_2}{q_2} \right] \\
 &= J_X(p, q) - (p_1+p_2) J_{X''}(p'', q'').
 \end{aligned}$$

Remarques :

* On montre que les propriétés ci-dessus sont vérifiées pour la divergence symétrique. Il en est de même en général pour les mesures de divergence définies sous forme continue.

* Par ailleurs, contrairement à la mesure d'information de Shannon, on montre que la mesure de divergence de Kullback (symétrique ou non) ne vérifie pas toujours la propriété de sous-additivité. En effet, si l'on considère les variables X et Y munies des tableaux des probabilités conjointes:

$$P = \begin{bmatrix} 1/4 & 1/4 \\ 1/4 & 1/4 \end{bmatrix} \text{ et } Q = \begin{bmatrix} 1/8 & 1/4 \\ 1/4 & 3/8 \end{bmatrix}$$

on vérifie que : $J_{XY}(P, Q) > J_X(p, q) + J_Y(p, q)$.

* On montre que lorsque deux variables X et Y sont munies des distributions $P = (p_{ij})$ et $Q = (q_{ij})$ telles que: $q_{ij} = p_i \cdot p_j$, alors:

$$J_{XY}(P, Q) = I(X:Y)$$

Cette expression peut être interprétée de la façon suivante: Si P et Q désignent respectivement la distribution vraie et la distribution de X,Y supposées indépendantes, alors la divergence de P par rapport à Q n'est autre que l'information transmise entre X et Y.

* On montre que pour une variable continue X, munie de deux lois gaussiennes :

$$p(x) = 1/\sqrt{2\pi\sigma_1} \exp [-(x - m_1)^2/2\sigma_1^2]$$

$$q(x) = 1/\sqrt{2\pi\sigma_2} \exp [-(x - m_2)^2/2\sigma_2^2]$$

On a :

$$J_X(p,q) = 1/2 [\log \sigma_1/\sigma_2 + (\sigma_1^2/\sigma_2^2 - 1) + (m_1 - m_2)^2/\sigma_2^2].$$

De même, pour la divergence symétrique, on a :

$$S_X(p,q) = \frac{1}{2\sigma_1^2\sigma_2^2} * [(\sigma_1^2 - \sigma_2^2)^2 + (\sigma_1^2 + \sigma_2^2)(m_1 - m_2)^2]$$

En particulier, si les deux lois sont :

—> de même variance ($\sigma_1 = \sigma_2 = \sigma$), on a :

$$J_X(p,q) = 1/2 (m_1 - m_2)^2/\sigma^2 \quad \text{et} \quad S_X(p,q) = (m_1 - m_2)^2/\sigma^2$$

—> de même moyenne ($m_1 = m_2 = m$), on a :

$$J_X(p,q) = 1/2 [\log \sigma_1/\sigma_2 + (\sigma_1^2/\sigma_2^2 - 1)]$$

et

$$S_X(p,q) = 1/2 (\sigma_1^2 - \sigma_2^2)^2/\sigma_1^2\sigma_2^2$$

On constate que les mesures de divergence de Kullback vérifient des propriétés similaires à celles de l'entropie de Shannon. Cependant, elles ne vérifient pas la propriété de sous-additivité (cf p16).

De telles mesures fournissent des critères qui s'appliquent à la sélection et à la reconnaissance des formes [Chen 76]. [McLean 85], ainsi qu'à la validation d'un modèle et à la détection des défaillances dans un système [Soderstrom 85].

III .3. Autres approches.

Dans ce paragraphe, nous considérons d'autres approches en théorie de l'information. Celles-ci se présentent comme des généralisations du concept de l'entropie de Shannon ou de la divergence de Kullback sous une formulation paramétrique. Pour plus de détails, voir [Ponomorenko 81].

III .3 1 Mesures d'entropie paramétrique.

Entropie d'ordre α [Rényi 61].

Soit une variable X munie d'une distribution de probabilité (p_i) , $i = 1, \dots, n$.

On définit l'entropie d'ordre α par :

$$H_\alpha(X) = \frac{1}{1-\alpha} \sum_{i=1}^n \log p_i^\alpha ; \alpha > 0 , \alpha \neq 1 .$$

On montre que l'entropie d'ordre α ne possède pas toutes les propriétés de l'entropie de Shannon. En particulier, on a:

$$* \quad \lim_{\alpha \rightarrow 1} H_\alpha(X) = H(X) .$$

$$* \quad \lim_{\alpha \rightarrow +\infty} H_\alpha(X) = - \log \text{Max}(p_i) .$$

* $H_\alpha(X)$ est décroissante par rapport à α .

- * $H_\alpha(X)$ n'est pas sous-additive.
- * $H_\alpha(X, Y) = H_\alpha(X) + H_\alpha(Y)$ si X et Y sont indépendantes.

La réciproque n'est pas généralement vérifiée. En effet: Pour $\alpha = 2$ et X, Y munies de la distribution:

$$P = \begin{pmatrix} 0 & 1-\sqrt{2}/2 \\ 1-\sqrt{2}/2 & \sqrt{2}-1 \end{pmatrix}$$

$H_\alpha(X, Y) = H_\alpha(X) + H_\alpha(Y) = -\log(6-4\sqrt{2})$ et X, Y ne sont pas indépendantes .

Entropie du type β [Daroczy 70].

Soit une variable X munie d'une distribution de probabilité (p_i) , $i = 1, 2, \dots, n$.

On définit l'entropie du type β par :

$$H^\beta(X) = \frac{1}{1-2^{1-\beta}} \left(1 - \sum_{i=1}^n p_i^\beta \right); \beta > 0, \beta \neq 1$$

L'entropie du type β vérifie des propriétés similaires à l'entropie de Shannon. On montre en particulier que:

$$* \quad \lim_{\beta \rightarrow 1} H^\beta(X) = H(X).$$

$$* \quad H^\beta(X) \leq \frac{1-n^{1-\beta}}{1-2^{1-\beta}}$$

Il y a égalité si X suit une distribution équiprobable.

$$* \quad H^\beta(X, Y) = H^\beta(X) + H^\beta(Y) + (2^{1-\beta} - 1)H^\beta(X) * H^\beta(Y) \text{ si } X \text{ et } Y \text{ sont indépendantes.}$$

La réciproque n'est pas vérifiée pour $\beta = 2$ et X, Y munies de la distribution:

$$P = \begin{bmatrix} 0 & 1-\sqrt{2}/2 \\ 1-\sqrt{2}/2 & \sqrt{2}-1 \end{bmatrix}$$

$$* \quad H^\beta(X, Y) \leq H^\beta(X) + H^\beta(Y) .$$

III.3.2. Mesures de divergence paramétriques.

Divergence d'ordre α [Rényi 61].

On définit pour une variable X munie de deux distributions p et q la divergence d'ordre α par :

$$J_X^\alpha(p, q) = \frac{1}{\alpha - 1} \log \sum_{i=1}^n \frac{p_i^\alpha}{q_i^{\alpha-1}} ; \alpha = 1.$$

Cette définition généralise celle de la divergence de Kullback, puisque l'on a :

$$\lim_{\alpha \rightarrow 1} J_X^\alpha(p, q) = J_X(p, q) .$$

On montre par ailleurs que :

* Si X et Y sont indépendantes, on a :

$$J_{XY}^\alpha(P, Q) = J_X^\alpha(p, q) + J_Y^\alpha(p, q) .$$

La réciproque n'est pas toujours vraie. Ceci peut être prouvé par le contre-exemple suivant :

$\alpha = 2$ et X, Y munies des deux distributions :

$$P = \begin{bmatrix} a/3 & 2a/3 \\ b/3 & 2b/3 \end{bmatrix} \quad Q = \begin{bmatrix} 1/6 & 1/3 \\ 1/3 & 1/6 \end{bmatrix}$$

avec $a = (7 - \sqrt{14})/5$, $b = (\sqrt{14} - 2)/5$.

Divergence du type β . [Rathie 72] .

On définit la divergence du type β par l'expression :

$$J_X^\beta(p, q) = \frac{1}{1-2^{\beta-1}} \left(1 - \sum_{i=1}^n \frac{p_i^\beta}{q_i^{\beta-1}} \right) \quad \beta \neq 1.$$

On montre que :

$$\lim_{\beta \rightarrow 1} J_X^\beta(p, q) = J_X(p, q).$$

On montre par ailleurs que si X et Y sont indépendantes, alors :

$$J_{XY}^\beta(p, q) = J_X^\beta(p, q) + J_Y^\beta(p, q) - (1-2^{\beta-1}) J_X^\beta(p, q) * J_Y^\beta(p, q).$$

La réciproque n'est pas vraie comme cela peut être prouvé à partir du contre-exemple ci-dessus pour $\beta = 2$.

On constate que l'ensemble des mesures paramétriques vérifient des propriétés similaires aux cas non paramétriques. Cependant, elles ne vérifient pas toutes la propriété de sous-additivité, et ne peuvent ainsi fournir des indices appropriés de couplage.

Par ailleurs, l'ensemble de ces mesures - parmi d'autres - constitue un éventail intéressant de critères pour la sélection et la reconnaissance des formes [Chen 76].

Signalons aussi que ces mesures dépendent d'un paramètre arbitraire et que dans tous les formalismes mathématiques, aucune signification concrète n'a été donnée pour un tel paramètre si bien que l'utilisateur se trouve confronté au choix de la mesure la mieux adaptée à son analyse.

III.3.3. Entropie du Max.

Soit une variable X munie d'une distribution de probabilité (p_i) , $i=1,2,\dots,n$.

On définit l'entropie du Max par:

$$H_M(X) = 1 - \max_{1 \leq i \leq n} (p_i).$$

On montre que l'entropie du Max vérifie des propriétés similaires à celles de l'entropie de Shannon, en particulier:

- * H_M est sous-additive : $H_M(X,Y) \leq H_M(X) + H_M(Y)$.
- * $H_M(X,Y) \geq \max (H_M(X) , H_M(Y))$.
- * Si Y est une variable obtenue par regroupement de modalités de X , On a:

$$H_M(Y) \leq H_M(X) .$$

- * $H_M(X) = 0$ si et seulement si X est constante.
- * X, Y indépendantes n'implique pas que:

$$H_M(X,Y) = H_M(X) + H_M(Y) , \text{ et réciproquement.}$$

III.3.4. Mesure d'inexactitude. [Kerridge 61]

Introduite par Kerridge, la mesure d'inexactitude peut être considérée comme une généralisation formelle de l'entropie de Shannon et de la divergence de Kullback.

Etant donnée une variable X , munie de deux distributions de probabilité notées: $p = (p_i)$ et $q = (q_i)$, $i = 1 \dots n$.

On définit l'inexactitude de p par rapport à q relative à X par:

$$A_X(p,q) = - \sum_{i=1}^n p_i \log q_i.$$

Bien entendu, une telle expression peut être étendue au cas continu.

Compte tenu des définitions de l'entropie de Shannon et de la divergence de Kullback, on a:

$$\begin{aligned} A_X(p,q) &= \sum_{i=1}^n p_i \log p_i + \sum_{i=1}^n p_i \log p_i - \sum_{i=1}^n p_i \log q_i. \\ &= H(X) + J_X(p,q). \end{aligned}$$

On montre que la mesure d'inexactitude possède des propriétés similaires à celles de l'entropie de Shannon et de la divergence de Kullback [Ponomorenko 81]. En particulier:

* $A_X(p,q)$ est positive, nulle si et seulement si X est constante.

* $A_X(p,q) = H(X)$ si et seulement si $p = q$.

Dans le cas où la distribution q est une estimation de la distribution p , $A_X(p,q)$ peut être interprétée comme mesure de l'incertitude totale due à l'indétermination de X d'une part, et à l'erreur d'estimation de p par q d'autre part.

Nous verrons ultérieurement l'application d'une telle mesure dans le cadre de la détermination de l'ordre d'un système.

IV. CONCLUSION.

Avant de procéder à l'analyse d'un système, le praticien doit disposer d'une connaissance préalable au sujet du système, et d'un certain nombre d'outils et de méthodes lui permettant de mener au mieux les différentes phases de l'analyse.

Dans ce chapitre, nous nous sommes intéressés d'abord à la description du système et à la génération des données. Nous avons présenté ensuite des éléments de la théorie de l'information issus des approches de Shannon et de Kullback.

Dans un premier temps, nous avons rappelé les définitions et les propriétés concernant l'entropie et la transinformation dans le cas discret et dans le cas continu, ainsi que les mesures de divergence de Kullback. On montre en particulier que les mesures de Kullback présentent des propriétés similaires à celles de l'entropie de Shannon, cependant elles ne vérifient pas la propriété de sous-additivité.

Dans un deuxième temps, nous nous sommes intéressés aux mesures d'information dites paramétriques. On montre que celles-ci ne vérifient pas toujours les mêmes propriétés que les mesures de Shannon et de Kullback et ne peuvent ainsi constituer des indices de couplage appropriés. Dans la suite de ce travail, nous verrons, que les éléments issus de l'information de Shannon et de Kullback, permettent de disposer d'outils intéressants pour l'analyse structurale des systèmes complexes, l'évaluation de la qualité d'un modèle, ainsi que pour la détermination de l'ordre d'un système.

CHAPITRE II :

**ANALYSE DES SYSTEMES STATIQUES
PAR LA THEORIE DE L'INFORMATION.**

I. INTRODUCTION.

Un des objectifs de l'analyse structurale des systèmes complexes consiste à rechercher une partition du système en sous-systèmes faiblement couplés. Dans ce cas, on doit disposer de moyens permettant d'analyser le couplage entre variables ou groupes de variables.

La T I de Shannon répond à ce besoin. En effet : elle génère des indices qui s'appliquent à tout type de variable (statiques, dynamiques, quantitatives, qualitatives, à ensemble de modalités structuré ou non) et qui permettent d'évaluer l'intensité de liaisons quelconques (linéaire, ou non linéaire, etc...) entre variables quantitatives ou qualitatives, ce qui n'est pas toujours le cas des indices issus de la statistique.

Ainsi, dans le cadre de la décomposition d'un système, la prise en compte d'indices informationnels appliqués aux variables par paires conduit à une matrice de couplage ou graphe d'interconnexion dont l'analyse fait apparaître une partition du système (approche binaire). Une telle approche peut s'avérer imprécise, d'où la nécessité de considérer des indices appliqués aux groupes de variables. Ceux-ci permettent d'élaborer des algorithmes pour la détermination de partitions "optimales" parmi toutes les partitions du système (approches multidimensionnelles).

Une autre approche consiste à rechercher dans quelle mesure, on peut établir des relations fonctionnelles entre les variables du système. Cette approche, essentiellement orientée dans un but de modélisation, peut permettre aussi

une décomposition du système. Pour une telle approche (Approche d'explication), on suppose que les variables du système peuvent se présenter en deux groupes: un groupe de variables à expliquer et un groupe de variables explicatives. L'approche d'explication consiste à déterminer pour chaque variable à expliquer, un sous ensemble de variables explicatives qui fournissent une information maximale. On considérera 2 cas selon que le système est déterministe ou non. Dans le cas non déterministe, l'approche d'explication conduit à séparer les modalités mal expliquées, à cause du bruit, ou dont l'explication nécessite l'introduction de variables supplémentaires.

II. ANALYSE DU COUPLAGE STATIQUE PAR LA T I.

II.1. Indices de couplage statique.

Dans un système complexe, il est généralement vérifié que les variables ne sont pas toutes couplées, ou si elles le sont, qu' il existe une certaine hiérarchie dans l'intensité des couplages [Simon 62].

Dans ce qui suit, nous considérons des indices issus de la Théorie de l'Information de Shannon pour mesurer l'intensité de couplage entre les variables.

Soient X_i et X_j deux variables d'un système statique.

Pour mesurer le couplage entre X_i et X_j , on peut utiliser la transinformation $I(X_i:X_j)$. Cependant cette mesure ne permet pas de reconnaître le sens de la liaison entre X_i et X_j : en effet pour une valeur maximale de $I(X_i:X_j)$, on peut avoir une liaison du type:

$$X_j = f(X_i) \text{ ou } X_i = g(X_j).$$

Le sens de la liaison est alors à déterminer à partir de considérations physiques appliquées sur le système.

L'entropie conditionnelle $H(X_i/X_j)$ permet également de mesurer le couplage entre X_i et X_j . Une telle mesure présente l'avantage de préciser le sens de la liaison entre X_i et X_j puisque pour $H(X_i/X_j)$ nulle, on a une liaison du type :

$$X_i = f(X_j).$$

D'autres mesures de couplage sont définies à partir de la transinformation, tels le coefficient de Conant [Conant 72] :

$$C_{ij} = \frac{I(X_i : X_j)}{H(X_j)}$$

Ou le coefficient de Richetin [Richetin 75]:

$$T_{ij} = \frac{I(X_i : X_j)}{\sqrt{H(X_i)H(X_j)}}$$

On peut aussi définir par analogie avec les indices de similarité, des coefficients de couplages tel l'indice de Dussauchoy [Dussauchoy 80] :

$$D_{ij} = \frac{I(X_i : X_j)}{H(X_i, X_j)}$$

Ou l'indice de Lopez de Mantaras [L. de Mantaras 77]:

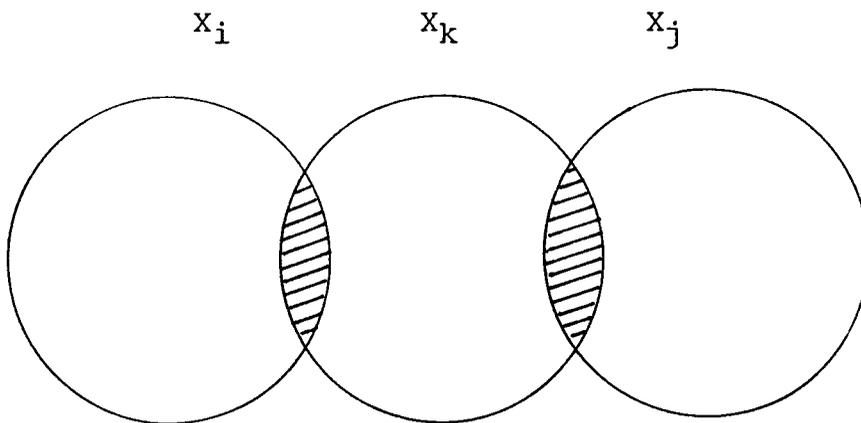
$$L_{ij} = \frac{H(X_i/X_j) + H(X_j/X_i)}{H(X_i, X_j)}$$

A partir de la transinformation conditionnelle, on peut également mesurer le couplage entre deux variables, dû à la présence d'une troisième. En effet, soient X_i , X_j , X_k , trois variables du système; le couplage entre X_i et X_j dû à l'élimination de l'influence de la variable X_k peut être mesuré par:

$$I_{X_k}(X_i:X_j) = H(X_i/X_k) - H(X_i/X_j, X_k)$$

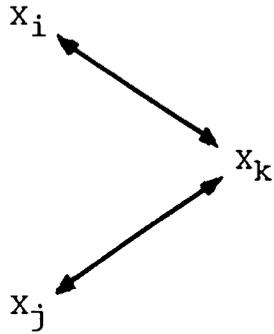
Lorsque $X_k = \Sigma \setminus \{X_i, X_j\}$, $I_{X_k}(X_i:X_j)$. I_{X_k} mesure le couplage direct entre X_i et X_j dû à l'élimination de l'influence des variables restantes [Rinck 73].

Notons que la mesure de couplage par la transinformation conditionnelle peut ne pas être satisfaisante; en effet, dans le cas illustré par le schéma suivant:



On a : $I_{X_k}(X_i:X_j) = 0$, $I(X_i:X_j) \neq 0$ et $I(X_j:X_k) \neq 0$.

les variables X_i et X_j étant couplées par l'intermédiaire de X_k , alors que la transinformation conditionnée par X_k est nulle. Dans ce cas, on peut dire qu'il y a une liaison indirecte entre X_i et X_j due à X_k :



II.2. Relations fondamentales.

A partir de la transinformation, Ashby [Ashby 69] a démontré que :

Si S_1, S_2, \dots, S_L forment une partition en sous-systèmes de Σ , alors on a :

$$I(X_1 : X_2 : \dots : X_N) = \sum_{i=1}^L I(S_i) + I(S_1 : S_2 : \dots : S_L)$$

Cette relation montre que l'information totale dans un système est la somme de l'information dans chaque sous-système plus la transinformation transmise entre tous les sous-systèmes. Une telle relation révèle le caractère multidimensionnel de la T I de Shannon et permet de rechercher une décomposition du système en sous-systèmes faiblement couplés par minimisation de l'indice $I(S_1 : S_2 : \dots : S_L)$.

Par ailleurs, Mc Gill [Mc Gill 54] a démontré la relation suivante :

$$I(X_1:X_2:\dots:X_N) = \sum_{i < j} I(X_i:X_j) + \sum_{i > j > k} Q(X_i, X_j, X_k) + \dots + Q(X_1, \dots, X_N).$$

où :

$$Q(X_i, X_j, \dots, X_1) = Q_{X_i}(X_j, \dots, X_1) - Q(X_j, \dots, X_1)$$

et :

$$Q_{X_i}(X_j, X_k) = I_{X_i}(X_j:X_k) - I(X_j:X_k).$$

Cette relation montre que l'information totale dans un système est égale à l'information transmise entre les variables prises par paires, plus la somme d'interactions à l'intérieur de groupes d'au moins trois variables. Elle met aussi en évidence le fait que si les interactions sont relativement élevées, une analyse de couplage par paires de variables peut s'avérer insuffisante; on peut alors envisager une analyse multidimensionnelle par le biais d'indices de couplage appliqués à des variables prises par triplets, quadruplets ...

III. DECOMPOSITION D'UN SYSTEME.

La décomposition d'un système complexe en sous-systèmes faiblement couplés répond à un besoin de réduire sa complexité apparente. Celle-ci peut être due à un grand nombre de variables qui décrivent le système et/ou à une multitude d'interactions entre elles.

Dans ce paragraphe, nous présentons quelques approches de décomposition et nous soulignons leurs limites respectives.

III.1. Approches binaires.

Il s'agit de déterminer l'intensité de couplage entre les variables prises deux à deux au moyen des indices décrits précédemment.

Ces indices permettent de construire une matrice de couplage, et un graphe d'interconnexion. L'analyse du graphe conduit à une partition du système [Conant 72] , [Richetin 75].

