

50376
1986
229

50376
1986
229

N d'ordre : 78

THÈSE

présentée à

**L'UNIVERSITE DES SCIENCES ET TECHNIQUES DE LILLE
FLANDRES ARTOIS**

pour l'obtention du titre de

DOCTEUR

Spécialité : Automatique

par

AYGALINC Pascal

Ingénieur EUDIL

**APPLICATION DE LA RECONNAISSANCE DES FORMES A L'AIDE
AU DIAGNOSTIC MEDICAL
SELECTION MULTICRITERE DE VARIABLES EXPLICATIVES**



Soutenue le 12 Décembre 1986 devant la Commission d'Examen

MM.	P.	VIDAL	Président
	B.	DUBUISSON	Rapporteur
	J.G.	POSTAIRE	Rapporteur
	M.	STAROSWIECKI	Directeur de Recherche
	O.	DUBOIS	Invité

AVANT - PROPOS

Le travail présenté dans ce mémoire est le résultat d'une collaboration suivie entre le Centre d'Automatique de l'Université des Sciences et Techniques de Lille Flandres-Artois et le Centre de Pathologie Néo Natale d'Arras.

Je tiens à exprimer ma plus grande reconnaissance à Monsieur le Professeur P. 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 aussi à exprimer ma profonde gratitude à Monsieur le Docteur O. DUBOIS, Médecin Chef au Centre de Pathologie Néo Natale d'Arras pour l'intérêt constant qu'il a porté à ce travail dont il est l'initiateur, et pour les conseils et les encouragements qu'il m'a dispensés.

Je voudrais également remercier Monsieur B. DUBUISSON, Professeur à l'Université de Technologie de Compiègne pour avoir accepté d'examiner mes travaux et de faire partie de ce Jury.

Que Monsieur J.C POSTAIRE, Professeur à l'Université des Sciences et Techniques de Lille Flandres-Artois trouve ici l'expression de ma profonde gratitude pour les conseils qu'il m'a prodigués durant cette étude, et pour avoir bien voulu apporter sa contribution au jugement de ce travail.

Mes plus vifs remerciements vont à Monsieur M. STAROSWIECKI, Professeur à l'Université des Sciences et Techniques de Lille Flandres-Artois, qui m'a dirigé et suivi tout au long de ce travail. Qu'il trouve dans ce mémoire le témoignage de ma reconnaissance pour l'encadrement éclairé, enthousiaste, généreux, et amical qui me permet de présenter cette thèse.

Enfin, je ne saurais terminer cet avant-propos sans remercier tous ceux qui de près ou de loin, par leur compétence et leur amitié, m'ont soutenu dans l'élaboration de ce travail et plus particulièrement Monsieur B. CEURSTEMONT pour son aide technique, Madame A. PIGNON qui m'a aidé par ses conseils pour la présentation de ce mémoire, et le Service de Reprographie.

- S O M M A I R E -

- SOMMAIRE -

INTRODUCTION GENERALE

CHAPITRE I :

1. INTRODUCTION	5
2. LES METHODES ET L'ENVIRONNEMENT DU MEDECIN HOSPITALIER	6
2.1 Rôles du médecin dans un service de pathologie néo-natale	6
2.2 Le diagnostic	7
2.3 Analyse de données : aide au diagnostic ...	9
3. DE L'OBSERVATION A LA DECISION	10
3.1 Modèle de la Reconnaissance des Formes	12
3.2 Modèle Système-Expert	12
4. CONSTRUCTION DE LA POPULATION D'APPRENTISSAGE ...	15

CHAPITRE II :

1. INTRODUCTION	17
2. LES DONNEES	18
2.1 Représentation de la population d'apprentissage	18
2.2 Partition de Ω	19
2.3 Mesure de probabilité	21
3. RECODAGE DES DONNEES	23
3.1 Formulation du problème	23
3.1.1 Cas particulier : X variable quantitative.....	24
3.2 Caractérisation de $P_X^*(j)$	25
3.2.1 Critères portant sur les caractéris- tiques des classes	25
3.2.2 Critères dépendant du type de variable	26
3.3 Partitions optimales par niveaux	28
3.4 Recherche des optimums locaux	29
3.4.1 Recherche ascendante	30
3.4.2 Recherche descendante	30
3.4.3 Recherche par la programmation dynamique	31
4. RECONSTRUCTION DE DONNEES MANQUANTES	34
4.1 Présentation	34
4.2 Recherche des plus proches voisins	35
4.2.1 Proximité	35
4.2.2 Procédure de sélection	37

4.3	Algorithmes de reconstruction	39
4.3.1	Reconstruction d'une donnée manquante	39
4.3.2	Reconstruction de K données manquantes	42
4.4	Une voie possible dans la recherche d'un individu représentatif d'une classe	45
4.4.1	Notion d'individu le mieux reconstitué.....	45
4.4.2	Application à la Reconnaissance des Formes	45
5.	CONCLUSION	46

CHAPITRE III :

1.	INTRODUCTION	48
2.	RECONNAISSANCE DES FORMES : PROBLEME STATIQUE ...	51
2.1	Présentation du modèle statique	51
2.2	Données du modèle de la R.D.F	52
3.	DECISIONS SUCCESSIVES	56
3.1	Décision séquentielle	56
3.2	Décision dynamique	57
4.	EVALUATION DES PERFORMANCES	58
4.1	Evaluation des performances de $\hat{\Phi}(S)$	58
4.1.1	Qualité de la discrimination	60
4.1.2	Efficacité de la discrimination	64
4.2	Evaluation des performances de S	66
4.2.1	Mesures probabilistes	66
4.2.2	Mesures informationnelles	68
5.	CONCLUSION	71

CHAPITRE IV :

1.	INTRODUCTION	73
2.	PROCESSUS D'AIDE AU DIAGNOSTIC	75
2.1	Processus de segmentation	75
2.2	Les différents types de noeuds	75
2.3	Aspect conversationnel	76
2.3.1	Les différents comportements du médecin	76
2.3.2	Analyse des différentes situations .	77
3.	PROBLEME DE DECISION	79
3.1	Ensemble des décisions admissibles	79
3.2	Mimisation de la probabilité d'erreur	80
3.3	Minimisation d'une fonction de risque	84
3.4	Validité des résultats	88

4. SELECTION DES EXAMENS COMPLEMENTAIRES	89
4.1 Mesure de la contribution de S_t à l'affinement et/ou la confirmation de D ...	90
4.2 Caractérisation multicritère de S_t	92
5. EVALUATION FIABLE DES DIFFERENTS CRITERES	95
6. RESOLUTION DU PROBLEME D'OPTIMISATION	97
6.1 Cas multicritère	97
6.1.1 Fonction d'utilité	97
6.1.2 Hiérarchisation des critères	98
6.1.3 Relation de surclassement	99
6.2 Simplifications	101
6.2.1 Recherche ascendante	101
6.2.2 Recherche descendante	102
7. CONCLUSION	103

CONCLUSION GENERALE

ANNEXES

BIBLIOGRAPHIE

- I N T R O D U C T I O N G E N E R A L E -

- INTRODUCTION GENERALE -

L'Informatique pénètre, lentement parfois, mais sûrement tous les domaines de la Santé. Cette pénétration est d'autant plus évidente (et d'autant plus efficace) que le domaine considéré se prête à un certain "formalisme".

Si historiquement, les premières applications informatiques furent destinées à des activités de gestion (paie du personnel hospitalier, facturation des actes médicaux, gestion des équipements et des matières, ...), celles qui suivirent eurent des finalités beaucoup plus médicales, allant de l'informatisation du dossier médical, de sa gestion jusqu'au contrôle de processus biologiques (monitoring, système d'évaluation des électrocardiogrammes, ...) /RAJA,79/ /COUV,81/ /VASS,82/ /LECE,83/...

A partir du moment où l'on a su maîtriser correctement les techniques du recueil et du traitement de l'information médicale, permettant ainsi la construction de banques de données, il était normal d'envisager la création de systèmes plus élaborés destinés à apporter au médecin une aide dans l'exercice de sa fonction (recherche, aide au diagnostic, à la prescription, à l'enseignement, ...).

Dans ce sens, l'application de programmes d'exploitation statistique (statistique au sens large : analyse factorielle, typologie, corrélations multiples, ...) sur les banques de données déboucha sur des études épidémiologiques, des études de morbidité, des études sur les influences réciproques éventuelles de plusieurs facteurs, ... /DOMB,76/ /NAKA,77/ /TITR,79/ /DAMI,84/ /FACY,84/ ...

Dans le domaine plus délicat de l'aide au diagnostic (domaine qui nous intéresse particulièrement dans le sens où il constitue le cadre de notre travail), les premières approches développées utilisèrent aussi ces programmes statistiques. Ils permirent ainsi la construction de tables de décisions diagnostiques probabilisées, la description visuelle d'une pathologie donnée (analyse discriminante), la prescription séquentielle d'examens complémentaires .../TERR,76/ /HABB,81/ /MINO,80/ /ZIGH,84/ ...

Plus récemment les techniques basées sur l'Intelligence Artificielle et l'approche Système Expert ont connu un développement important comme en témoigne la multiplicité des applications publiées, /SHOR,74/ /DONN,84/ /AUVE,86/ /DUPO,86/

Le Centre de Pathologie Néonatale d'Arras avec lequel nous avons travaillé à mis en oeuvre depuis quelques années une informatisation de ses dossiers, constituant ainsi une banque de données qu'il était logique d'exploiter. Le but du présent travail est de proposer la structure d'un système d'aide au diagnostic, basée sur les connaissances extraites à partir de ces données.

La population d'apprentissage étant constituée d'un ensemble de nouveau-nés prématurés admis au Centre pour des troubles respiratoires, nous proposons une approche qui permet de déterminer le système relationnel qui mène au diagnostic, en utilisant le modèle de la Reconnaissance des Formes, pour la résolution du problème de décision relatif à l'individu sous diagnostic. La prescription d'examen complémentaires fait appel à la Théorie de l'Information et à des critères liés aux conditions d'obtention des variables pour évaluer les ensembles explicatifs admissibles. Le choix d'un ensemble de paramètres pour la décision est alors le résultat d'une optimisation multicritère.

Le premier chapitre permet au lecteur non initié de se familiariser avec le milieu médico hospitalier. Nous présentons les différents objectifs recherchés, et nous discutons des différents modèles qui peuvent les satisfaire. Enfin, nous précisons le modèle retenu et les problèmes qu'il soulève au niveau du recueil des données sur un horizon d'observation approprié.

Dans le second chapitre, nous nous intéressons aux traitements préliminaires à l'analyse proprement dite. Ils concernent essentiellement le recodage des variables et la reconstruction de données manquantes, et formulent ainsi les problèmes évoqués ci dessus.

Le troisième chapitre traite du modèle de la Reconnaissance des Formes et de ses extensions possibles à des problèmes de décisions successives. Nous exposons les différents indices qui permettent d'évaluer les performances de la discrimination obtenue, aussi bien au niveau de sa "qualité" que de son "efficacité". Certains de ces critères seront repris par la suite pour établir le système d'aide au diagnostic.

Enfin, nous présentons au quatrième chapitre le processus interactif d'aide au diagnostic pour lequel nous avons développé des procédures de décision et de sélection de variables spécifiques. La procédure proposée est calquée sur la démarche médicale telle qu'elle a été analysée (chapitre I), c'est ce qui explique son caractère conversationnel très développé.

- C H A P I T R E I -

1 . INTRODUCTION

La description d'un service hospitalier, des méthodes de travail du médecin et de ses aspirations peuvent faire l'objet d'un développement entier qui sort du cadre de ce travail. Néanmoins, il est important de définir le contexte de notre étude et de préciser les buts recherchés.

A cette fin, ce chapitre a d'abord pour but de familiariser le lecteur au milieu médico-hospitalier, plus précisément au milieu de la médecine d'urgence et d'intervention pour lequel une aide à la décision médicale peut être envisagée (Service de Pathologie Néonatale d'Arras, dirigé par le professeur O. Dubois). Les propos tenus tout au long de cet exposé souligneront les différents objectifs que doit satisfaire un système d'aide dans un tel service.

De multiples réalisations à support informatique ont déjà été conçues /MATH, 82/, en ce qui concerne le domaine de l'aide au diagnostic. Nous rappellerons les différentes approches qui ont été utilisées pour faciliter le travail du médecin dans cet acte médical. Enfin, nous présenterons celle que nous avons retenue et les différents problèmes qu'elle soulève, plus précisément ceux liés à la construction d'une population d'apprentissage.

2 . LES METHODES ET L'ENVIRONNEMENT DU MEDECIN HOSPITALIER

2.1 ROLES DU MEDECIN DANS UN SERVICE DE PATHOLOGIE NEO-NATALE

L'objet de notre travail nous a amené à côtoyer pendant quelques temps les femmes et les hommes qui animent le centre de Pathologie Néo-Natale d'Arras. Lors de notre premier contact avec ce domaine, il est impressionnant de voir s'affairer ces personnes autour des incubateurs où se trouvent les nouveau-nés prématurés.

Dans un tel service, suivant que le patient présente ou non des signes pathologiques, différentes actions sont menées :

- S'il s'agit d'un nouveau-né prématuré ne présentant pas des signes de souffrances, de grandes malformations, , le personnel soignant intervient de manière à placer l'enfant dans les conditions similaires à celles de la gestation, jusqu'à ce qu'il ait atteint un poids normal.

- Par contre, si le patient présente des signes de souffrance pathologique, le médecin en service tente de déterminer les origines de ces troubles afin d'intervenir efficacement pour préserver la vie de l'enfant.

Il apparait alors que le médecin a un triple rôle : observer, diagnostiquer, traiter ; ces trois actions, qui doivent être effectuées d'autant plus rapidement que le cas est urgent, correspondent à trois domaines importants qui définissent l'exercice de l'acte médical.

a) l'*éthiologie* qui correspond au domaine de la recherche des causes des maladies.

b) la *pathologie* qui correspond à la description des maladies.

c) la *thérapie* qui correspond aux actions à mener pour conduire le malade à la guérison.

Ces trois domaines sont parfaitement interdépendants comme on peut l'imaginer. Les traitements à prescrire dépendent bien évidemment des maladies à combattre, maladies appartenant au champ de la pathologie associée aux symptômes observés.

En sa qualité d'observateur, le médecin soumet son malade à un certain nombre d'examens parmi lesquels on distingue :

- *l'examen clinique* : il permet l'observation par exemple de troubles fonctionnels (troubles de différentes fonctions organiques : respiration, digestion, ...), de troubles généraux (température, pouls, ...), de troubles physiques observés à l'auscultation ou à la palpation. Sa caractéristique principale est qu'il est simple et rapide à réaliser.

- *les examens paracliniques* : à titre d'illustration, on peut citer les examens biologiques, les examens radiologiques, les examens électriques (ECG, EEG, ...). Ils exigent souvent la mise en oeuvre de techniques d'analyse élaborées ; ce qui entraîne un délai d'obtention du résultat important. Ils sont donc limités par l'urgence des décisions et actions à mener, voire même par leur pénibilité et éventuellement par leur coût financier. Il revient alors au médecin de les sélectionner sérieusement.

L'ensemble de ces différentes informations médicales recueillies sur le patient constitue le dossier médical, et c'est à partir de la synthèse et de l'analyse de ce dossier que le médecin établit le diagnostic.

2.2 LE DIAGNOSTIC

Il est intéressant de façon générale d'essayer de décrire le cheminement intellectuel que suivent les médecins pour établir un diagnostic. L'analyse de la stratégie adoptée permet de dégager quelques caractéristiques simples à partir desquelles il est possible d'énoncer les différentes propriétés qu'un système d'aide doit vérifier.

Le diagnostic consiste à faire le choix entre plusieurs hypothèses de maladies. Pour effectuer ce choix, le médecin fait appel à son savoir, ses connaissances livresques, l'expérience acquise au cours de ses observations et interventions passées, ainsi qu'aux différentes informations dont il dispose sur l'état de santé du patient.

Hormis certains cas évidents, la quantité d'informations recueillies dès les premiers instants de la vie d'un nouveau-né prématuré ne permet pas au médecin d'établir avec une complète certitude un diagnostic. La démarche qu'il adopte dans cette première phase est assimilable au mode progressif, c'est à dire que la stratégie du raisonnement est guidée par les différents signes cliniques observés et les résultats des examens déjà

réalisés. Cette stratégie conduit le praticien à une ou plusieurs hypothèses de diagnostic.

Le problème qui se pose alors est celui du choix des examens complémentaires, de leur ordre de réalisation et de l'opportunité de poursuivre les investigations au vu des résultats des examens déjà réalisés. La démarche adoptée ici est celle du mode régressif (raisonnement guidé par les buts). En effet, selon l'alternative diagnostique dans laquelle il se trouve, le médecin prescrit des examens soit pour confirmer une hypothèse envisagée à l'analyse des symptômes observés, soit pour augmenter la certitude attachée à l'une des hypothèses de maladie entre lesquelles il hésite encore.

Le caractère itératif d'une telle procédure tend à minimiser l'erreur de diagnostic. Soulignons que, compte tenu de l'urgence des actions à entreprendre en vue de l'amélioration des conditions de santé du patient, il existe à tout instant de cette procédure une adéquation entre les propositions diagnostiques et les traitements que prescrit le médecin. En effet, l'intervention d'un spécialiste, le transfert dans un centre spécialisé, les soins urgents sont autant de facteurs dont la vie du nouveau-né dépend.

Nous avons tenté, dans ce paragraphe, de montrer quelle était la démarche du médecin lors de l'établissement d'un diagnostic. Ainsi si l'on veut prétendre aider le médecin dans cette tâche, il convient de développer un système d'aide au diagnostic qui présente au moins les caractéristiques suivantes :

- être capable, à tout instant t , de fournir une ou des propositions de diagnostic à partir des données relevées avant cette date ; ceci afin d'aider le plus tôt possible le médecin dans sa démarche thérapeutique.

- proposer la prescription d'examens supplémentaires dont la mesure permettrait d'affiner ou de confirmer l'hypothèse diagnostique émise en t , ceci tout en tenant compte de leurs conditions d'obtention.

- effectuer ces deux fonctions dans un temps relativement rapide compte tenu de l'urgence des décisions thérapeutiques à prendre.

Enfin, un tel système doit être évidemment conversationnel, c'est à dire qu'à tout instant, le médecin reste le seul maître du choix définitif, et peut influencer sur les résultats fournis par la machine.

2.3 ANALYSE DES DONNEES : AIDE AU DIAGNOSTIC

Avec la mise au point des techniques nouvelles en matière d'instrumentation médicale, le médecin s'est vu contraint de synthétiser un nombre d'informations de plus en plus important.

Dans ce but, différents travaux faisant appel à des méthodes classiques de statistique et d'analyse des données lui ont permis d'appréhender cette quantité, sans cesse croissante, d'informations. Néanmoins, ces traitements en temps différé (les résultats des différentes mesures et analyses de laboratoire étant connus) conduisent en général à des études prévisionnelles ou informatives, ou à l'observation de l'évolution d'une population d'une manière épidémiologique...

Si ces méthodes restent utiles d'un point de vue descriptif et préventif (évolution d'une pathologie, évaluation des fréquences de certaines maladies dans une région donnée, ...), elles se prêtent mal aux exigences d'une aide au diagnostic dans son aspect opérationnel. En effet, elles ne prennent pas en compte le caractère évolutif du dossier médical, l'urgence des actions à mener, les conditions d'obtention des données qui caractérisent chaque malade (ordre de réalisation des examens, fiabilité des mesures, délai d'obtention, ...). C'est pourquoi, nous ne discuterons dans ce qui suit que des différents approches qui répondent aux objectifs exposés précédemment.

3 . DE L'OBSERVATION A LA DECISION

Il est clair que si l'on veut analyser un processus, il convient tout d'abord de l'observer. Cette communication qui s'établit en quelque sorte entre l'observateur et le processus couvre deux finalités essentielles.

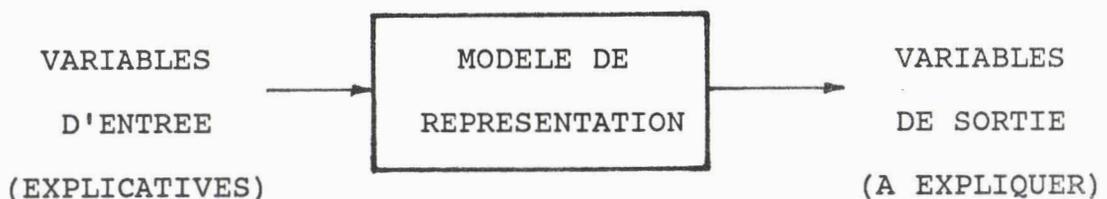
1) l'approfondissement des connaissances sur le processus.

2) la sélection d'une action à entreprendre sur le processus.