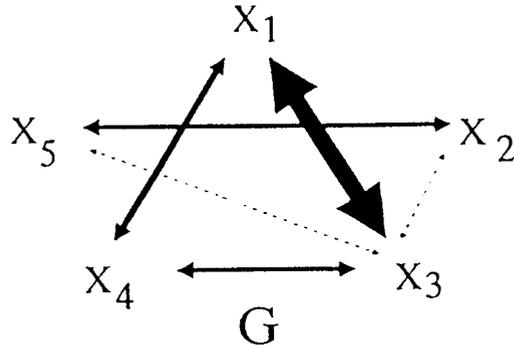
A titre d'exemple, considérons le système défini par l'ensemble des variables : $\Sigma = \{X_1, X_2, X_3, X_4, X_5\}$, et dont le tableau est donné comme suit :

$\Omega \backslash \Sigma$	X_1	X_2	X_3	X_4	X_5
w_0	1	1	1	1	1
w_1	1	1	3	2	1
w_2	3	2	3	2	2
w_3	3	2	3	1	1
w_4	3	1	3	1	1
w_5	2	1	3	1	1
w_6	2	1	2	2	1
w_7	1	2	2	2	2
w_8	1	2	1	2	1
w_9	2	2	3	1	2
w_{10}	2	1	2	2	2
w_{11}	1	1	2	2	1
w_{12}	1	2	1	2	2
w_{13}	2	2	3	2	1

Tableau des données de Σ .

Le calcul de la transinformation normalisée telle le coefficient T_{ij} conduit à la matrice de couplage et au graphe d'interconnexion suivants : [Richetin]

$$\begin{array}{c}
 X_1 \\
 X_2 \\
 X_3 \\
 X_4 \\
 X_5
 \end{array}
 \left(
 \begin{array}{ccccc}
 X_1 & X_2 & X_3 & X_4 & X_5 \\
 1 & 0 & 0.34 & 0.08 & 0 \\
 0 & 1 & 0.06 & 0.02 & 0.16 \\
 0.34 & 0.06 & 1 & 0.21 & 0.02 \\
 0.08 & 0.02 & 0.21 & 1 & 0.05 \\
 0 & 0.16 & 0.02 & 0.05 & 1
 \end{array}
 \right)$$



L'examen du graphe G conduit à la partition du système en deux sous-systèmes:

$$\Sigma_1 = \{X_1, X_3, X_4\} \quad \text{et} \quad \Sigma_2 = \{X_2, X_5\}.$$

Soulignons que les approches binaires sont intéressantes puisque leur mise en oeuvre est simple, cependant elles ne peuvent conduire en général qu'à des décompositions approximatives. C'est le cas par exemple d'un système de N variables où l'information échangée entre les N variables prises ensemble est positive, alors que celle échangée entre variables deux à deux est nulle [Toro 82](Chap I.§III.1.2.3).

Il convient alors, pour la décomposition des systèmes complexes, de procéder suivant des approches multidimensionnelles.

III.2. Approches multidimensionnelles.

Soit $P(\Sigma) = \{S_1, S_2, \dots, S_L\}$ une partition de Σ .

Considérons la relation fondamentale:

$$I(X_1:\dots:X_N) = \sum_{i=1}^L I(S_i) + I(S_1:S_2:\dots:S_L)$$

(1) (2) (3)

Minimiser (3) par rapport aux partitions de Σ revient à maximiser (2); Cette minimisation conduit à une partition triviale $\{ \Sigma \}$, puisque pour toute partition $\{S_1, S_2, \dots, S_L\}$ de Σ , on a $I(\Sigma) \leq I(S_1:S_2:\dots:S_L)$. Afin de lever cette trivialité, on considère une décomposition du système en minimisant (3) par niveaux. Celle-ci consiste à prendre en compte à chaque niveau les partitions qui ont un même nombre de classes donné.

Dans ce qui suit, nous présentons quelques méthodes de décomposition.

III.2.1. Méthodes hiérarchiques.

III.2.1.1. Algorithme hiérarchique ascendant: (A.A).

A partir d'une partition de niveau k , on recherche une partition de niveau $k-1$, moins fine, par regroupement de deux sous-systèmes dont la transinformation est maximale.

Procédure:

- Initialisation : $P_0 = \{\{X_1\}, \dots, \{X_N\}\}$.

- Itération $k-1$:

on dispose de $P_{k-1} = \{S_1, S_2, \dots, S_{k-1}\}$

- Itération k :

$P_k = \{S_1, S_2, \dots, S_\alpha \cup S_\beta, \dots, S_{k-1}\} \setminus \{S_\alpha, S_\beta\}$ tel que :

$$I(S_\alpha : S_\beta) = \text{Max}_{\substack{1 \leq i \leq k-1 \\ 1 \leq j \leq k-1}} I(S_i : S_j)$$

III.2.1.2. Algorithme hiérarchique descendant: (A.D).

L'algorithme descendant consiste à rechercher à partir d'une partition d'un niveau k une partition de niveau k+1 par éclatement de l'une de ses classes en deux, de façon à ce que la transinformation entre les classes obtenues, soit minimale.

Procédure:

- Initialisation : $P_0 = \{ \Sigma \}$

- Itération k-1 :

on dispose de $P_{k-1} = \{S_1, S_2, \dots, S_{k-1}\}$

- Itération k :

$P_k = \{S_1, S_2, \dots, S_\alpha, S_* \setminus S_\alpha, \dots, S_{k-1}\} \setminus \{S_*\}$ tel que :

$$I(S_* \setminus S_\alpha : S_\alpha) = \text{Min}_{\substack{1 \leq i \leq k-1 \\ S \subset S_i}} I(S_i \setminus S : S_i)$$

III.2.1.3 Algorithme des approximations successives: (A.A.S.)

L'algorithme des approximations successives se présente comme composition des deux algorithmes ascendant et descendant de façon alternée. Il consiste de passer d'une partition de niveau r à une partition de niveau $r+1$ par l'algorithme descendant et de repasser à une partition de niveau r par l'algorithme ascendant, ou vice-versa.

Procédure :

- Initialisation : Pour un niveau r donné, on pose :

$$P_0 = \{S_1, S_2, \dots, S_r\}$$

- Itération $k-1$:

$$\text{on dispose de } P_{k-1} = \{S_{1,k-1}, \dots, S_{r,k-1}\}$$

- Itération k :

$$P_k = \{S_{1,k}, \dots, S_{r,k}\} \text{ tel que :}$$

$$P_k = AD \circ AA(P_{k-1}) \text{ ou } P_k = AA \circ AD(P_{k-1})$$

où :

$$AA(\{S_1, S_2, \dots, S_r\}) = \{S_1, S_2, \dots, S_\alpha \cup S_\beta, S_r\} \setminus \{S_\alpha, S_\beta\}$$

avec :

$$I(S_\alpha : S) = \text{Max } I(S_i : S_j) \\ 1 \leq i \leq r \\ 1 \leq j \leq r$$

et :

$$AD(\{S_1, S_2, \dots, S_r\}) = \{S_1, S_2, \dots, S_\alpha, S_* \setminus S_\alpha, \dots, S_r\} \setminus \{S_*\}$$

avec :

$$I(S_* \setminus S_\alpha : S_\alpha) = \text{Min } I(S_i \setminus S : S) \\ 1 \leq i \leq r \\ S \subset S_i$$

III.2.2. Méthode de programmation dynamique.

La méthode de programmation dynamique permet d'obtenir des optima locaux à chaque niveau de façon itérative. Elle s'appuie sur la proposition suivante :

Proposition.:

Pour $S \subset \Sigma$ tel que $\text{card}(S) \geq r$, on considère $P_r(S)$ une partition de S tel que :

$$I(P_r(S)) = \min_{P_r(S)} I(S_1 : \dots : S_r)$$

On montre que :

$$I(P_r(S)) = \min_{\substack{\sigma \subset S \\ \phi \neq \sigma \neq S}} (I(\sigma : S \setminus \sigma) + I(P_{r-1}(S \setminus \sigma)))$$

Procédure:

- Initialisation: $P_0 = \{ \Sigma \}$

- Itération k-1 :

On dispose de $P_{k-1} = P_{k-1}(\Sigma)$;

- Itération k :

On considère $P_k = P_k(\Sigma)$ tel que:

$$I(P_k(\Sigma)) = \min_{\substack{\sigma \subset \Sigma \\ \phi \neq \sigma \neq \Sigma}} (I(\sigma : \Sigma \setminus \sigma) + I(P_{k-1}(\Sigma \setminus \sigma)))$$

III.2.3 Optimalité des méthodes :

Les algorithmes hiérarchiques ascendant et descendant ne sont pas optimaux . Cependant, pour avoir des optima locaux, on arrête la procédure à l'itération qui donne un nombre de sous-systèmes satisfaisant compte tenu par exemple de la minimisation du critère:

$$I(S_1:S_2:\dots:S_L)$$

$$I(X_1:X_2:\dots:X_N)$$

défini à partir de la relation fondamentale.

L'algorithme des approximations successives élabore des partitions dites stables dans le sens où elles constituent des points fixes pour les opérateurs AAOAD et ADOAA. Pour les niveaux 1,2, N-1 et N, les partitions stables sont des optima globaux, cependant ce n'est pas toujours le cas pour les niveaux 3,...,N-2.

L'algorithme de programmation dynamique parcourt le treillis $P(\Sigma)$ en élaborant à chaque niveau r , une partition optimale en r classes à partir des partitions en $r-1$ classes des sous-ensembles de S , obtenues au niveau précédent.

Si les méthodes de décomposition, ci-dessus, s'avèrent simples à mettre en oeuvre, elle restent néanmoins peu efficaces lorsqu'il s'agit de traiter des systèmes très complexes. Une approche, dans ce cas, consiste à utiliser le fait que les variables ne sont pas toutes liées et que souvent ,elles peuvent être classées en deux groupes; un

groupe de variables explicatives accessibles à la mesure, à l'observation, etc... Et un groupe de variables à expliquer dont la mesure est difficile, coûteuse... On peut alors rechercher dans quelle mesure une variable peut être expliquée à partir d'autres. Cette approche d'explication permet de mettre en évidence une structure d'interconnexion pouvant conduire à :

--> Une partition du système en sous-systèmes faiblement couplés.

--> Un modèle du système sous forme de relations fonctionnelles (Aspect Modélisation).

IV. APPROCHE D'EXPLICATION :

IV.1. Le concept de modélisabilité :

Dans le but d'étudier dans quelle mesure une variable est liée, fonction, ou explique une ou plusieurs variables, on considère des indices informationnels. Ceux-ci par leurs propriétés et leurs interprétations intuitives, permettent de définir le concept de modélisabilité d'un système [Staroswiecki 84].

IV.1.1 Définitions.

Soient X une variable de Σ et $S \subset \Sigma \setminus \{X\}$, on définit par: [Conant 72]

$$d^{\circ}(X, S) = \frac{I(X:S)}{H(X)}$$

le degré de modélisabilité de la variable X par S .

$$\bar{d}^{\circ}(X, S) = \frac{H(X/S)}{H(X)}$$

le degré de non modélisabilité de X par S .

Pour $i = 1, \dots, N$; soient $X_i \subset \Sigma$ et $S_i \subset \Sigma \setminus \{X_i\}$ tels que:

$$I(X_i:S_i) = \text{Max}_{S \subset \Sigma \setminus \{X_i\}} I(X_i:S).$$

$$\text{On définit par : } d^{\circ}(S_1, \dots, S_N) = \frac{\sum_{i=1}^N I(X_i : S_i)}{\sum_{i=1}^N H(X_i)}$$

le degré de modélisabilité du système par S_1, S_2, \dots et S_N .

IV.1.2 Remarques:

- Une valeur voisine de 1 du degré de modélisabilité de X par S indique la possibilité d'obtenir une relation fonctionnelle du type $X=f(S)$; par contre si cette valeur est proche de 0, les variables contenues dans S ne suffisent pas pour expliquer ou modéliser S , ou bien X et S sont soumis à un bruit important.

- Une valeur voisine de 1 du degré de modélisabilité du système indique dans quelle mesure il est possible de décrire le système par le modèle :

$$\left\{ \begin{array}{l} X_1 = f_1(S_1) \\ X_2 = f_2(S_2) \\ \cdot \\ \cdot \\ \cdot \\ X_N = f_N(S_N) \end{array} \right.$$

En effet,

$$\frac{\sum_{i=1}^N I(X_i : S_i)}{\sum_{i=1}^N H(X_i)} = 1 \quad \leftrightarrow \quad \sum_{i=1}^N I(X_i : S_i) - \sum_{i=1}^N H(X_i) = 0$$

\leftrightarrow pour tout $i \leq N$; $I(X_i : S_i) = H(X_i)$

\leftrightarrow pour tout $i \leq N$; il existe $f_i : M_{S_i} \rightarrow M_{X_i}$ telle que : $X_i = f_i(S_i)$

IV.2. Formulation de l'approche d'explication.

Pour chaque variable X_i de Σ , il s'agit de trouver le ou les sous-ensembles explicatifs S_i de $\Sigma \setminus \{X_i\}$, qui fournissent le maximum d'information au sujet de X_i ; autrement dit :

$$I(X_i : S_i) = \max_{S \subset \Sigma \setminus \{X_i\}} I(X_i : S)$$

Du fait que l'information est croissante par rapport à l'inclusion, il est clair que $\Sigma \setminus \{X_i\}$ est un sous ensemble explicatif optimal. Pour pallier cette trivialité, il convient de considérer des sous-ensembles explicatifs de cardinal le plus faible possible. La recherche de solutions du problème considéré peut alors être conduite par niveaux (où chaque niveau est caractérisé par des sous-ensembles explicatifs de même cardinal).

Ainsi le problème ci-dessus peut être formulé comme suit:

Pour chaque variable X_i et chaque entier r , trouver les sous-ensembles S_i^r tels que :

le cardinal de S_i^r est égal à r et $I(X_i:S_i^r)$ est maximale.

IV.3. Cas déterministe.

Dans le cas d'un système déterministe, on a :

$$I(X_i:\Sigma\setminus\{X_i\}) = H(X_i)$$

Le sous ensemble explicatif optimal au niveau r de X_i est tel que :

$$I(X_i:S_i^r) = H(X_i).$$

Pour déterminer les sous-ensembles explicatifs optimaux, on considère une approche hiérarchique descendante. Celle-ci consiste à générer à partir de sous-ensembles explicatifs à r éléments des sous-ensembles à $r-1$, $r-2$, ... éléments.

A l'itération r , la quantité $I(X_i:S_i^r)$ étant connue, on cherche des sous-ensembles explicatifs à $r-1$ éléments par extraction d'une variable X de S_i^r de façon à ce que la quantité $I(X_i:S_i^r\setminus\{X\})$ soit maximale. Ce passage étant assuré par la proposition suivante:

Proposition:

Soit $S \subset \Sigma \setminus \{Y\}$ et $X \notin S$, $X \notin Y$.

Si $I(X:S\setminus\{X\}) = H(X)$ et $I(Y:S) = H(Y)$

alors: $I(Y:S\setminus X) = H(Y)$.

Preuve :

$$I(X:S\setminus X) = H(X) \iff H(S) = H(S\setminus X)$$

$$I(Y:S) = H(Y) \iff H(Y,S) = H(S).$$

$$\text{Donc : } H(Y,S) = H(S\setminus X).$$

Par ailleurs, puisque $I(X:S\setminus X) = H(X)$, on a:

$$I(X:Y \cup S\setminus X) = H(X) \quad \text{C'est à dire : } H(Y,S\setminus X) = H(Y,S).$$

Il en résulte que: $H(Y,S\setminus X) = H(S\setminus X)$, autrement dit:

$$I(Y:S\setminus X) = H(Y).$$

Si on pose $S = S_i^r$ et $Y = X_i$, on obtient:

$$I(X:S_i^r \setminus X) = H(X)$$

$$\text{et} \qquad \qquad \qquad \implies I(X_i:S_i^r \setminus X) = H(X_i)$$

$$I(X_i:S_i^r) = H(X_i)$$

Pour tout $X \subset \Sigma$ tel que $X \subset S_i^r$ et $X \neq X_i$

Ainsi les sous-ensembles explicatifs optimaux au niveau $r-1$ sont alors définis par:

$$S_i^{r-1} = S_i^r \setminus \{X\} \text{ tel que } X \in S_i^r \text{ et } I(X:S_i^r \setminus X) = H(X).$$

Algorithme :

Début

$r \leftarrow N-1$

tant que $r \geq 1$ faire

$i \leftarrow 1$

 tant que $i \leq N$ faire

$j \neq i$

 tant que $j \leq N$ faire

 Si $I(X_i:S_i^r)$ pour $X_i \in S_i^r$ n'est pas connu,
 le calculer connaissant $I(X_j:S_i^r \cup X_i)$;

 déduire $I(X_j:S_i^r)$.

 Si $I(X_i:S_i^r) < H(X_i)$ et $I(X_j:S_i^r \cup \{X_i\}) = H(X_j)$

 calculer $I(X_j:S_i^r)$

$j \leftarrow j+1$

 ftq

$i \leftarrow i+1$

 ftq

$r = r+1$

ftq

écrire pour chaque X_i , $i = 1, 2, \dots, N$; $\{S_{ij}^r\} = S_i$,

l'ensemble de tous les sous-ensembles explicatifs de X_i .

Fin.

Optimalité de l'algorithme.

A chaque niveau r et pour chaque variable X_i , l'algorithme recherche parmi les sous-ensembles à r éléments de S_i^{r+1} , ceux qui vérifient $I(X_i:S_i^r)$ maximale. Cette recherche exhaustive assure l'optimalité de l'algorithme.

Après avoir déterminé pour chaque X_i , un sous-ensemble explicatif S_i , on peut procéder à la décomposition du système après analyse du graphe d'interconnexion ci-dessous :

$$G = (\Sigma, U_R) \quad \text{où} \quad U_R = \{(X_i, X_j) / X_i \in \Sigma, X_j \in S_i\}$$

L'arc (X_i, X_j) peut être interprété de la façon suivante :

" La variable X_j contribue seule ou avec d'autres variables à expliquer ou déterminer X_i "

Notons que l'intérêt d'un tel graphe réside dans son caractère multidimensionnel, qui tient compte des aspects échappant aux graphes classiques, construits à partir d'indices de couplage binaires.

IV.4. Cas non déterministe.

La méthode décrite ci-dessus s'avère insuffisante dans le cas non déterministe, du fait que les variables ne peuvent être expliquées totalement, autrement dit:

$$I(X_i : \Sigma \setminus X_i) < H(X_i).$$

Cette absence d'explication peut être due:

--> A un bruit important qui vient se superposer aux modalités des variables du système.

--> A un manque de variables dans la description du système (Système-Source), ou bien au fait que le système est trop complexe pour être appréhendé par un modèle déterministe.

Dans la suite, nous décrirons une approche d'explication [Staroswiecki, Barboucha 87] - [Staroswiecki - Haouat 90]. Celle-ci vise à diviser l'ensemble des modalités d'une variable à expliquer X , en

un sous-ensemble de modalités bien expliquées et un sous-ensemble de modalités mal expliquées à cause du bruit ou dont l'explication requiert l'introduction de variables supplémentaires.

IV.4.1. Séparation des modalités mal expliquées.

Pour séparer les modalités mal expliquées des modalités bien expliquées, on teste les deux hypothèses suivantes :

H_0 : X est non expliquée à cause de la présence de bruit (Absence apparente d'explication).

H_1 : X est non expliquée à cause d'un manque de variables dans la description du système (Absence réelle d'explication).

Dans la suite, on dispose à partir du tableau des données du système, d'un tableau de contingence des probabilités conditionnelles de la variable X sachant un sous-ensemble explicatif S:

X	x ₁	...	x _j	...	x _n
S					

s ₁					
·			·		
·			·		
s _i		...	p _{j/i}	...	
·			·		
·			·		
s _m					

Les tests statistiques :

Pour utiliser les tests statistiques, on considère un modèle de comportement de bruit. Si on adopte un modèle de comportement uniforme, le profil de chaque ligne du tableau de contingence relatif à X et à un sous-ensemble explicatif S de X peut être estimée et représentée par:

$$P_0 = \left\{ \frac{1-p_i}{n-1}, \dots, p_i, \dots, \frac{1-p_i}{n-1} \right\} \text{ où } p_i = \text{Max}_j p_{j/i}$$

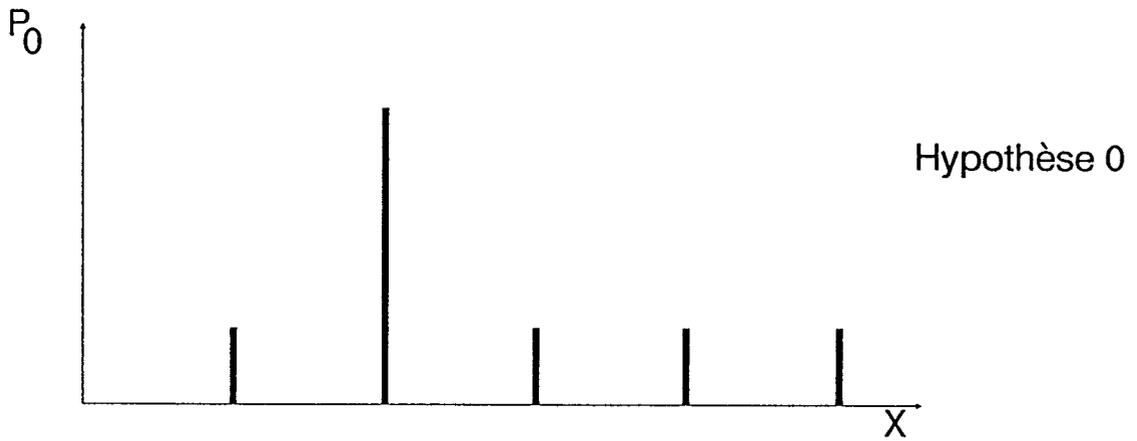


Fig II.1

Cette distribution correspond à l'hypothèse H_0 où X est non expliquée à cause du bruit.

La comparaison de P_0 à la distribution réelle :

$$P_1 = \{p_{1/i}, \dots, p_i, \dots, p_{n/i}\}:$$

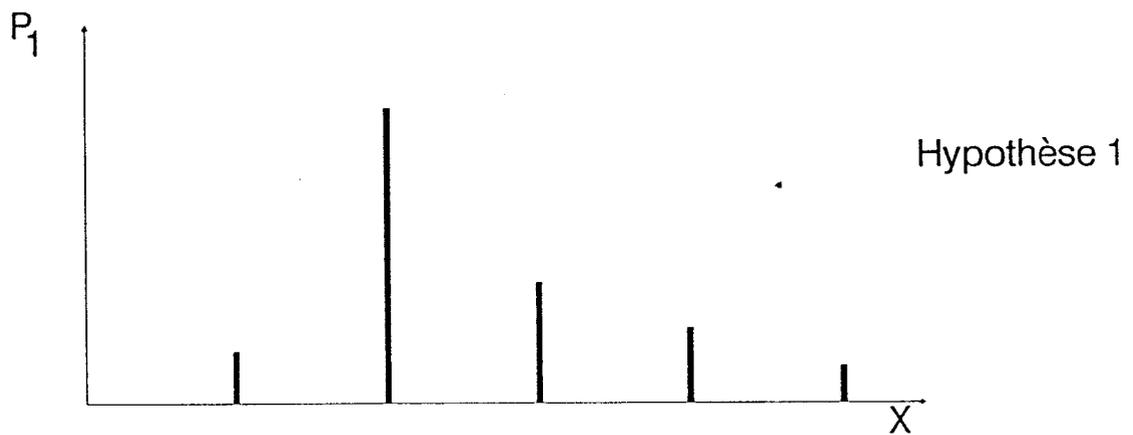


Fig II.2

au moyen de tests statistiques [Benzecri 80], tel le KHI^2 , permet de décider dans quelle mesure l'hypothèse du bruit H_0 peut être acceptée ou rejetée aux dépens de H_1 . Autrement dit, il s'agit d'appliquer la règle de décision suivante :

$$KHI^2(P_0, P_1) = \sum_{j=1}^n \frac{(p_{j/i} - (1-p_i)/(n-1))^2}{p_{j/i}}$$

$$KHI^2(P_0, P_1) \begin{cases} \leq \tau & X \text{ est non expliquée à cause du bruit } (H_0). \\ > \tau & X \text{ est non expliquée à cause d'un manque de} \\ & \text{variables explicatives } (H_1). \end{cases}$$

Où τ est un seuil en fonction duquel on peut déterminer la probabilité ($\text{pr}(KHI^2 \geq \tau)$) d'accepter ou de rejeter l'hypothèse H_0 . Cette quantité est extraite de la table du 2 en tenant compte du nombre des degrés de liberté n .