A titre d'illustration, pour 1), elle correspond à une étude statistique permettant d'indiquer si le tabac est une des causes du cancer des voies aériennes ; pour 2), c'est l'émission d'un diagnostic à partir des symptômes observés sur un patient afin de prescrire une thérapie efficace le conduisant à la guérison.

Néanmoins, cette observation reste plus ou moins objective dans le sens où l'observateur sélectionne par le biais des capteurs utilisés, les informations qui présentent un intérêt. En effet, il est facile d'imaginer que les données recueillies dans le dossier médical dépendent du cas pathologique considéré.

Chaque réalisation du processus (un malade pour notre application) est alors décrit par un ensemble de variables qui peuvent prendre un certain nombre de modalités. On suppose qu'il a été fixé d'avance une division de cet ensemble en deux groupes : les variables d'entrée du système, soit les variables explicatives, et les variables de sortie, soit les variables à expliquer, c'est à dire celles qui pourront être reconstruites à partir de la connaissance de l'état des variables d'entrée. Le problème que l'on se pose est d'expliquer les variables de sortie à partir des variables d'entrée, soit en d'autres termes, bâtir un modèle représentant le processus.



Le choix d'un modèle de représentation peut être empirique (modèles de conduite) ou bien dicté par les connaissances dont on dispose sur le processus (modèle de connaissance). Ces connaissances peuvent provenir :

a) soit des lois physiques qui régissent le processus.

Le modèle s'écrit de façon analytique (équations reliant les sorties et les entrées : $S = f(E)$), ou de façon symbolique (règles de transition).

b) soit de l'analyse de données statistiques.

Dans le cas numérique, on se donne plusieurs représentations mathématiques susceptibles de modéliser le processus ; l'expérimentation permettra d'une part de déterminer les valeurs des coefficients apparaissant dans l'écriture des modèles de conduite et d'autre part, de sélectionner celui qui représente le mieux (au sens d'un certain critère) le processus.

Dans le cas non numérique, on peut faire appel au modèle de la Reconnaissance Des Formes (RDF). Ce dernier est décrit par des formes types, elles mêmes décrites par un ensemble de prototypes. La réunion de tous les prototypes en présence constitue ce que l'on appelle une population d'apprentissage. L'affectation d'une nouvelle observation (n'appartenant pas à la population d'apprentissage) à l'une des formes types s'opère à l'aide d'une procédure de décision établie à partir de l'analyse des prototypes (par exemple, règle des plus proches voisins, procédure Bayésienne, ...).

c) soit d'experts du domaine.

Le modèle Système-Expert (SE) est constitué d'un ensemble de règles conceptualisant l'expérience professionnelle et le savoir d'un expert du domaine. Ces règles peuvent exprimer des lois physiques, ou des relations statistiquement prouvées (expérience de l'expert). Elles formulent ainsi les relations d'entrées/sorties et déterminent un arbre relationnel. La conclusion est obtenue grâce à des procédures de déplacement dans l'arbre qui activent des règles selon les faits (les modalités prises par les entrées).

Compte tenu de la structure algébrique de la variable diagnostic, nous discuterons dans ce qui suit des modèles RDF et SE qui peuvent s'appliquer à notre problème, sans pour cela faire de conclusions définitives quant à leurs validités respectives.

3.1 MODELE DE LA RECONNAISSANCE DES FORMES

Ce modèle s'appuie sur la donnée d'une population d'apprentissage et il va de soi que son emploi se limite à la résolution de problèmes où il est possible de disposer de données statistiques suffisamment importantes.

Cette population qui correspond au Système-Données (cf II.1.1) est supposée contenir l'ensemble des relations entrées/sorties nécessaires à la résolution du problème. Une première étape dite de discrimination consiste à extraire cette connaissance de cette population et permet ainsi de passer au Système-Générateur, c'est à dire à l'ensemble des relations génératrices permettant de connaître (exactement ou approximativement) le comportement des variables de sortie en fonction des variables d'entrée. L'application de ces relations ainsi déduites à une entrée particulière permet alors de résoudre le problème de décision (de l'individu sous diagnostic pour notre étude).

Il est clair que l'efficacité de ce modèle dépend de la richesse du Système-Données puisque celle-ci est liée à la quantité d'informations contenue dans les données. On peut remarquer que les connaissances extraites sont des plus objectives dans la mesure où elles ne sont pas entachées d'erreurs subjectives pouvant provenir du jugement et de la mémoire d'un expert humain.

Quant aux limites de ce modèle, nous avons vu qu'il nécessite obligatoirement la collection d'un certain nombre de réalisations du processus (données statistiques qui ne sont pas toujours disponibles dans certain cas). Par ailleurs, il faut noter aussi son incapacité à expliquer son processus de raisonnement puisque ce modèle s'appuie sur des méthodes mathématiques, statistiques, informationnelles, ... différentes de celles employées par l'expert humain. Ce qui constitue un obstacle majeur pour gagner la confiance des utilisateurs.

3.2 MODELE SYSTEME EXPERT

Si l'approche statistique (RDF) est principalement numérique, l'approche système-expert (Intelligence Artificielle) est de nature symbolique provenant de l'expérience professionnelle conceptualisée des experts du domaine.

Notons que ce modèle de décision compte de plus en plus d'applications dans le domaine médical ; citons par exemple MYCIN pour les maladies infectieuses, INTERNIST pour la médecine interne, PIP pour les maladies rénales, TROPICAID pour les maladies susceptibles d'être rencontrées dans les pays en voie de développement, Il semble d'après /BONN,81/ que la

popularité des systèmes experts dans ce domaine est due à la difficulté que l'on rencontre à formaliser le raisonnement qui mène au diagnostic.

La base de connaissance doit contenir dans un système idéal toutes les connaissances du domaine nécessaires pour qu'un système déductif agisse comme un expert (connaissances de l'état de l'art dans le domaine et connaissances de la façon de l'appliquer à une classe de problèmes) /PINS,81/.

L'approche intelligence artificielle est donc une procédure d'inclusion ; on résout le problème général qui consiste à établir les relations qui lient les sorties aux entrées, et ceci pour toutes les "valeurs" prises par les entrées et on applique certaines de ces relations au problème particulier connu d'après les faits.

L'avantage de ces systèmes est qu'ils utilisent des heuristiques acquises auprès des experts humains, ce qui leur permet de "tracer" le raisonnement utilisé en des termes compréhensibles par les utilisateurs. Cette capacité d'explicitation du mécanisme de raisonnement est un facteur primordial de confiance des utilisateurs et de convivialité ; De plus, cette propriété leur donne de grandes possibilités pédagogiques.

Néanmoins, leurs performances sont liées à la "transférabilité" du savoir des experts pour laquelle aucune méthodologie générale n'existe encore. En effet, peu d'experts humains sont capables d'exprimer de façon exhaustive toutes leurs connaissances liées à leurs domaines ; de ce fait, les erreurs et omissions peuvent être fréquentes (quoiqu'il est toujours possible de les corriger ultérieurement grâce à la propriété d'explicitation du raisonnement suivi).

Par ailleurs, si la modélisation des "inférences faibles" (règles auxquelles on associe des coefficients de crédibilité, de plausibilité ... /WONG,86/ /PASS,86/ /MORE,86/), part capitale du raisonnement humain, peut nous permettre de mieux comprendre la nature de ces dernières, différents problèmes se posent quant à l'évaluation des coefficients apparaissant dans ces règles. En effet, différentes modélisations du raisonnement approximatif sont aujourd'hui disponibles (logique floue, théorie bayésienne, ...) mais il est impossible de donner une justification théorique de la meilleure modélisation par l'une ou les autres voies choisies ; une justification ne pourra être donnée qu'expérimentalement car si les règles sont censées être vraies, on doit pouvoir les retrouver dans la réalité. Par exemple, la règle " a => b plausibilité 70 " ne veut pas dire que dans un échantillon représentatif de la population concernée, le cas " a \bar{b} " ne se rencontrerait pas, mais signifie que la cas " ab " est "plus vrai". Si la plausibilité est mesurée par une probabilité, on aurait alors :

$$\text{probabilité (b/a)} = 0.7 \qquad \text{probabilité}(\bar{b}/a) = 0.3$$

Dans ce sens, on peut alors penser que la connaissance d'un expert n'est autre qu'un ensemble de relations observées depuis des millénaires sur une population d'apprentissage de taille infinie, et formalisées de façon déclarative.

$$a \Rightarrow b \qquad \Leftrightarrow \qquad \text{probabilité}(b/a) = 1$$

Il apparaît alors que si le processus que l'on veut modéliser ne vérifie pas la propriété de stationarité, les règles (voire les coefficients de plausibilité, de crédibilité ... qui peuvent intervenir dans l'écriture de ces règles) sont sujettes à être sans cesse modifiées avec le temps. Les questions qui se posent alors sont de savoir comment constater cette évolution du processus, et quelles modifications doit on apporter au modèle. Celles-ci ne pouvant être ni déduites, ni inventées, elles ne pourront qu'être observées et c'est pourquoi, il semble raisonnable de dire que la combinaison de traitements statistiques et de techniques d'intelligence artificielle est souhaitable. En effet, même si les crédibilités fournies par l'expert ne sont pas forcément des probabilités d'occurrences de tel diagnostic face à tels symptômes, leurs valeurs mises à priori dans le système peuvent quand même être facilitées par la connaissance de statistiques si ces dernières sont disponibles. De plus, même si l'expert se fie à son expérience professionnelle pour définir la liste des faits à prendre en compte pour établir tel diagnostic, une analyse discriminante ne pourra que l'aider dans cette tâche, voire éventuellement fiabiliser ou confirmer ses intuitions. Il en résultera alors la construction d'une machine très conviviale et dont la fiabilité ne fera pas défaut avec le temps.

Dans ce sens, comme nous disposons d'une banque de données au centre de pathologie néo-natale /COUV,81/, il nous a paru opportun de l'utiliser. Notre travail portera principalement sur l'étude de méthodes permettant d'extraire de cette base de connaissance le système relationnel qui mène au diagnostic. L'approche retenue est donc celle qui fait intervenir le modèle de la reconnaissance des formes. Ce modèle qui prend en compte naturellement les probabilités ne pourra qu'aider le médecin pour expliquer son "raisonnement approximatif" car comme on peut l'imaginer, dans un tel domaine le praticien est souvent amené à manipuler des inférences faibles.

4 . CONSTRUCTION DE LA POPULATION D'APPRENTISSAGE

Bien que la banque de données ait été constituée depuis plus d'une dizaine d'années, on ne peut envisager de construire une population d'apprentissage à partir de la collection de dossiers médicaux s'étalant sur une telle période. Les progrès techniques de l'instrumentation médicale, l'évolution de la pathologie, de l'environnement (conditions de vie, ...) en sont principalement les raisons. De façon plus précise, ceux-ci ont provoqué :

- une amélioration de la fiabilité des données (analyseurs de plus en plus précis, dotés de systèmes de contrôle de qualité, ...)
- une évolution du dossier médical (présence de nouveaux types d'analyses remplaçant éventuellement d'autres)
- évolution de certaines caractéristiques (morphologie, ...)

Ainsi, pour satisfaire l'homogénéité des dossiers sur ces différents points, on ne prendra que des dossiers relativement récents, soit sur une période d'environ trois années (période fixée par le corps médical).

Par ailleurs, le nombre de nouveau-nés prématurés par an admis au centre pour des troubles respiratoires n'est guère élevé. On le restreint d'autant plus que l'on ne retiendra que ceux qui peuvent être considérés comme des cas "purs" pour la pathologie considérée ; ceci afin de s'assurer de la représentativité de l'échantillon choisi.

L'approche retenue fait appel comme on pourra le voir au calcul des probabilités. Celles-ci pourront être estimées à partir des fréquences relatives observées sur la population d'apprentissage si celle-ci est de cardinal suffisant. Le maintien d'un effectif raisonnable impose pour la construction de l'échantillon la non élimination des dossiers incomplets. D'autre part, afin d'accélérer le calcul des probabilités multidimensionnelles, on sera amené à limiter le nombre de modalités prises par les variables explicatives. Une première étape de notre travail consistera donc à étudier des algorithmes de reconstruction de données manquantes et de recodage de variable. Ces traitements préliminaires font l'objet du chapitre suivant.

- C H A P I T R E I I -

1 . INTRODUCTION

Toute analyse par les données suppose deux opérations successives ; d'abord le recueil des données et ensuite leur traitement.

Le recueil des données permet la construction d'une population d'apprentissage qui sera par la suite la seule connaissance disponible du système.

Il est clair qu'une excellente analyse effectuée sur des données "fausses" n'a évidemment aucune valeur. De ce fait, c'est bien la première des deux opérations qui est la plus importante. A la limite, on peut dire que le traitement est purement mécanique alors que le recueil suppose jugement, choix, discernement et bon sens.

Il y a de nombreuses règles à respecter pour effectuer un bon échantillonnage et être à peu près sûr de sa représentativité. De plus, l'évaluation de paramètres statistiques ou informationnels (moyenne, probabilité, entropie) que nécessitent les traitements futurs, conditionne la taille de l'échantillon.

Tout cela est fondamental mais ne constitue pas l'objet de ce chapitre. Il existe sur ce genre de question de nombreux (et excellents) ouvrages auxquels nous ne pouvons que renvoyer /BENZ,80/, /LEBA,82/, /PICA,82/, /SAPO,78/...

Dans ce chapitre, on considère les traitements préliminaires aux traitements d'analyse proprement dits. Leur rôle est de définir une population d'apprentissage utilisable à partir d'une collection de données brutes. Ils concernent essentiellement le recodage des variables et la reconstruction de données manquantes.

2 . LES DONNEES

2.1 REPRESENTATION DE LA POPULATION D'APPRENTISSAGE

Soit $\Omega = \{w_1, w_2, \dots, w_n\}$ une population d'apprentissage. On possède un ensemble de paramètres $\Sigma = \{X_1, X_2, \dots, X_p\}$ qui peuvent être mesurés sur chacun des individus de Ω . Ces paramètres présentent les caractéristiques suivantes :

- certains ne peuvent être évalués qu'à partir d'une date bien déterminée (le $j^{\text{ème}}$ jour de vie, par exemple pour un prématuré).
- certains sont mesurés de façon systématique à des intervalles de temps donnés.
- certains sont quantitatifs, d'autres qualitatifs.

Dans ces conditions, le dossier médical, c'est à dire l'ensemble des paramètres décrivant chaque individu w , se présente sous la forme suivante :

$$X(w, K) = \begin{bmatrix} X_1(w, 1) & \dots & X_1(w, K) \\ \vdots & & \vdots \\ \cdot & & \cdot \\ \cdot & X_i(w, t) & \cdot \\ \vdots & & \vdots \\ \vdots & & \vdots \\ X_p(w, 1) & \dots & X_p(w, K) \end{bmatrix}$$

dans lequel $X_i(w, t)$ est la mesure de ce paramètre au jour t pour l'individu w supposé avoir séjourné K jours au centre. Soit M_{X_i} , l'ensemble des modalités de ce paramètre, X_i peut être considéré comme une application de Ω vers M_{X_i} ; M_{X_i} contient évidemment la modalité "non mesuré".

$$\begin{array}{ccc} X_i : \Omega & \longrightarrow & M_{X_i} \\ w & \longmapsto & X_i(w) \end{array}$$

Soit $\mathcal{P}(\Sigma)$, l'ensemble des parties non vides de Σ , on notera par $\mathcal{P}_t \in \mathcal{P}(\Sigma)$, le sous-ensemble de paramètres dont la mesure est possible au jour t , $S_t \in \mathcal{P}(\mathcal{P}_t)$ un sous-ensemble de paramètres effectivement mesurés au jour t (c.à.d tel que leurs valeurs sont disponibles pour la décision en $t+1$), et M_{S_t} ,

l'ensemble des modalités pouvant être prises par ces paramètres.

Au fur et à mesure que le temps s'écoule, l'ensemble des paramètres décrivant chaque individu w devient de plus en plus grand.

$$\begin{array}{lll}
 t = 1 & : & w \longmapsto S(w) = S_0(w) \quad S \in \mathcal{P}(\mathcal{I}_0) \\
 t = 2 & : & w \longmapsto S(w) = S_0(w) \cup S_1(w) \quad S \in \mathcal{P}(\mathcal{I}_0) \times \mathcal{P}(\mathcal{I}_1) \\
 & \vdots & \\
 & \vdots & \\
 t = K & : & w \longmapsto S(w) = \bigcup_{i=0}^{K-1} S_i(w) \quad S \in \prod_{i=0}^{K-1} \mathcal{P}(\mathcal{I}_i)
 \end{array}$$

où \mathcal{I}_0 désigne le sous ensemble de paramètres décrivant l'historique de l'enfant avant son arrivée au Centre.

Lorsqu'un système est connu par un ensemble de variables, et un ensemble de modalités associé à chaque variable, il définit le Système-Source au sens de la hiérarchie épistémologique des systèmes proposée par J.KLIR /KLIR,69,75,76,77/. Il est le niveau le plus bas de connaissance de cette hiérarchie (niveau 0). Il est clair que nous travaillerons sur le niveau 1, soit le Système-Données. Il est obtenu à partir des données relevées sur la population jusqu'au jour K . Il se présente alors sous la forme d'un tableau initial de données et constitue la seule connaissance disponible pour toute notre analyse.

Observations	\mathcal{I}_0	\mathcal{I}_1	\mathcal{I}_t	\mathcal{I}_{K-1}
w_1	$S_t(w_j)$					
·						
·						
w_j						
·						
w_n						

- TABLEAU INITIAL DES DONNEES -

2.2 PARTITIONS DE Ω .

Soit $\mathcal{P}(\Omega) = \{C, C \subseteq \Omega\}$, l'ensemble des parties de Ω , on dit que $F = \{F_1, F_2, \dots, F_k\}$ est une partition en k classes de Ω si et seulement si :

$$* F_i \in \mathcal{P}(\Omega) \text{ et } F_i \neq \emptyset \text{ pour } i = 1, 2, \dots, k$$

$$* \text{ pour tout } i \neq j, F_i \cap F_j = \emptyset$$

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

Chaque variable X_i (respectivement un sous ensemble de variables S) définit par elle même une partition P_{X_i} (resp. P_S) de l'ensemble Ω . Chaque classe est formée des w identiques du point de vue de X_i (resp. S).

$$X_i \in \mathcal{P}(\Sigma), \quad X_i : \Omega \longrightarrow M_{X_i}$$

$$P_{X_i} = \{ w : X_i(w) = m, m \in M_{X_i} \}$$

En notant $P(\Omega)$, l'ensemble des partitions de Ω , on peut dire que Σ (resp. $\mathcal{P}(\Sigma)$) $\subseteq P(\Omega)$.

Par ailleurs, on peut établir dans $P(\Omega)$ la relation \leq : " être plus fine que " :

$$R = \{ R_1, R_2, \dots, R_i, \dots, R_r \} \in P(\Omega)$$

$$Q = \{ Q_1, Q_2, \dots, Q_j, \dots, Q_q \} \in P(\Omega)$$

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

Cette relation \leq est un ordre partiel ; $P(\Omega)$ muni de cette relation constitue un treillis.

Cette possibilité d'associer univoquement à chaque variable une partition de Ω , nous permet d'établir un ordre et une équivalence entre variables /TORO,82/.

Pour cela, considérons deux variables définies sur Ω :

$$\begin{array}{ccc} X : \Omega & \longrightarrow & M_X & & Y : \Omega & \longrightarrow & M_Y \\ & & w \longmapsto & X(w) & & & w \longmapsto & Y(w) \end{array}$$

On dit que:

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

Dans ce cas, il est facile de montrer qu'il existe une fonction $f : M_X \rightarrow M_Y$ bijective telle que $Y=f(X)$ et $X=f^{-1}(Y)$. X n'est alors qu'un recodage univoque de Y et vice-versa.

$$ii) \quad X \text{ est plus fine que } Y \quad \Leftrightarrow \quad P_X \leq P_Y$$

Dans ce cas, on peut montrer qu'il existe une fonction $f : M_X \rightarrow M_Y$ telle que $Y=f(X)$.

Rappelons que la théorie de l'Information ne distingue pas entre variables équivalentes.

2.3 MESURE DE PROBABILITE

Dans sa forme la plus générale /DACU,82/, une mesure de probabilité sur un ensemble fini Ω est une fonction :

$$\begin{array}{ccc} \text{Pr} : \mathcal{P}(\Omega) & \longrightarrow & \mathbb{R} \\ A & \longmapsto & \text{Pr}(A) \end{array}$$

qui vérifie :

$$- 1 - \quad \forall A \in \mathcal{P}(\Omega), \quad \text{Pr}(A) \geq 0$$

$$- 2 - \quad \forall i = 1, k, \quad A_i \in \mathcal{P}(\Omega) \text{ et } i \neq j \quad A_i \cap A_j = \emptyset \quad \checkmark$$

$$\text{Pr}\left(\bigcup_{i=1}^k A_i\right) = \sum_{i=1}^k \text{Pr}(A_i)$$

$$- 3 - \quad \text{Pr}(\Omega) = 1$$

Ainsi, si $X : \Omega \rightarrow M_X$ est une variable quelconque définie sur Ω , on définit pour toute modalité m_X de X une probabilité d'occurrence notée par $\text{Pr}(X=m_X)$.