On peut également considérer l'information de Kullback pour accepter ou rejeter l'hypothèse H_0 au dépend de H_1 .

On a :

$$\begin{aligned} J(P_0, P_1) &= \sum_{j=1}^n q_i \log \frac{q_i}{p_{j/i}} \quad \text{où } q_i = \frac{1-p_i}{n-1} \\ &= (1-p_i) \log \left[\frac{1-p_i}{n-1} \right] - \left[\sum_{j=1}^n \log p_{j/i} \right] \left(\frac{1-p_i}{n-1} \right). \end{aligned}$$

$$J(P_0, P_1) \begin{cases} \leq \tau & X \text{ est non expliquée à cause du bruit } (H_0). \\ > \tau & X \text{ est non expliquée à cause d'un manque de} \\ & \text{variables explicatives } (H_1). \end{cases}$$

Cependant, l'utilisation de tels tests ne permet pas de conclure; c'est le cas par exemple où la ligne étudiée est totalement bruitée et ne correspond à aucun phénomène réel:

En effet, si l'on estime par $P_1 = (\frac{1}{n}, \dots, \dots, \frac{1}{n})$

le profil réel d'une telle ligne, on obtient :

$$\chi^2(P_0, P_1) = \frac{\sum_{j=1}^n \left[\frac{1}{n} - \frac{1 - 1/n}{n-1} \right]^2}{\frac{1}{n}} = 0$$

De même : $J(P_0, P_1) = 0$

Ce qui signifie que l'hypothèse H_0 est à retenir alors que l'on est manifestement dans H_1 . Pour pallier cet inconvénient, on évalue la variation entropique au niveau de chaque ligne du tableau de contingence.

Variation entropique :

Considérons les éléments de la distribution $P(X/S=s_j)$ suivant l'ordre décroissant et appelons X' la variable X après réunion de ses deux premières modalités :

Le calcul de l'entropie de X avant et après regroupement donne :

$$H(X/S=s_i) = \sum_{j=1}^n p_{j/i} \log p_{j/i}.$$

$$H(X'/S=s_i) = -(p_{1/i}+p_{2/i}) \log(p_{1/i}+p_{2/i}) - \sum_{i=1}^n p_{j/i} \log p_{j/i}$$

En notant $p_{2/i} = a \cdot p_{1/i}$ $0 \leq a \leq 1$, la variation entropique s'écrit :

$$\delta H = p_{1/i} ((1+a) \log(1+a) - a \log a).$$

Une telle fonction de a et de $p_{1/i}$, après normalisation peut être mise sous la forme :

$$f(a) = \delta H / 2p_{1/i} = 1/2 (1+a) \log(1+a) - a \log a .$$

Elle a l'allure suivante :

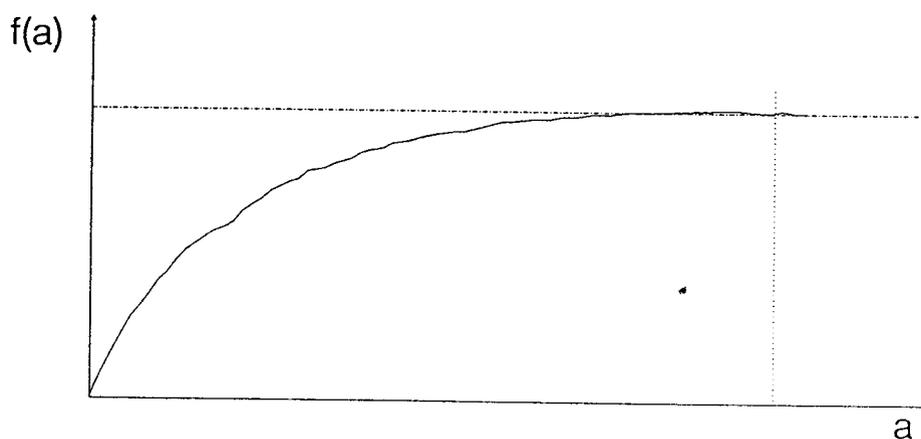


Fig II.3

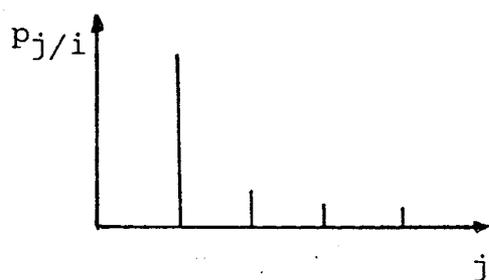
REMARQUES :

* Si a est voisin de zéro, la variation entropique tend vers zéro et les éléments de probabilités $p_{j/i}$ ($j > 1$) sont très faibles par rapport à $p_{1/i}$. Seul $p_{j/i}$ est largement supérieur à zéro. On peut dire que la modalité s_i de S explique bien la première modalité de X (Fig:A).

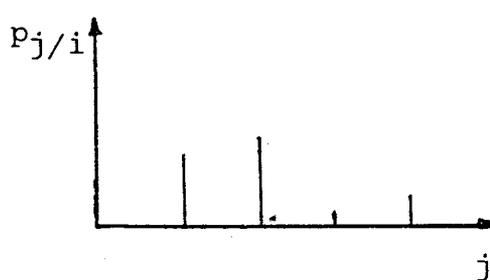
* Si a prend des valeurs non négligeables ($0 << a < 1$), en particulier lorsque $p_{1/i}$ et $p_{2/i}$ sont de même grandeur, alors deux cas sont envisageables conduisant à l'hypothèse H_1 :

-> $p_{1/i}$ et $p_{2/i}$ prennent des valeurs élevées (Fig:B); les deux premières modalités de X correspondent à deux phénomènes réels, la variable explicative S ne peut expliquer à elle seule X (Hypothèse H_1).

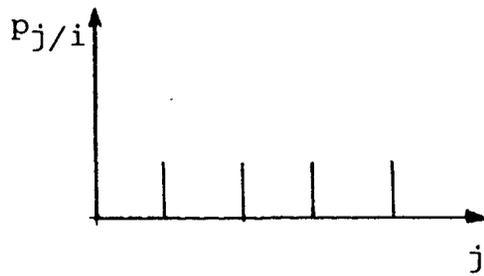
-> $p_{1/i}$ et $p_{2/i}$ prennent des valeurs faibles (Fig:C), il en est alors de même pour $p_{j/i}$ ($j > 3$). La ligne étudiée est totalement bruitée. Comme la variation entropique δH est non nulle, on peut décider de déclarer l'hypothèse H_1 vraie, compte tenu d'une valeur limite de δH .



(Fig:A)



(Fig:B)



(Fig:C)

Compte tenu des différents cas envisagés et du critère δH , on peut énoncer la règle de décision suivante:

$$\delta H \begin{cases} \leq \tau'' & X \text{ est non expliquée à cause du bruit } (H_0) \\ & \text{si } p_{1/i} \approx 0 \\ & X \text{ est expliquée par } S \text{ si } p_{1/i} \gg 0 \\ > \tau'' & X \text{ est non expliquée à cause d'un manque} \\ & \text{de variables explicatives } (H_1). \end{cases}$$

Ainsi , à l'issue de l'une des trois règles de décision considérées pour chaque ligne du tableau de contingence, on peut séparer l'ensemble des modalités de S en deux sous-ensembles :

$$M_S^1 = \{ s_i \in M_S / H_0 \text{ est accepté } \}$$

$$M_S^2 = \{ s_i \in M_S / H_1 \text{ est rejeté au dépend de } H_1 \}$$

M_S^1 est le sous-ensemble de modalités explicatives de S qui ne peuvent expliquer bien X, à cause du bruit.

M_S^2 est le sous-ensemble des modalités explicatives de S qui ne suffisent pas, à elles seules, à expliquer la partie non expliquée de X .

On peut alors rechercher des variables explicatives supplémentaires pour l'explication totale de X .

IV.4.2. Recherche de variables explicatives supplémentaires.

Dans la description du système au niveau 0 (système-source), seules les variables pertinentes sont prises en compte, alors que d'autres variables sont volontairement mises de côté (variables secondaires) pour diverses raisons: encombrement, coût élevé des mesures, augmentation de la dimension du système...

Si l'on suppose que les relevés des variables secondaires sont disponibles, on peut rechercher parmi ces variables celles qui permettent d'améliorer l'explication.

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

Pour obtenir des sous-ensembles explicatifs optimaux, on procède comme suit:

1 - Pour $s_i \in M_S^2$, $i=1, \dots, |M_S^2|$ on définit l'ensemble S_i des sous-ensembles explicatifs de X sachant que $S = s_i$. Autrement dit :

$$S_i = \{ R \subset Z / \delta H(X/ S=s_i, R=r_j) \leq \mu, \text{ pour tout } r_j \in M_R \}$$

Chaque sous-ensemble R de S_i contribue à réduire la variation de l'entropie de X sachant $S=s_i$.

Si $S_i = \phi$, aucun sous-ensemble de variables secondaires de Z ne contribue à expliquer la variable X sachant $S = s_i$

Si $S_i \neq \phi$, il existe un ou plusieurs sous-ensembles explicatifs de X sachant $S = s_i$.

2 - Pour chaque $i=1, \dots, |M_S^2|$, on considère un sous-ensemble explicatif de X de cardinal minimal, soit R .

3 - Un sous-ensemble explicatif S' de X est défini par la réunion des sous-ensembles R_i et de S . Autrement dit:

$$S' = S \cup R_1 \cup \dots \cup R_{|M_S^2|}.$$

V. CONCLUSION.

Dans ce chapitre, nous avons rappelé des indices issus de la théorie de l'information pour évaluer l'intensité de couplage entre les variables. De tels indices ont l'avantage de s'appliquer à tout type de variables (quantitatives, qualitatives, ...) et à tout type de relations (linéaires, non linéaires ...).

Nous avons décrit des méthodes de décomposition suivant des approches binaires et multidimensionnelles. Dans le premier cas, la mesure de couplage entre les variables deux à deux permet d'aboutir à un graphe d'interconnexion dont l'analyse conduit à une partition du système. Une telle approche se trouve limitée du fait qu'elle ne considère pas les liaisons entre plus de deux variables. Dans le deuxième cas, la prise en compte de couplage entre groupes de variables permet, à travers des algorithmes du type hiérarchique ou de programmation dynamique, de rechercher des partitions du système en sous-systèmes faiblement couplés. Cette approche nécessite un volume de calcul important, particulièrement lorsqu'il s'agit de traiter des systèmes de grande dimension.

Nous avons ensuite étudié une approche d'explication. Celle-ci consiste à rechercher pour chaque variable, un ou plusieurs sous-ensembles explicatifs. Une telle approche se trouve limitée dans le cas des systèmes soumis à des perturbations d'origines diverses. Dans ce cas, l'approche consiste d'abord à séparer pour chaque variable, les modalités bien expliquées, des modalités mal expliquées à cause du bruit ou dont l'explication nécessite l'introduction de variables explicatives supplémentaires.

CHAPITRE III :
ANALYSE DES SYSTEMES DYNAMIQUES PAR
LA T.I.
DETERMINATION DE L'ORDRE D'UN SYSTEME.

I. INTRODUCTION.

Ce chapitre porte sur l'analyse des systèmes dynamiques par la Théorie de l'Information.

Nous présentons d'abord quelques indices informationnels dynamiques, et nous étudions leurs propriétés de couplage.

De tels indices permettent d'aboutir à des partitions du système et à évaluer sa modélisabilité en fonction du nombre de périodes de l'horizon d'observation prises en compte.

Par ailleurs, dans le cadre de la détermination de l'ordre d'un système, nous exposerons quelques approches classiques et nous proposerons une approche basée sur un indice informationnel dynamique.

II. ANALYSE DES SYSTEMES DYNAMIQUES PAR LA T.I.

L'analyse des systèmes dynamiques nécessite la connaissance de l'information fournie par les variables au cours de leur évolution temporelle. Cette information est d'autant plus importante que l'horizon d'observation est plus large. Par ailleurs, dans la mesure où les liaisons entre variables sont traduites par une transmission d'information, il convient, pour évaluer l'intensité des liaisons, de considérer des mesures de couplage informationnel qui prennent en compte le caractère dynamique des variables:

II.1. Mesures de couplage dynamique.

On suppose que les variables du système sont connues par leur évolution sur un horizon de temps de p périodes; par ailleurs, on note respectivement par \bar{X}^p et X^p , la variable X observée successivement sur les intervalles :

$$[k-1, \dots, k-p]$$

$$[k, k-1, \dots, k-p]; \text{ pour } p \leq k \leq L.$$

Les coefficients de couplage décrits auparavant (Chap II.§2.1), peuvent être étendus au cas dynamique en considérant les valeurs prises par les variables sur un horizon décalé d'au moins une période.

Etant données X et Y deux variables quelconques de Σ , On obtient:

- le coefficient de couplage dynamique de Conant :

$$C^1(Y, X) = \frac{I(Y: X^1)}{H(Y)}$$

Ce coefficient mesure le degré de liaison dynamique entre deux variables considérées à deux instants successifs. Lorsque $C^1(Y, X)$ est nul, la variable Y , à un instant donné k , est indépendante de X considérée à l'instant $k-1$. D'autre part, lorsque $C^1(Y, X)$ est égal à un, il existe une liaison dynamique univoque de la forme : $Y(k) = f(X(k-1))$. [Conant.72]

- le coefficient de couplage dynamique de Richetin:

$$T^1(Y, X) = \frac{I(Y: \bar{X}^1)}{\sqrt{H(Y) H(\bar{X}^1)}}$$

Comme le coefficient $C^1(Y, X)$, $T^1(Y, X)$ mesure le degré de liaison dynamique entre deux variables. De plus, lorsque $T^1(Y, X)$ est égal à un, il existe une relation dynamique biunivoque de la forme $Y(k) = f(X(k-1))$.[Richetin.75]

- le coefficient de couplage dynamique de Rinck:

$$R^1(Y, X) = \frac{I_{\Sigma \setminus \{Y\}}(Y: \bar{X}^1)}{H(Y / \Sigma \setminus \{Y\})}$$

Le coefficient de couplage $R^1(Y, X)$ mesure le degré de liaison entre deux variables lorsque toutes les autres variables du système sont considérées constantes. Lorsque $R^1(Y, X)$ est égal à un, il existe une relation univoque telle que $Y(k) = f(S(k), X(k-1))$; (où $S = \Sigma \setminus Y$). Par contre, si $R^1(Y, X)$ est nul, Y et X ne sont pas nécessairement indépendants aux instants $k, k-1$.

Si l'on tient compte de l'évolution des variables sur p périodes, ces coefficients prennent la forme générale :

$$C^P(Y, X) = \frac{I(Y: X^P)}{H(Y)}$$

$$T^P(Y, X) = \frac{I(Y: X^P)}{\sqrt{H(Y) H(\bar{X}^P)}}$$

$$R^P(Y, X) = \frac{I_{\Sigma \setminus \{Y\}}(Y: \bar{X}^P)}{H(Y / \Sigma \setminus \{Y\})}$$

D'autre part, on peut considérer les coefficients :

$$AP(Y) = \frac{I(Y:\bar{Y}^P)}{H(Y)}$$

$$SP(Y,X) = \frac{I(Y:\bar{Y}^P, X^P)}{I(Y:Y^P)}$$

$$CP(Y,X) = \frac{I(Y:\bar{Y}^P, X^P)}{H(Y)}$$

Dans ce qui suit, nous donnons quelques propriétés de tels coefficients, ainsi que leurs interprétations.

II.2 Propriétés.

On montre que :

P_1 $0 \leq CP(Y,X) \leq 1$ de plus ;

$CP(Y,X) = 0$ si et seulement si X^P et Y sont indépendantes .

$CP(Y,X) = 1$ si et seulement si il existe une relation f telle que: $Y = f(X^P)$.

Les démonstrations découlent des propriétés de la transinformation.

Le coefficient $CP(Y,X)$ indique la possibilité d'obtenir un modèle d'ordre p de la forme :

$$Y(k) = f(X(k), X(k-1), \dots, X(k-p)).$$

On peut dire que le coefficient $CP(Y,X)$ permet de mettre en évidence des relations dynamiques du type moyenne mobile (M A).

$$P2 \quad CP(Y,X) \leq CP^{+1}(Y,X)$$

La démonstration résulte du fait que:

$$I(Y:XP) \leq I(Y:XP^{+1})$$

Cette propriété montre que lorsque p augmente, le couplage mesuré entre X et Y augmente.

P3 $0 \leq TP(Y,X) \leq 1$, de plus, On a :

$TP(Y,X) = 0$ si et seulement si XP et Y sont indépendantes.

On a : $TP(Y,X) = 1$ si et seulement si il existe une relation f telle que:

$$Y = f(XP).$$

Les démonstrations découlent des propriétés de la transinformation.

Le coefficient $TP(Y,X)$ indique la possibilité d'obtenir un modèle d'ordre p de la forme:

$$Y(k) = f(X(k), X(k-1), \dots, X(k-p)).$$

Ce qui correspond à un modèle du type M A.

Notons que les coefficients $CP(Y,X)$ et $TP(Y,X)$ permettent de mettre en évidence des relations similaires.

P₄ Si pour p suffisamment grand, $H(X^p)$ est constant, alors on montre que $T^p(Y, X)$ est croissant par rapport à p .

Démonstration:

Puisque $I(Y: X^p) \leq I(Y: X^{p+1})$ et pour p assez grand, $H(X^p)$ est constante par rapport à p (cas par exemple où la distribution de probabilité de X^p ne varie pas à partir d'un certain rang), on a :

$$H(X^p) = H(X^{p+1}); p > \bar{p} .$$

Par conséquent :

$$\frac{I(Y: X^p)}{\sqrt{H(Y) H(\bar{X}^p)}} \leq \frac{I(Y: X^{p+1})}{\sqrt{H(Y) H(\bar{X}^{p+1})}}$$

P₅ $0 \leq R^p(Y, X) \leq 1$ de plus,

$$R^p(Y, X) = 0 \text{ ssi } H(Y / \Sigma \setminus \{Y\}) = H(Y / X^p, \Sigma \setminus \{Y\})$$

$R^p(Y, X) = 1$ ssi il existe une relation f telle que:

$$Y = f(X^p, \Sigma \setminus \{Y\})$$

Lorsque $R^p(Y, X) = 0$, on peut dire que la variable X , seule, ne contribue pas à l'obtention d'une relation dynamique qui explique Y sur un horizon de p périodes.

On constate aussi que le coefficient généralisé de Rinck ne met en évidence que des liaisons statiques avec le reste des variables de Σ .

P₆ $R^p(Y, X)$ est croissant.

Du fait que $H(Y:\bar{X}^p, \Sigma \setminus \{Y\}) \geq H(Y:\bar{X}^{p+1}, \Sigma \setminus \{Y\})$, on a :

$$I_{\Sigma \setminus \{Y\}}(Y, \bar{X}^p) \leq I_{\Sigma \setminus \{Y\}}(Y, \bar{X}^{p+1}).$$

Par conséquent : $R^p(Y, X) \leq R^{p+1}(Y, X)$

P₇ $0 \leq A^p(Y) \leq 1$ de plus , on a :

$A^p(Y) = 1$ si et seulement si il existe une relation f telle que : $Y = f(\bar{Y}^p)$.

Un tel coefficient indique la possibilité d'obtenir un modèle d'ordre p de la forme $Y(k) = f(Y(k-1), \dots, Y(k-p))$, traduisant ainsi le comportement autonome de la variable Y .

Autrement dit, ce coefficient permet de mettre en évidence des liaisons du type auto-régressif (A R).

P₈ $A^p(Y)$ est croissant

La démonstration résulte du fait que:

$$I(Y: X^p) \leq I(Y: X^{p+1}) .$$

P₉ $1 \leq S^p(Y, X) \leq 1/A^p(Y)$ de plus,

$$S^p(Y, X) = 1 \text{ si et seulement si } I(Y: \bar{Y}^p, X^p) = I(Y: \bar{Y}^p)$$

$$S^p(Y, X) = 1/A^p(Y) \text{ si et seulement si } S^p(Y, X) = 1$$

si et seulement si il existe une relation f telle que:

$$Y = f(\bar{Y}^p, X^p)$$

On peut dire alors que $SP(Y,X)$ permet de mettre en évidence des liaisons du type auto-régressif à moyenne mobile (A R M A).

Démonstration :

puisque $I(Y:XP) \leq I(Y:\bar{Y}^P, XP)$, on a : $1 \leq SP(Y,X)$

Par ailleurs,

$$SP(Y,X) = \frac{I(Y:\bar{Y}^P, XP)}{H(Y)} * \frac{H(Y)}{I(Y:\bar{Y}^P)} \leq \frac{1}{AP(Y)}$$

Il y a égalité si et seulement si $SP(Y,X) = 1$.

si et seulement si il existe une relation f telle que :

$$Y = f(\bar{Y}^P, XP)$$

P₁₀ $SP(Y,X)$ est croissant si $AP(Y)$ est constant par rapport à p .

Démonstration :

Si $AP(Y)$ est constant par rapport à p , (cas où la distribution de \bar{Y}^P ne change pas),

$$\text{on a : } I(Y:\bar{Y}^P) = I(Y:\bar{Y}^{P+1})$$

$$\text{Comme } I(Y:\bar{Y}^P, XP) \leq I(Y:\bar{Y}^{P+1}, XP^{+1}),$$

$$\text{on obtient : } SP(Y,X) \leq SP^{+1}(Y,X).$$

P₁₁ $CP(Y,X) \leq SP(Y,X)$ et $AP(Y) \leq SP(Y,X)$

Démonstration :

$$c^p(Y, X) = \frac{I(Y: X^p)}{H(Y)} \leq \frac{I(Y: \bar{Y}^p, X^p)}{H(Y)} \leq \frac{I(Y: \bar{Y}^p, X^p)}{I(Y: \bar{Y}^p)} = s^p(Y, X)$$

De même:

$$A^p(Y) = \frac{I(Y: \bar{Y}^p)}{H(Y)} \leq \frac{I(Y: \bar{Y}^p, X^p)}{H(Y)} \leq \frac{I(Y: \bar{Y}^p, X^p)}{I(Y: \bar{Y}^p)} = s^p(Y, X)$$

$\overline{P_{12}}$ $c^p(Y, X)$ est croissant par rapport à p .

La démonstration découle du fait que:

$$I(Y: \bar{Y}^p, X^p) \leq I(Y: \bar{Y}^{p+1}, X^{p+1})$$

II.4. Décomposition des systèmes dynamiques.

L'analyse structurale des systèmes dynamiques apparaît plus complexe que celle des systèmes statiques. En effet, une extension directe des méthodes utilisées dans le cas statique ne conduit à une solution satisfaisante que si l'on dispose de suffisamment d'information au sujet du système, et d'indices appropriés pour évaluer les interactions dynamiques.

Dans le cadre d'une analyse binaire, la prise en compte d'indices dynamiques appliqués aux variables décalées d'une période seulement peut conduire à une partition non satisfaisante. Ceci est dû au fait que de tels indices ne

mesurent qu'une part des couplages dynamiques entre les variables comme le montre l'exemple suivant.

Exemple de décomposition.

Soit $\Sigma = \{X_1, X_2, X_3, X_4, X_5\}$ un système dynamique représenté comme suit :

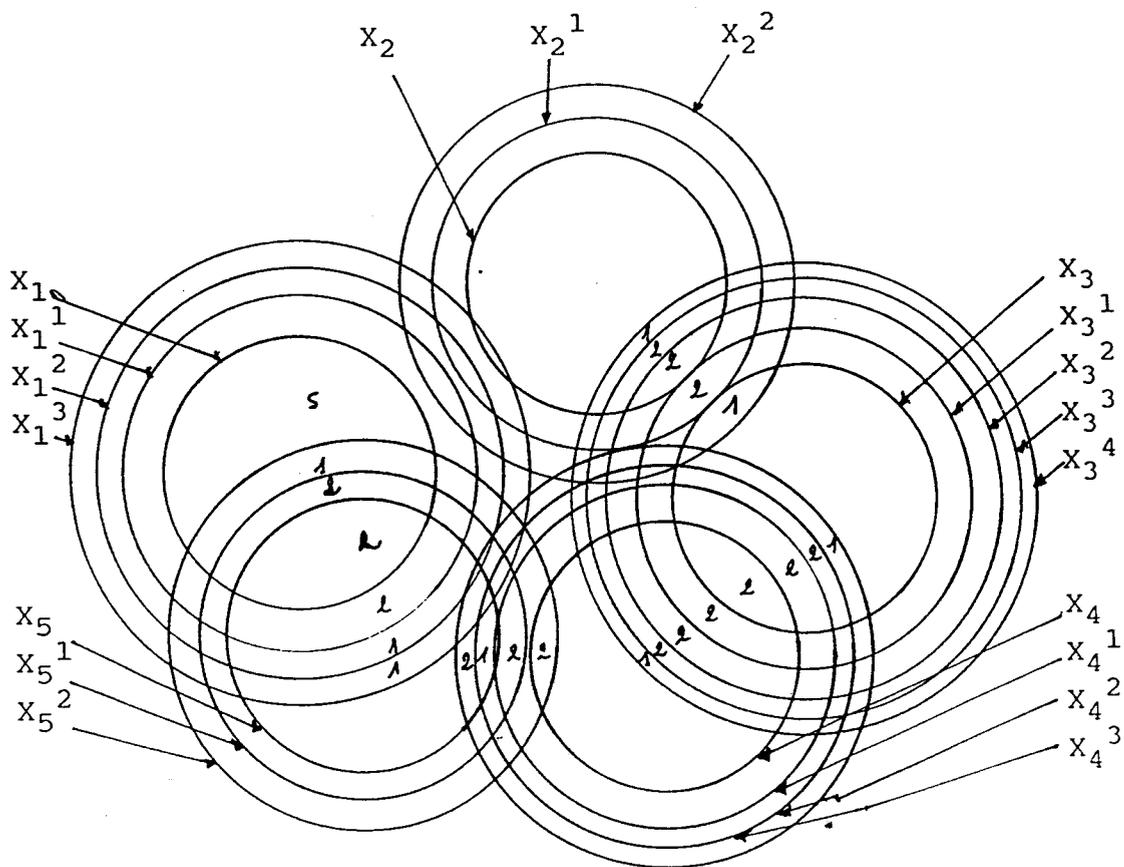


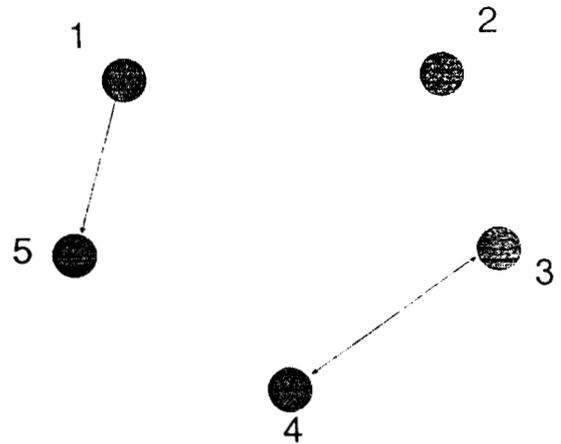
Fig III.1

Les différents cercles concentriques représentent les valeurs de l'entropie en fonction du nombre de périodes pris en compte. Une telle représentation peut être justifiée

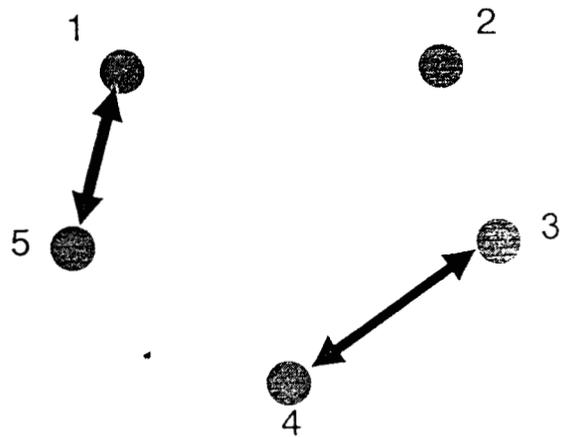
par le fait que l'entropie est monotone par rapport à l'inclusion ($H(X^p) \leq H(X^{p+1})$).

Considérons le coefficient dynamique d'ordre p : $CP(X_i, X_j)$. Les matrices de couplage associées pour $p = 0, 1, 2$, ainsi que leur graphe d'interconnexion sont représentés par:

	X_1	X_2	X_3	X_4	X_5
X_1	1	0	0	0	0.2
X_2	0	1	0	0	0
X_3	0	0	1	0.2	0
X_4	0	0	0.2	1	0
X_5	0.2	0	0	0	1



	X_1	X_2	X_3	X_4	X_5
X_1	1	0	0	0	0.4
X_2	0	1	0	0	0
X_3	0	0	1	0.4	0
X_4	0	0	0.4	1	0
X_5	0.4	0	0	0	1



	X_1	X_2	X_3	X_4	X_5
X_1	1	0	0	0	0.5
X_2	0	1	0.2	0	0
X_3	0	0.1	1	0.6	0
X_4	0	0	0.6	1	0
X_5	0.5	0	0	0.1	1

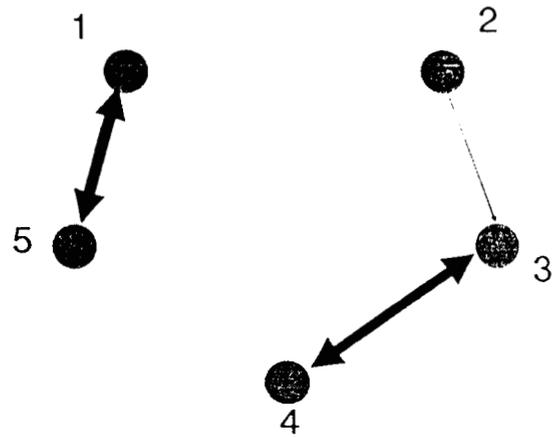


Fig III.2

On constate que les coefficients c^0 et c^1 ne rendent pas compte de l'interaction dynamiques entre X_2 et X_3 contrairement à c^2 , et engendrent ainsi une partition, différente de celle engendrée par c^2 . Cela s'explique par le fait que le degré des interactions dynamiques augmente à cause de l'augmentation de l'information fournie sur un intervalle de période plus grande.

II.5 Indices de modélisabilité dynamique.

Avant de procéder à la modélisation d'un système dynamique, il convient de disposer d'outils permettant de juger la faisabilité d'une telle phase. Pour cela, on considère des indices informationnels dynamiques ; ceux-ci, compte tenu de leurs propriétés, permettent de définir le concept de modélisabilité d'un système dynamique.

II.5.1 Définitions.

Soient $Y \in \Sigma$ et $S \in \Sigma \setminus \{Y\}$. On définit par :

$$c^p(Y, S) = \frac{I(Y:SP)}{H(Y)}$$

le degré de modélisabilité d'ordre p de Y par S .

$$A^p(Y) = \frac{I(Y:\bar{Y}^p)}{H(Y)}$$

le degré de modélisabilité d'ordre p de Y par elle même.

$$\bar{c}^p(Y, S) = \frac{I(Y:\bar{Y}^p, SP)}{H(Y)}$$

le degré de modélisabilité d'ordre p de Y par elle même et par S .

En particulier, $\bar{c}^p(Y, \Sigma \setminus \{Y\})$ est le degré de modélisabilité d'ordre p de Y .

$$\bar{c}^p(\Sigma) = \frac{1}{N} * \frac{\sum_{i=1}^N c^p(Y_i, \Sigma \setminus \{Y_i\})}{\sum_{i=1}^N H(Y_i)}$$

est l'indice moyen de modélisabilité d'ordre p du système.

$$\bar{e}^p(\Sigma) = \frac{\sum_{i=1}^N I(Y_i: \bar{Y}_i^p, S_i^p)}{\sum_{i=1}^N H(Y_i)} \quad \text{où } S_i = \Sigma \setminus \{Y_i\}$$

est le degré de modélisabilité du système à l'ordre p.

En particulier, si Y_i désigne une sortie du système et S_i un sous-ensemble explicatif de Y_i , formé à partir de variables d'entrées et de sorties du système, $\bar{e}^p(\Sigma)$ est le degré de modélisabilité du système par des relations du type A R M A d'ordre p.

II.5.2 Propriétés :

* Une valeur voisine de 1 de $CP(Y,S)$ indique la possibilité d'obtenir une relation entre Y et S de la forme:

$$Y(k) = f(S(k), S(k-1), \dots, S(k-p))$$

* Si $AP(Y) \approx 1$, alors il existe une relation de récurrence g telle que :

$$Y(k) = g(Y(k-1), \dots, Y(k-p))$$

* Si $\bar{e}^p(\Sigma) \approx 1$, le système peut être modélisé par une relation de la forme :

$$\left\{ \begin{array}{l} Y_1(k) = f_1(Y_1(k-1), \dots, Y_1(k-p), S_1(k), S_1(k-1), \dots, S_1(k-p)) \\ Y_2(k) = f_2(Y_2(k-1), \dots, Y_2(k-p), S_2(k), S_2(k-1), \dots, S_2(k-p)) \\ \cdot \\ \cdot \\ \cdot \\ Y_N(k) = f_N(Y_N(k-1), \dots, Y_N(k-p), S_N(k), S_N(k-1), \dots, S_N(k-p)) \end{array} \right.$$

Dans le cas où les variables Y_i désignent les sorties du système et S_i les sous-ensembles explicatifs de Y_i , on obtient un modèle du système défini à partir d'un ensemble de relations partielles (A R M A). Celles-ci forment des modèles des sous-systèmes: $\{Y_1, S_1\}$, $\{Y_2, S_2\}$, \dots , $\{Y_S, S_S\}$ (s : nombre des sorties du système).

II.5.3. Exemple d'application.

Considérons à titre indicatif, le système illustré dans l'exemple ci-dessous :

$$\Sigma = \{X_1, X_2, X_3, X_4\}.$$

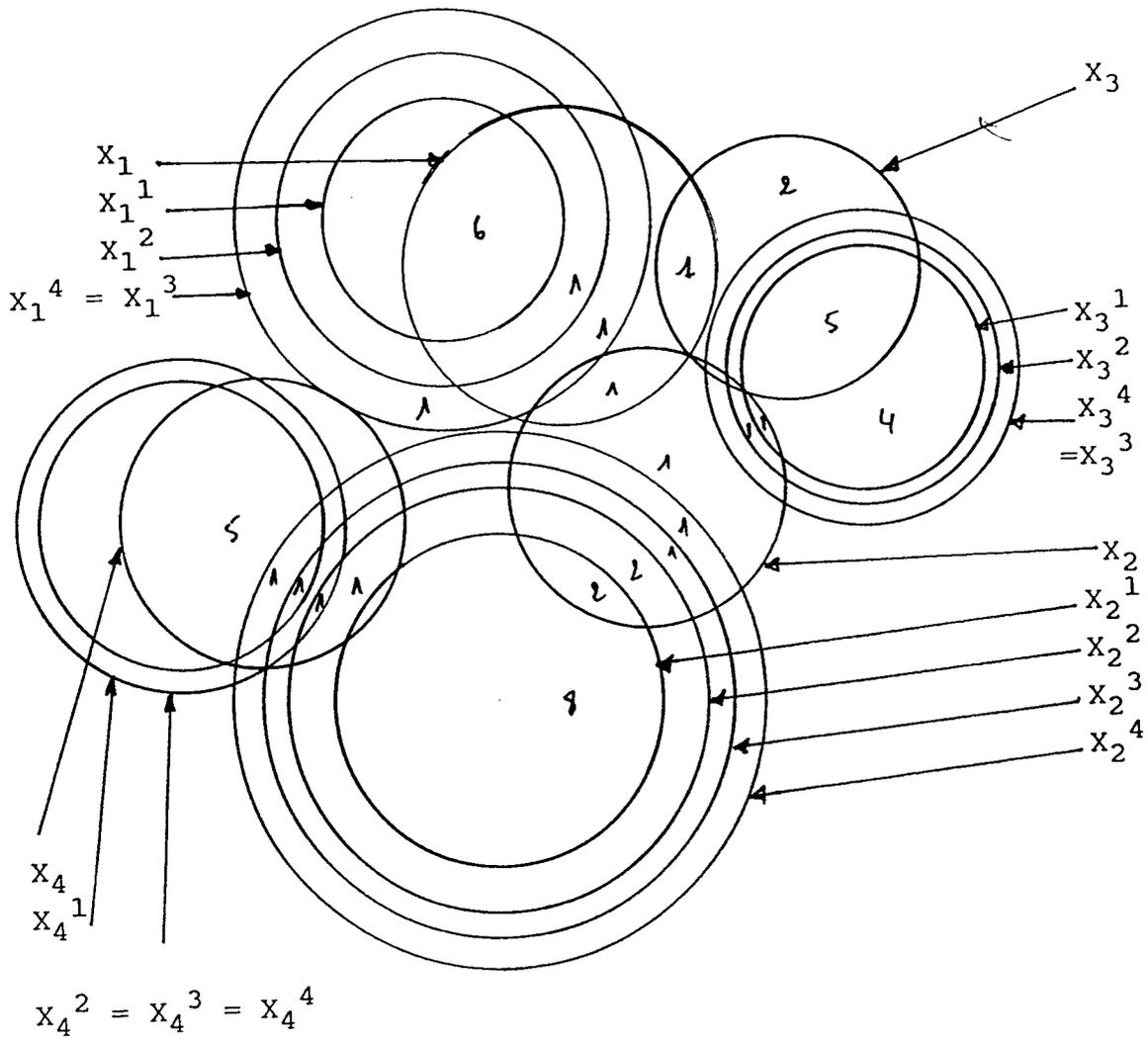


Fig III.3

On a : $H(X_1) = H(X_2) = H(X_3) = H(X_4) = 10$

Le degré de modélisabilité de chaque variable et du système sont représentés dans le tableau suivant :

$\Sigma \backslash p$	$c^0(X_i, S_i)$	$c^1(X_i, S_i)$	$c^2(X_i, S_i)$	$c^3(X_i, S_i)$	$c^4(X_i, S_i)$
X_1	20%	60%	70%	80%	80%
X_2	10%	40%	70%	90%	90%
X_3	20%	80%	90%	100%	100%
X_4	0%	60%	90%	90%	90%
$CP(\Sigma)$	15%	62.5%	80%	90%	90%
$CP(\Sigma)$	15%	62.5%	80%	90%	90%

Les sous-ensembles S_i sont tels que : $S_i = \Sigma \setminus \{X_i\}$

Le système Σ est modélisable à 90% à l'ordre 3. Le degré de non modélisabilité s'évalue à 10% , cela est dû soit à un manque de variable pour la modélisation en particulier de variables X_1, X_2, X_4 ,soit au bruit.

$$\begin{aligned} \text{Comme: } c^3(X_1, S_1) &= c^3(X_1, X_3) \\ c^3(X_2, S_2) &= c^3(X_2, X_1, X_3) \\ c^3(X_3, S_3) &= c^3(X_3, X_1, X_2) \\ c^2(X_4, S_4) &= c^2(X_4, X_2) \end{aligned}$$

$\{X_3\}$, $\{X_1, X_3\}$ et $\{X_1, X_2\}$ constituent des sous-ensembles explicatifs de X_1, X_2 et X_3 à l'ordre 3. $\{X_2\}$ un sous-ensemble explicatif de X_4 à l'ordre 2.

le système peut alors être modélisé sous la forme :

$$\left\{ \begin{array}{l}
 X_1(k) = f_1(X_1(k-1), X_1(k-2), X_1(k-3), \\
 \quad X_3(k), X_3(k-1), X_3(k-2), X_3(k-3), \delta_1) \\
 \\
 X_2(k) = f_2(X_2(k-1), X_2(k-2), X_2(k-3), X_1(k), X_1(k-1), \\
 \quad X_1(k-2), X_1(k-3), X_3(k), X_3(k-1), \\
 \quad X_3(k-2), X_3(k-3), \delta_2) \\
 \\
 X_3(k) = f_3(X_3(k-1), X_3(k-2), X_3(k-3), X_1(k), X_1(k-1), \\
 \quad X_1(k-2), X_1(k-3), X_2(k), \\
 \quad X_2(k-1), X_2(k-2), X_2(k-3)) \\
 \\
 X_4(k) = f_4(X_4(k-1), X_4(k-2), X_2(k), X_2(k-1), X_2(k-2), \delta_3) .
 \end{array} \right.$$

où δ_1 , δ_2 et δ_3 représentent les variables manquantes ou le bruit.

III. DETERMINATION DE L'ORDRE D'UN SYSTEME.

III.1 Introduction.

Dans cette partie, nous nous intéressons à la détermination de l'ordre d'un système au moyen de la Théorie de l'Information. Nous exposerons brièvement quelques approches utilisant un modèle probabiliste ou un modèle analytique dont il faut identifier les paramètres. Nous proposons ensuite une approche basée sur l'utilisation d'un coefficient informationnel dynamique.

III.2 Approches utilisant un modèle.

III.2.1 Approche d'Akaike. [Akaike 73.81]

L'approche proposée par Akaike pour identifier l'ordre d'un système suppose que le système est représenté par un modèle connu sous forme de densité de probabilité: $f(x, \theta^k)$, où x représente le vecteur des valeurs prises par les variables du système et θ^k , un vecteur de paramètres inconnus, de dimension k inconnue.

Pour identifier l'ordre du système, Akaike considère une estimation $\tilde{f}(x, \tilde{\theta}^k)$ de $f(x, \theta^k)$ à partir de L observations. Cette fonction de densité est comparée à $f(x, \theta^k)$ au moyen de la mesure d'inexactitude (Chap I.§III.3.4):

$$A(f, \tilde{f}) = - \int_{M_X} f(x, \theta^k) \log \tilde{f}(x, \tilde{\theta}^k) dx.$$

$$= E_f[-\log \tilde{f}(x, \tilde{\theta}^k)]$$

où E_f est l'espérance mathématique relative à f et $\tilde{\theta}^k$ un sous-vecteur de θ^k ($\tilde{k} \leq k$).

L'ordre du système est l'entier \tilde{k} qui minimise le critère ci-dessus.

Dans la suite, nous décrirons brièvement deux variantes pour la mise en oeuvre de cette approche. Celles-ci s'appuient sur le choix d'une estimation du critère $A(f, \tilde{f})$.

Estimation par la fonction de vraisemblance.

Compte-tenu de la 2° égalité ci-dessus, Akaike propose l'estimation définie par:

$$\tilde{A}(f, \tilde{f}) = - \frac{1}{L} \sum_{i=1}^L \log \tilde{f}(x_i, \tilde{\theta}^k)$$

où x_i est la $i^{\text{ème}}$ valeur observée de X .

Cette estimation peut être interprétée comme mesure d'inexactitude discrète de la distribution $(\tilde{f}(x_i, \tilde{\theta}^k))_{1 \leq i \leq L}$ par rapport à celle du système, considérée uniforme.

Une telle estimation ne peut conduire à la détermination d'un ordre satisfaisant que si les données suivent une loi uniforme; ce qui n'est pas le cas en général. Afin d'améliorer la procédure, Akaike propose une estimation de $A(f, \tilde{f})$ à partir de la mesure de Kullback:

Estimation à partir de la mesure de Kullback :

Compte-tenu de la relation définie auparavant (Chap I. §III.3.4) :

$$A(f, \tilde{f}) = H(f) + J(f, \tilde{f})$$

où $H(f) = - \int f(x, \theta^k) \log f(x, \tilde{\theta}^k) dx$ est l'entropie du système

$$\text{et } J(f, \tilde{f}) = \int_{M_X} f(x, \theta^k) \log \frac{f(x, \theta^k)}{f(x, \tilde{\theta}^k)} dx$$

est la mesure de Kullback.

Minimiser $A(f, \tilde{f})$ par rapport à $\tilde{\theta}^k$ et \tilde{k} revient à minimiser $J(f, \tilde{f})$.