Sous certaines hypothèses concernant le système (ergodicité, stationnarité), et le nombre d'observations de Ω /CONA,72//CRON,63/, ces quantités peuvent être estimées à partir des fréquences évaluées sur Ω .

$$m_X \in M_X, \quad \text{Pr}(X = m_X) = \frac{|\{w \in \Omega : X(w) = m_X\}|}{|\Omega|}$$

Les procédures de décisions ainsi que les différents indices informationnels utilisés par la suite nécessitent l'estimation des probabilités multidimensionnelles. Le calcul de celles-ci à l'aide des fréquences relatives est un problème simple mais dont le nombre d'opérations (donc le temps de calcul) croît très vite avec le nombre de variables et/ou le nombre de modalités prises par ces variables.

Nous nous intéresserons dans le prochain paragraphe au cas où l'on est amené à limiter le nombre de modalités pour réduire ce temps de calcul.

3 . RECODAGE DES VARIABLES

Réduire le nombre de modalités d'une variable consiste à regrouper plusieurs modalités en une seule. Bien sûr, ce recodage ne se fait pas sans perte d'information. Cette perte traduisant une diminution de la capacité d'observation de la variable en question /DUFO,76/. Dans la mesure du possible, il faut toujours chercher à rendre plus ou moins équiprobables toutes les modalités de la variable /CONA,76/. Cependant, cette option peut conduire à une extrémité regrettable ; celle de placer des individus arbitrairement dans une même classe pour des questions de masses et non par rapport à la description des données elles-mêmes. Il est certain que pour effectuer un meilleur découpage, il convient de tenir compte de l'information sur les variables (valeurs caractéristiques, domaine de variation, variable à expliquer, variable explicative) /JAMB,78/.

3.1 FORMULATION DU PROBLEME

Comme il est possible d'associer à chaque variable X définie sur Ω , une partition de Ω , ce recodage détermine une nouvelle partition de Ω .

En effet, si $P_X(k) = \{F_1, F_2, \dots, F_k\}$ désigne la partition en k classes de X induite par les données, il existe $P_X(j) \leq P_X(k)$ telle que chaque classe C_i de $P_X(j)$ soit justement l'union des classes F_i correspondant au regroupement des modalités de X .

Moyennant une renumérotation, cela peut s'écrire :

$$P_X(k) = \{ \underbrace{F_1, F_2, \dots, F_a}_{C_1}, \underbrace{F_{a+1}, \dots, F_b}_{C_2}, \dots, \underbrace{F_{b+1}, \dots, F_u}_{C_i}, \dots, \underbrace{F_k}_{C_j} \}$$

$$P_X(j) = \{ C_1, C_2, \dots, C_i, \dots, C_j \}$$

Ainsi, $P_X(j)$ détermine une partition des classes de $P_X(k)$.

Le problème s'énonce alors comme : parmi toutes les partitions plus grosses que $P_X(k)$ (partition "naturelle" associée à la variable X), trouver la meilleure au sens d'un certain critère donné.

$$P_X^*(j) = \{ C_1, C_2, \dots, C_j \}, \quad C_i \in P_X^*(j) \Rightarrow C_i = \bigcup_{q: F_q \in C_i} F_q$$

et telle que : $P_X^*(j) \not\leq P_X(j) \quad \forall P_X(j) \in P_{X,j}(\Omega)$

où $P_{X,j}(\Omega)$ désigne l'ensemble des partitions de Ω associées à la variable X en j classes ($j \leq k$),

et \succ : préférer à.

Une contrainte supplémentaire apparaît dans la formulation lorsque X est une variable ordinale. La relation d'ordre existante sur les classes F_i doit être conservée sur les classes C_i de $P_X^*(j)$. D'où la restriction suivante pour construire C_i :

En notant \mathcal{R} la relation d'ordre,

$$\forall F_i, F_{i+1} \in P_X(k) \quad \text{on a} \quad F_i \mathcal{R} F_{i+1}$$

$$\forall C_i, C_{i+1} \in P_X^*(j) \quad \text{on doit avoir} \quad C_i \mathcal{R} C_{i+1}$$

On peut remarquer que ceci entraîne en particulier :

$$C_i = \bigcup_{r: F_r \subset C_i} F_r \quad \text{tel que} \quad F_r \subset C_i, F_r \mathcal{R} C_{i+1} \text{ et } C_{i-1} \mathcal{R} F_r$$

Ce problème s'énonce alors comme un problème de classification sous contrainte d'ordre /LECH,76/.

3.1.1 CAS PARTICULIER X : VARIABLE QUANTITATIVE

Si X prend ses modalités dans R , soit sur le segment $[a, b[$ on peut associer à chaque classe de C_i de $P_X^*(j)$ un intervalle v_i, v_{i+1} .

Le rangement en classes des individus de Ω permettant d'établir $P_X^*(j)$ s'énonce alors :

$$\forall w \in \Omega, w \in C_i \Leftrightarrow X(w) \in [v_i, v_{i+1}[, [v_i, v_{i+1}[\subset [a, b[$$

Dans ce cas, trouver la partition $P_X^*(j)$ revient à découper de façon optimale le segment $[a, b[$ en j intervalles tels que :

$$\bigcup_{i=1}^j [v_i, v_{i+1}[= [a, b[$$

$$[v_i, v_{i+1}[\cap [v_k, v_{k+1}[= \emptyset \quad k \neq i$$

En d'autres termes, ceci consiste à déterminer les $j-1$ valeurs séparatrices des classes puisque $v_1=a$ et $v_{j+1}=b$.

3.2 CARACTERISATION DE $P_X^*(j)$

La détermination de $P_X^*(j)$ dépend aussi bien de la nature algébrique de X que du critère choisi pour mesurer la qualité de $P_X^*(j)$ et par voie de conséquence celle du recodage effectué.

Nous allons tout d'abord présenter des critères portants sur les caractéristiques des classes de $P_X^*(j)$. Leur emploi se limite au cas où $P_X^*(j)$ est une partition associée à une variable quantitative /MARF,83/. Par la suite, on s'intéressera à d'autres critères, indépendants de la nature algébrique des variables. Ils seront exposés en fonction du type de la variable (variable à expliquer, variable explicative).

3.2.1 CRITERES PORTANT SUR LES CARACTERISTIQUES DES CLASSES

a) CLASSES EQUIDISTANTES

La partition $P_X^*(j)$ en j classes équidistantes est obtenue en découpant le domaine des valeurs de X en j intervalles de même longueur.

Soit $(a, b) \in \mathbb{R}^2$, les bornes inférieure et supérieure de ce domaine, la longueur unitaire associée à chaque classe C_i est égale à :

$$L = (b - a) / j$$

Les valeurs séparatrices des classes sont déterminées par la récurrence suivante :

$$i = 1, j \quad v_{i+1} = v_i + L \quad \text{avec} \quad v_1 = a$$

b) CLASSES EQUIPROBABLES

On se propose de chercher la partition $P_X^*(j)$ qui présente des classes de même "poids", c'est à dire qui contiennent le même nombre d'observations de Ω .

Soit f, la densité de probabilité de X sur Ω telle que :

$$\begin{aligned} f(X) &> 0 \\ \int_a^b f(X) dX &= 1 \end{aligned}$$

Les classes C_i de $P_X^*(j)$ seront équiprobables si leurs valeurs séparatrices vérifient l'expression suivante :

$$i = 1, j \quad \int_{v_i}^{v_{i+1}} f(X) dX = 1/j \quad \text{avec } v_1 = a$$

Pour un nombre de classes donné, la partition ainsi obtenue est celle qui permet de tirer le plus d'information (au sens de la Théorie de l'Information de Shannon) de la variable X . En effet, l'entropie de X (resp. P_X) est maximale lorsque chaque modalité de X (resp. chaque classe de P_X) possède la même probabilité d'occurrence.

$$H(P_X^*(j)) = \sum_{i=1}^j \Pr(C_i) \log_2 \Pr(C_i)$$

$$H(P_X^*(j)) = H_{\max} = \log_2 j \quad \Leftrightarrow \quad \Pr(C_i) = 1/j \quad \forall i = 1, j$$

Cependant, ces deux critères ne nous renseignent pas sur la valeur optimale de " j ", le nombre de modalités à retenir pour le nouveau codage de X . Il devra alors être choisi en fonction des connaissances a priori sur le système.

Par ailleurs, signalons que ces codages ne détériorent pas la structure algébrique de la variable X .

3.2.2 CRITERES DEPENDANT DU TYPE DE VARIABLE

a) X : VARIABLE A EXPLIQUER

Dans le cadre du problème d'expliquer une variable à partir d'autres, on peut être amené à choisir comme partition de la variable à expliquer, celle qui soit le plus difficilement explicable. On se place alors dans le cas le plus défavorable pour résoudre ce problème. On espère par ce choix arbitraire assurer une plus grande fiabilité aux résultats trouvés.

Dans ces conditions, la qualité de P_X^* s'exprime par référence à une mesure d'incertitude, par exemple son entropie.

On cherche donc :

$$P_X^* \quad \text{tel que} \quad H(P_X^*) \geq H(P_X)$$

b) X : VARIABLE EXPLICATIVE DE Y

On suppose que la partition associée à la variable à expliquer Y est connue. On désire maintenant déterminer la partition P_X^* qui permet de " faciliter " la résolution du problème : expliquer Y à partir de X. En d'autres termes, que P_X^* mette en valeur ou conserve le caractère explicatif de X.

Les critères d'optimalité s'expriment par référence à des indices mesurant la contribution de X à l'explication de Y.

. des distances entre les distributions statistiques de X et de Y (distance du χ^2 ,

. des indices informationnels tels que:

I(X:Y) : quantité d'information commune entre X et Y.

H(Y/X) : quantité d'information restant à connaître sur Y quand on connaît déjà celle fournie par X.

T(Y:X) : les différents coefficients de couplage entre variables.

$$T_1 = I(Y:X)/H(Y) \quad /CONA,72/$$

$$T_2 = I(Y:X)/(H(X).H(Y))^{1/2} \quad /RICH,75/$$

...

. des indices mesurant le pouvoir décisionnel que la partition P_X permet d'obtenir pour déterminer P_Y (probabilité d'erreur, mesure de séparabilité des classes, ...)

c) (X:Y) : X VARIABLE EXPLICATIVE DE Y

Nous allons maintenant nous intéresser au couple (variable à expliquer, variable explicative). On veut déterminer à la fois P_Y et P_X sous les contraintes suivantes :

P_Y : soit explicable par P_X

P_X : conserve son caractère explicatif

La qualité des deux partitions s'exprime par référence à l'indice $H(P_Y/P_X)$. La minimisation de cet indice peut s'effectuer par une procédure de type axe par axe : minimiser par rapport à P_X , puis minimiser par rapport à P_Y avec itération de cette procédure jusqu'à stabilisation.

3.3 PARTITIONS OPTIMALES PAR NIVEAUX

Compte tenu de la monotonie des différents indices caractérisant P_X^* par rapport à la relation \leq , la meilleure partition n'est autre que la solution triviale $P_X^*(k)$, soit la partition induite par les données.

$$\begin{aligned}
 P_X \leq P_X' & \Rightarrow & H(P_X) & \geq H(P_X') \\
 & & H(Y/P_X) & \leq H(Y/P_X') \\
 & & \text{Pr d'erreur}(P_X) & \leq \text{pr d'erreur}(P_X') \\
 & & \cdot & \cdot \\
 & & \cdot & \cdot \\
 & & \cdot & \cdot
 \end{aligned}$$

Pour remédier à ce problème, la recherche de P_X^* s'effectue par niveaux.

On note par $\mathbb{P}_X(\Omega)$, l'ensemble des partitions de Ω associées aux différents codages de X. $\mathbb{P}_X(\Omega)$ muni de la relation \leq constitue un treillis.

Dans ce treillis, on définit le niveau r par l'ensemble $P_{X,r}(\Omega)$:

$$P_{X,r}(\Omega) = \left\{ \begin{array}{l} \text{partitions de } \Omega \text{ associées au codage de X en } r \\ \text{modalités} \end{array} \right\}$$

Il est clair que si X a un codage initial en k modalités (codage induit par les données sur Ω), on a :

$$P_{X,k}(\Omega) = \{ F_1, \dots, F_i, \dots, F_k \} \text{ avec } F_i = \{ w, X(w)=m_i, m_i \in M_X \}$$

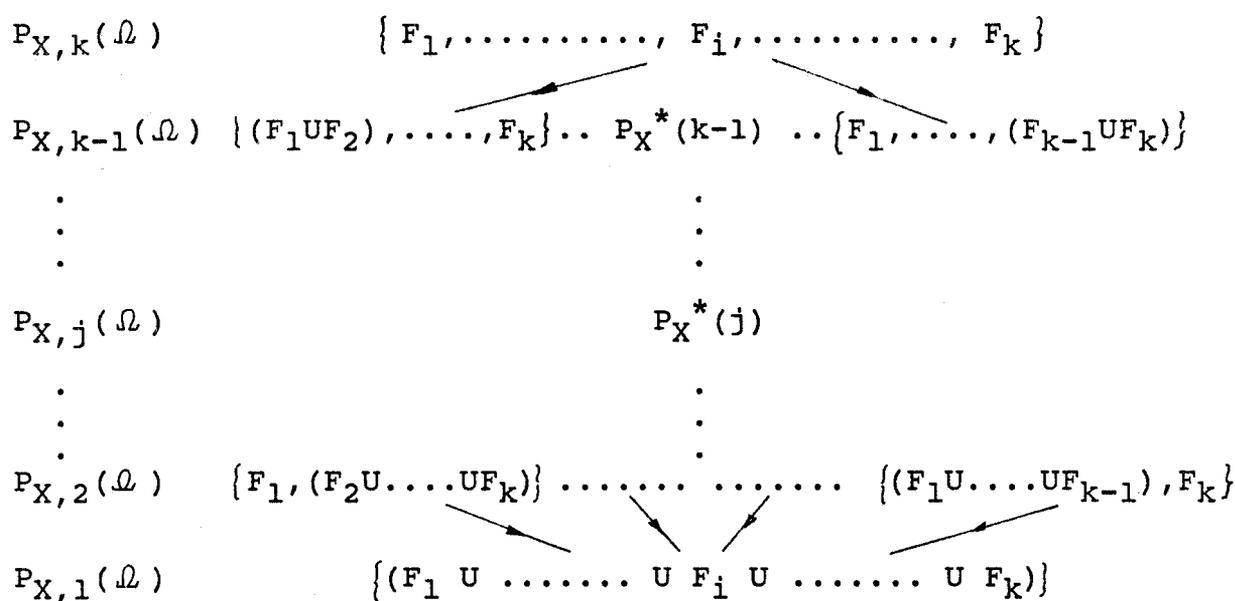
$$P_{X,1}(\Omega) = \{ (F_1 \cup \dots \cup F_i \cup \dots \cup F_k) \}$$

$$P_{X,q}(\Omega) = \emptyset \text{ pour tout } q = k+1, k+2, \dots$$

$$P_{X,j}(\Omega) \cap P_{X,i}(\Omega) = \emptyset \text{ pour tout } i \neq j$$

$$\bigcup_{j=1}^k P_{X,j}(\Omega) = \mathbb{P}_X(\Omega)$$

La représentation graphique de $\mathbb{P}_X(\Omega)$ peut se faire comme une hiérarchie par niveaux ; les arcs entre niveaux voisins correspondant à la relation \leq .



Soit J , le critère à minimiser, exprimant la qualité du recodage, le problème s'énonce alors :

Pour $j = 1, k$, déterminer la (les) partition(s) $P_X^*(j) = \{C_1^*, C_2^*, \dots, C_j^*\} \in P_{X,j}(\Omega)$ telle que :

$$\forall P_X(j) \in P_{X,j}(\Omega) \quad J(P_X^*(j)) \leq J(P_X(j))$$

Les $P_X^*(j)$ seront appelés Optimums locaux. On voit immédiatement que :

$$P_X^*(1) = \{(F_1 U \dots U F_i U \dots U F_k)\}$$

$$P_X^*(k) = \{F_1, \dots, F_i, \dots, F_k\}$$

puisque aussi bien $P_{X,1}(\Omega)$, que $P_{X,k}(\Omega)$ ne contiennent qu'une seule partition, qui est forcément l'optimum local.

J vérifiant la propriété de monotonie par rapport à \leq , les optimums locaux vérifient :

$$J(P_X^*(r)) \leq J(P_X^*(r-1))$$

Il faut cependant noter que cette propriété n'implique pas forcément que $P_X^*(r) < P_X^*(r-1)$. En effet, les optimums locaux ne constituent pas malheureusement une chaîne ordonnée des partitions. Cela rend plus difficile leur détermination comme on va le voir.

3.4 RECHERCHE DES OPTIMUMS LOCAUX

Dans l'ensemble $P_X(\Omega)$, nous ne possédons pas, bien entendu, les structures algébriques nécessaires pour les

méthodes classiques d'optimisation (dérivées, gradients, multiplicateurs de Lagrange). Il faudra que la recherche des optimums locaux se fasse aux moyens d'algorithmes qui se déplacent dans le treillis $\mathbb{P}_X(\Omega)$.

L'optimalité des différents algorithmes présentés a été discutée dans /TORO,82/.

3.4.1 RECHERCHE ASCENDANTE

Pour passer d'une partition $P_X(r) = \{C_1, C_2, \dots, C_i, \dots, C_r\}$ du niveau r à une partition du niveau $r-1$, il suffit de regrouper en une seule, deux de ses classes C_i, C_j . On se propose d'étudier un algorithme qui partant de la solution triviale en k classes, $P_X^*(k)$, traverse les différents niveaux du treillis de $\mathbb{P}_X(\Omega)$, en établissant à chaque étape le meilleur regroupement possible, au sens du critère d'optimalité.

ALGORITHME AGREGATIF : AA

A partir d'une partition en r classes, on cherche parmi toutes celles du niveau $r-1$ moins fines, celle qui minimise le critère d'optimalité. Ceci revenant à dire qu'on agrège les deux classes qui occasionnent la plus petite augmentation du critère.

Notons que si la variable est de type ordinaire, les deux classes dont on effectue le regroupement doivent être voisines.

L'algorithme n'est évidemment pas optimal. Une application de cet algorithme est donnée dans /ROMB,82/, /SBAI,83/.

3.4.2 RECHERCHE DESCENDANTE

La méthode que nous présentons maintenant est en quelque sorte l'inverse de celle décrite auparavant. Il s'agit d'un algorithme de segmentation, procédant par fractionnements successifs des classes.

En partant de la solution triviale $P_X^*(1)$, l'algorithme traverse les différents niveaux du treillis $\mathbb{P}_X(\Omega)$ par éclatement à chaque étape, d'une classe en deux.

ALGORITHME DESAGREGATIF : AD

A partir d'une partition de r classes, on cherche parmi celles du niveau $r+1$, celle qui minimise le critère

d'optimalité. Ceci s'obtient en recherchant la classe qui après éclatement en deux, provoque la plus petite augmentation du critère.

Aussi bien que AA, l'algorithme AD n'est pas optimal.

3.4.3 RECHERCHE DES OPTIMUMS LOCAUX PAR LA PROGRAMMATION DYNAMIQUE

Si les critères d'optimalité peuvent s'exprimer sous une forme additive, la recherche des optimums locaux pourra être faite à l'aide de la programmation dynamique. Ce paragraphe est inspiré d'une approche proposée par /LECH,76//KESK,84/. Elle ne concerne que les partitions associées aux variables ordinales.

Notons, toutefois qu'il est possible d'appliquer ce genre d'algorithme pour des partitions associées à des variables nominales. Mais ceci conduit à des temps de calcul très longs comme on peut le voir dans /TORO,82/, dans le cadre d'une décomposition par niveaux d'un système.

Partant de la solution triviale $P_X^*(k) = \{F_1, F_2, \dots, F_k\}$, on se propose de déterminer les partitions :

$$P_X^*(j) = \{C_1^*, C_2^*, \dots, C_j^*\} \quad \text{avec } j = 1, 2, \dots, k-1$$

telle que :

$$C_i^* \in P_X^*(j) : C_i^* = \{F_r \cup F_{r+1} \cup \dots \cup F_{r+p}\}$$

$$P_X(j) \in P_{X,j}(\Omega) \quad J(P_X^*(j)) \leq J(P_X(j))$$

où $J(P_X(j))$ caractérise la qualité de $P_X(j)$.