En estimant d'une part, $J(f, \tilde{f})$ par la moyenne du rapport de vraisemblance évalué à partir des L observations, autrement dit :

$$\tilde{J}(f, \tilde{f}) = \frac{1}{L} \sum_{i=1}^L \log \frac{f(x_i, \theta^k)}{f(x_i, \tilde{\theta}^k)}$$

et en considérant d'autre part que $2N\tilde{J}(f, \tilde{f})$ est asymptotiquement distribué suivant la loi du KHI^2 à $k - \tilde{k}$ degrés de liberté, Akaike propose d'estimer $A(f, \tilde{f})$ par l'expression :

$$\bar{A}(f, \tilde{f}) = - \frac{1}{L} \sum_{i=1}^L \log \tilde{f}(x_i, \tilde{\theta}^k) - \frac{k - \tilde{k}}{2L}$$

Cette estimation a la particularité d'être sans biais et conduit à déterminer un ordre satisfaisant par rapport à l'estimation $\tilde{A}(f, \tilde{f})$.

Signalons que l'efficacité de l'approche d'Akaike dépend de la qualité de l'estimation du critère $A(f, f)$, prise en compte. Comme l'indique Kashyap [Kashyap 80], pour obtenir une approximation correcte de l'ordre du système, il faut que le biais de l'estimation considérée soit de la forme $k \times \phi(N)$ où ϕ est une fonction de L vérifiant les conditions suivantes :

$$\begin{aligned} & \cdot \phi(L) > 0 \\ & \quad \phi(L) \\ & \cdot \lim_{L \rightarrow \infty} \frac{\phi(L)}{L} = 0 \end{aligned}$$

Notons par ailleurs que l'approche d'Akaike utilise un modèle statistique sous forme de loi de probabilité paramétrique et que pour déterminer l'ordre du système, il est nécessaire d'identifier les paramètres du modèle.

III.2.2. Approche de Rissanen [Rissanen 76].

L'approche de Rissanen, pour identifier l'ordre d'un système peut être formulée de la façon suivante :

Considérons $X = F(\theta_n)$ un vecteur de variables observables du système liées à un vecteur de paramètres θ_n , où n est inconnu. Les paramètres θ_n et X sont supposés prendre leurs valeurs suivant une distribution donnée.

On note $\theta_k = (\theta_1, \theta_2, \dots, \theta_k)$ où $k \leq n$, un sous-vecteur de k paramètres.

A partir de L observations, on considère une estimation $\tilde{\theta}_k$ de $\bar{\theta}_k$ et \hat{X} de X .

L'ordre du système suivant l'approche de Rissanen est l'entier k minimisant l'entropie conjointe:

$$H(\hat{X}, \tilde{\theta}_k) = - \int_{M_X} f(\hat{x}, \tilde{\theta}_k) \log f(\hat{x}, \tilde{\theta}_k) d\hat{x}$$

où $f(\hat{x}, \tilde{\theta}_k)$ est la densité de la variable $\hat{X} = X - \tilde{X}$.

En supposant que la fonction de densité $f(\hat{x}, \tilde{\theta}_k)$ suit une loi normale, on montre que : minimiser l'entropie conjointe par rapport à \tilde{k} , revient à minimiser l'expression:

$$K(\hat{X}, \tilde{\theta}_k) = \log(\det V_k(\hat{X})) + \frac{1}{L} \sum_{i=1}^{\tilde{k}} \log \sigma_{\theta_i}^2 + \frac{\tilde{k}}{L} (1 + \log 2\pi)$$

où $V_k(\hat{X})$ est la matrice de variance-covariance de \hat{X} , et $\sigma_{\theta_i}^2$ est la variance de θ_i estimées à partir de L observations.

Signalons que dans cette approche, les distributions de probabilité de X et de θ_n sont supposées connues; dans le cas contraire, Rissanen propose de choisir une distribution de X et de θ_n qui engendre une entropie maximale.

Notons par ailleurs que pour déterminer l'ordre du système, il est nécessaire d'identifier les paramètres $\tilde{\theta}_k$.

Un exemple d'application.

Considérons un système monovariante d'entrée X et de sortie Y , représenté par un modèle linéaire d'ordre k de la forme:

$$X(t) = \sum_{i=1}^k \theta_i Y(t-i) + \theta_0 Y(t).$$

On suppose vérifiées les conditions suivantes:

Le système est stationnaire.

X est une variable gaussienne de variance égale à 1.

$Y(t-i)$ et $Y(t-j)$ sont indépendantes pour $i \neq j$, $i \neq 0$, $j \neq 0$.

$Y(t-i)$ et $X(t)$ sont indépendantes pour $i > 0$.

$$Y(t) \text{ est de moyenne nulle et } \|\theta_k\|^2 = \sum_{i=1}^k \theta_i^2 < 1, \theta_0 \neq 0$$

Notons $V_0 = \text{cov}(Y(t), X(t)), t > t_0$ et $V_k = E(Y, Y)$, la matrice de variance-covariance de Y sur k périodes.

Du fait de la linéarité du modèle, Y suit une loi gaussienne ($f(y, \theta_k)$) de moyenne nulle et de variance σ^2 , dépendant des paramètres θ_i et telle que:

$$\sigma^2 = \frac{\theta_0^2}{1 - \|\theta_k\|^2}$$

Considérons un horizon d'observation $[1, L]$. A partir d'un échantillon $\{ Y_1, Y_2, \dots, Y_L \}$, on estime un sous-vecteur de paramètres $\theta_k = (\theta_1, \dots, \theta_k)$ où $k < k$, soit θ_k . De même, on estime la variance σ^2 par:

$$\tilde{\sigma}^2 = \frac{1}{L-1} \sum_{i=1}^L Y_i^2$$

Par ailleurs, si le modèle était d'ordre \tilde{k} , Y suivrait une loi gaussienne ($f(y, \tilde{\theta}_k)$) de variance:

$$\tilde{\sigma}^2 = \frac{\tilde{\theta}_0^2}{1 - \|\tilde{\theta}_k\|^2}$$

L'estimation de $A(f, \tilde{f})$ à partir de la fonction de vraisemblance d'une part, et de la mesure de Kullback d'autre part, s'expriment respectivement par:

$$\tilde{A}(f, \tilde{f}) = \log \sqrt{2\pi\sigma} + 1/2$$

et

$$\bar{A}(f, \tilde{f}) = \log \sqrt{2\pi\sigma} + 1/2 - (k-k)/2L.$$

L'erreur entre $A(f, \tilde{f})$ et $\tilde{A}(f, \tilde{f})$ s'écrit:

$$A(f, \tilde{f}) - \tilde{A}(f, \tilde{f}) = 1/2 (\sigma^2 / \tilde{\sigma}^2 - 1)$$

De même:

$$A(f, \tilde{f}) - \bar{A}(f, \tilde{f}) = 1/2 [\sigma^2 / \tilde{\sigma}^2 - (k - k)/L].$$

Par ailleurs, le calcul de l'entropie conduit à:

$$H(Y, \theta_k) = \log \sqrt{2\pi e\sigma}$$

L'ordre du système est le plus petit entier k , rendant minimale l'erreur (Approches d'AKAIKE) ou l'entropie (Approche de RISSANEN).

III.2.3. Approche de Dufour: [Dufour 85]

L'approche de Dufour pour la détermination de l'ordre d'un système suppose que celui-ci est monovariante et représenté par un modèle exprimé sous la forme suivante :

$$Y(k) = \sum_{i=1}^n a_i Y(k-i) + b_n Y(k).$$

Le système est supposé stationnaire et l'entrée gaussienne.

Dufour utilise le calcul de l'entropie et la transinformation pour identifier les paramètres et la variance de la sortie d'un modèle d'ordre donné n . L'ordre



du système est obtenu par comparaison des variances des sorties système-modèle.

Cette approche peut être résumée comme suit:

Supposons que le système est d'ordre n_0 .

A partir des données du système, on calcule successivement :

1 - La variance σ_Y^2 de la sortie-système en fonction de l'entropie de Y.

2 - Pour une suite de modèles de la forme ci-dessus et d'ordre croissant: 1,2, ..., n_0 , n_0+1 , ..., on estime la variance σ_n^2 de la sortie-modèle en fonction des paramètres (ceux-ci étant identifiés en fonction de l'entropie et de la transinformation).

3 - On compare σ_Y et σ_n . L'ordre à attribuer au système étant l'entier minimal pour lequel le rapport:

$$\frac{\sigma_Y}{\sigma_n} \text{ ---- prend une valeur voisine de 1.}$$

Notons que cette approche s'applique seulement aux systèmes monovariabiles représentés par un modèle linéaire. De plus, elle nécessite l'identification des paramètres du modèle au moyen de la transinformation en considérant que l'entrée du système est gaussienne.

III.3. Approche utilisant l'indice dynamique \bar{C}^p

L'approche que nous présentons dans ce paragraphe diffère de celles citées auparavant par le fait qu'elle n'utilise pas une classe particulière de modèles

analytiques ou statistiques. De plus, elle s'applique à des systèmes multivariés sans hypothèses particulières au sujet des variables (gaussiennes par exemple).

III.3.1. Présentation de l'approche :

On considère que le système est stationnaire et que l'ensemble de ses variables est divisé en deux groupes : un groupe de variables à expliquer, \mathcal{Y} , et un groupe de variables explicatives, \mathcal{S} .

On suppose que les variables prennent leurs modalités sur un horizon d'observation $[0, T, 2T, \dots, LT]$.

Pour chaque variable à expliquer Y , on considère un sous-ensemble explicatif optimal de cardinal minimal, obtenu à partir de l'indice:

$$\bar{C}^p(Y) = \frac{I(Y; \bar{Y}^p, S^p)}{H(Y)}$$

On suppose que $p \in [0, p^u]$, où p^u est un entier largement inférieur à L .

Pour déterminer l'ordre du système, on considère une approche qui s'appuie sur le fait que les indices $\bar{C}^p(Y)$ sont croissants par rapport à p et tendent vers une valeur stationnaire égale à $\bar{C}^{p^u}(Y)$. (Fig III.4.)

L'ordre du système est l'entier p^* défini par:

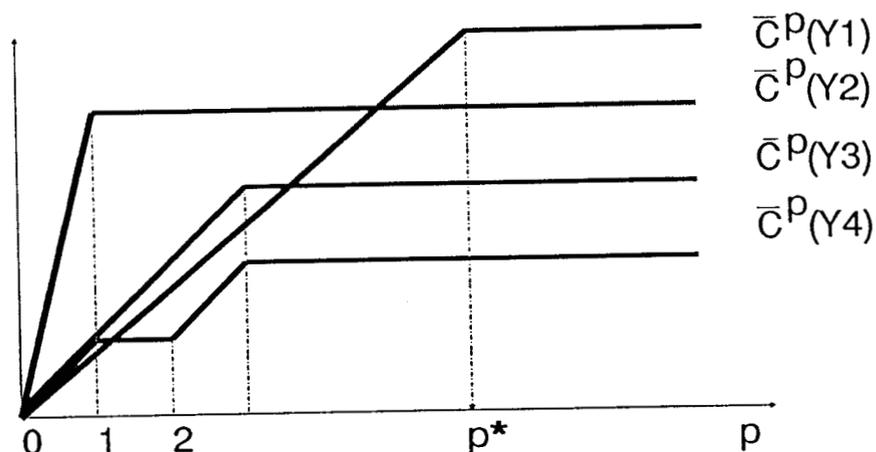
$$\text{Max}_{Y \in \mathcal{Y}} \text{Min}_{p \leq N} \bar{C}^p(Y).$$

En effet, pour chaque variable Y , une valeur maximale de $\bar{C}^p(Y)$ indique que les valeurs de Y peuvent être

déterminées à partir de p périodes antérieures à un instant donné. L'ordre du système est justement le nombre de périodes nécessaires à la détermination de ces valeurs.

Illustration:

Y_1, Y_2, \dots, Y_m désignent m variables explicatives dont l'indice dynamique correspondant évolue comme suit:



FigIII.4

L'ordre du système est égal à p^* , le plus petit entier pour lequel $\bar{C}^P(Y_i)$ est maximal pour tout $i=1, 2, \dots, m$.

III.3.2. Procédures de détermination de l'ordre d'un système.

Désormais, on notera $\bar{C}(Y)$ au lieu de $\bar{C}^{p^*}(Y)$.

Il s'agit de comparer pour chaque rang p , les coefficients $\bar{C}^P(Y)$ à $\bar{C}(Y)$. L'ordre du système est l'entier p^* , à partir duquel les écarts:

$$E^P(Y) = \bar{C}(Y) - \bar{C}^P(Y)$$

sont stationnaires.

Notons que $E^P(Y)$ peut être interprété comme la perte d'information au sujet de Y due à l'attribution au système de l'ordre p . Ainsi l'ordre du système est l'entier minimum, acceptable, pour lequel la perte d'information au sujet de chaque variable Y de \mathcal{Y} est minimale.

III.3.2.1. Algorithme I.

Début

$i \leftarrow 1$

Tant que $i < \text{card}(\mathcal{Y})$ faire

$p \leftarrow 0$

Tant que $p < p_m$ faire

Calculer $\bar{C}^P(Y_i)$ et en déduire $E^P(Y_i)$.

Si $E^P(Y_i) > \mu$ alors $p = p+1$

Sinon $p = p^u$

Ftq

Ecrire $p = p[i]$

$i = i+1$

Ftq

Calculer le plus grand entier p^* parmi

$p[1], p[2], \dots, p[\text{card}(\mathcal{Y})]$

L'ordre du système est égal à p^* .

Fin.

Pour chaque variable Y_i , l'algorithme détermine le plus petit entier $p[i]$ pour lequel la perte de l'information au sujet de Y_i est minimale.

Une autre procédure pour déterminer l'ordre d'un système consiste à minimiser la variation de l'information

fournie au sujet d'une variable Y , sur deux entiers successifs. Cette variation peut être évaluée par:

$$e^p(Y) = \bar{c}^p(Y) - \bar{c}^{p-1}(Y).$$

Notons que:

$$\begin{aligned} e^p(Y) &= \frac{I(Y:\bar{Y}^p, S^p) - I(Y:\bar{Y}^{p-1}, S^{p-1})}{H(Y)} \\ &= \frac{H(Y/\bar{Y}^{p-1}, S^{p-1}) - H(Y/\bar{Y}^p, S^p)}{H(Y)} \\ &= \frac{H(Y/\bar{Y}^p, S^p) - H(Y/\bar{Y}^{p-1}, S^{p-1}, Y(p), S(p))}{H(Y)} \end{aligned}$$

Donc:

$$e^p(Y) = \frac{I(Y:Y(p), S(p))}{\bar{Y}^{p-1}, S^{p-1} H(Y)}$$

Où $Y(p), S(p)$ sont les variables Y et S décalées de p périodes.

Compte tenu de ce critère, on peut déterminer l'ordre d'un système à partir de l'algorithme suivant:

III.3.2.2. Algorithme II.

Début

$i \leftarrow 1$

Tant que $i < \text{card}(Y)$ faire

$p \leftarrow 0$

Tant que $p < p^u$ faire

Calculer $e^p(Y_i)$.

Si $e^p(Y_i) > \mu'$ alors $p = p+1$

Sinon $p = [i-1]$

Ftq

Ecrire $p = p[i]$

$i = i+1$

Ftq

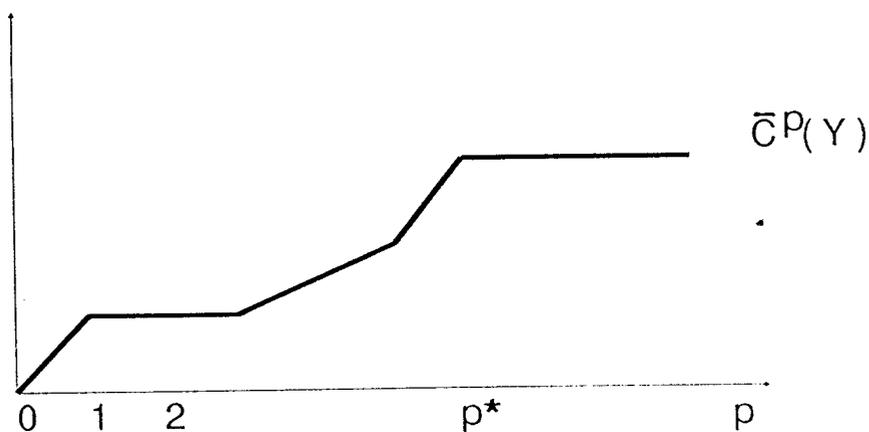
Calculer le plus grand entier p^* parmi

$p[1], p[2], \dots, p[\text{card}(Y)]$

L'ordre du système est égal à p^*-1 .

Fin.

L'algorithme II détermine pour chaque variable Y_i , le plus petit entier pour lequel l'information reste inchangée. A ce rang, il n'est pas toujours sûr que l'information au sujet de Y_i soit maximale, comme l'illustre la figure ci-dessous:



FigIII.5

On a: $e^1(Y)=0$; $e^2(Y)=e^3(Y)=0$.

L'algorithme II peut fournir dans ce cas seulement une solution sous-optimale. Pour remédier à cela, on peut envisager de calculer successivement $e^P(Y_i)$ à partir de p^u .

Considérons alors l'algorithme suivant:

III.3.2.3. Cas particulier (Algorithme III).

Début

$i \leftarrow 1$

Tant que $i < \text{card}(\mathcal{Y})$ faire

$p \leftarrow p_m$

Tant que $0 \leq p < p_m$ faire

Calculer $e^P(Y_i)$.

Si $e^P(Y_i) < \mu^n$ alors $p = p-1$

Sinon, l'ordre du système est p^u .

Ftq

Ecrire $p = p[i]$

$i = i+1$

Ftq

Calculer le plus grand entier p^* parmi

$p[1], p[2], \dots, p[\text{card}(\mathcal{Y})]$

L'ordre du système est égal à p^*-1 .

Fin.

Remarque

Compte tenu des propriétés des indices informationnels, les critères $E^P(Y)$ et $e^P(Y)$ pourront être calculés de façon itérative si on dispose d'hypothèses restrictives au sujet des variables (variables gaussiennes, indépendantes sur deux instants successifs par exemple).

Notons que l'approche utilisant l'indice dynamique, que nous venons de décrire, nécessite un tableau de données dynamiques. Ce tableau peut être généré à partir de l'observation du système. Dans ce qui suit, nous allons supposer que les données dont on dispose sont issues d'un système linéaire d'ordre k , défini par les équations suivantes:

$$X_{t+1} = AX_t + BU_t.$$

$$Y_t = CX_t + E_t.$$

Où:

$X \in \mathbb{R}^n$ est le vecteur d'état du système.

$U \in \mathbb{R}^q$ est le vecteur d'entrée.

$Y \in \mathbb{R}$ est la sortie.

$E \in \mathbb{R}$ est le bruit.

C'est une variable aléatoire de fonction d'autocorrélation nulle ($\text{Cov}(E_{t-i}, E_{t-j}) = \delta_{ij} e$)

$A \in L(\mathbb{R}^n, \mathbb{R}^n)$ est la matrice d'état du système.

$B \in L(\mathbb{R}^n, \mathbb{R}^q)$ est la matrice d'entrée.

Le but visé dans ce paragraphe, est de montrer que l'approche étudiée plus haut permet de retrouver l'ordre du système. Cela revient à montrer que le coefficient \bar{C}^p vérifie:

$$\dots \leq \bar{C}^{k-2} \leq \bar{C}^{k-1} < \bar{C}^k = \bar{C}^{k+1} = \dots$$

Les équations ci-dessus, écrites aux instants $t, t-1, \dots, t-p$, donnent:

$$\begin{bmatrix} Y_{t-p} \\ \cdot \\ \cdot \\ \cdot \\ Y_t \end{bmatrix} = \begin{bmatrix} C & 0 & 0 \\ CA & CB & \\ \cdot & \cdot & \cdot \\ \cdot & X_{t-p} + \cdot & \cdot \\ \cdot & \cdot & \cdot \\ CAP & CAP^{-1}B \dots CB & 0 \end{bmatrix} \begin{bmatrix} U_{t-p} \\ \cdot \\ \cdot \\ \cdot \\ U_t \end{bmatrix} + \begin{bmatrix} E_{t-p} \\ \cdot \\ \cdot \\ \cdot \\ E_t \end{bmatrix}$$

Ce qui peut être mis sous la forme:

$$Y_t^P = O^P X_{t-p} + C^P U_t^P + E_t^P.$$

Compte tenu du théorème de Cayley-Hamilton, on peut mettre Y_t sous la forme:

$$Y_t = \tilde{\alpha} \bar{Y}_t^k + \tilde{\beta} U_t^k + \tilde{\Gamma} E_t^k.$$

Où k est l'ordre du système, $\tilde{\alpha}$, $\tilde{\beta}$ et $\tilde{\Gamma}$ sont des vecteurs non nuls.

Pour montrer que le coefficient dynamique devient stationnaire, on considère 2 cas:

Cas déterministe: $E = 0$

On a: $Y_t = \tilde{\alpha} \bar{Y}_t^k + \tilde{\beta} U_t^k$

Une telle relation implique que:

$H(Y_t/\bar{Y}_t^k, U_t^k) = 0$, Autrement dit: $\bar{C}^k(Y) = 1$

Par ailleurs, on a:

$H(Y_t/\bar{Y}_t^{k-1}, U_t^{k-1}) > 0$.

Démonstration:

$H(Y_t/\bar{Y}_t^k, U_t^k) = H(Y_t/\bar{Y}_t^{k-1}, U_t^{k-1}) + \delta_t^k$.

Où $\delta_t^k = I \bar{Y}_t^{k-1} U_t^{k-1} (Y_t, Y_{t-k})$.

On a: $\delta_t^k \neq 0$

En effet:

$\delta_t^k = 0$ si et seulement si Y_t et Y_{t-k} sont indépendantes (pour \bar{Y}^{k-1} et U^{k-1} fixés).

Ceci est contradictoire avec le fait que le système est d'ordre k .

D'autre part, $H(Y_t/\bar{Y}_t^k, U_t^k) = 0$

Il en résulte que $H(Y_t/\bar{Y}_t^{k-1}, U_t^{k-1}) > 0$. Autrement dit:

$$\bar{c}^{k-1} < \bar{c}^k.$$

Cas non déterministe: $E \neq 0$

On a:

$$Y_t = \tilde{\alpha} \bar{Y}_t^k + \tilde{\beta} U_t^k + \tilde{\gamma} E_t^k.$$

Comme $H(Y_t/\bar{Y}_t^k, U_t^k) = H(E_t^k)$

$$H(Y_{t+1}/\bar{Y}_{t+1}^{k+1}, U_{t+1}^{k+1}) = H(E_{t+1}^{k+1}),$$

on obtient compte tenu de la stationnarité de E et du fait que $\text{Cov}(E_{t-i}, E_{t-j}) = \delta_{ij} \text{var}(E)$:

$$H(Y_t/\bar{Y}_t^k, U_t^k) = k H(E)$$

$$H(Y_{t+1}/\bar{Y}_{t+1}^{k+1}, U_{t+1}^{k+1}) = (k+1)H(E), \text{ pour } t > \bar{t}$$

$$\text{Ainsi } \bar{c}^k - \bar{c}^{k+1} = \frac{(k+1)H(E) - kH(E)}{H(Y)} \geq \frac{H(E)}{H(Y)}$$

puisque \bar{c}^p est croissant par rapport à p on a: $\bar{c}^k = \bar{c}^{k+1}$.

Par ailleurs, on montre de façon similaire au cas déterministe que:

$$\bar{c}^{k-1} < \bar{c}^k.$$

IV. CONCLUSION.

Dans ce chapitre, nous nous sommes intéressés à l'analyse des systèmes dynamiques au moyen de la T.I.

Après avoir présenté un certain nombre d'indices informationnels dynamiques, nous avons étudié leurs propriétés. De tels indices permettent de traduire le comportement dynamique des variables et d'évaluer le couplage dynamique entre elles, en fonction du nombre de périodes pris en compte.

Les propriétés d'indices informationnels dynamiques ont permis également d'envisager :

--> La décomposition des systèmes dynamiques en fonction du nombre de périodes.

--> D'étudier le concept de modélisabilité dynamique.

--> Et enfin de procéder à la détermination de l'ordre d'un système. Dans cette partie, nous avons rappelé quelques approches classiques et proposé ensuite une procédure informationnelle pour déterminer l'ordre d'un système.

CHAPITRE IV :
EVALUATION DE LA QUALITE D'UN MODELE
PAR LA THEORIE DE L'INFORMATION.

I. INTRODUCTION.

Une des phases essentielles de l'analyse des systèmes consiste à rechercher le ou les modèles qui sont les plus représentatifs du système.

Dans la plupart des cas, la pluralité des critères et leur diversité, font que le système peut être représenté par différents modèles, si bien que l'analyste se trouve confronté au choix du "meilleur" modèle à retenir pour représenter le système.

Dans ce chapitre, nous nous intéressons à deux approches informationnelles pour évaluer la qualité d'un modèle :

-la première approche consiste à comparer l'information extraite du modèle à celle connue au sujet du système.

-la deuxième consiste à évaluer au moyen d'indices informationnels la divergence entre les données fournies par le modèle et celles du système.

Pour ce faire, nous définissons les partitions induites par un modèle qui s'exprime sous forme de relations d'entrée-sortie, et nous étudions leurs propriétés. De telles partitions peuvent être rendues similaires au modèle dans le sens où l'information extraite du modèle est celle contenue dans la partition qu'il induit.

Nous considérons ensuite des indices issus de la T.I de Shannon et de Kullback appliqués aux partitions pour

évaluer la qualité d'un modèle, ainsi que la performance d'un modèle par rapport à un autre.

Enfin, nous étudions les deux approches dans le cas d'un modèle linéaire, que nous illustrons à travers un exemple.

II. VARIABLES ET PARTITIONS.

Rappelons que Σ est l'ensemble des variables du système et que Ω est l'ensemble des observations.

II.1. Partition associée à une variable.

II.1.1. Définition.

Soit $\mathcal{P}(\Omega)$ l'ensemble des parties de Ω , supposé fini, de cardinal L .

On dit que $\mathcal{P}_k(\Omega) = \{P_1, P_2, \dots, P_k\}$ est une partition en k classes de Ω ssi :

$$- P_i \in \mathcal{P}(\Omega) , P_i \neq \emptyset, i = 1, 2, \dots, k$$

$$- P_i \cap P_j = \emptyset, i \neq j.$$

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

On note par $\mathcal{P}_k(\Omega)$, l'ensemble des partitions de Ω en k classes et par $\mathcal{P}(\Omega)$ l'ensemble de toutes les partitions de Ω . On a :

$$\mathcal{P}(\Omega) = \bigcup_{i=1}^L \mathcal{P}_i(\Omega).$$

Chaque variable X (resp sous-ensemble S) d'un ensemble Σ définit une partition de Ω , notée P_X (resp P_S) et telle que :

$$P_X = \{X^{-1}(x)/x \in M_X\} = \{\{w \in \Omega/X(w) = x\}, x \in M_X\}.$$

Cette partition est obtenue par regroupement dans une même classe de tous les w auxquels X attribue une même modalité.

Notons aussi que P_X ne dépend pas des modalités de X , mais seulement du fait qu'elles sont distinctes.

II.1.2. Quelques propriétés.

* A toute variable de Σ , on peut associer de façon univoque une partition de Ω . Ce qui permet formellement d'identifier Σ à une partie de $\mathcal{P}(\Omega)$ d'une part, et d'autre part, de considérer $\mathcal{P}(\Omega)$ comme l'ensemble de toutes les variables définies sur Ω .

* On peut établir dans $\mathcal{P}(\Omega)$ la relation: " plus fine que " (α) telle que:

$$P = \{P_1, P_2, \dots, P_k\}, Q = \{Q_1, Q_2, \dots, Q_l\}$$

$P \alpha Q$ si et seulement si pour tout $Q_i \in Q$, il existe $P_j \in P$ tel que : $P_j \alpha Q_i$.

Cette relation qui est un ordre partiel, permet de munir $\mathcal{P}(\Omega)$ d'une structure de treillis. [Birk 64].

* A partir d'une telle relation et de la correspondance entre variable et partition, on peut définir un ordre et une équivalence entre les variables de Σ ; en effet:

$$\begin{array}{ll} X: \Omega \text{ -----} > M_X & Y: \Omega \text{ -----} > M_Y. \\ w \text{ -----} > X(w) & w \text{ -----} > Y(w). \end{array}$$

On a :

X est équivalente à Y si et seulement si $P_X = P_Y$.

Dans ce cas, il est facile de montrer qu'il existe une relation biunivoque f de M_X vers M_Y telle que:

$$Y = f(X) \text{ et } X = f^{-1}(Y).$$

Ainsi, deux variables sont équivalentes si et seulement si l'une s'obtient par transformation biunivoque de l'autre.

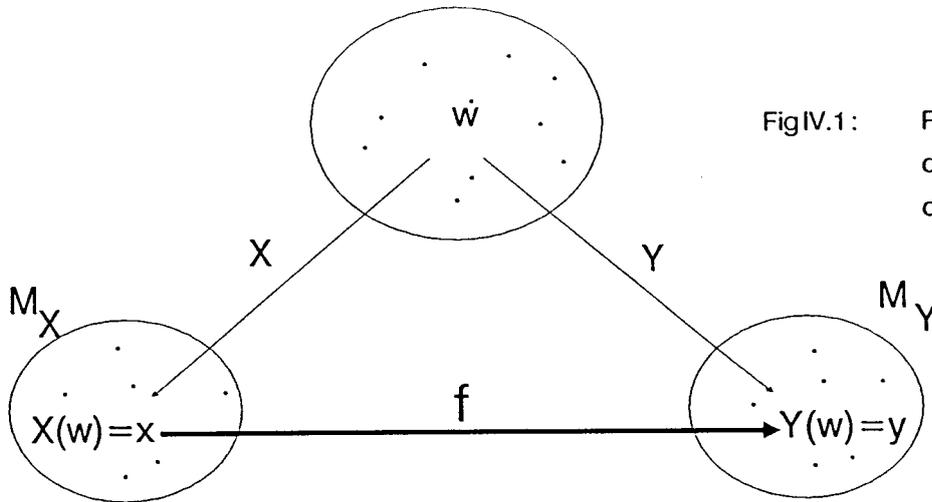
* X est plus fine que $Y \iff P_X \alpha P_Y$.

Dans ce cas, on peut montrer qu'il existe une fonction f de M_X vers M_Y telle que : $Y = f(X)$.

II.2 Partition associée à une relation: $f(X)=Y$.

Soient X, Y deux variables de Σ et f une relation fonctionnelle définie de M_X vers M_Y .

X, Y et f peuvent être représentées comme suit :



FigIV.1: Représentation
d'une relation
du type: $f(X)=Y$

II.2.1 Partition de l'ensemble d'observation.

A chaque relation f , on peut associer une partition de Ω définie par :

$$P_f = \{ (f \circ X)^{-1}(y), y \in f(M_X) \} = \{ \{ w / f(X(w)) = y \}, y \in f(M_X) \}$$

Chaque classe de P_f est obtenue à partir d'une modalité y de Y , par regroupement de toutes les observations des antécédants de y .

Notons qu'à toute relation f est associée une partition P_f unique et que si une relation g est obtenue par transformation biunivoque de f , alors g engendre la même partition que f .

On peut également remarquer que si $f(M_X) = M_Y$, la partition P_f contient autant de classes qu'il y a de modalités de Y .

Proposition:

Soit f une relation définie de M_X vers M_Y et P_f la partition induite par f , alors:

P_f est moins fine que P_X .

De plus, P_f et P_X sont équivalentes si et seulement si f est une relation biunivoque.

Démonstration:

Soit $X^{-1}(x)$ une classe de P_X , on va montrer qu'il existe une classe P^j de P_f telle que: $X^{-1}(x) \subset P^j$.

Posons $y = f(x)$ et $P^j = \{w \in \Omega / f(X(w)) = y\}$.

$w \in X^{-1}(x) \rightarrow X(w) = x \rightarrow f(X(w)) = f(x) = y \rightarrow w \in P^j$

Donc : $X^{-1}(x) \subset P^j$.

II.2.2 Partition de l'ensemble des modalités de Y

Pour définir une partition de M_Y associée à f , on considère une partition quelconque de M_X que l'on note:

$$P(M_X) = \{M_X^1, \dots, M_X^n\}.$$

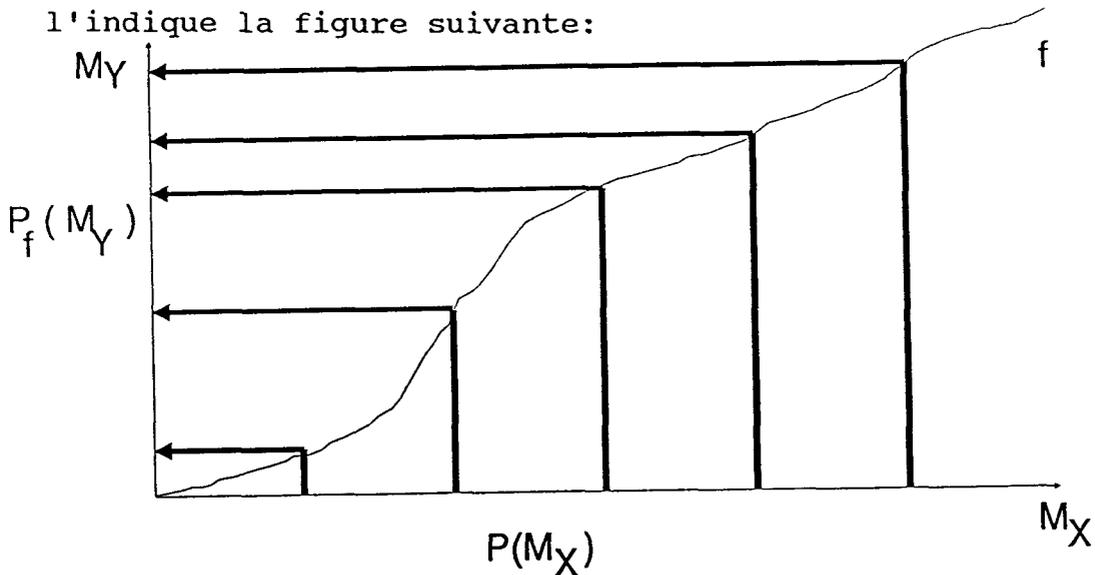
On distingue deux cas, suivant que f est biunivoque ou non:

-Si f est biunivoque, la partition induite s'écrit:

$$P_f(M_Y) = \{ f(M_X^1), f(M_X^2), \dots, f(M_X^n) \}.$$

Chaque classe de $P_f(M_Y)$ regroupe les images des modalités de X qui appartiennent à une même classe de $P(M_X)$.

Une telle partition peut être schématisée comme l'indique la figure suivante:



FigIV.2

Notons que c'est justement l'hypothèse "f biunivoque" qui assure que $P_f(M_Y)$ est une partition, en effet:

- 1°: Montrons que $f(M_X^i) \cap f(M_X^j) = \emptyset$ pour tout $i \neq j$:

$$\text{Si } f(M_X^i) \cap f(M_X^j) = \emptyset,$$

il existe $y \in f(M_X^i) \cap f(M_X^j)$. Par conséquent,

il existe $x' \in M_X^i$ et $x'' \in M_X^j$ tels que:

$$y = f(x') = f(x'').$$

Comme f est biunivoque, on a : $x' = x''$.

Donc: $M_X^i \cap M_X^j \neq \emptyset$, ce qui est contradictoire.

- 2° : Montrons que $M_Y = \bigcup_{i=1}^n f(M_X^i)$

Si $y \in M_Y$, il existe $x \in M_X$ tel que: $y = f(x)$.

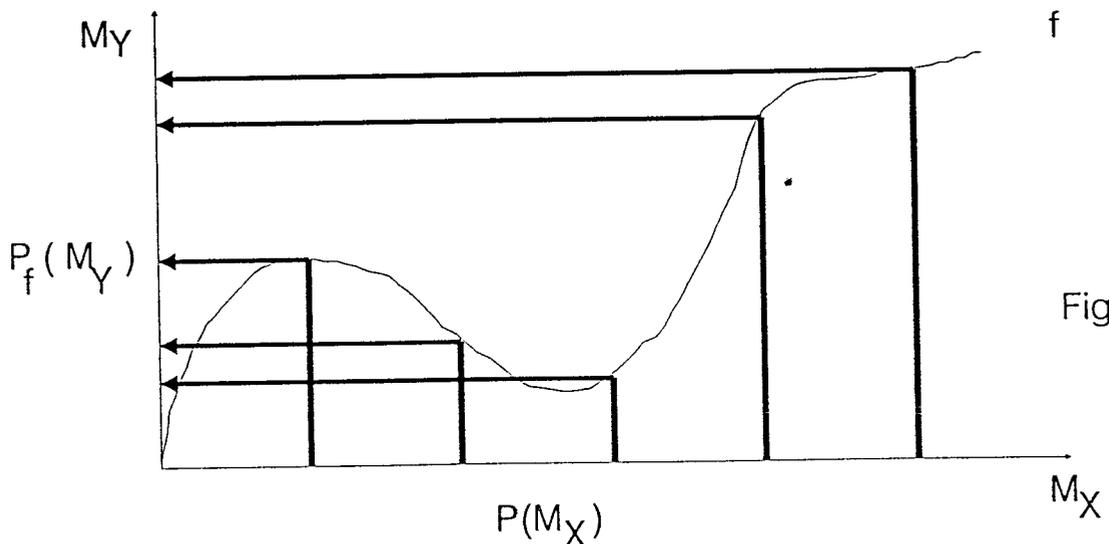
Puisque $M_X = \bigcup_{i=1}^n M_X^i$, il existe i tel que $x \in M_X^i$.

Donc: $y \in f(M_X^i)$.

Il en résulte que : $M_Y = \bigcup_{i=1}^n f(M_X^i)$.

-Si f n'est pas biunivoque, la partition de M_Y induite par f peut être définie par extraction d'une partition du recouvrement $\{f(M_X^i)\}$ de M_Y . A titre indicatif, on peut considérer la partition la plus fine de ce recouvrement. (Rappelons q'un recouvrement de M_Y est une famille de classes de M_Y non nécessairement disjointes).

Une telle partition peut être schématisée comme suit:



Proposition:

$P_f(M_Y)$ et $P'_f(M_Y)$ sont deux partitions induites par f respectivement à partir des partitions $P(M_X)$ et $P'(M_X)$ de M_X .

Si $P(M_X) \alpha P'(M_X)$ alors, $P_f(M_Y) \alpha P'_f(M_Y)$

Démonstration:

Soit $P'^j_f(M_Y)$ est une classe de $P'_f(M_Y)$:

$$P'^j_f(M_Y) = \{ y \in M_Y / f(y) \in P'^j(M_X) \}$$

Comme $P(M_X) < P'(M_X)$, on a: $P'^j(M_X) = \bigcup_k P^{jk}(M_X)$

$$\begin{aligned} \text{Alors, } P'^j_f(M_Y) &= \{ y \in M_Y / f(y) \in \bigcup_k P'^{jk}(M_X) \} \\ &= \bigcup_k \{ y \in M_Y / f(y) \in \bigcup_k P^{jk}(M_X) \} \\ &= \bigcup_k P_f^{jk}(M_Y) \end{aligned}$$

Autrement dit: $P_f(M_Y) \alpha P'_f(M_Y)$.

III.2.3. Partition de l'ensemble des modalités de X

Considérons une partition $P(M_Y)$ de M_Y que l'on note:

$$P(M_Y) = \{M_Y^1, \dots, M_Y^m\}.$$

On définit la partition de M_X associée à f par :

$$P_f(M_X) = \{ f^{-1}(M_Y^1), \dots, f^{-1}(M_Y^m) \}.$$

Une telle partition peut être représenté par le schéma suivant :

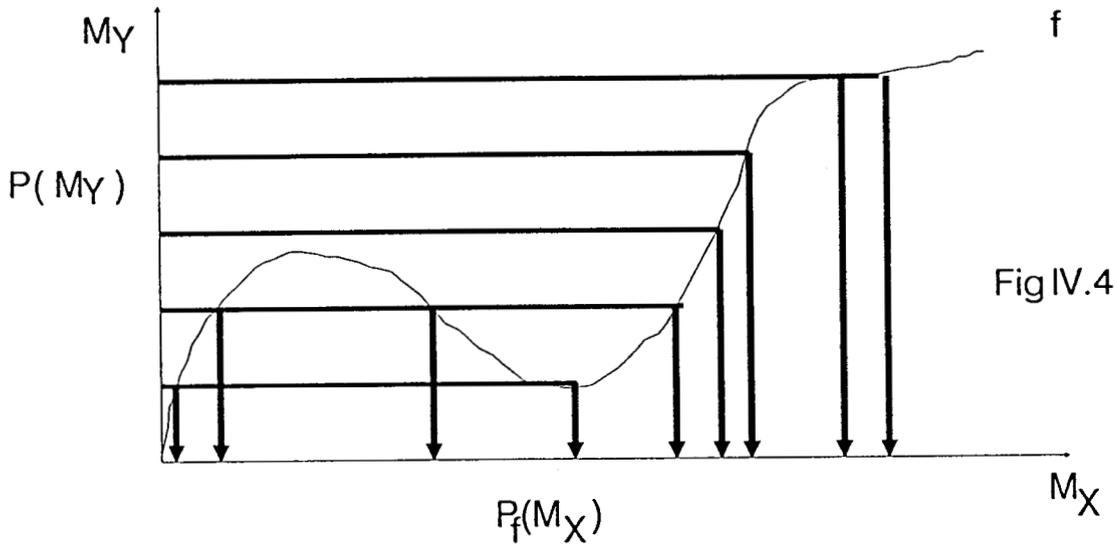


Fig IV.4

Notons que lorsque $P(M_Y)$ est la partition atomique de M_Y , $P_f(M_X)$ est la partition dont chaque classe regroupe les modalités de X liées par f à une modalité de Y .

Cette partition peut être considérée équivalente à la partition $P_f(\Omega)$ puisque par définition, à chaque classe de $P_f(M_X)$ correspond une classe de $P_f(\Omega)$ et réciproquement.

II.3 Pourquoi s'intéresse-t-on aux partitions induites?

Nous venons de voir qu'à partir d'un modèle f défini de M_X vers M_Y , on peut définir une partition de l'espace d'observation ou des modalités. Cette correspondance permet de rendre similaire le modèle f à la partition qu'il induit.

Ainsi en termes informationnels, la quantité d'information que l'on extrait d'un modèle peut être considérée égale à celle fournie par la partition qu'il engendre, ce qui autorise à évaluer la qualité d'un modèle à partir d'indices informationnels appliqués aux partitions.

III. QUALITE D'UN MODELE.

Dans le cadre de la modélisation d'un système, l'analyste pose le problème de l'évaluation de la qualité du modèle à retenir pour représenter le système. Ce problème nécessite la donnée d'un critère permettant de comparer le modèle au système.

Dans ce travail, nous considérons des critères issus de la théorie de l'information.

On désigne respectivement par : P_X , P_Y , P_f , les partitions induites par X, Y et f ; où f est un modèle défini de M_X vers M_Y .

III.1. Indices informationnels de Shannon.

Pour évaluer la qualité du modèle f , on considère l'indice défini par:

$$d_I(f) = I(P_Y : P_f).$$

$d_I(f)$ peut être interprété comme la quantité d'information extraite de la variable X par le modèle au sujet de Y . Cette quantité est inférieure ou égale à l'information transmise entre X et Y .

Notons que lorsque le système est déterministe et que $d_I(f)$ est égal à $H(Y)$, f est un modèle parfait, en effet: l'information transmise entre X et Y est maximale et égale à celle extraite du système.

Pour deux modèles f et g , on peut comparer la qualité de l'un par rapport à celle de l'autre, en comparant $d_I(f)$ et $d_I(g)$.

Ainsi f peut être considéré plus performant que g si:

$$d_I(f) > d_I(g).$$

Proposition

Soient f et g deux modèles, l'expression $D_I(f,g)$ définie par :

$$D_I(f,g) = \frac{|I(P_Y:P_f) - I(P_Y:P_g)|}{H(P_Y)}$$

est un écart normalisé.

Preuve:

- * $D_I(f,g) \geq 0$
- * $f = g \rightarrow I(P_Y:P_f) = I(P_Y:P_g) \rightarrow D_I(f,g) = 0.$
- * $D_I(f,g) = D_I(g,f).$
- * Soient f,g,h trois modèles, on a:

$$\begin{aligned} D_I(f,g) &= \frac{|I(P_Y:P_f) - I(P_Y:P_h)|}{H(P_Y)} \\ &= \frac{|I(P_Y:P_f) - I(P_Y:P_g) + I(P_Y:P_g) - I(P_Y:P_h)|}{H(P_Y)} \\ &\leq \frac{|I(P_Y:P_f) - I(P_Y:P_g)| + |I(P_Y:P_g) - I(P_Y:P_h)|}{H(P_Y)} \end{aligned}$$

$$\leq D_Y(f,g) + D_Y(g,h).$$

Notons que $D_I(f,g) = 0$ n'implique pas que $f = g$. Par conséquent, $D_I(f,g)$ n'est pas une distance.

Si f^* est un modèle optimal ($d_Y(f^*) = I(P_Y:P_X)$), la quantité :

$$D_I(f, f^*) = \frac{I(P_Y:P_X) - I(P_Y:P_f)}{H(P_Y)}$$

mesure l'information non encore extraite du modèle f au sujet du système.

III.2. Indice informationnel de Kullback.

A partir de l'information de Kullback, on peut évaluer la qualité d'un modèle f par l'expression :

$$d_J(f) = J(P_Y:P_f).$$

Une telle expression peut être interprétée comme la mesure de divergence entre les données du système et celles fournies par le modèle.

Notons que lorsque $d_J(f) = 0$, les données du modèle sont réparties suivant la même distribution que celles des données du système.

Ainsi un modèle parfait, compte tenu de ce critère, vérifie $d_J(f) = 0$.