Soit $J(C_i)$, le coût associé à la classe C_i , $J(P_X(j))$ est tel que :

$$J(P_X(j)) = \sum_{i=1}^j J(C_i)$$

Le principe d'optimalité de Bellman nous permet d'écrire :

$$\left(\sum_{i=1}^j J(C_i) \right)^* = \min_{\substack{q \\ C_1 = \bigcup_{i=1}^q F_i \\ q : \bigcup_{i=q}^k F_i = \bigcup_{i=2}^j C_i}} \left(J(C_1) + \left(\sum_{i=2}^j J(C_i) \right)^* \right)$$

D'où le problème plus général qui consiste à déterminer :

$$\left(\sum_{i=m}^j J(C_i) \right)^* \text{ pour tout } m = 1, 2, \dots, j$$

sachant que
$$\bigcup_{i=m}^j C_i = \bigcup_{i=1}^k F_i \setminus \bigcup_{i=1}^{m-1} C_i$$

Pour la valeur particulière $m = j$, la solution n'est autre que la solution triviale :

$$J^*(C_j) = J(C_j) \quad C_j = \bigcup_{i=q}^k F_i \quad q = k-j, \dots, k$$

En conséquence, les classes C_i^* constituant $P_X^*(j)$ seront déterminées à partir de la récurrence suivante :

$$m = 1, j-1 \quad \left(\sum_{i=m}^j J(C_i) \right)^* = \text{Min} \left(J(C_m) + \left(\sum_{i=m+1}^j J(C_i) \right)^* \right)$$

$$C_m = \bigcup_{i=m}^q F_i$$

$$q : \bigcup_{i=q}^k F_i = \bigcup_{i=m+1}^j C_i$$

a) CAS PARTICULIER $X \in \mathbb{R}$

Comme nous l'avons vu, déterminer la partition $P_X^*(j)$ en j classes revient à découper de façon optimale le domaine des valeurs probables de X , soit le segment $[a, b[$, en j intervalles.

Le problème plus général consiste à découper de façon optimale le segment $[v_i, b[$ en m intervalles avec $v_i \in [a, b[$.

Soit $J(v_i, v_{i+1})$, le coût associé à l'intervalle v_i, v_{i+1} représentant la classe C_i , soit $J_j(a, b)$, le coût associé au découpage en j intervalles du segment a, b :

$$J_j(a, b) = \sum_{i=1}^j J(v_i, v_{i+1}) \quad \text{avec } v_1 = a, v_{j+1} = b$$

La récurrence fondamentale permettant de déterminer les

"j-1" valeurs séparatrices des classes s'écrit alors :

$$m = 2, j \quad J_m^*(v_i, b) = \text{Min} (J(v_i, v_{i+1}) + J_{m-1}^*(v_{i+1}, b))$$

$$v_i \in [v_{j-m}, v_{i+1} [\quad \text{si } m=j$$

$$v_i = v_1 \quad \text{si } m=j$$

avec $J_1^*(v_i, b) = J_1(v_i, b)$ pour tout $v_i \in [a, b[$

4 . RECONSTRUCTION DE DONNEES MANQUANTES

4.1 PRESENTATION

L'élaboration de tout programme permettant la reconstruction de données manquantes passe dans un premier temps par une phase d'apprentissage. Celle-ci consiste à collecter parmi les dossiers complets ceux qui s'apparentent le mieux possible au sens de certains critères à celui qui est incomplet. On attribue généralement le nom de " plus proches voisins " à cette famille de dossiers.

A partir de cette base de connaissance, on élabore une méthodologie qui permet de déterminer la modalité prise par la donnée manquante. On distingue à ce niveau différentes approches possibles :

- 1 : On suppose qu'il existe suffisamment de redondance d'informations dans le dossier médical ; la reconstruction de la donnée s'effectue à l'aide des caractéristiques connues ou une partie de celles-ci. On se pose le problème d'expliquer une variable (le paramètre non mesuré) à partir d'autres (ceux effectivement mesurés). Selon la nature des variables, on doit alors résoudre :

a) dans le cas non numérique, un problème de reconnaissance des formes avec apprentissage (sous entendant une sélection de variables explicatives).

b) dans le cas numérique, un problème de modélisation et un problème d'identification.

La modélisation d'un système est le choix d'une représentation mathématique traduisant les effets de l'entrée (les paramètres mesurés) sur la sortie du système (le paramètre manquant). Ce choix peut être empirique ou dicté par les connaissances à priori du système (linéaire, non linéaire, stationnaire ou non, entrée simple ou multiple,...). On parle de modèle de conduite dans le premier cas, de modèle de connaissance dans le second /LAUR,85/.

Quant à l'Identification, elle est la recherche des valeurs numériques des coefficients apparaissant dans le modèle. Elle est souvent obtenue par optimisation d'un critère, choisi de façon à traduire la similitude de comportement du système et du modèle vis à vis des entrées sorties.

Le choix de la méthode d'identification est lié à la représentation mathématique du modèle.

Si le modèle est de type linéaire, il existe une importante diversité de méthodes (méthodes déterministes, stochastiques, paramétriques /POVY,74//CUEN,73/).

Dans le cas non linéaire, l'identification des paramètres est plus complexe. Il existe beaucoup moins de méthodes adaptées à de tels systèmes. Celles les plus fréquemment utilisées sont la méthode du gradient, la méthode axe par axe /RICH,71//PCHE,77/. Cependant, il est possible dans certain cas, d'émettre l'hypothèse de linéarité autour d'un point de fonctionnement ; ceci simplifie considérablement le problème car on dispose ainsi de toutes les méthodes du cas linéaire /TITR,79/.

- 2 . Par l'intermédiaire des plus proches voisins du dossier incomplet, on connaît plusieurs réalisations du paramètre non mesuré. On peut considérer qu'elles déterminent l'ensemble des modalités candidates pour être l'estimation de la donnée manquante. La reconstruction est alors le choix de l'une d'entre elles. Ce dernier est obtenu par optimisation d'un critère exprimant la qualité de l'estimation.

Mais avant d'énoncer les différents critères qui nous permettront de résoudre ce problème de décision, nous allons nous intéresser à la construction du voisinage du dossier incomplet.

4.2 RECHERCHE DES PLUS PROCHES VOISINS

Il s'agit de déterminer, parmi un ensemble de dossiers complets, ceux qui ressemblent le plus au sens d'un certain critère à celui qui est incomplet. Pour cela, il nous faut définir :

i) pour chaque dossier complet, une fonction mesurant sa "proximité" avec le dossier incomplet et ceci aux moyens des caractéristiques effectivement mesurées sur chacun des dossiers.

ii) une procédure de sélection permettant de construire le voisinage en question.

4.2.1 PROXIMITE

On se propose de chercher une application P qui associe, à tout couple (w_i, w_j) d'un ensemble \mathbb{L}^2 , la valeur d'un indice $P(w_i, w_j)$ réel, positif permettant d'apprécier la proximité de w_i avec w_j .

$$\text{On cherche donc : } P : \Omega \times \Omega \longrightarrow \mathbb{R}^+$$

$$(w_i, w_j) \longmapsto P(w_i, w_j)$$

$$\text{vérifiant } P(w_i, w_j) = P(w_j, w_i)$$

$$P(w_i, w_j) = 0 \iff w_i = w_j$$

L'utilisation de distance ou indices de distance répond parfaitement à ce problème ; Ces deux applications vérifient par définition ces propriétés (Annexe I).

L'évaluation de ces indices est établie à partir des caractéristiques qui décrivent chaque élément w_i de Ω . Selon la nature des variables, on note par :

$d_p(w_i, w_j)$: l'indice $P(w_i, w_j)$ calculé à partir de p variables quantitatives (à valeurs dans \mathbb{R}).

$d_q(w_i, w_j)$: l'indice $P(w_i, w_j)$ lorsqu'il s'agit de q variables qualitatives.

a) CAS NUMERIQUE : $d_p(w_i, w_j)$

Si les p variables (X_1, X_2, \dots, X_p) décrivant l'individu w_i prennent leurs modalités dans \mathbb{R} , l'individu peut être considéré comme un point de l'espace \mathbb{R}^p .

$$\Omega \longrightarrow \mathbb{R}^p$$

$$w_i \longmapsto X(w_i) = (X_1(w_i), X_2(w_i), \dots, X_p(w_i))$$

Dans ces conditions, la quantité $d_p(w_i, w_j)$ est en général donnée par l'expression suivante :

$$d_p(w_i, w_j) = (X(w_i) - X(w_j))^t M (X(w_i) - X(w_j))$$

où M est une métrique sur l'espace \mathbb{R}^p .

Il existe un nombre important de métriques et une énumération exhaustive est pratiquement impossible. Nous ne citerons que les plus usuelles à savoir :

$M = I$: métrique identité

$M = 1/\text{Var}(X_i)$: métrique dite normée

$M = V^{-1}$: métrique de Mahalanobis (Inverse de la matrice de variance-covariance)

b) CAS NON NUMERIQUE : $d_q(w_i, w_j)$

Il existe aussi un nombre considérable de formulations possibles pour l'évaluation de $d_q(w_i, w_j)$ (JACCARD, RUSSEL & KAO, KULCHINSKI, SOKAL & MICHENER, ...). Mais celles-ci ne s'appliquent que sur des variables booléennes, c'est à dire qui ne présentent que deux modalités (0 ou 1).

Certains des paramètres qualitatifs décrivant un prématuré ne possèdent pas cette propriété de dichotomie. On sera alors amené à recoder ces variables par un codage disjonctif complet pour pouvoir appliquer ces différents indices (méthode utilisée pour transformer un tableau de réponses à un questionnaire en tableau logique /JAMB,78/).

Signalons avant de passer à l'étude de la procédure de sélection que ces transformations n'affectent pas le calcul de $d_q(w_i, w_j)$.

4.2.2 PROCEDURE DE SELECTION

La procédure de sélection s'exprime au moyen des indices définis précédemment. Etant donné Ω_c , l'ensemble des dossiers complets et w , l'individu sur lequel on a mesuré $p-r$ paramètres quantitatifs au lieu de p , $q-s$ paramètres qualitatifs au lieu de q , $w_i \in \Omega_c$ est voisin de w s'il vérifie :

$$\begin{cases} d_{p-r}(w_i, w) \leq \varepsilon_{p-r} \\ d_{q-s}(w_i, w) \leq \varepsilon_{q-s} \end{cases} \quad (\varepsilon_{p-r}, \varepsilon_{q-s}) \in \mathbb{R}^2$$

Par ailleurs, on pourra être amené à introduire des contraintes supplémentaires que devra vérifier chaque w_i retenu. Par exemple, que w_i et w présentent la même modalité diagnostic ; soit en d'autres termes, que l'indice $d(w_i, w)$ calculé sur la composante diagnostic soit nul.

$\varepsilon_{p-r}, \varepsilon_{q-s}$ sont choisis de façon à obtenir un nombre suffisant de voisins pour pouvoir permettre l'application des algorithmes de reconstruction.

Chaque voisin pourra être doté d'un poids établi en fonction de sa ressemblance avec le dossier incomplet. Ceci permet de munir le voisinage de w , V_w , d'une mesure de probabilité .

$$\begin{array}{ccc} V_w & \longrightarrow & R \\ w_i & \longleftarrow & p_i = f(d_{p-r}, d_{q-s}) \end{array}$$

tel que

$$\sum_{w_i \in V_w} p_i = 1$$

Remarques :

La détermination du voisinage de w nécessite de nombreux calculs de distances et une grande place mémoire du point de vue informatique. Cependant, il existe différents moyens pour accélérer cette recherche.

- Certains travaux établissent une première sélection de dossiers sur un espace de représentation de faible dimension. On diminue ainsi le nombre de calculs de distances dans l'espace réel /FRIE,75/.

- D'autres consistent, soit à rechercher des métriques pour lesquels les calculs sont simples et rapides /RICH,80/, soit une représentation plus simple des données /JAMB,78/.

4.3 ALGORITHMES DE RECONSTRUCTION

Nous avons vu que compléter un dossier revient à prendre comme estimation du paramètre manquant, une des modalités prises par ce paramètre sur l'ensemble des plus proches voisins. Si l'on recherche la meilleure, il faut pouvoir comparer toutes les modalités candidates et déterminer l'optimale. Bien sur, ceci est conditionné par la manière d'évaluer les performances de la reconstruction.

Nous allons tout d'abord étudier le problème élémentaire qui consiste à reconstruire une donnée manquante. Par la suite, on envisagera une extension de ce problème à k données manquantes.

4.3.1 RECONSTRUCTION D'UNE DONNEE MANQUANTE

Soit X , le paramètre non mesuré sur l'individu w , et $M_X = \{m_1, m_2, \dots, m_p\}$ l'ensemble des modalités induit par l'ensemble des plus proches voisins de w , V_w .

Il s'agit de déterminer $\hat{X}^*(w)$, estimation de X sur l'individu w tel que :

$$J(\hat{X}^*(w)) \geq J(\hat{X}(w)) \quad \text{pour tout} \quad \hat{X}(w) \in M_X$$

où $J(\hat{X}(w))$ mesure la qualité de la reconstruction de $\hat{X}(w)$.

Parmi les différentes mesures, on distingue :

a) VARIANCE MINIMALE

On désire sélectionner parmi toutes les estimations possibles de X , celle qui est le centre de la distribution statistique de X sur l'ensemble des plus proches voisins. La performance de l'estimation est évaluée à partir de la variance de X .

$$\text{VAR}(X) \leq \text{VAR}(X) \quad \forall \hat{X}(w) \in M_X$$

$$\hat{X}^*(w) \quad \hat{X}(w)$$

$$\text{avec} \quad \text{VAR}(X) = \sum_{i=1}^p \text{Pr}(m_i) (m_i - \hat{X}(w))^2$$

$$X(w)$$

La solution de ce problème n'est autre que la valeur moyenne de X sur les plus proches voisins.

$$X^*(w) = \sum_{i=1}^p \text{Pr}(m_i) \cdot m_i$$

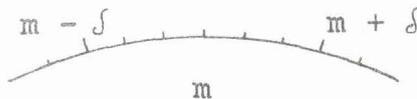
b) VRAISEMBLANCE MAXIMUM

Il s'agit de prendre pour estimation de la donnée manquante celle qui apparaît avec la plus grande probabilité sur l'ensemble des plus proches voisins, soit le mode.

$$\hat{X}^*(w) \text{ tel que } \text{Pr}(\hat{X}^*(w)) \geq \text{Pr}(\hat{X}(w)) \quad \forall \hat{X}(w) \in M_X$$

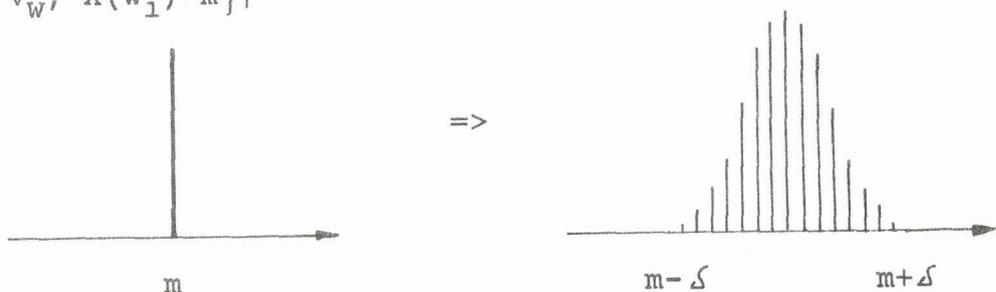
Jusqu'à présent, les deux estimateurs, dont la simplicité de mise en oeuvre est évidente, ont une précision infinie. Cependant, en réalité les données issues des différents dossiers sont évaluées avec une certaine précision (erreurs de mesure, précision des appareils, ...). On peut alors imaginer que les individus présentant la modalité m pour le paramètre X pourraient être distribués dans l'intervalle $[m-\delta, m+\delta]$ avec δ dépendant de la précision de la mesure de X .

Dans cet intervalle, le nombre de modalités à retenir est lié à la précision absolue (par exemple, la précision de lecture).



On peut alors penser : dans la réalité, il existe une distribution des individus dans cet intervalle, dont m est la valeur moyenne. Ceci conduit alors à décomposer (suivant un modèle plus ou moins sophistiqué) la fréquence observée sur V_w relative à la modalité m en un certain nombre de raies dans l'intervalle $[m-\delta, m+\delta]$.

$$|\{w_i \in V_w, X(w_i)=m\}|$$

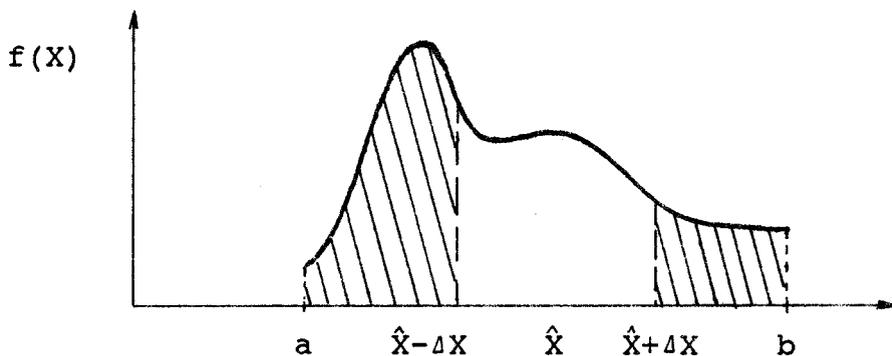


Dans ces conditions, on se propose d'étudier un estimateur dont la fiabilité dépend de la précision avec laquelle le paramètre à reconstituer est mesuré.

c) PROBABILITE D'ERREUR

La définition de la fiabilité donnée par la Commission Electrotechnique Internationale /PAGE,80/ permet de mettre en évidence la notion de probabilité de succès (ou d'erreur) pour évaluer cette quantité.

Dans le cadre de reconstruction de donnée, pour une estimation \hat{X} et une précision ΔX données, la probabilité d'erreur définit un degré de fiabilité à accorder à la grandeur reconstituée. Elle varie dans le sens contraire de la fiabilité, c'est à dire que plus sa valeur est petite, meilleure est l'estimation. Sa représentation graphique est donnée par l'aire hachurée de la figure ci-dessous.



- DENSITE DE PROBABILITE DE X SUR V_w -

Calculons sa valeur,

$$\begin{aligned} \text{Pr d'erreur}(\hat{X}) &= \int_a^{\hat{X}-\Delta X} f(x) dx + \int_{\hat{X}+\Delta X}^b f(x) dx = 1 - \int_{\hat{X}-\Delta X}^{\hat{X}+\Delta X} f(x) dx \\ &= 1 - \text{Pr}(\hat{X} \in [\hat{X}-\Delta X, \hat{X}+\Delta X[) \\ &= \text{fonction}(\hat{X}, \Delta X) \end{aligned}$$

En conséquence, la formulation du problème d'estimation optimale de $X(w)$ est donnée par :

pour un ΔX fixé, déterminer $\hat{X}^*(w)$ tel que :

$$\text{Pr d'erreur}(\hat{X}^*(w)) \leq \text{Pr d'erreur}(\hat{X}(w)) \quad \forall \hat{X}(w) \in M_X$$

Quant à la précision à donner à la reconstruction de X , soit ΔX , elle sera choisie en fonction des conditions normales

d'obtention du paramètre manquant (erreur de mesure, précision de l'appareil...)

$$X = g_1(\Delta X) \quad \Delta X : \text{erreur absolue de la mesure.}$$

$$X = g_2(\Delta X/X) \quad \Delta X/X : \text{erreur relative.}$$

$$X = g_1(\Delta X) + g_2(\Delta X/X)$$

Cependant, cette méthode de reconstruction nécessite la "connaissance" de la loi de distribution de la variable manquante sur l'ensemble des plus proches voisins. On ne dispose en général que des fréquences relatives. Le passage de celles-ci à la loi de densité /HABB,79/ alourdira au niveau temps de calcul la procédure de reconstruction.

Dans l'annexe II, différents exemples de reconstitution sont donnés.

4.3.3 RECONSTRUCTION DE K VARIABLES MANQUANTES

a) RECONSTRUCTION SIMULTANEE

Soit w , un individu sur lequel on a mesuré $n-k$ paramètres au lieu de n , V_w , l'ensemble des plus proches voisins de w établi sur l'espace Σ^{n-k} , le but que l'on se fixe maintenant est de reconstituer de manière optimale le vecteur $X(w) = (X_1(w), X_2(w), \dots, X_k(w))$.

Dans le cas où les composantes de $X(w)$ sont indépendantes, le problème se trouve grandement simplifié puisqu'il peut se décomposer en k problèmes élémentaires.

Malheureusement, cette hypothèse d'indépendance ne peut pas être faite dans bien des cas. On peut alors envisager la recherche de $X^*(w)$ à partir d'algorithmes tels que la méthode du gradient, la méthode axe par axe si les structures algébriques le permettent.