Pour deux modèles f et g , on peut comparer la qualité de l'un par rapport à celle de l'autre en comparant $d_J(f)$ à $d_J(g)$.

De même, on peut dire que f est plus performant que g si $d_J(f) < d_J(g)$.

proposition.

Soient f et g deux modèles, l'expression $D_J(f,g)$ définie par :

$$D_J(f,g) = |J(P_Y:P_f) - J(P_Y:P_g)| \text{ est un écart .}$$

Cette proposition peut être démontrée de la même façon que la proposition précédente.

Notons que :

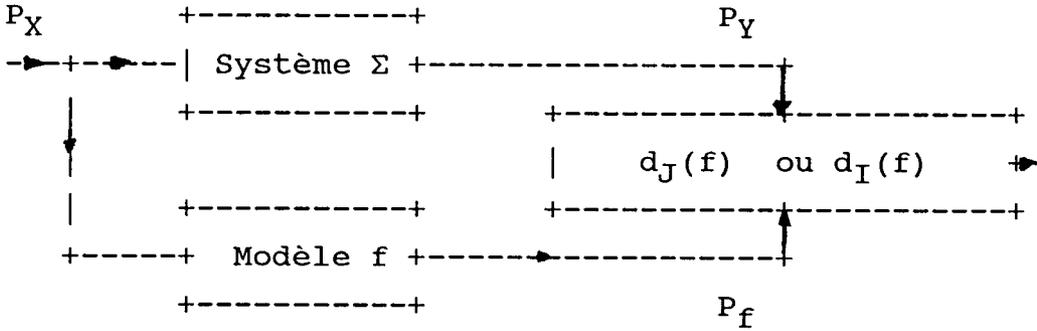
$$\begin{aligned} D_J(P_f:P_g) &= \left| \sum_{i=1}^m P_R(P_Y^i) \log \frac{P_R(P_Y^i)}{P_R(P_f^i)} - \sum_{i=1}^m P_R(P_Y) \log \frac{P_R(P_Y^i)}{P_R(P_g^i)} \right| \\ &= \left| \sum_{i=1}^m P_R(P_Y^i) \log \frac{P_R(P_f^i)}{P_R(P_g^i)} \right| \end{aligned}$$

$D_J(f,g)$ peut alors être interprété comme la divergence entre les données fournies par f d'une part et par g d'autre part, et ce par rapport à celles du système.

Remarques :

-Pour calculer d_J , il faut disposer d'un tableau ($2m$) de probabilité de P_Y et P_f , alors que pour le critère d_I , il faut un tableau ($m \times m$) de probabilité conjointe de P_Y et P_f . Le critère d_J nécessite moins de données que d_I .

-L'évaluation de la qualité d'un modèle par des critères informationnels se ramène à la comparaison de partitions, celle associée à la variable sortie du système et celle engendrée par le modèle à partir de l'entrée X. Ceci peut être schématisé par:



(Fig.IV.5)

-Les indices cités ci-dessus ne s'expriment pas en fonction des modalités prises par les sorties du modèle et du système, mais seulement de leurs répartitions.

-On peut également apprécier la qualité d'un modèle biunivoque f en comparant la partition induite par f^{-1} à celle engendrée par X . Dans ce cas, on a:

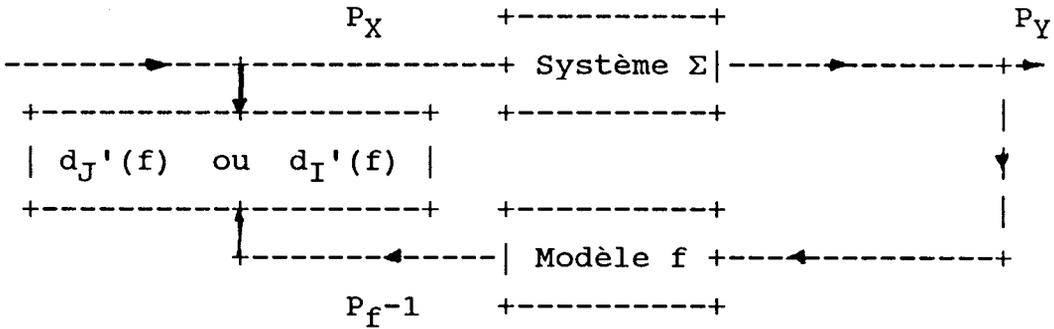
$$d'_I(f) = I(P_X:P_f^{-1}).$$

et

$$d'_J(f) = J(P_X:P_f^{-1}).$$

Ceci peut être schématisé par:





(Fig.IV.6)

III.3. Autres indices.

Notons qu'il est également possible de considérer d'autres critères tels l'entropie résiduelle, la variation entropique, les critères informationnels issus d'indices de similarité.

En considérant l'entropie résiduelle, on obtient un critère complémentaire à $d_I(f)$ puisque l'on a :

$$H(P_Y/P_f) = H(P_Y) - I(P_Y:P_f).$$

Ce critère s'interprète comme l'information non extraite du modèle au sujet de Y.

On peut considérer aussi le critère défini par :

$$\delta H(f) = H(P_Y) - H(P_f).$$

Ce critère exprime la variation de l'entropie entre les sorties modèle-système. Ce qui s'interprète comme l'information totale non fournie par le modèle f.

Enfin, les indices informationnels issus de similarité, compte tenu de la diversité dans leurs interprétations et leurs propriétés de distance, pourraient également fournir

des critères intéressants pour l'évaluation de la qualité d'un modèle.

III.4. Cas de variables continues.

Lorsque les variables du système et le modèle f sont considérés sous forme continue, on peut évaluer la qualité de f à partir des indices décrits auparavant, et ce, après une discrétisation appropriée des variables.

Toutefois, il est possible de formuler la qualité d'un modèle, sous forme continue, à partir d'indices équivalents à d_I et d_J :

$$C_I(f) = \int_{M_Y * M_Y} p_f(y, z) \log \frac{p_f(y, z)}{q(y)p_f(z)} dydz.$$

et :

$$C_J(f) = \int_{M_Y} q(y) \log \frac{q(y)}{q_f(y)} dy.$$

où $q(y)$ est la densité de probabilité de la sortie-système et $q_f(y)$, $q_f(y, z)$ sont les densités de probabilité induites par le modèle f . De telles densités peuvent être déterminées comme suit: [Saporta 78]

On considère $D_z = \{x \in M_X / f(x) \leq z\}$.

L'événement $f(x) \leq z$ équivaut à ce que l'événement x soit dans le domaine D_z , ce qui se traduit par:

$$p_r(f(x) \leq z) = \text{pr}(x \in D_z) = Q(z).$$

Connaissant la densité de probabilité de x , soit $p(x)$, celle induite par f s'écrit :

$$p_f(z) = \frac{d}{dz} Q(z) = \frac{d}{dz} p(x) dx.$$

En particulier, lorsque f est strictement monotone, on a :

$$p_f(z) = \frac{p(f^{-1}(z))}{|d/dz f^{-1}(z)|}$$

Notons que les indices informationnels peuvent s'avérer intéressants dans le cas continu, particulièrement lorsque le modèle est paramétrique. En effet, l'optimisation de ces indices peut conduire à des méthodes d'identification [Baram & Sandell. 78]. [Durgaryan. 82]. [Henning. 82]

IV APPLICATION AU MODELE LINEAIRE .

On propose dans ce paragraphe, d'évaluer la qualité d'un modèle linéaire. Pour cela, on suppose que les variables X et Y prennent leurs modalités dans des intervalles que l'on note respectivement par :

$$\begin{aligned} [x_i, x_{i+1}] [i = 0, 1, \dots, n-1 \\ [y_j, y_{j+1}] [j = 0, 1, \dots, m-1. \end{aligned}$$

S'il existe une relation déterministe entre X et Y , la matrice de contingence relative à P_X, P_Y prend la forme suivante :

P_Y	$[Y_0, Y_1[, \dots [Y_j, Y_{j+1}[, \dots [Y_{m-1}, Y_m]$		
P_X			
$[x_0, x_1[$	p_0	$\delta_{0,j}$	$\delta_{0,m-1}$
.			
.			
.			
$[x_i, x_{i+1}[$	$\delta_{i,0}$	p_i	$\delta_{i,m-1}$
.			
.			
.			
$[x_{n-1}, x_n[$	$\delta_{n-1,0}$	$\delta_{n-1,j}$	p_{n-1}

(Fig.IV.7)

Cette matrice possède un seul élément à probabilité dominante par ligne. Dans ce cas, l'information transmise entre P_X et P_Y est voisine de $H(P_Y)$.

Nous allons montrer que si les $\delta_{i,j}$ prennent des valeurs faibles, alors $H(P_Y) - I(P_Y:P_X)$ est voisine de 0.

$$I(P_Y:P_X) = H(P_Y) + H(P_X) - H(P_X, P_Y)$$

$$= - \sum_{j=0}^{m-1} (p_j + \sum_{i \neq 0} \delta_{i,j}) \log(p_j + \sum_{i=0}^{n-1} \delta_{i,j})$$

$$- \sum_{i=0}^{n-1} (p_i + \sum_{j \neq 0} \delta_{i,j}) \log(p_i + \sum_{j=0; j \neq i}^{m-1} \delta_{i,j})$$

$$+ \sum_{i=0}^{n-1} \sum_{j=0; \neq i}^{m-1} (p_i + \delta_{i,j}) \log(p_i + \delta_{i,j})$$

En considérant le développement limité de $-x \log x$ à l'ordre 1 en p_i , on obtient:

$$I(P_Y:P_X) = H(P_Y) - \sum_{i=0}^{n-1} \sum_{j=0; \neq i}^{m-1} \delta_{i,j} (\log p_i + 1)$$

$$+ \sum_{i=0}^{n-1} \sum_{j=0; \neq i}^{m-1} \delta_{i,j} \log \delta_{i,j} + o(\delta_{i,j}).$$

Posons $0 \leq \delta_{i,j} \leq \delta$, on obtient:

$$0 \leq I(P_Y:P_X) - H(P_Y) = \sum_{i=0}^{n-1} \sum_{j=0; \neq i}^{m-1} \delta_{i,j} (\log p_i + 1)$$

$$- \sum_{i=0}^{n-1} \sum_{j=0; \neq i}^{m-1} \delta_{i,j} \log \delta_{i,j} + o(\delta_{i,j}).$$

Comme $\delta_{i,j} \ll 1/2$ et $-x \log x$ est croissante sur $[0, 1/2]$, on a:

$$- \sum_{i=0}^{n-1} \sum_{j=0; \neq i}^{m-1} \delta_{i,j} \log \delta_{i,j} \leq -n(m-1)\delta \log \delta.$$

Par ailleurs, $\log p_i + 1 > 0$ puisque $1/2 \ll p_i < 1$, donc:

$$\sum_{i=0}^{n-1} \sum_{j=0; \neq i}^{m-1} \delta_{i,j} (\log p_i + 1) \leq n(m-1)\delta$$

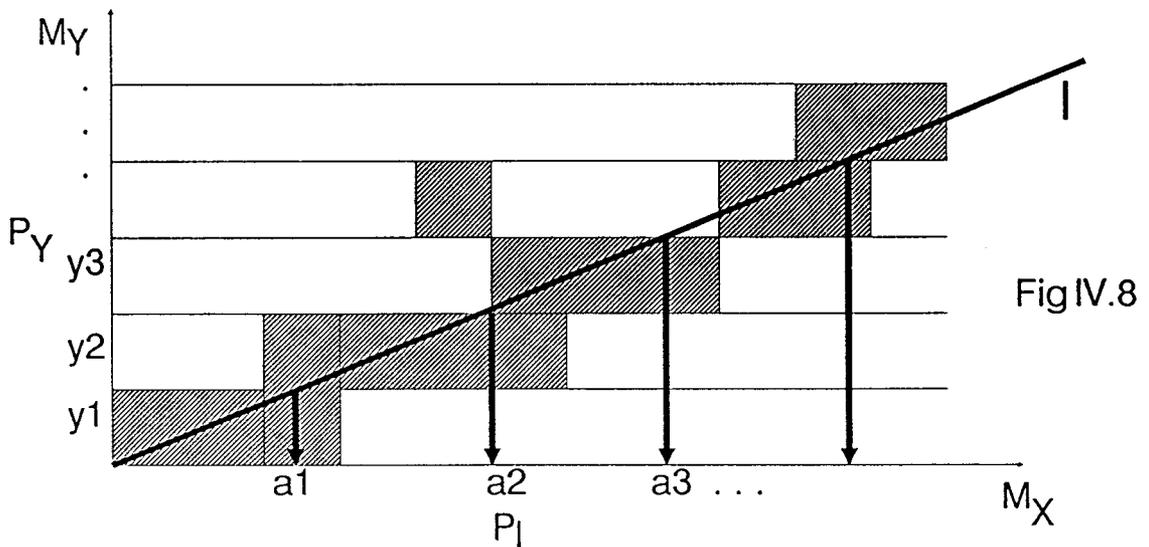
Il en résulte que:

$$0 \leq I(P_Y:P_X) - H(P_Y) \leq n(m-1)\delta(1 - \log \delta).$$

On peut alors conclure que plus les valeurs de $\delta_{i,j}$ sont proches de zéro, plus l'information $I(P_Y:P_X)$ est proche de $H(P_Y)$

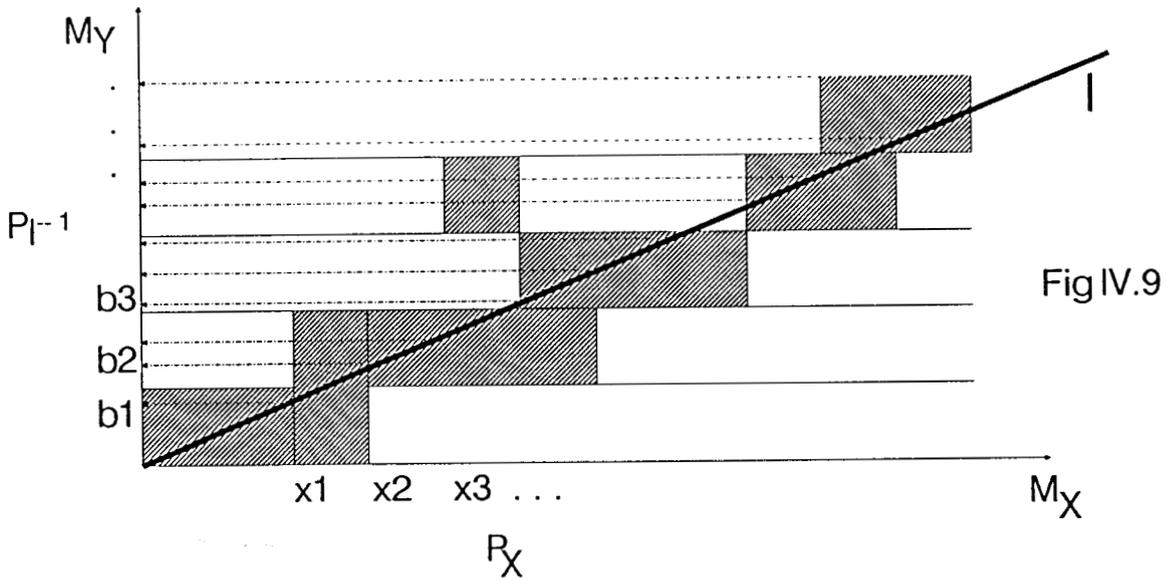
Notons que l'information $I(P_Y:P_X)$ reste inchangée si l'on regroupe les lignes du tableau précédent, dont les éléments non nuls se trouvent sur un même colonne. Ce qui revient à considérer une partition de X , qui apporte la même quantité d'information que P_X .

Par ailleurs, considérons un modèle linéaire l entre X et Y . Si l'on s'intéresse à l'évaluation de la qualité d'un tel modèle, il convient de définir sa partition induite. Celle-ci peut être obtenue par projection de la partition P_Y sur M_X ou de la partition P_X sur M_Y , comme il est indiqué dans les schémas suivants:



Partition induite par l

Les valeurs séparatrices a_j de P_X sont telles que:
 $y_j = l(a_j)$



Partition induite par l^{-1} .

Les valeurs séparatrices b_i de M_Y étant telles que :
 $b_i = l(x_i)$

IV.1. Calcul des partitions induites:

Les partitions induites P_l et P_l^{-1} s'écrivent respectivement :

$$P_l = \{ \{ w/X(w) \quad l^{-1}[y_j, y_{j+1}[\quad j=0, 1, \dots, m-1 \} \}.$$

$$= \{ \{ w/X(w) \quad [a_j, a_{j+1}[\quad j=0, 1, \dots, m-1 \} \}.$$

De même :

$$P_l^{-1} = \{ \{ w \in \Omega/Y(w) \quad [b_i, b_{i-1}[\quad i=0, 1, \dots, n-1 \} \}.$$

Lorsque les classes sont équidistantes, on a :

$$P_l = \{ \{ w/j * dX + x_0 \leq X(w) < (j+1) * dX + x_0, j=0, 1, \dots, m-1 \} \}$$

et

$$P_1^{-1} = \{ \{ w/i * dY + Y_0 \leq Y(w) < (i+1) * dY + Y_0 \}, i=0, 1, \dots, m-1 \}$$

Où $dX = a_{j+1} - a_j$ et $dY = b_{i+1} - b_i$ pour $i=0, \dots, n-1, j=0, \dots, m-1$.

En considérant l'intersection des classes de P_1 et de P_Y , ainsi que celles de P_1^{-1} et de P_X , on peut générer deux tableaux de probabilités conjointes, lesquels permettent d'évaluer les critères ci-dessous:

$$d_I(1) = I(P_Y : P_1) = \sum_{i=1}^n \sum_{j=1}^m \text{pr}(P_Y^j, P_1^i) \log \frac{\text{pr}(P_Y^j, P_1^i)}{\text{pr}(P_Y^j) \text{pr}(P_1^i)}$$

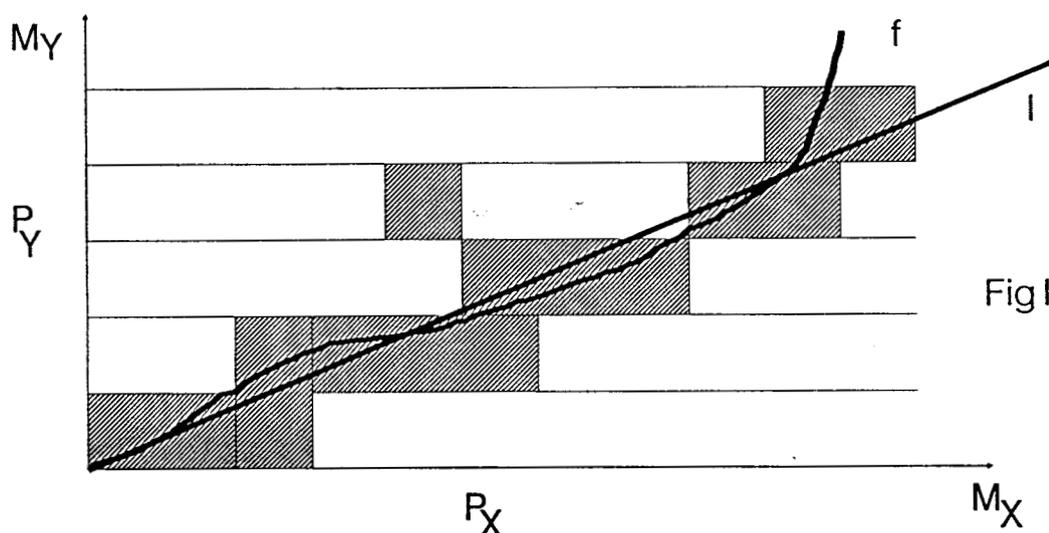
$$d_J(1) = J(P_Y : P_1) = \sum_{i=1}^n \text{pr}(P_Y^i) \log \frac{\text{pr}(P_Y^i)}{\text{pr}(P_1^i)}$$

$$d_I'(1) = I(P_Y : P_1^{-1}) = \sum_{i=1}^n \sum_{j=1}^m \text{pr}(P_Y^j, P_1^{i-1}) \log \frac{\text{pr}(P_Y^j, P_1^{i-1})}{\text{pr}(P_Y^j) \text{pr}(P_1^{i-1})}$$

$$d_J'(1) = J(P_Y : P_1^{-1}) = \sum_{i=1}^n \text{pr}(P_Y^i) \log \frac{\text{pr}(P_Y^i)}{\text{pr}(P_1^{i-1})}$$

IV .2. Performance d'un modèle linéaire.

Il s'agit de comparer un modèle linéaire l à un autre modèle quelconque (linéaire ou non) que l'on note f . Pour cela, on considère les partitions induites par l et par f , que l'on compare au moyen des indices ci-dessus:



En d'autres termes, cela revient à mesurer l'écart défini par : $D(l,f) = d(l) - d(f)$ où d désigne un indice d'évaluation de qualité.

Notons que $D(l,f)$ peut être positif, négatif, ou nul.

Dans le cas où d désigne le critère de Shannon, on peut dire que :

Si $D(l,f) > 0$, le modèle l est plus performant que f dans le sens où il fournit une quantité d'information plus grande.

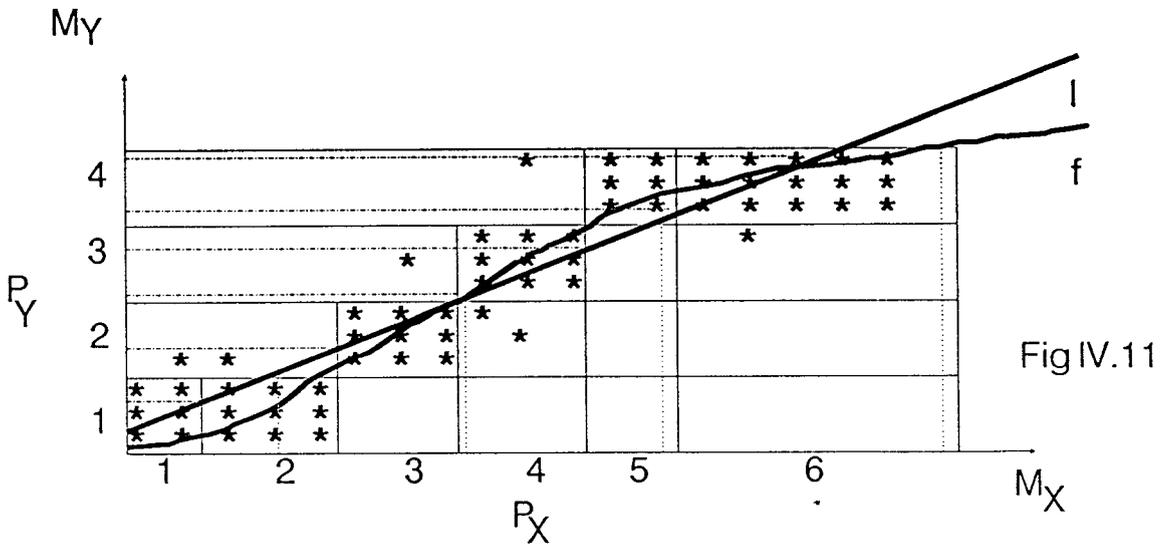
Si $D(l,f) < 0$, le modèle l est moins performant que f .

Si $D(l,f) = 0$, les modèles l et f sont similaires

Dans le cas où d désigne le critère de Kullback, on peut dire que si $D(l,f) < 0$, alors le modèle l est plus performant que f dans le sens où les données de l divergent moins que celles de f par rapport à celles du système .

IV.3. Exemple d'application.

Considérons un système monovarié d'entrée X et de sortie Y et un modèle entrée-sortie f . On suppose que l'ensemble des données du système et du modèle sont représentées par le schéma suivant:



Compte tenu de ces données et des partitions associées à X et Y , le tableau de contingence relatif à P_X et P_Y s'écrit:

$P_X \backslash P_Y$	1	2	3	4	$P_r(P_X)$
1	0.1	0.02	0	0	0.12
2	0.15	0.02	0	0	0.17
3	0	0.15	0.01	0	0.16
4	0	0.01	0.15	0.02	0.18
5	0	0	0	0.15	0.15
6	0	0	0.02	0.2	0.22
$P_r(P_Y)$	0.25	0.2	0.18	0.37	

Ce qui conduit à :

$H(P_X)$	$H(P_Y)$	$H(P_Y, P_X)$	$I(P_Y: P_X)$	$I(P_Y: P_X)/H(P_Y)$
2.5	2.0	3.0	1.5	0.75

De même, les données représentées ci-dessus permettent d'exprimer les tableaux de contingence relatifs à P_Y , P_f et P_X , P_f^{-1} :

$P_f \backslash P_Y$	1	2	3	4	$P_r(P_f)$
1	0.25	0.03	0	0	0.28
2	0	0.15	0.02	0	0.17
3	0	0.02	0.15	0.02	0.19
4	0	0	0.01	0.35	0.36
$P_r(P_Y)$	0.25	0.2	0.18	0.37	

$P_X \setminus P_f^{-1}$	1	2	3	4	5	6	$P_r(P_X)$
1	0.03	0.06	0.03	0	0	0	0.12
2	0.05	0.1	0.02	0	0	0	0.17
3	0	0	0.15	0.01	0	0	0.16
4	0	0	0.02	0.15	0.01	0	0.18
5	0	0	0	0	0.1	0.05	0.15
6	0	0	0	0.02	0.14	0.06	0.22
$P_r(P_f^{-1})$	0.08	0.16	0.22	0.18	0.25	0.11	

Précisons que f peut être un modèle quelconque du système, obtenu à partir d'une méthode d'identification par exemple.

Les tableaux de contingence ci-dessus permettent d'obtenir :

$H(P_f)$	$H(P_Y, P_f)$	$I(P_Y: P_f)$	$I(P_Y: P_f)/H(P_Y)$	$J(P_Y, P_f)$
1.9	2.4	1.5	0.75	0.03

$H(P_f^{-1})$	$I(P_X: P_f^{-1})$	$I(P_X: P_f^{-1})/H(P_X)$	$J(P_X, P_f^{-1})$
2.5	1.5	0.75	0.13

On constate que l'information extraite du modèle f est égale à celle connue au sujet du système, autrement dit que le modèle f est optimal. Par ailleurs, l'indice de Kullback montre une forte similarité entre la répartition des données des sorties système-modèle

On peut également remarquer que la variation entropique $H(P_Y) - H(P_f)$ donne un bon moyen pour apprécier la qualité du modèle f .

Notons que dans cet exemple, $J(P_X, P_f^{-1})$ prend une valeur non négligeable, ce qui implique que les distributions $\text{Pr}(P_X)$ et $\text{Pr}(P_f^{-1})$ ne sont pas identiques. Ce critère ne permet donc pas d'indiquer si le système peut être convenablement représenté par la relation $X=f^{-1}(Y)$.

Considérons le modèle linéaire l entre X et Y obtenu au moyen de la méthode de régression à partir des données (Fig IV.11) ($Y = 0.67 X + 0.74$, avec $\text{corr}(X, Y) = 0.91$).

Pour évaluer la qualité du modèle l à l'aide des critères définis auparavant, on détermine les tableaux de contingence relatifs à P_Y, P_l et P_X, P_l^{-1} .

$P_l \setminus P_Y$	1	2	3	4	$P_r(P_l)$
1	0.15	0.04	0	0	0.19
2	0.1	0.15	0.02	0	0.27
3	0	0.01	0.15	0.17	0.33
4	0	0	0.01	0.2	0.21
$P_r(P_Y)$	0.25	0.2	0.18	0.37	

$P_X \setminus P_1^{-1}$	1	2	3	4	5	6	$P_r(P_X)$
1	0.06	0.03	0.03	0	0	0	0.12
2	0.15	0.02	0	0	0	0	0.17
3	0	0.06	0.1	0	0	0	0.16
4	0	0	0.02	0.1	0.04	0.02	0.18
5	0	0	0	0	0.05	0.1	0.15
6	0	0	0	0.02	0.06	0.14	0.22
$P_r(P_f^{-1})$	0.21	0.11	0.15	0.12	0.15	0.26	

Ce qui conduit à :

$H(P_1)$	$I(P_Y, P_1)$	$I(P_Y: P_1)/H(P_Y)$	$J(P_Y, P_1)$
2.0	1.0	0.50	0.16

$H(P_1^{-1})$	$I(P_X: P_1^{-1})$	$I(P_X: P_1^{-1})/H(P)$	$J(P_X, P_1^{-1})$
2.5	1.3	0.52	0.08

On constate que l'information extraite du modèle 1 est sensiblement inférieure à celle extraite du système, ce qui signifie qu' au sens de l'information de Shannon, le modèle 1 n' est pas représentatif du système pris en compte. Il en est de même lorsque l'on considère le critère $I(P_Y: P_1^{-1})$.

Cependant, on peut remarquer que l'indice $J(P_X, P_1^{-1})$ prend une valeur voisine de zéro, ce qui implique que les distributions $Pr(P_X)$ et $Pr(P_1^{-1})$ sont similaires. Compte

tenu de ce critère, le système peut être convenablement représenté sous la forme : $X = l^{-1}(Y)$.

V. CONCLUSION .

Ce chapitre est consacré à l'évaluation de la qualité d'un modèle au moyen d'indices informationnels.

Nous avons défini d'abord ce qu'est une partition induite par un modèle sous forme de relation fonctionnelle entrée-sortie, et nous avons étudié ses propriétés. une telle partition peut être similaire au modèle dans le sens où l'information extraite du modèle est celle contenue dans la partition qu'il induit.

Nous nous sommes ensuite à l'évaluation de la qualité d'un modèle suivant deux approches informationnelles utilisant respectivement l'indice de Shannon et de Kullback. Ces approches ont été appliquées en particulier au cas d'un modèle linéaire monovarié.

CONCLUSION GENERALE

Dans ce travail, nous avons abordé différentes approches en Théorie de l'Information, telles l'approche de Shannon, de Kullback et les approches paramétriques. Nous avons souligné que l'information de Shannon s'applique d'une part à des variables quelconques, qualitatives ou quantitatives, logiques ou floues, à ensemble de modalités structuré ou non. D'autre part, contrairement aux indices issus de la statistique, elle permet de traduire l'existence ou non de couplage de nature quelconque (linéaire, non linéaire ...) dans un système.

Par ailleurs, nous avons montré que contrairement à l'entropie de Shannon, la divergence de Kullback ne permet pas d'avoir la propriété de sous-additivité. Il en est de même pour les mesures d'information du type paramétriques.

Dans le cadre de l'analyse structurale et la décomposition des systèmes statiques, nous avons étudié des mesures de couplage informationnelles, et signalé leurs limites respectives, Nous avons décrit ensuite des méthodes de décomposition:

L'évaluation du couplage entre les variables du système prises deux à deux conduit, à travers l'analyse d'une matrice ou d'un graphe d'interconnexion, à une partition du système. Cette partition, simple à mettre en oeuvre, n'est cependant pas toujours satisfaisante dans le cas de systèmes où les interactions sont trop importantes.

La prise en compte du couplage entre groupe de variables, à partir d'algorithmes du type hiérarchique ou de programmation dynamique, permet une partition du système en sous-systèmes faiblement couplés. Ces méthodes ont l'inconvénient de ne pas s'appliquer aux systèmes de grande taille, particulièrement lorsque les données du systèmes sont fortement bruitées.

Nous avons présenté ensuite une approche qui s'appuie sur le concept d'explication d'une variable à partir d'autres: la mise en oeuvre d'une telle approche, dans le cas de données bruitées, nécessite pour chaque variable, la détection des modalités bien expliquées, et des modalités mal expliquées à cause du bruit, ou dont l'explication requiert l'introduction de variables supplémentaires.

Le troisième volet de ce travail, a porté sur l'analyse des systèmes dynamiques et la détermination de l'ordre d'un système. Après avoir décrit différentes mesures de couplage dynamique, nous avons proposé des mesures de couplage dynamique faisant intervenir la longueur de l'horizon d'observation du système.

La prise en compte de tels coefficients, permet d'obtenir des partitions du système en fonction de la longueur de l'horizon d'observation.

Dans le cadre de la détermination de l'ordre d'un système, nous avons décrit des approches informationnelles. Celles-ci nécessitent en général la donnée d'un modèle et l'identification de ses paramètres. A partir d'un indice informationnel dynamique, nous avons proposé une approche qui s'applique aux systèmes multivariables, et qui ne nécessite pas la donnée d'un modèle particulier du système. Une telle approche conduit à la détermination de l'ordre d'un système à travers deux algorithmes simples à mettre en oeuvre.

En ce qui concerne la modélisation structurale, nous nous sommes intéressés à l'évaluation de la qualité d'un modèle au moyen d'indices informationnels. Nous avons établi une relation de similitude entre un modèle du type entrée-sortie et une partition de l'espace d'observation du système. Une telle similitude permet de considérer

l'information que l'on peut extraire du modèle, égale à celle contenue dans la partition qu'il engendre.

Nous avons ensuite exposé deux approches pour évaluer la qualité d'un modèle : d'une part en comparant l'information extraite du modèle à celle du système au moyen d'indices issus de la transinformation de Shannon. D'autre part, en évaluant la divergence entre les données fournies par le modèle et celles du système à partir de la mesure de Kullback.

L'application de ces approches a permis en particulier d'apprécier la qualité de représentation d'un système au moyen d'un modèle linéaire.

Loin d'être épuisé, le domaine de l'analyse des systèmes complexes par la T I, que nous avons voulu explorer dans cette thèse, nécessite encore bien des recherches. Outre l'expérimentation et la mise en oeuvre des approches proposées, on peut discerner quelques directions de recherche qui nous semblent intéressantes:

--> Développer de nouveaux indices informationnels (paramétriques ou non) pour la décomposition des systèmes dynamiques. De tels indices doivent tenir compte du bruit, et fournir des partitions qui prennent en considération l'aspect évolutif du système.

--> Appliquer les approches informationnelles dans le domaine du diagnostic et de la détection des pannes dans un système défaillant. Les méthodes de recherche des sous-ensembles explicatifs pourront en particulier, fournir des outils intéressants dans ce domaine.

REFERENCES BIBLIOGRAPHIQUES

H. AKAIKE .

Information theory and extension of the maximum likelihood principle.

Proc.2Int.Symp.on Information Theory; Tsahkadsor.America, 2-8 sept 1973. Edité par:B.N.Petrov.F.Csaki.Budapest. p:267-281.

H. AKAIKE .

Modern development of statistical methods. Trends and progress in System Identification.

Edité par P.Eykhoff. Oxford. Pergamon.IFAC Series for Graduates, Research Workers and Practicing Engineers Vol.I p.169-184. 1981

R. ASH.

Information Theory.

John Wiley & Sons, 1965

W.R. ASHBY.

Two tables of Identities governing Information flow within large Systems.

Communication of the American Society of Cybernetics. Vol I.n°2. p 2-8. 1969.

Y.BARAM - N.R. SANDELL.

An Information Theoretic Approach to Dynamic Systems Modelling and Identification.

I.E.E.E Transactions on Automatic Control. Vol. AC.23.
N°1. 1978. p: 61-66.

M. BARBOUCHA.

Modélisation Structurale des Systèmes Complexes.
Extraction et Validation des règles d'un système expert.

Thèse d'état. Centre d'Automatique de l'Université de
Lille I. France. Juin 1987.

M. BARBOUCHA.- M. STAROSWIECKI.

Modélisation Structurale des Systèmes Complexes.

Sixth Inter Symposium on Modelling, Identification and
Control. IASTED. Grindelwald. Suisse. February 1987.

J.P. BENZECRI.

Analyse des données.

Dunod 1980.

C. BIRKHOFF.

Lattice theory.

American Mathematical Society. 4th edition. 1964.

C.H. CHEN.

On information and distance measures, error bounds, and
feature selection.

Information Sciences .Vol 10.1976 p 159-173.

R.C. CONANT.

Detecting Subsystems of a Complex System.

I.E.E.E. Trans, on Systems, Man and Cybernetics.Vol SMC-2
n°4.p 550-553 September 1972.

R.C. CONANT.

Laws of information which govern Systems.

I.E.E.E. Trans. on Systems, Man and Cybernetic.Vol SMC-6
n°4. p 240-255 April 1976.

Z. DAROCZY.

Generalized information functions.

Information & Control. Vol 16.p 36-51 1970.

J. DUFOUR.

Méthodes et Methodologies d'analyse de systèmes complexes. Application aux procédés industriels et aux systèmes macroéconomiques.

Thèse d'Etat. Université Claude Bernard, Lyon. Mai 1979.

I.S. DURGARYAN - F.F PASHCHENKO.

An Information Theory Approach to Identification.

6th I.F.A.C Symposium. June 82. Washington DC, U.S.A.
Vol I. p: 634-638.

A. DUSSAUCHOY.

Generalised information theory on some ordered sets and structure of systems.

14th Annual Conf, on Information Sciences in Systems. Princeton University 1980.

A. HAOUAT.

An information theoretic approach of coupling analysis and décomposition of dynamic systems.

Octobre 1986. I.E.E.E International Conference on Systems, Man, and Cybernetic. Atlanta. USA.

A. HAOUAT - M. STAROSWIECKI.

Fine modelling by means of information measures.

Soumis à I.A.S.T.E.D International Conference on Control. Juin 1990.

K. HENNING.

On the use of Entropy functions for Identification on Dynamic Systems.

6th I.F.A.C Symposium. June 82. Washington DC, U.S.A. Vol I. p: 629-633.

R.L. KASHYAP.

Inconsistency of the AIC rule for estimating the order of auto-regressive models.

I.E.E.E. Trans. on Automatic Control. Vol AC-25. p 996-998 1980.

D.F. KERRIDGE.

Inaccuracy and Inference.

Journal of Royal Statistical Society. Serie B. Vol 23.
p:184-194.

G. KLIR.

An approach to general systems theory.

Van Nostrand Reinhold. New-York. 1969.

G. KLIR.

On the representation of activity arrays.

International Journal of general systems. Vol 2. 1975.

G. KLIR.

Identification of generative structure in empirical data.

International Journal of General Systems. Vol 3. 1976.

G. KLIR.

On the problem of computed aided structure identification :some experimental observation and resulting guidelines.

International Journal of Man-Machine Studies. Vol 9 1977.

S. KULLBACK.

Information theory and statistics.

Wiley Publications in Statistics New-York 1959.

R. LOPEZ DE MANTARAS.

Auto-apprentissage d'une partition. Application au classement itératif de données multidimensionnelles.

Thèse de doctorat de spécialité. Université Paul Sabatier. Toulouse 1977.

W.J. MC GILL.

Multivariate Information Transmisssion.

Psychométrie, Vol 19 p 97-116. 1954.

G.F. MCLEAN - K. PALANIAPPAN - M.E. JERNIGAN.

Information measures for shape description.

Publié in Departement of System Design Engieniring. University of Waterloo . Ontario . Canada. 1985.

G.A. MILER.

What is information measurement?.

American Psychologist, Vol 8, n°2 p 50-51. 1963.

A. NAAT.J. DUFOUR.

Procédure de caractérisation de certaines classes de systèmes linéaires au moyen de la transinformation.

A.P.I.I. n°19. p 575-594. 1985.

E. PARZEN.

Modern probability theory and its applications.

John Wiley and Sons. New-York. 1960.

M.F. PONOMORENKO.

Information theory and identification.

Report 81-E-122 in Eindhoven University of Technology.
Netherland. October 1981.

P.N. RATHIE-P.L KANNAPAN.

A direct-divergence function of type β .

Information and Control. Vol 20. p 38-45. 1972.

A. Rényi.

On the measure of entropy and information.

Proc.4th Berkeley Symp on mathematical statistic and
probability. Vol 1 . p 547-561. Juin 1960

M. RICHETIN.

Analyse structurale des systèmes complexes en vue d'une
commande hiérarchisée.

Thèse d'Etat. Université Paul Sabatier. Toulouse. Juillet
1975.

R.E. RINK.

Information theoretic methods for modelling and analysing large systems.

5th IFIP Conference on optimisation techniques. Rome 1973.
Publié in Lecture Notes on Computer Sciences. n°3 Springer
verlang. Allemagne 1973.

J. RISSANEN.

Mini max entropy estimation of models for vector processes.

Mathematics in Science and Engineering vol 126. p 97-119.
1976.

G. SAPORTA.

Théorie et Méthodes de la statistique.
Paris-Technip. 1978.

M. SBAI.

Analyse structurale des systèmes complexes : Méthodes
d'explication et de partition.

Thèse de 3^e cycle. Université de Lille I. 1983.

C.E. SHANNON.

A mathematical theory of communication.

Bell System Technical Journal; Vol 27. 1948.

H.A. SIMON.

The architecture of complexity.

Proc. of the American Philosophical Society. Vol 106. n°6.
1962.

T. SODERSTROM - K. KUMAMARU .

On the use of Kullback discrimination index for model
validation and fault detection.

Automatic Control and System Analysis Group . Departement
of Technology . Uppsala Univrsity . Sweden.1985.

M. STAROSWIECKI.

Analyse structurale des systèmes complexes.

RAIRO Automatique. Vol 18. n°2. 1984.

M. STAROSWIECKI. V. TORO.

Structural analysis of complex systems by means of
canonical analysis.

VII International conference on systems science. Wroclaw.
Pologne. Sept 1981.

M. STAROSWIECKI. M. BARBOUCHA.

Modélisation structurale des systèmes complexes.

6° Congrès international de la Wosc. Paris. Sept 1984.

M. STAROSWIECKI - A. HAOUAT. - M. BARBOUCHA.

Coarse to fine modelling by a Structural Approach.
Février 1990. I.A.S.T.E.D International Conference on
Modelling, Identification and Control. Innsbruck. Autriche.

V. TORO. M. STAROSWIECKI.

Méthodes heuristiques et décomposition optimale des systèmes complexes.

Congrès AFCET Automatique. Nantes. Oct 1981.

V. TORO.

Contribution à l'analyse structurale des systèmes complexes à l'aide de l'entropie et de ses généralisations.

Thèse de 3° Cycle. Université de Lille I. Mars 1982.

B. YVINEC. M. MILGRAM. B. DUBUISSON.

Estimation de l'entropie dans le cas continu en vue de la décomposition des systèmes complexes.

RAIRO Automatiques. Vol 12 n°1. 1978.

MOTS CLES:

Systemes complexes - Analyse structurale - Modélisation
Partition - Théorie de l'Information.

RESUME:

Le travail présenté dans cette thèse a pour objet: l'analyse structurale des systèmes complexes, l'évaluation de la qualité d'un modèle et la détermination de l'ordre d'un système au moyen de la Théorie de l'Information (T I).

Dans un premier temps, on présente les principales définitions et propriétés de l'entropie et de la transinformation de Shannon, des mesures de divergence de Kullback, et de certaines mesures paramétriques.

La deuxième partie de ce travail est consacrée à l'analyse structurale et la décomposition des systèmes statiques. On présente des approches binaires et multi-dimensionnelles. Celles-ci examinent l'information échangée par groupes de deux variables ou plus, et conduisent soit à l'analyse d'un graphe d'interconnexion, soit à des partitions du système à travers des algorithmes du type hiérarchique ou de programmation dynamique. On présente ensuite une approche qui s'appuie sur le concept de modélisabilité. Celle-ci consiste à déterminer pour une variable à expliquer, un sous-ensemble explicatif optimal dans le sens où l'information est maximale. Dans le cas où le système est soumis à des perturbations d'origines diverses, on considère une procédure qui consiste à séparer pour chaque variable à expliquer, les modalités bien expliquées des modalités mal expliquées à cause du bruit, ou dont l'explication nécessite l'introduction de variables supplémentaires.

On s'intéresse dans la troisième partie de ce travail à l'analyse structurale des systèmes dynamiques et à la

détermination de l'ordre d'un système. Dans un premier temps, on étudie des indices de couplage dynamique qui prennent en compte l'information véhiculée entre les variables au cours de l'évolution du système. Ces indices s'appliquent à la décomposition des systèmes dynamiques, ainsi qu'à l'évaluation de leur degré de modélisabilité. Après avoir présenté quelques approches classiques pour la détermination de l'ordre d'un système, on propose une approche qui utilise un indice informationnel dynamique.

Le quatrième volet de ce mémoire est consacré à l'évaluation de la qualité d'un modèle statique du type entrée-sortie au moyen de la T I. On propose deux approches pour apprécier la qualité d'un modèle : par l'évaluation de l'information extraite de la partition-modèle au sujet du système; par la mesure de la divergence entre les données fournies par le modèle et celles du système. Ces approches sont illustrées dans le cas d'un modèle linéaire.

SUMMARY

The aim of this thesis is about structural analysis of complex systems, modelling valuation, and determination of system's order, by the use of Information Theory.

In first, we present definitions and proprieties of Shannon's entropy, Kullback's information measures, and some parametric information indices.

The second part of this work treats structural analysis of static systems. We present binary and multivariate approaches which allow to a partition of the system through interconnexion matrix analysis, hierarchic or dynamic programming. An other approach is presented. It is based on the concept of "Explaining a set of variables by means of an other one". In this approach, for each variable to explain, well explained values are separated from the set of values for which explanation requires to introduce supplementary explicative variables.

The third part of this work is about structural analysis of dynamic system, and determination of system's order. General informational dynamic indices are studied. They can be used to find a partition into weakly coupled subsystems. At least, a dynamic indice is applied to evaluate the order of a system through some algorithms.

In the last part, we propose two approaches for modelling valuation, by the use of informational indices. the first approach takes account of the Shannon's information which is extracted from model data about the system. The second one uses the Kullback's discrimination measure, to compare data which is derived from the system and its model. Each of these approaches are illustrated for a linear model.