Toutes ces méthodes nécessitent de nombreux calculs (estimations des fonctions de répartition multidimensionnelles, probabilités conjointes ...) ; ce qui les rend trop pénalisantes au niveau du temps de calcul si k a une valeur élevée. Pour palier à ces problèmes, nous proposons deux méthodes :

a) La première consiste à rechercher un nouvel espace de représentation des individus dans lequel le problème peut se décomposer en k problèmes élémentaires. Cette solution est optimale comme on pourra le voir mais son emploi se limite au

cas où les k paramètres non mesurés sont quantitatifs, et où on envisage que les dépendances ou indépendances linéaires entre variables.

b) La deuxième reconstitue $X(w)$ composante par composante. Elle est évidemment sous optimale mais n'impose aucune structure algébrique sur les variables manquantes. Elle définit la reconstruction séquentielle.

RECHERCHE D'UN NOUVEL ESPACE DE REPRESENTATION

Au lieu de travailler sur les variables manquantes, on se propose de travailler dans l'espace de leurs combinaisons linéaires, muni du produit scalaire défini par la covariance. Ceci permet d'exprimer la notion d'indépendance entre variables par la notion d'orthogonalité. Dans ces conditions, on désire que ce nouvel espace soit muni d'une base orthonormée.

Ceci se formule :

$$\exists k \leq p, \quad i = 1, k \quad Y_i = \sum_{j=1}^p a_{ji} X_j \quad \text{tel que} \quad E(Y_i) = 1$$

$$\text{et } j \neq i \quad E(Y_i, Y_j) = 0$$

Ceci revient à déterminer :

$$Y = A X \quad \text{avec } A = [a_{ij}] \quad \text{tel que } E(Y^t Y) = I$$

$$\text{avec } E(Y^t Y) = A^t V A$$

où V désigne la matrice de variance-covariance dont le terme général V_{ij} est égal à la covariance des variables X_i et X_j , et A^t la matrice transposée de A .

En multipliant à gauche par A , et à droite par A^t le terme $E(Y^t Y)$, on obtient :

$$A E(Y^t Y) A^t = A A^t V A^t A = A A^t$$

Supposons que AA^t soit inversible, l'expression devient :

$$V A^t A = I \quad \Leftrightarrow \quad A^t A = V^{-1}$$

Ainsi, toute matrice A vérifiant cette condition permet d'obtenir un nouvel espace de représentation où le problème de reconstruction s'énonce plus simplement.

Une fois les variables Y_i reconstruites, le retour à l'espace initial consiste à résoudre le système suivant :

$$Y^* = A X^*$$

b) RECONSTRUCTION SEQUENTIELLE

Dans le cas où les structures algébriques ne permettent pas d'effectuer un changement d'espace de représentation et/ou que k est trop élevé pour appliquer une recherche exhaustive, la reconstruction séquentielle définit une heuristique simplificatrice pour résoudre le problème. L'algorithme correspondant est le suivant :

```

Initialisation  $i = 0$ 
tant que  $i < k$ 
  Rechercher  $V_w$  sur l'espace  $\Sigma^{p-k+i}$ 
  Pour tout  $X_i \in X$ ,
    déterminer son estimation optimale
  fin du pour;
  Accepter la meilleure estimation  $X^*_j$ 
   $X = X \setminus X^*_j$ 
   $i = i + 1$ 
fin du tant que;
Ecrire  $X^*$ 
fin:

```

4.4 UNE VOIE POSSIBLE DANS LA RECHERCHE D'UN INDIVIDU REPRESENTATIF D'UNE CLASSE

Une approche classique consiste à prendre pour élément représentatif d'une classe le centre de gravité de la classe, soit l'individu "moyen". L'approche proposée ici consiste à remplacer cette notion d'individu moyen par celle d'individu le mieux reconstitué.

4.4.1 NOTION D'INDIVIDU LE MIEUX RECONSTITUE

L'individu le mieux reconstitué est obtenu à partir d'un individu "vide" pour lequel on a reconstruit toutes ses composantes, sauf bien entendu sa classe d'appartenance qui est à priori connue.

Notons, que si la reconstruction de ses paramètres est établie de façon simultanée à partir de l'estimateur à variance minimale, l'individu le mieux reconstitué est le centre de gravité de la classe considérée. Pour les autres estimateurs, nous avons vu qu'il ne l'était pas forcément (cf annexe II).

4.4.2 APPLICATION A LA RECONNAISSANCE DES FORMES

La reconnaissance des formes est constituée par une catégorie de problèmes dont l'objet est de déterminer la forme (ou la classe) à laquelle appartient un élément, et ceci à partir de paramètres qui le représentent. Sa formulation sera donnée de façon plus précise dans le chapitre suivant.

Parmi les règles d'affectation les plus couramment utilisées, on distingue celles à caractère géométrique. L'approche de Sébestyen /ROME,73/ consiste à affecter l'individu à la classe qui lui est le plus proche au sens d'une certaine métrique définie sur l'espace des individus.

Dans le cas où la classe est représentée par un seul individu, soit par exemple l'individu le mieux reconstitué, la règle d'affectation s'écrit alors :

$$w \in \text{Classe } i \iff d(w, \bar{w}_i) < d(w, \bar{w}_j) \text{ pour tout } \bar{w}_i \neq \bar{w}_j$$

où \bar{w}_i désigne l'individu le mieux reconstitué, représentatif de la classe i .

5 . CONCLUSION

Dans ce chapitre, nous nous sommes intéressés aux différents problèmes qui sont liés à la construction d'une population d'apprentissage à partir de la Banque de Données élaborée au Centre. Différents critères et méthodes ont été proposés en ce qui concerne le recodage des variables et la reconstruction de données manquantes. Celles-ci nous ont conduits à la définition d'un "individu fictif" (individu le mieux reconstitué), dont les caractéristiques peuvent être utilisées pour la discrimination et le classement.

Ces traitements préliminaires étant formulés, nous allons nous intéresser au chapitre suivant à l'exploitation de cette base de connaissance, c'est à dire au modèle de la Reconnaissance des Formes.

- C H A P I T R E I I I -

1 . INTRODUCTION

La reconnaissance des formes /SIMO,84/ est constituée par une catégorie de problèmes dont l'objet est l'identification automatique d'un être physique à partir des données qui le caractérisent.

Les domaines d'application de cette théorie, et par conséquent l'aspect que peuvent revêtir de tels problèmes, sont très variés, allant de l'analyse de la parole, la reconnaissance de figures géométriques sur une image, le contrôle de processus et le suivi de leur fonctionnement /SMOL,82/ /GREN,85/, jusqu'à l'élaboration de diagnostics médicaux /KULI,70/.

Lorsque l'on aborde un problème de classement, on constitue en général une population d'apprentissage. Cet ensemble est formé d'un certain nombre de classes que l'on connaît à priori, ou bien que l'on a mises en évidence par des méthodes de classification et qui correspondent à des états connus du phénomène étudié. On dispose alors de prototypes représentatifs de chacune des classes en présence.

Munis de cette population d'apprentissage, le but de la discrimination est de trouver, à partir de l'analyse de ces prototypes une ou plusieurs règles de décision permettant d'affecter des individus nouveaux à l'une des classes répertoriées.

Suivant la nature du phénomène étudié, la structure algébrique des variables qui décrivent les individus, on a recours à des procédures à caractère géométrique ou relevant de la théorie de la décision statistique.

a) PROCEDURES A CARACTERE GEOMETRIQUE

On peut citer, par exemple, les techniques de discrimination linéaire ou non linéaire. Elles consistent à déterminer une ou plusieurs hypersurfaces qui servent de frontières entre les classes d'apprentissage. Si les classes sont linéairement associées, ces surfaces sont des hyperplans. Les figures suivantes illustrent les deux cas, linéaire, non linéaire :

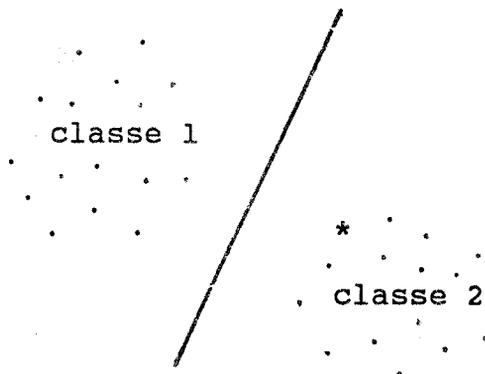


fig A : Discrimination
linéaire

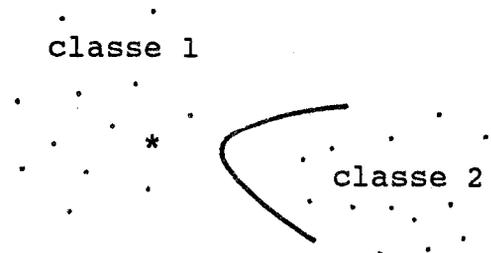


fig B : Discrimination
non linéaire

L'affectation d'un individu nouveau se fait en fonction de sa position par rapport aux frontières déterminées auparavant. Ainsi l'individu représenté par un point "*" sera affecté à la classe 2 dans le cas de la figure A, et à la classe 1 dans le cas de la figure B. Précisons enfin que les hypersurfaces sont calculées à l'aide des individus de l'ensemble d'apprentissage.

Il existe également des techniques ne faisant pas appel à un découpage de l'espace de représentation. Ce sont par exemple les règles de type k plus proches voisins /COVE,67/. Un individu est affecté à la classe C_i si parmi ses k voisins les plus proches (au sens d'une métrique), la majorité de ses voisins appartient à la classe C_i . Notons que cette règle a donné lieu à quelques variantes : pondération des individus /BRON,84/, modulation des distances de l'individu à classer à chacun de ses voisins /DUDA,76/, morphologie de la succession des classes représentées par les plus proches voisins /GREN,85/.

b) PROCEDURES STATISTIQUES

On suppose dans ce cas que les lois de probabilité des classes sont connues (hypothèse paramétrique) ou que les différentes mesures de probabilités peuvent être estimées à partir de fréquences observées sur la population d'apprentissage (hypothèse non paramétrique). On utilise alors des procédures qui font appel à ces quantités pour déterminer de façon optimale, au sens de la théorie de la décision statistique /ARQU,79/, la classe d'appartenance des nouveaux individus. Soulignons que cette façon de procéder permet d'évaluer la qualité du classement opéré et qu'elle n'impose pas (dans le cas non paramétrique) de contrainte particulière sur la structure algébrique des variables qui caractérisent les individus à identifier.

Compte tenu de la nature des données relevées dans le dossier médical, nous aborderons dans ce chapitre, le problème de la reconnaissance des formes comme un problème de décision statistique sous l'hypothèse non paramétrique. Nous présenterons tout d'abord le modèle statique, puis nous envisagerons les extensions de ce modèle à des situations plus complexes sur le plan temporel.

Enfin, nous préciserons comment il est possible d'évaluer les performances d'une discrimination. Certains indices exposés seront utilisés par la suite lorsque nous formulerons le problème de la reconnaissance des formes relatif à l'aide au diagnostic médical.

2 . RECONNAISSANCE DES FORMES : PROBLEME STATIQUE

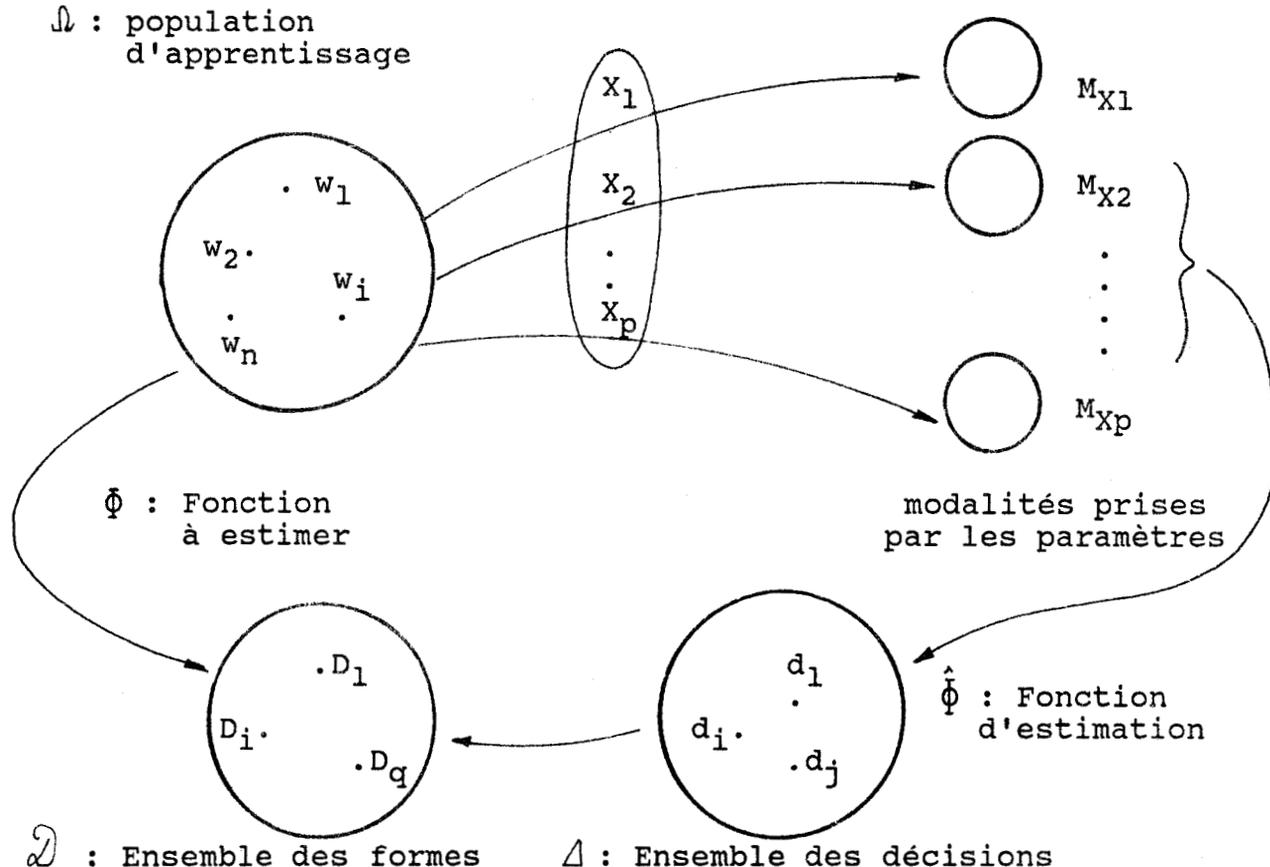
2.1 PRESENTATION DU MODELE

On suppose que l'on dispose d'une population d'apprentissage $\Omega = \{w_1, w_2, \dots, w_n\}$. Ces individus se classifient dans des formes ou classes, chaque classe étant associée à une modalité de la variable à expliquer. Pour cela, on possède un ensemble de paramètres X_1, X_2, \dots, X_p qui peuvent être mesurés sur chacun des individus de Ω .

Le diagramme suivant résume cette formulation.

S : Ensemble des paramètres

Ω : population d'apprentissage



- PROBLEME DE LA RECONNAISSANCE DES FORMES -

L'objectif est alors d'estimer la fonction $\hat{\Phi}$, qui à chaque individu w fait correspondre sa forme $\hat{\Phi}(w)$, à l'aide des paramètres X_1, X_2, \dots, X_p . On suppose bien entendu que sur la population d'apprentissage, la fonction Φ est connue.

En d'autres termes, on veut prendre pour chaque modalité prise par les paramètres X_1, X_2, \dots, X_p , une décision d_i concernant le choix d'une des formes de l'ensemble \mathcal{D} .

L'ensemble des modalités prises par les paramètres, M_S , l'ensemble des formes à reconnaître, \mathcal{D} , et l'ensemble des décisions, Δ , constituent la structure décisionnelle du problème. Dans cette structure, il existe une relation entre \mathcal{D} et Δ que nous précisons par la suite (cf 2.2).

Dans le modèle statique, cette structure est invariante dans le temps. La décision (c.à.d la reconnaissance) est unique. Chaque individu de Ω est décrit par un ensemble de paramètres qui est utilisé en partie ou en totalité par le traitement. Si le principe de causalité est respecté, la reconnaissance ne peut intervenir que postérieurement à la mesure de ces paramètres.

Etant données les notations définies en II.2.1, si la reconnaissance s'effectue après une période d'observation de K jours, il s'agit de :

i) trouver le sous-ensemble S de paramètres qui fournit le plus d'information pour déterminer quelle est la forme (ou la classe) à laquelle appartient chaque individu (problème de sélection de variables discriminantes).

ii) résoudre le problème de décision, soit construire une fonction $\hat{\phi}$, estimation de ϕ telle que :

$$\begin{array}{l} \hat{\phi} : M_S \longrightarrow \Delta \\ m_S \longmapsto \hat{\phi}(m_S) \quad \text{avec } S \in \prod_{t=0}^{K-1} \mathcal{P}(\mathcal{I}_t) \end{array}$$

où m_S désigne une des modalités prises par S : $m_S \in M_S$

Δ , l'ensemble des décisions

S, le sous ensemble de paramètres décisionnels effectivement mesurés au jour K

2.2 DONNEES DU MODELE DE LA R.D.F

Le diagramme du problème de la reconnaissance des formes présenté au paragraphe précédent est construit sur les données suivantes :

A : une structure décisionnelle (\mathcal{D}, Δ, M_S)

a) \mathcal{D} : l'ensemble des formes

Il se confond avec l'ensemble des modalités de la variable à expliquer.

Soit D , cette variable, D_i , $i=1, 2, \dots, q$ ses modalités on notera:

$$\mathcal{D} = M_D = \{ D_1, \dots, D_i, \dots, D_q \}$$

Cet ensemble est probabilisé ; les probabilités à priori pouvant être connues ou non. Lorsqu'elles sont connues, elles le sont en général à partir des fréquences relevées sur la population d'apprentissage, et peuvent être utilisées dans les méthodes de décision (règle de Bayes). Lorsqu'elles ne sont pas connues, on est amené à mettre en oeuvre des procédures de décision ne les utilisant pas, ou à les remplacer par leurs valeurs les plus défavorables (entropie maximale, règle du minimax, modèle pseudo-bayésien /ARQU,79//SAPO,78/).

b) M_S : l'ensemble des modalités prises par les paramètres décrivant chaque individu de Ω

Il est probabilisé par l'intermédiaire de Ω . La mesure de probabilité de S dépend du paramètre D ; c'est la loi de probabilité de S conditionnelle à D : $\Pr(S/D)$.

c) Δ : l'ensemble des décisions

Il est dans la majorité des applications en correspondance bijective avec l'ensemble des formes.

$$\Delta = \{d_1, \dots, d_i, \dots, d_q\} \text{ tel que } d_i \Leftrightarrow D_i \quad \forall D_i \in \mathcal{D}$$

Cependant, d'autres ensembles de décisions peuvent être considérés, définissant d'autres approches possibles.

i) ETUDE LOCALE

De façon générale, une décision peut être associée à l'une des parties de \mathcal{D} . L'étude locale consistera à prendre pour Δ :

$$\Delta = \Delta_i = \{d_i, \bar{d}_i\} \quad \text{tel que} \quad d_i \Leftrightarrow D_i$$

$$i \in \{1, \dots, q\} \quad \bar{d}_i \Leftrightarrow \bigcup_{j \neq i} D_j$$

Par ce biais, la variable à expliquer est considérée comme une variable présentant deux modalités ; l'une étant celle que l'on veut étudier, l'autre regroupant toutes celles qui restent.

Il y aura autant d'études locales de ce type que de modalités prises par la variable à expliquer. On peut remarquer que leurs résultats ne sont pas forcément cohérents entre eux.

ii) ETUDE GLOBALE

L'ensemble des décisions est l'ensemble des parties de l'ensemble des modalités de la variable à expliquer.

$$\Delta = \mathcal{P}(M_D)$$

Une telle définition permet la prise en compte d'informations partielles, par la souplesse qu'elle autorise au niveau des décisions : par exemple $\phi(m_S) = D_1 \cup D_5$.

Compte tenu de cette souplesse, nous distinguerons dans Δ :

1) les décisions simples ; d_i avec $i=1, q$

Elles concernent le choix d'une seule forme, et sont donc associées aux parties de \mathcal{D} à un seul élément ; $d_i \in M_D$.

2) les décisions composées (ou multiples) ; d_i avec $i = q+1, 2^q-1$

Elles ne permettent pas la reconnaissance d'une forme de façon unique ; elles sont associées aux parties de \mathcal{D} à plus d'un élément ; $d_i \in \mathcal{P}_2(M_D)$, ensemble des parties de M_D de cardinal supérieur ou égal à 2.

Nous reviendrons sur l'intérêt de cet ensemble de décisions lorsque nous aborderons le problème de l'aide au diagnostic (problème de décisions successives). Nous regarderons aussi les conséquences qu'une décision multiple peut apporter sur le choix du vecteur de paramètres explicatifs.

B : un ensemble fonctionnel : ϕ

Pour la construction de la fonction d'estimation de ϕ , l'observateur dispose d'applications ϕ , appelées stratégies ou

procédures de décision statistiques parmi lesquelles on distingue /ARQU,79/:

- 1) $\hat{\phi}_{na}$: les stratégies non aléatoires (applications de M_S dans Δ).
- 2) $\hat{\phi}_a$: les stratégies aléatoires (applications de M_S dans un ensemble de lois de probabilité sur Δ).

Toutefois, une solution sous la forme de fonction de décision aléatoire peut paraître parfois moins pratique pour l'observateur qu'une stratégie non aléatoire (si l'action à associer à la décision doit être déterministe).

Ainsi, pour l'aide au diagnostic, on se contentera de rechercher $\hat{\phi} \in \hat{\phi}_{na}$ puisque la connaissance du diagnostic permet au médecin de prendre une décision thérapeutique. $\hat{\phi}$ sera donc une application de M_S dans Δ , qui associe à toute modalité m_S de S une décision d_i .

$$\begin{array}{ccc} \hat{\phi} : M_S & \longrightarrow & \Delta \\ m_S & \longmapsto & \hat{\phi}(m_S) = d_i \end{array}$$

Comme chacune des modalités m_S induit sur Δ une loi de probabilité par l'intermédiaire de D et de \mathcal{A} , il sera possible d'associer à chaque estimation un indice de qualité.

C : un ou des indices de performance

La résolution du problème de la R.D.F nécessite le choix d'une procédure de décision et d'un sous-ensemble de variables discriminantes. Les choix étant faits, ils conditionnent très fortement le résultat. Il existe dans la littérature un grand nombre d'indices qui permettent d'évaluer les performances d'un système de reconnaissance ; une énumération exhaustive est impossible, et nous nous limiterons à exposer au paragraphe 4 de ce chapitre ceux que nous avons étudiés.

Le dossier médical évoluant au cours du temps, il est clair que le problème de décision relatif à une aide au diagnostic dans le contexte d'une médecine d'intervention ne peut être statique. C'est pourquoi nous présenterons tout d'abord les différents problèmes de décisions successives.

3 . DECISIONS SUCCESSIVES

Jusqu'à présent, le modèle de la reconnaissance des formes présenté correspond au cas statique. La décision conduisant à une reconnaissance est unique ; elle est prise à un instant fixé par l'observateur par rapport à la tranche d'observation.

On se propose maintenant d'appliquer ce modèle à des situations plus complexes sur le plan temporel où l'observateur prend une suite de décisions sur un ensemble d'instantanés donnés, $t = 1, 2, \dots, K$. On considère dans cet esprit la décision séquentielle et la décision dynamique.

3.1 DECISION SEQUENTIELLE

Le problème de la décision séquentielle correspond au cas où il y a acquisition d'informations jusqu'à la prise de décision effective. Celle-ci est prise dès que possible, sur un ensemble d'instantanés de décision fixés à priori, sous la contrainte de réaliser une performance considérée comme suffisante, voire optimale en un certain sens.

La structure décisionnelle du problème possède les propriétés suivantes :

- \mathcal{D} : l'ensemble des formes est indépendant du temps.

- M_S : il varie dans le temps par l'acquisition de mesures supplémentaires

- Δ : l'ensemble des décisions est indépendant du temps. Il lui est adjoint (sauf dans le cas de l'étude globale) une décision supplémentaire "reconnaissance non performante" à tous les instantanés de décision autres que celui qui termine éventuellement la tranche d'observation. La décision "reconnaissance non performante", notée par la suite x , ou une décision multiple (qui ne permet pas la reconnaissance d'une forme de façon unique) entraîne l'acquisition de mesures supplémentaires .

On doit alors bâtir autant de fonctions d'estimation que nécessaire pour arriver à une décision unique, et ceci le plus tôt possible.

$$\begin{array}{llll}
 t = 1 & \hat{\phi}_1 & : M_S \longrightarrow \Delta^* & s \in \mathcal{P}(\mathcal{I}_0) \\
 t = 2 & \hat{\phi}_2 & : M_S \longrightarrow \Delta^* & s \in \mathcal{P}(\mathcal{I}_0) \cdot \mathcal{P}(\mathcal{I}_1) \\
 \vdots & \vdots & & \vdots \\
 t = K & \hat{\phi}_K & : M_S \longrightarrow \Delta & s \in \prod_{t=0}^{K-1} \mathcal{G}(\mathcal{I}_t)
 \end{array}$$

avec $\Delta^* = \Delta \cup \{x\}$

Ce problème peut être résolu par décomposition en K problèmes statiques pour chacun desquels S est fonction du temps. Pour les $K-1$ premiers problèmes, la décision supplémentaire "reconnaissance non performante" est adjointe à l'ensemble des décisions Δ .

3.2 DECISION DYNAMIQUE

Le problème de décision dynamique correspond au cas où il y a acquisition d'informations et contrainte d'effectuer une reconnaissance à chaque instant d'un ensemble d'instant de prise de décision. C'est a priori le problème global. La structure décisionnelle $(M_S, \mathcal{D}, \Delta)$ varie dans le temps.

S peut s'accroître de façon analogue au problème de décision séquentielle. Il peut être soumis à la contrainte d'oublier l'information la plus ancienne.

Là aussi, ce problème peut être décomposé en une suite de problèmes de décision statique, pour chacun desquels Δ, \mathcal{D}, M_S sont fonctions du temps.

On peut alors dire : dans un contexte de médecine d'urgence, une aide au diagnostic dans son aspect opérationnel s'apparente à ce type de problème de décision. En effet, on doit, compte tenu de l'information disponible (les paramètres effectivement mesurés à l'instant de décision), émettre des propositions de diagnostic à tout instant t afin d'aider le plus tôt possible le médecin dans sa démarche thérapeutique. L'évaluation de ces décisions fait l'objet de ce qui suit.

4 . EVALUATION DES PERFORMANCES

Une fois $\hat{\phi}$ et S déterminés, le problème de la R.D.F est résolu. Il est clair que le choix de $\hat{\phi}(S)$, estimation de ϕ dépend aussi bien de la nature algébrique des variables que de l'information disponible à priori sur le phénomène concerné par la décision (le diagnostic). Cette information est essentiellement de nature statistique, relative aux lois de probabilité des paramètres aléatoires S et éventuellement de D.

La variable "diagnostic" étant qualitative, nous ne parlerons pas du cas numérique qui relève de la modélisation et de l'identification.

Pour notre étude, définir la fonction d'estimation $\hat{\phi}(S)$ revient à définir une procédure de décision $\hat{\phi} \subset \hat{\phi}_{na}$, qui associe à toute modalité m_S de la variable S une décision $\hat{\phi}(m_S)$. Dans ces conditions, évaluer la qualité de cette estimation revient à calculer une "distance" entre les deux applications :

$$\begin{array}{ccc} \phi : \Omega & \longrightarrow & \mathcal{D} \\ w & \longmapsto & \phi(w) \\ \\ \hat{\phi} : M_S & \longrightarrow & \Delta \\ m_S & \longmapsto & \hat{\phi}(m_S) \end{array}$$

Par ailleurs, il est possible de déterminer des indices qui permettent de mesurer uniquement les performances de S sans pour cela considérer $\hat{\phi}(S)$.

4.1 EVALUATION DES PERFORMANCES DE $\hat{\phi}(S)$

En appliquant la fonction d'estimation $\hat{\phi}$ sur chaque élément de la partition de Ω associée à S, $P_S(\Omega)$, il nous est possible de construire la partition de Ω associée à l'estimation $\hat{\phi}(S)$, soit $P_{\hat{\phi}}(\Omega)$.

$$\begin{array}{ccc} \Omega & \xrightarrow{\text{mesures : S}} & m_S \quad P_S(\Omega) \\ \downarrow \phi & & \downarrow \hat{\phi} \\ P_{\phi}(\Omega) & & P_{\hat{\phi}}(\Omega) \end{array}$$

Ainsi, tout indice de comparaison des deux partitions, $P_\phi(\Omega)$ et $P_{\hat{\phi}}(\Omega)$, permet d'évaluer les performances de l'estimation $\hat{\phi}(S)$. Selon les ensembles de décisions, les partitions à comparer $P_\phi(\Omega)$, $P_{\hat{\phi}}(\Omega)$ s'écrivent de la façon suivante :

$$1. \text{ Etude classique : } \Delta = \{d_1, \dots, d_i, \dots, d_q\} \Leftrightarrow M_D$$

$$P_\phi(\Omega) = \{D_1, \dots, D_i, \dots, D_q\} \quad \text{avec } D_i = \{w \in \Omega, \phi(w) = D_i\}$$

$$P_{\hat{\phi}}(\Omega) = \{d_1, \dots, d_i, \dots, d_q\} \quad \text{avec } d_i = \{w \in \Omega, \hat{\phi}(m_S(w)) = d_i\}$$

$$2. \text{ Etude locale : } \Delta = \Delta_i = \{d_i, \bar{d}_i\} \quad i \in \{1, \dots, q\}$$

$$P_{\phi_i}(\Omega) = \{D_i, \bar{D}_i\} \quad \text{avec } \bar{D}_i = \bigcup_{j \neq i} D_j$$

$$P_{\hat{\phi}_i}(\Omega) = \{d_i, \bar{d}_i\}$$

$$3. \text{ Etude globale : } \Delta = \{d_1, \dots, d_q, d_{q+1}, \dots, d_{2q-1}\} \Leftrightarrow \mathcal{P}(M_D)$$

$$P_\phi(\Omega) = \{D_1, \dots, D_i, \dots, D_q\} \quad P_{\hat{\phi}}(\Omega) = \{d_1, \dots, d_q, d_{q+1}, \dots, d_{2q-1}\}$$

Si à l'ensemble des décisions est adjointe la décision supplémentaire "reconnaissance non performante", les partitions $P_{\hat{\phi}}(\Omega)$ possèdent une classe supplémentaire (notée x). Cette classe regroupe tous les individus de Ω pour lesquels on a différé la prise de décision.

$$\Delta^* = \Delta \cup \{x\} \quad \Rightarrow \quad P_{\hat{\phi}^*}(\Omega) = P_{\hat{\phi}}(\Omega) \cup \{x\}$$

Au niveau des performances, nous distinguerons :

1. Qualité de la discrimination

Elle exprime la plus ou moins grande similitude existante entre les deux partitions $P_\phi(\Omega)$ et $P_{\hat{\phi}}(\Omega)$ (ou $P_{\hat{\phi}^*}(\Omega)$).

2. Efficacité de la discrimination

Elle mesure l'incertitude résiduelle connaissant l'estimation donnée par $\hat{\phi}(S)$.

4.1.1 QUALITE DE LA DISCRIMINATION

a) POURCENTAGE DE MAL CLASSES

Une façon de mesurer la qualité de la discrimination est de s'intéresser aux erreurs commises par la fonction d'estimation $\hat{\phi}(S)$, soit le %MAL CLASSES. Selon les ensembles de décisions, cet indice est défini par les différentes expressions suivantes :

$$1 : \Delta = M_D$$

Pour l'étude classique, cet indice correspond au nombre d'individus qui ont fait l'objet d'une décision fautive ($\phi(w) \neq \phi(m_S(w))$) rapporté à l'effectif total de la population d'apprentissage, soit le cardinal de Ω . Ce qui nous donne :

$$\%MALCLASSES = \frac{|\{w \in \Omega / \phi(w) \neq \hat{\phi}(m_S(w))\}|}{|\Omega|}$$

$$2 : \Delta^* = \Delta \cup \{x\}$$

Dans ce cas, cet indice est défini par l'expression suivante :

$$\%MALCLASSES = \frac{|\{w \in \Omega / \phi(w) \neq \hat{\phi}(m_S(w)) \text{ et } \hat{\phi}(m_S(w)) \neq x\}|}{|\{w \in \Omega / \hat{\phi}(m_S(w)) \neq x\}|}$$

Il s'agit du nombre d'individus ayant fait l'objet d'une décision fautive mais qui permet la reconnaissance d'une des formes de \mathcal{D} ($\hat{\phi}(m_S(w)) \neq x$), rapporté au nombre total d'individus pour lesquels on a reconnu une des formes de \mathcal{D} .

Notons toutefois que cet indice apporte peu d'information lorsque la procédure $\hat{\phi}$ minimise la probabilité d'erreur et ne retient que les décisions les plus fiables /HABB,79/. Dans ces conditions, la valeur du %MALCLASSES s'exprime directement à partir des seuils intervenant dans la procédure de décision, qui constituent une donnée initiale dans la formulation du problème.

$$3 : \Delta = \mathcal{P}(M_D)$$

Pour l'étude globale, par analogie avec la définition donnée en 2, une formulation possible du %MALCLASSES est donnée par :

$$\%MALCLASSES = \frac{|\{w \in \Omega / \phi(w) \neq \hat{\phi}(m_S(w)) \text{ et } \hat{\phi}(m_S(w)) = d_i \text{ avec } i \leq q\}|}{|\{w \in \Omega / \hat{\phi}(m_S(w)) = d_i \text{ avec } i \leq q\}|}$$

Il s'agit du nombre d'individus ayant fait l'objet d'une décision "simple" fautive, rapporté au nombre d'individus ayant fait l'objet d'une décision simple. Rappelons que les décisions simples (d_i , $i=1, q$) permettent la reconnaissance d'une des formes de \mathcal{D} puisqu'elles sont associées aux parties de \mathcal{D} à un seul élément.

Dans cette approche, on considère que la réunion des différentes populations ayant fait l'objet d'une décision composée (décisions associées aux parties de \mathcal{D} à plus d'un élément) est équivalente à la population pour laquelle on n'a pas pu reconnaître de façon fiable leur classe d'appartenance lorsque l'ensemble des décisions était $\Delta^* = \Delta \cup \{x\}$.

Cependant, on peut faire les remarques suivantes au niveau des décisions composées :

* $\phi(w) \in \hat{\phi}(m_S(w))$: l'individu w ne fait pas l'objet d'une erreur de classement, néanmoins celui-ci n'est pas d'une efficacité maximale.

* $\phi(w) \notin \hat{\phi}(m_S(w))$: on peut affirmer que l'individu w est mal classé.

Dans ce sens, une mesure plus "fine" du %MALCLASSES est donnée par :

$$\%MALCLASSES = \frac{|\{w \in \Omega / \phi(w) \notin \hat{\phi}(m_S(w))\}|}{|\Omega|}$$

b) RISQUE

Ce mode d'évaluation suppose la donnée d'une fonction de perte C ; C est une application de $\Delta \times \mathcal{D}$ dans \mathbb{R}^+ qui associe à tout couple (d_i, D_j) un coût C_{ij} .

$$C : (\Delta, \mathcal{D}) \longrightarrow \mathbb{R}^+$$

$$(d_i, D_j) \longmapsto C(d_i, D_j) = C_{ij}$$

Selon la situation (d_i, D_j) , ce coût traduit une pénalité ou une bonification à donner au système de reconnaissance. Il est clair que cette notion de coût peut s'étendre à tout type de décision ($\phi(m_S(w)) =$ décision multiple, non décision (x)).

L'indice, appelé communément risque, est construit à partir de ces coûts C_{ij} ; il a pour expression :

$$R = \sum_{D_j \in \mathcal{D}} \sum_{d_i \in \Delta} \frac{C_{ij} |\{w \in \Omega, \phi(m_S(w)) = d_i \text{ et } \phi(w) = D_j\}|}{|\Omega|}$$

On peut alors remarquer, que dans le cas d'une étude globale $\Delta = \mathcal{P}(M_D)$, ou bien encore $\Delta^* = \Delta \cup \{x\}$, le risque est à la fois une mesure de qualité et d'efficacité. Soulignons aussi que son expression est équivalente à celle du %MALCLASSES pour les valeurs particulières des coûts C_{ij} définis ci après :

- i) pour $\Delta = M_D$ $C_{ii} = 0$ et $C_{ij} = 1$ pour $i \neq j$
- ii) pour $\Delta = \mathcal{P}(M_D)$ $C_{ij} = 0$ pour $D_j \in d_i$
 $C_{ij} = 1$ pour $D_j \notin d_i$

L'application de ces deux approches au diagnostic médical sera envisagée de façon plus détaillée au chapitre suivant.

c) COMPARAISON DES PARTITIONS $P_\phi(\Omega)$ ET $P_{\hat{\phi}}(\Omega)$

La similitude entre les fonctions d'affectation ϕ et $\hat{\phi}(S)$ peut être mesurée au moyen d'indices s'appliquant aux partitions induites par ces fonctions. L'utilisation de distances ou indices de distance répond parfaitement à ce problème ; les propriétés à vérifier étant :

$$\begin{array}{ccc} \mathbb{P}(\Omega) \times \mathbb{P}(\Omega) & \longrightarrow & \mathbb{R}^+ \\ (P_\phi(\Omega), P_{\hat{\phi}}(\Omega)) & \longmapsto & d(P_\phi(\Omega), P_{\hat{\phi}}(\Omega)) \end{array}$$

tel que :

$$\begin{aligned} d(P_\phi(\Omega), P_{\hat{\phi}}(\Omega)) &= d(P_{\hat{\phi}}(\Omega), P_\phi(\Omega)) \\ d(P_\phi(\Omega), P_{\hat{\phi}}(\Omega)) &= 0 \quad \Leftrightarrow \quad P_\phi(\Omega) = P_{\hat{\phi}}(\Omega) \end{aligned}$$

Différentes mesures de distance entre partitions ont été proposées dans la littérature /LOPE,73//SBAI,83/. Cependant, il faut noter que l'égalité des partitions n'implique pas forcément l'égalité des fonctions d'affectation. En effet, on obtient la même partition $P_\phi(\Omega)$ pour tout recodage de la fonction $\hat{\phi}$. Il est facile de montrer que l'égalité des partitions implique forcément l'égalité des contenus sémantiques si la procédure d'estimation minimise la probabilité d'erreur, et ceci quelle que soit l'étude menée (étude classique $\Delta = M_D$, étude locale $\Delta_i = \{d_i, \bar{d}_i\}$, étude globale $\Delta = \mathcal{P}(M_D)$) (c.f Annexe III).

Remarquons enfin que cet indice permet d'évaluer à la fois l'efficacité et la qualité de la discrimination.

4.1.2 EFFICACITE DE LA DISCRIMINATION

Compte tenu de la définition donnée en III.4.1 (page 59), il est clair que ce critère ne mesure pas l'efficacité thérapeutique d'une décision, mais seulement l'incertitude résiduelle de la variable diagnostic attachée à cette décision. Dans ce sens, cette mesure n'intervient que dans le cas où l'ensemble des décisions est soit $\Delta^* = \Delta \cup \{x\}$ avec Δ celui de l'étude locale ou classique, ou bien encore $\Delta = \mathcal{P}(M_D)$.

a) POURCENTAGE DE NON CLASSES

L'efficacité de la discrimination est d'autant plus grande que l'effectif de la population sur laquelle aucune des formes de \mathcal{D} n'a pu être reconnue de façon précise est faible.

Pour $\Delta^* = \Delta \cup \{x\}$ (: étude classique ou locale), cette population correspond à l'élément "x" de la partition $P_{\hat{\phi}^*}(\Omega)$. Pour $\Delta = \mathcal{P}(M_D)$, soit l'étude globale, elle est la réunion des différentes populations associées aux éléments de $P_{\hat{\phi}}(\Omega)$ qui ont fait l'objet d'une décision composée.

Le rapport de l'effectif de ces populations à celui de Ω constitue une mesure d'efficacité. L'indice défini est le %NONCLASSES.

Cependant, on peut dire que l'incertitude laissée par une décision composée est d'autant plus faible que cette décision est associée à une partie de \mathcal{D} comprenant peu d'éléments de \mathcal{D} . Dans ce sens, il apparaît nécessaire d'établir une mesure d'efficacité faisant intervenir ces différents degrés d'incertitude. Elle fait l'objet de ce qui suit.

b) ENTROPIE RESIDUELLE MOYENNE

Cet indice permet la prise en compte des distributions de probabilités relatives aux estimations, lorsque celles-ci ne correspondent pas à une forme unique.

On note par H_w , l'entropie résiduelle relative à l'individu w :

$$H_w = H (\phi(w) / \hat{\phi}(m_G(w)))$$

L'entropie résiduelle moyenne sera :

$$H = \frac{1}{|\Omega|} \sum_{w \in \Omega} H_w$$

Pour chacun des ensembles de décisions considérés, il est possible de majorer cette quantité :

$$1 : \Delta = \Delta_i = \{d_i, \bar{d}_i, x\} \quad i = 1, 2, \dots, q$$

$$w \text{ tel que } \hat{\phi}(m_S(w)) = d_i : H_w = 0$$

$$" \quad " \quad \hat{\phi}(m_S(w)) = \bar{d}_i : H_w = 0$$

$$" \quad " \quad \hat{\phi}(m_S(w)) = x : 0 < H_w \leq \log_2 2 = 1$$

$$\text{alors } H = \frac{1}{|\Omega|} \sum_{w \in X} H_w \leq \frac{1}{|\Omega|} \sum_{w \in X} 1 = \%NONCLASSES$$

$$2 : \Delta = \{d_1, \dots, d_i, \dots, d_q, x\}$$

$$w \text{ tel que } \hat{\phi}(m_S(w)) = d_i : H_w = 0 \quad i = 1, \dots, q$$

$$" \quad " \quad \hat{\phi}(m_S(w)) = x : 0 < H_w \leq \log_2 q$$

$$\text{alors } H = \frac{1}{|\Omega|} \sum_{w \in X} H_w \leq \frac{1}{|\Omega|} \sum_{w \in X} \log_2 q = \%NONCLASSES \cdot \log_2 q$$

$$3 : \Delta = \{d_1, \dots, d_i, \dots, d_q, d_{q+1}, \dots, d_{2^q-1}\}$$

$$w \text{ tel que } \hat{\phi}(m_S(w)) = d_i : H_w = 0 \quad i = 1, \dots, q$$

$$" \quad " \quad \hat{\phi}(m_S(w)) = d_j = D_i U D_j : 0 < H_w \leq \log_2 2$$

$$" \quad " \quad \hat{\phi}(m_S(w)) = d_k = D_i U D_j U D_k : 0 < H_w \leq \log_2 3$$

⋮
⋮
⋮

$$" \quad " \quad \hat{\phi}(m_S(w)) = d_{2^q-1} : 0 < H_w \leq \log_2 q$$

Il est facile de montrer que dans ce cas, l'entropie résiduelle moyenne est majorée de la même façon que précédemment :

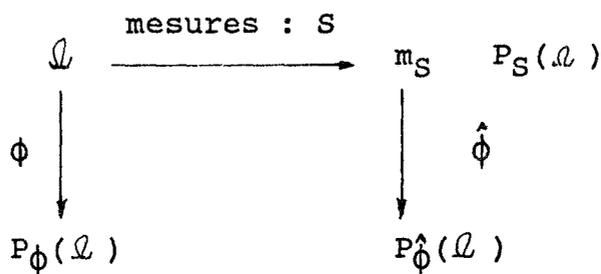
$$0 < H \leq \%NONCLASSES \cdot \log_2 q$$

Jusqu'à présent, nous avons exposé les différents indices qui permettent d'évaluer les performances de l'estimation $\hat{\phi}(S)$. Néanmoins, on peut dire de façon intuitive que plus les classes sont séparées dans l'ensemble de représentation, meilleures sont les chances de succès du système de reconnaissance. Dans ce sens, différents indices, appelés mesures de séparabilité de classes, ont été proposés dans la littérature, permettant ainsi d'évaluer le pouvoir de discrimination de S.

4.2 EVALUATION DES PERFORMANCES DE S

La traduction mathématique d'une mesure de séparabilité de classes peut être une "distance" entre individus ou entre classes. Etant donnée la structure algébrique des données du dossier médical (qualitatives et quantitatives), nous n'exposerons pas en détail les distances telles que : les distances entre centres de gravité, ou entre individus frontières... Quoique simples à mettre en oeuvre, ces distances sont généralement peu employées car elles traduisent mal le recouvrement des classes. Nous nous sommes donc intéressés aux mesures probabilistes.

L'évaluation des performances de S consiste à comparer les partitions $P_{\phi}(\Omega)$ et $P_S(\Omega)$.



Notons que si ce mode d'évaluation est retenu pour effectuer le choix de S, l'observateur ne prend pas en compte la perte d'informations occasionnée par la transformation de S par le filtre $\hat{\phi}$; ces indices apparaissent alors comme des mesures d'efficacité.

4.2.1 MESURES PROBABILISTES

Par l'utilisation de ces indices, toute l'information sur la structure en probabilités des individus et des classes est exprimée. Cette information provient des densités de probabilités conditionnelles et des probabilités à priori des classes.

Ces indices sont généralement des bornes de la probabilité d'erreur (qui est une mesure de qualité). Une revue détaillée

est donnée par /CHEN,76/ /KANA,70/. Pour de plus récents travaux, on pourra se reporter à /COPA,85/.

Parmi les plus connues, citons :

- la borne de CHERNOFF

$$J_C(D_i, D_j) = E_S (\Pr(D_i/S=m_S)^{1-\alpha} \cdot \Pr(D_j/S=m_S)^\alpha)$$

où $\Pr(D_i/S=m_S)$ désigne la probabilité de la forme D_i sachant que S prend la modalité m_S ,

et E est l'opérateur espérance mathématique.

Pour la valeur particulière de $\alpha = 1/2$, on obtient le coefficient de BHATTACHARYYA :

$$J_B(D_i, D_j) = E_S ((\Pr(D_i/S=m_S) \cdot \Pr(D_j/S=m_S))^{1/2})$$

- la divergence de KULLBACK

$$J_K(D_i, D_j) = E_S \left[\left(\frac{\Pr(D_i/S=m_S)}{\Pr(D_i)} - \frac{\Pr(D_j/S=m_S)}{\Pr(D_j)} \right) \log \frac{\Pr(D_j) \cdot \Pr(D_i/S=m_S)}{\Pr(D_i) \cdot \Pr(D_j/S=m_S)} \right]$$

- la divergence de χ^α

$$J_{\chi^\alpha}(D_i, D_j) = E \left[\left| 1 - \frac{\Pr(D_i/S=m_S)}{\Pr(D_j/S=m_S)} \right|^\alpha \Pr(D_j/S=m_S) \right] \quad \alpha > 1$$

Ces différents indices ne s'appliquent que dans le cas de deux classes ou pour toute étude locale. Néanmoins, une indication raisonnable de la séparabilité de q classes est obtenue en évaluant la distance moyenne entre les classes prises deux à deux /FU,68/. On a alors :

$$J = \sum_{i=1}^q \sum_{j=1}^q \Pr(D_i) \cdot \Pr(D_j) J_I(D_i, D_j)$$

Cependant, comme il a été remarqué dans /CHEN,76/, la majoration plus ou moins grande de la probabilité d'erreur n'est pas forcément un bon critère pour l'observateur, pour déterminer l'indice qui s'adapte le mieux à son problème. Par ailleurs, certaines mesures ne permettent pas de mettre en évidence la notion d'indépendance statistique entre les variables, et seule une indication raisonnable de la séparabilité de q classes peut être établie.

Ces différentes remarques nous ont conduits à étudier les mesures informationnelles.

4.2.2 MESURES INFORMATIONNELLES

La Théorie de l'Information (TI) est née en 1948 avec les travaux de SHANNON et de WIENER /SHAN,48/. Cette théorie eut un développement rapide et fécond. Différentes mesures d'information sont maintenant proposées (entropie de Shannon, entropie de Reyni, entropies paramétriques /PONO,81/ /DEVI,74/ /DARO,70/, entropies pondérées /GUIA,77/ /EMPT,76/).

Nous nous sommes plus particulièrement intéressés à la théorie de l'information de Shannon car celle-ci propose des indices dont l'interprétation est très naturelle (dans la plupart des cas, il est possible de leur donner une représentation graphique /MILL,63/). Ceci est du essentiellement à sa construction axiomatique, et plus précisément de l'axiome de sous-additivité. De plus, ces indices sont très sensibles à n'importe quel type de relations entre variables (linéaire, non linéaire,...), et permettent ainsi de mesurer l'indépendance statistique des variables. Enfin, la mesure d'information mutuelle entre deux variables n'impose pas de contrainte particulière sur le nombre de modalités prises par ces variables.

Si l'on prétend expliquer une variable D (le diagnostic) à partir d'autres, issues d'un ensemble \mathcal{S} (résultats d'examens) : $S \subset \mathcal{S}$, on a intérêt à choisir un S qui possède le maximum d'information commune avec D, c'est à dire que $I(D:S)$ soit aussi grand que possible (cf ANNEXE IV).

Cependant une affirmation telle que $I(D:S) = 10$ est ambiguë pour l'observateur du fait que rien n'est dit sur la possibilité d'améliorer le choix de S. C'est pourquoi, il est plus intéressant d'exprimer les performances de S par rapport à celles que l'on pourrait obtenir avec \mathcal{S} . On définit alors, les indices suivants /TORO,82/, /STAR,84/.

* POURCENTAGE ENCORE NON EXPLIQUÉ PAR S DE LA PARTIE EXPLICABLE : %ENEPE(D/S)

$$\%ENEPE(D/S) = \begin{cases} H(D/S) - H(D/\mathcal{S}) / I(D:\mathcal{S}) & \text{si } I(D:\mathcal{S}) > 0 \\ 0 & \text{si } I(D:\mathcal{S}) = 0 \end{cases}$$

Interprêtons intuitivement cet indice :

$$\%ENEPE(D/S) = \frac{\left[\begin{array}{l} \text{Partie de D qui reste} \\ \text{à expliquer sachant S} \end{array} \right] - \left[\begin{array}{l} \text{Partie non explicable} \\ \text{de D} \end{array} \right]}{\left[\text{Partie explicable de D} \right]}$$

$$\%ENEPE(D/S) = \frac{\left[\text{Partie de D qu'il serait possible d'expliquer} \right]}{\left[\text{Partie explicable de D} \right]}$$

$$= \left[\begin{array}{l} \text{Pourcentage encore non expliqué par S de la} \\ \text{partie explicable de D} \end{array} \right]$$

Par ailleurs, si $I(D:\mathcal{S}) = 0$, la partie explicable de D est nulle, dont trivialement, il ne resterait rien à expliquer de cette partie explicable.

Cette mesure rejoint la notion d'inertie expliquée que l'on rencontre pour évaluer les performances de chaque axe factoriel dans une analyse discriminante (les variables explicatives étant obligatoirement quantitatives) /FOUC,82/.

Rappelons que la partie explicable, $I(D:\mathcal{S})$ permet de donner une idée de la faisabilité du problème : expliquer une variable (D) à partir d'autres (\mathcal{S}).

Afin de donner une interprétation simple de ce coefficient, il convient de le normaliser par rapport à la partie à expliquer, soit $H(D)$; l'indice obtenu mesure alors le pourcentage de la partie explicable de D, soit $\%PE(D/\mathcal{S})$.

$$\%PE(D/\mathcal{S}) = I(D:\mathcal{S}) / H(D)$$

Ainsi, si $\%PE(D/\mathcal{S})$ est proche de 1, on peut dire que la variable D peut être décrite à partir de \mathcal{S} (relations fortes entre D et \mathcal{S}). Par contre, une valeur de cet indice proche de 0 indique tout simplement que l'on a peu de chance d'établir un modèle de décision.

Par ailleurs, il faut noter que, si une approche classique consiste à déterminer un sous ensemble S constitué de variables indépendantes (analyse discriminante), une approche informationnelle ne conduit pas forcément à ce résultat.

En effet, supposons que l'on recherche de façon optimale (au sens de la TI) un sous ensemble S de cardinal égal à 2, inclus dans \mathcal{S} . Le problème est alors le suivant :

trouver $S = \{X_i^*, X_j^*\} \in \mathcal{G}(\mathcal{L})$ tel que :

$$I(D:S) = \max_{S' \in \mathcal{G}(\mathcal{L})} I(D:S') \quad \forall S' = \{X_i, X_j\} \in \mathcal{G}(\mathcal{L})$$

L'application du principe d'optimalité de Bellman donnerait l'algorithme suivant :

$$I(D:S) = \max_{X_i \in \mathcal{L} \setminus X_j^*} [I(D:X_i) + I_{X_i}(D:X_j^*)]$$

où $I_{X_i}(D:X_j^*)$ représente la quantité d'information commune aux deux variables D, X_j^* connaissant X_i /RICH,75/.

Si X_i, X_j^* sont indépendantes, on a : $I_{X_i}(D:X_j^*) = I(D:X_j^*)$.
Néanmoins rien ne nous permet d'affirmer que :

$$\nexists X_i' \in \mathcal{L} \setminus X_j^* \quad I(D:X_i') + I_{X_i'}(D:X_j^*) \geq I(D:X_i) + I(D:X_j^*)$$

Remarquons pour terminer que les indices, $I(D:S)$, $H(D/S)$, $\%ENEPE(D/S)$ conduisent au même résultat pour le choix de S .

$$\begin{aligned} \max_S I(D:S) &\Leftrightarrow \min_S H(D/S) \\ &\Leftrightarrow \min_S \%ENEPE(D/S) \end{aligned}$$

5 . CONCLUSION

Le problème de la reconnaissance des formes, vu sous l'aspect de la théorie de la décision statistique, nécessite la donnée d'une procédure de décision $\hat{\phi}$ et d'un ensemble de paramètres discriminants S . Leurs choix influencent par conséquent les performances de la solution aussi bien au niveau de sa qualité que de son efficacité. Nous verrons au chapitre suivant que ces deux critères sont incompatibles (une augmentation de la qualité se faisant au détriment de l'efficacité). Par ailleurs, soulignons qu'il n'existe aucune méthodologie générale pour la détermination de $\hat{\phi}$ et S . Leurs choix sont obtenus par l'optimisation d'un critère construit à partir des indices de performance exposés.

On développe dans le chapitre suivant des procédures de détermination de $\hat{\phi}$ et de S spécifiques au problème de l'aide au diagnostic médical. On constatera que dans une telle application, le choix d'un ensemble explicatif optimal ne peut être dicté par une approche ne faisant intervenir que les performances décisionnelles.

- C H A P I T R E I V -

1 . INTRODUCTION

On considère l'ensemble des nouveau-nés admis au Centre de Pathologie Néo-Natale d'ARRAS pour des troubles respiratoires.

Pour chaque prématuré, un dossier médical est constitué. Il évolue au cours du temps et comprend :

a) des données cliniques relatives à l'historique de l'enfant avant son arrivée au centre (âge gestationnel, données décrivant l'évolution de la grossesse, de l'accouchement, ..). On note S_0 l'ensemble de ces variables, M_{S_0} l'ensemble de leurs modalités et $S_0(w) \in M_{S_0}$ les modalités prises par un individu w .

b) des données paracliniques relevées durant son séjour au Centre (résultats d'examens biologiques, radiologiques, ..). On note \mathcal{J}_t l'ensemble des variables mesurables à l'instant t quantitatives ou qualitatives.

Par la suite, s'ajoutera au dossier le diagnostic résultat de l'analyse et de la synthèse des différentes données (variable D , de modalités D_i , $i = 1, \dots, q$).

L'objectif d'une aide au diagnostic est d'essayer de faciliter le travail du médecin en lui proposant, compte tenu des données décrivant un nouveau-né au centre, le ou les diagnostics les plus probables appartenant au champ de la pathologie associé au symptôme d'entrée.

Devant l'extrême abondance des données qui concernent chaque malade, qu'elles soient cliniques, biologiques, .., il est important sur le plan intellectuel et sur le plan économique de reconnaître celles qui suffisent à caractériser au mieux telle entité diagnostic ou pronostic. C'est pourquoi, on se pose en général, en plus du problème décisionnel, un problème de sélection de variables explicatives.

Pour cela, on dispose d'une population d'apprentissage, Ω , constituée de prématurés considérés comme des cas "purs" au sens médical, pour le diagnostic qui leur est attribué.

Cette population d'apprentissage permet de résoudre le problème de décision "statique" dans lequel toutes les variables caractérisant au mieux les individus sont supposées connues à l'instant de décision.

Cependant, dans son aspect opérationnel, une méthode d'aide au diagnostic ne peut se baser, à une date donnée, que sur la connaissance des variables mesurées avant cette date. Dans ce sens, il convient de développer à la date t , outre des méthodes d'aide à la décision à partir des données connues, des méthodes de prescription d'examens complémentaires pour confirmer, ou affiner, le diagnostic à la date $t+1$.

Un tel processus est naturellement arborescent. Nous développerons dans ce chapitre pour chaque noeud de l'arbre (terminal ou non) :

a) la formulation du problème de décision relatif à ce noeud

b) la formulation du problème de sélection de variables explicatives en ce noeud (formulation multicritère).

2 . PROCESSUS D'AIDE AU DIAGNOSTIC

2.1 PROCESSUS DE SEGMENTATION

On note \underline{w} un prématuré pour lequel une aide au diagnostic est désirée ($\underline{w} \notin \Omega$). Le processus est alors le suivant :

$t-1$ A l'instant t , on dispose de l'ensemble des données $\bigcup_{i=0}^{t-1} S_i(\underline{w})$ caractérisant l'individu \underline{w} . De la population d'apprentissage Ω , on peut extraire les individus identiques à \underline{w} relativement à ces variables. Soit $\Omega_t(\underline{w})$ ce sous ensemble, il peut être défini de façon récurrente par :

$$\Omega_t(\underline{w}) = \{ w \in \Omega_{t-1}(\underline{w}) \text{ tel que } S_{t-1}(w) = S_{t-1}(\underline{w}) \} \text{ pour } t > 1$$

$$\text{avec } \Omega_1(\underline{w}) = \{ w \in \Omega \text{ tel que } S_0(w) = S_0(\underline{w}) \}$$

Il apparait immédiatement que cette procédure de segmentation de Ω génère à chaque étape l'ensemble des "plus proches voisins" de \underline{w} dans Ω (i.e, ceux qui ont répondu pour le moment de façon identique) /TERR,76//EMPT,76/.

En considérant l'ensemble des réponses possibles, ce processus définit une structure arborescente. Sur cet arbre, différentes situations peuvent se produire définissant divers types de noeud.

2.2 LES DIFFERENTS TYPES DE NOEUDS

Dans le cadre d'une médecine d'urgence, où l'on doit émettre des propositions de diagnostic à tout instant t , on utilisera la distribution de la variable diagnostic sur $\Omega_t(\underline{w})$ afin de prescrire un traitement au malade \underline{w} .

Cependant, la quantité d'informations disponible dès les premières heures de vie d'un prématuré n'est généralement guère suffisante pour que les médecins puissent acquérir une complète certitude au niveau du diagnostic. Confrontés à cette même réalité, les situations suivantes peuvent alors se produire :

a) l'entropie de la variable diagnostic est faible sur $\Omega_t(\underline{w})$. On est sur un noeud terminal de l'arbre, et un diagnostic unique peut être établi avec une bonne fiabilité : $d_t(\underline{w}) \in M_D$.

b) l'entropie de la variable diagnostic est forte sur $\Omega_t(\underline{w})$. On est sur un noeud non terminal de l'arbre, et une

bonne fiabilité de la conclusion ne peut être obtenue qu'en établissant un diagnostic multiple $d_t(w) \in \mathcal{P}_2(M_D)$.

Cette approche suppose naturellement que le cardinal de $\Omega_t(w)$ ne soit pas trop faible. Cependant, cette condition risque de ne pas être toujours vérifiée (limite bien connue des processus d'interrogation (cf IV.5), ou encore w est un cas "rare" et admet peu de voisins dans la population d'apprentissage Ω). Il convient alors d'introduire un troisième type de noeud.

Le cardinal de $\Omega_t(w)$ est trop faible pour que la distribution de la variable diagnostic ait une signification statistique. Le système d'aide ne peut pas émettre dans de bonnes conditions une proposition diagnostique, et il précisera au médecin les raisons de cette incapacité.

On appellera noeud divisible tout noeud pour lequel le processus de segmentation peut être poursuivi et noeud indivisible, un noeud de cardinal trop faible pour que cela soit le cas.

Le traitement prescrit dépend bien évidemment de la nature du noeud et de même pour le problème de décision posé au praticien à l'étape $t+1$. Dans ce sens, un dialogue entre le système d'aide à la décision et le médecin s'avère nécessaire pour que la machine puisse prendre en compte les nouveaux objectifs à satisfaire pour la suite des opérations.

2.3 ASPECT CONVERSATIONNEL

2.3.1 LES DIFFERENTS COMPORTEMENTS DU MEDECIN

Selon la proposition diagnostique émise en t par le système d'aide, le médecin peut adopter différents comportements qui définiront le problème à résoudre en $t+1$.

a) dans le cas où $\Omega_t(w)$ est un noeud terminal correspondant au diagnostic unique $d_t(w) \in M_D$, il s'agit pour le praticien de garantir en $t+1$ ce résultat. Le problème posé est alors un problème de confirmation, et il s'applique sur un certain sous ensemble de Ω , $\Omega_t(w)$ incluant $\Omega_t(w)$. Il est évident que le sous ensemble de confirmation ne peut se réduire à $\Omega_t(w)$ puisque ce dernier ne contient (à peu d'exceptions près) que des individus qui présentent la même modalité diagnostic, il faut donc choisir $\Omega_t(w)$ parmi les noeuds pères de $\Omega_t(w)$.

b) dans le cas où $\Omega_t(w)$ est un noeud non terminal, le diagnostic formulé en t est multiple ($d_t(w) \in \mathcal{P}_2(M_D)$) et il s'agit pour le praticien d'affiner cette proposition

diagnostique en $t+1$: $d_{t+1}(\underline{w}) \subset d_t(\underline{w})$. Le problème d'affinement s'applique en $\Omega_t(\underline{w})$ (si celui-ci est divisible) ou sur un sous ensemble l'incluant s'il est indivisible.

c) on peut remarquer que, lorsque $\Omega_t(\underline{w})$ n'est pas un noeud terminal, le praticien peut désirer combiner les approches a) et b) : garantir le résultat déjà acquis tout en essayant de l'améliorer (compromis entre l'affinement et la confirmation).

Pour passer au noeud suivant, soit $\Omega_{t+1}(\underline{w})$, le système d'aide au diagnostic doit disposer en $t+1$ d'informations supplémentaires évaluées en t sur l'individu \underline{w} ; dans ce sens, il proposera au praticien les examens complémentaires qu'il juge optimaux compte tenu du problème à résoudre (confirmation, affinement, ou les deux).

Dans ces conditions, le passage au noeud suivant revient à résoudre un problème de reconnaissance des formes statique qui, comme nous l'avons vu en III.2.1, consiste à :

i) déterminer un ensemble de variables discriminantes (les examens complémentaires à prescrire).

ii) appliquer une procédure de décision qui permettra de résoudre le problème de décision relatif au noeud $\Omega_{t+1}(\underline{w})$.

La formulation de ces deux problèmes sera vue, au paragraphe 3 pour le problème de décision, et au paragraphe 4 pour le problème de sélection.

Il est clair que les différents comportements présentés sous-entendent une suite "normale" des événements. Cependant, cette suite peut être interrompue par des situations d'échec qu'il paraît sage d'étudier pour éventuellement y apporter des solutions.

2.3.2 ANALYSE DES DIFFERENTES SITUATIONS

1) les objectifs escomptés sont atteints

Pour le problème de confirmation, on a alors $d_{t+1}(\underline{w}) = d_t(\underline{w})$ et le processus d'aide est terminé ; éventuellement, d'autres confirmations peuvent être sollicitées de la part du praticien.

Pour les deux autres démarches ($d_{t+1}(\underline{w}) \subset d_t(\underline{w})$), le processus continue après avoir reçu les nouvelles consignes.

2) les objectifs escomptés ne sont pas atteints

Trois types de situations peuvent se produire :

a) la proposition diagnostique $d_{t+1}(\underline{w})$ n'est pas cohérente avec celle émise en t ; on a alors $d_{t+1}(\underline{w}) \neq d_t(\underline{w})$. Comme cette incohérence n'apparaît qu'à l'étape $t+1$, on peut penser que cette situation résulte d'une saisie incorrecte des paramètres $S_t(\underline{w})$, ou bien encore que les mesures de S_t sur \underline{w} sont entachées d'erreur. Dans ce sens, la machine propose dans un premier temps au médecin de vérifier l'acquisition des dernières mesures. Néanmoins, si cette incohérence est réelle (les mesures de S_t étaient bonnes), le processus d'aide s'avère inefficace et seul le praticien est à même de résoudre ce problème.

b) la proposition diagnostique $d_{t+1}(\underline{w})$ n'affine pas le résultat acquis en t : $d_{t+1}(\underline{w}) = d_t(\underline{w})$ avec $d_t(\underline{w}) \in \mathcal{D}_2(M_D)$. On peut alors imaginer que la machine ne disposait pas d'assez d'informations supplémentaires évaluées en t pour affiner $d_t(\underline{w})$. Dans ces conditions, le praticien est conduit à proposer un univers de choix plus important afin qu'une nouvelle sélection puisse être établie améliorant ainsi la situation antérieure.

c) le cardinal de $\Omega_{t+1}(\underline{w})$ est trop faible pour qu'une proposition de diagnostic puisse être émise dans de bonnes conditions. En effet, compte tenu des questions posées jusqu'ici, l'individu \underline{w} a très peu de voisins dans Ω , et il est possible qu'il constitue un cas particulier nécessitant un recours au praticien. Cependant, cette situation peut aussi résulter "d'accidents" dans la segmentation ayant conduit à isoler cet individu à tort. Il convient alors pour continuer la procédure d'inclure l'individu \underline{w} dans une population de cardinal plus important. Celle-ci peut être obtenue en créant un nouveau noeud par fusion (cf IV.5) ou, en retournant en un noeud antécédent du noeud indivisible considéré. Chaque retour en arrière faisant perdre les informations acquises grâce aux examens relatifs au noeud supprimé, on sera amené à limiter ces retours en arrière.

On dispose alors d'un certain nombre de populations candidates à être le noeud de redémarrage pour la procédure. Pour chacune de ces populations, le système est capable de proposer les examens complémentaires qui lui semblent les meilleurs pour assurer ce redémarrage. Le choix final appartient alors au praticien sachant que, si aucune des propositions ne lui convient, l'individu \underline{w} constitue selon toute vraisemblance un cas particulier qu'il devra traiter comme tel. Notons toutefois qu'il lui est toujours possible d'obtenir d'autres propositions en autorisant des retours en arrière supplémentaires.

Une autre façon de créer un noeud de redémarrage est obtenue en oubliant les informations les plus anciennes. Elle est exploitée de manière identique à ce qui a été vu ci-dessus.

3 . PROBLEME DE DECISION

L'approche proposée nous a permis de décomposer le problème de décision dynamique qu'est une aide au diagnostic dans un contexte de médecine d'intervention, en une suite de problèmes de décision statique ; à l'étape $t+1$, la structure décisionnelle du problème présente les caractéristiques suivantes:

\mathcal{D} : ensemble des formes. C'est l'ensemble des modalités prises par la variable diagnostic pour tout individu appartenant au champ de la pathologie associée au symptôme d'entrée (les troubles respiratoires). Par conséquent, \mathcal{D} est invariant dans le temps.

M_S : c'est l'ensemble des modalités prises par les examens complémentaires S_t sur la population $\Omega_t(\underline{w})$.

Δ : ensemble des décisions admissibles. Il dépend du problème posé en $t+1$ par le praticien (cf IV.3.1). Soulignons qu'afin de prendre en compte les informations partielles (i.e les informations qui ne permettent pas la reconnaissance d'un diagnostic de façon unique), nous avons choisi d'initialiser cet ensemble à $\Delta_0 = \mathcal{P}(M_D)$.

En s'inspirant des schémas classiques de la théorie de la décision statistique /SPAT,73/, nous développerons deux approches pour la construction de la procédure de décision.

i) une approche probabiliste. La procédure minimise la probabilité d'erreur.

ii) une approche bayésienne. Le critère de décision s'exprime sous la forme d'un risque à minimiser.

Cette procédure, appliquée aux modalités prises par S_t sur l'individu \underline{w} nous permettra d'émettre la proposition diagnostique relative au noeud $\Omega_{t+1}(\underline{w})$; proposition diagnostique choisie de façon optimale parmi un certain sous ensemble de $\mathcal{P}(M_D)$ constituant l'ensemble des décisions admissibles.

3.1 ENSEMBLE DES DECISIONS ADMISSIBLES

L'ensemble des décisions admissibles au temps $t+1$ dépend bien évidemment de la démarche adoptée par le praticien selon la décision émise par le système d'aide à l'instant t , soit $d_t(\underline{w})$.

a) Dans le cas où $d_t(\underline{w})$ correspond à un diagnostic unique ($d_t(\underline{w}) \in M_D$), le problème posé par le praticien en $t+1$

est un problème de confirmation, et l'ensemble des décisions admissibles se réduit alors à $\Delta_{t+1} = \{d_t(\underline{w}), \bar{d}_t(\underline{w})\}$.

b) Dans le cas où $d_t(\underline{w})$ correspond à un diagnostic multiple ($d_t(\underline{w}) \in \mathcal{P}_2(M_D)$), le problème en $t+1$ est :

i) soit un problème d'affinement, et l'ensemble des décisions admissibles est donné par $\Delta_{t+1} = \mathcal{P}(d_t(\underline{w}))$.

ii) soit un problème d'affinement/confirmation, et $\Delta_{t+1} = \mathcal{P}(d_t(\underline{w})) \cup \{\bar{d}_t(\underline{w})\}$.

La structure du problème étant définie, nous allons maintenant nous intéresser à la construction de la procédure de décision.

3.2 MINIMISATION DE LA PROBABILITE D'ERREUR

Une approche basée sur l'observateur idéal de KOTELNIKOV-SIEGERT /ASH,65/ minimise la probabilité d'erreur. Elle constitue un cas particulier de la théorie de BAYES /SPAT,73/.

La procédure correspondante s'exprime de la façon suivante:

$$\hat{\phi} : M_S \longrightarrow \Delta$$

$$m_S \longmapsto \hat{\phi}(m_S) = d_i \quad \text{tel que : } \Pr(d_i) = \max_{d_j \in \Delta} \Pr(d_j)$$

avec $\Pr(d_j) = \Pr(D_j/S_t=m_S)$

où $\Pr(D_k/S_t=m_S)$ désigne la probabilité à postériori de la forme D_k quand S_t prend la modalité m_S sur la population $\Omega_t(\underline{w})$.

Cette procédure s'applique pour $\Delta = M_D$ et de la même façon, pour tout ensemble de décisions $\Delta = \{\bar{d}_i, d_i\}, i=1, \dots, q$ (étude locale). Dans ces conditions, son emploi se limite au cas où le problème posé en $t+1$ est un problème de confirmation ; les décisions admissibles s'excluant mutuellement.

Par contre, pour un problème d'affinement, minimiser la probabilité d'erreur ne peut être le seul critère. En effet, pour $\Delta = \mathcal{P}(M_D)$, la décision $(D_1 \cup D_2 \cup \dots \cup D_q)$ ne permet la reconnaissance d'un diagnostic unique, tout en assurant une probabilité d'erreur nulle (solution triviale). En conséquence, il nous faut tenir compte de la quantité d'information qu'apporte l'estimation $\hat{\phi}(m_S)$ sur la connaissance de la

variable diagnostic lorsque l'ensemble des décisions admissibles comprend des propositions multiples /AYGA,86'/.

a) PRISE EN COMPTE DES CHOIX MULTIPLES

Le spectre du traitement prescrit au malade w étant d'autant plus large que l'incertitude laissée par la décision $d_t(w)$ est grande, on doit bâtir une procédure $\hat{\phi}$ qui permet d'obtenir :

i) des propositions diagnostiques fiables : nous avons choisi comme mesure de fiabilité, la probabilité d'erreur.

ii) des propositions diagnostiques efficaces, c'est à dire qui laissent le moins possible d'incertitude sur la variable diagnostic. Nous avons pris comme mesure d'efficacité, l'entropie résiduelle de D sachant $\hat{\phi}(m_S)$, soit $H(D/\hat{\phi}(m_S))$. Notons que cette notion d'efficacité rejoint celle que nous avons déjà rencontrée pour évaluer les performances d'une discrimination (III 4.1.2).

Il s'agit de résoudre un problème bicritère. Le concept d'optimalité conduit à rechercher les solutions non dominées. Les remarques suivantes nous permettront de préciser cet ensemble.

REMARQUES :

Sur $\mathcal{P}(M_D)$, la probabilité d'erreur ainsi que l'entropie résiduelle vérifient la propriété de monotonie.

Pour tout $d_i, d_j \in \mathcal{P}(M_D)$ tel que $d_i \subseteq d_j$

$$. 1 . \Pr \text{ d'erreur}(\hat{\phi}(m_S)=d_i) \geq \Pr \text{ d'erreur}(\hat{\phi}(m_S)=d_j)$$

$$. 2 . H(D/\hat{\phi}(m_S)=d_i) \leq H(D/\hat{\phi}(m_S)=d_j)$$

PREUVES :

Proposition n 1 :

$$\begin{aligned} \Pr \text{ d'erreur}(\hat{\phi}(m_S)=d_i) &= 1 - \Pr(d_i) \\ &= 1 - \sum_{D_k \in d_i} \Pr(D_k/S_t=m_S) \end{aligned}$$

$$\begin{aligned} \Pr \text{ d'erreur} &= \Pr \text{ d'erreur}(\hat{\phi}(m_S)=d_j) - \Pr \text{ d'erreur}(\hat{\phi}(m_S)=d_i) \\ &= \sum_{D_k \in d_j \setminus d_i} \Pr(D_k/S_t=m_S) \end{aligned}$$

Comme $\Pr(D_k/S_t=m_S) \geq 0$, $\Delta \Pr$ d'erreur ≥ 0

Proposition n 2 :

$$H(D/\hat{\phi}(m_S)=d_i) = - \sum_{D_k \in d_i} p_k \log_2 p_k \quad \text{avec } p_k = \Pr(D_k/S_t=m_S)/\Pr(d_i)$$

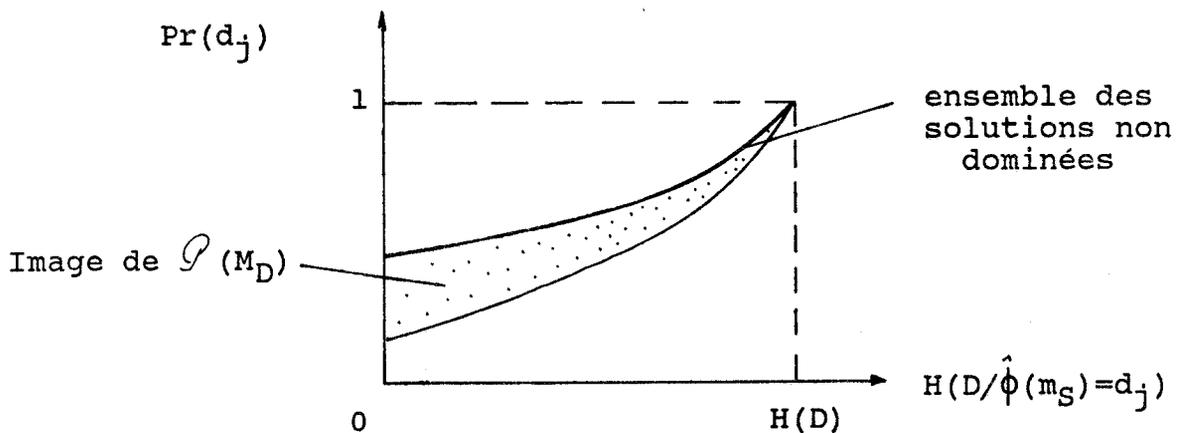
$$H(D/\hat{\phi}(m_S)=d_j) = - \sum_{D_k \in d_j} p_k' \log_2 p_k' \quad \text{avec } p_k' = \Pr(D_k/S_t=m_S)/\Pr(d_j) \\ \geq p_k$$

$$\begin{aligned} \Delta H &= H(D/\hat{\phi}(m_S)=d_j) - H(D/\hat{\phi}(m_S)=d_i) \\ &= - \sum_{D_k \in d_j} p_k' \log_2 p_k' + \sum_{D_k \in d_i} p_k \log_2 p_k \\ &= - \sum_{D_k \in d_i} (p_k' \log_2 p_k' + p_k \log_2 p_k) - \sum_{D_k \in d_j \setminus d_i} p_k' \log_2 p_k' \end{aligned}$$

$$\left\{ \begin{array}{l} (- \sum_{D_k \in d_j \setminus d_i} p_k' \log_2 p_k') \geq 0 \\ - \sum_{D_k \in d_i} (p_k' \log_2 p_k' + p_k \log_2 p_k) > \sum_{D_k \in d_i} p_k' \log_2 (p_k/p_k') \geq 0 \end{array} \right.$$

$$\Rightarrow \Delta H \geq 0$$

Les deux critères sont donc incompatibles. La figure suivante fait apparaître dans le plan (efficacité, fiabilité), l'image de l'ensemble des décisions associées à $\mathcal{P}(M_D)$, ainsi que l'ensemble des solutions non dominées :



- IMAGE DE $\mathcal{P}(M_D)$ DANS LE PLAN (Efficacité, Fiabilité) -

Dans ces conditions, une formulation du problème d'estimation optimale est donnée par :

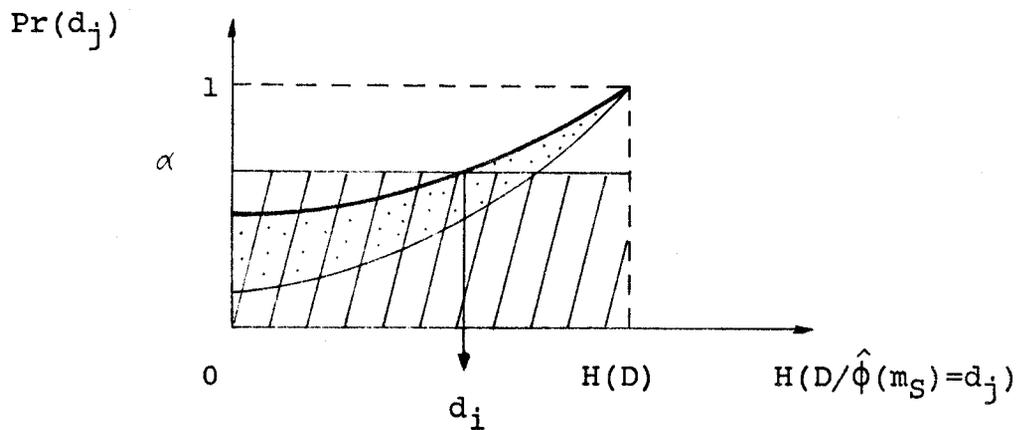
$$\hat{\phi} : M_S \longrightarrow \Delta = \mathcal{P}(M_D)$$

$$m_S \longmapsto \hat{\phi}(m_S) = d_i = \bigcup_{k \in K} D_k$$

tel que :

$$1 : H(D/\hat{\phi}(m_S)=d_i) = \min_{d_j \in \Delta} (H(D/\hat{\phi}(m_S)=d_j))$$

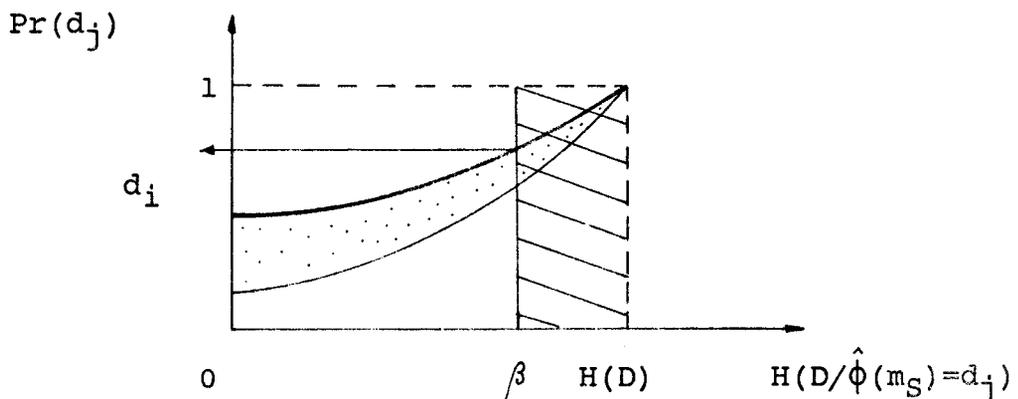
$$\Pr(d_i) \geq \alpha \quad ; \quad \alpha \in [0, 1[$$



ou bien encore :

$$2 : \Pr(d_i) = \max_{d_j \in \Delta} (\Pr(d_j))$$

$$H(D/\hat{\phi}(m_S)=d_i) \leq \beta \quad ; \quad \beta \in [0, H(D)[$$



La procédure développée ici vise à fiabiliser les propositions diagnostiques par la minimisation ou la majoration de la probabilité d'erreur. Cependant, dans le cadre d'une médecine d'urgence, où les risques de mortalité sont très importants, il convient le plus souvent de parer au plus *dangereux* qu'au plus *probable*. En effet, compte tenu de la vitesse de dégradation de l'état de certains cas pathologiques, le médecin est souvent amené à prendre une décision thérapeutique sans pour cela avoir acquis une complète certitude au niveau du diagnostic.

Ainsi, si l'on considère qu'il existe une correspondance entre l'ensemble des diagnostics et l'ensemble des traitements, choisir le diagnostic D_i conduit à prescrire le traitement T_i . Les propositions diagnostiques n'ont plus alors les mêmes efficacités thérapeutiques ; le traitement T_j peut convenir pour la maladie D_i mais le traitement T_i n'est pas forcément efficace pour combattre la maladie D_j . Dans ces conditions, le modèle de décision développé jusqu'à présent s'avère insuffisant et il apparaît que l'introduction d'un élément de nature non probabiliste soit nécessaire pour prendre en compte l'efficacité thérapeutique des propositions diagnostiques.

3.3 MINIMISATION D'UNE FONCTION RISQUE

Pour répondre à l'objection précédente, on propose pour résoudre ce problème l'utilisation d'une procédure Bayésienne.

Celle-ci suppose la donnée d'une fonction de perte C ; C est une application de $\Delta \times \mathcal{D}$ dans \mathbb{R}^+ qui associe à tout couple (d_i, D_j) un coût C_{ij} . Ce coût traduit l'efficacité thérapeutique de la proposition d_i pour combattre la maladie D_j .

$$C : (\Delta, \mathcal{D}) \longrightarrow \mathbb{R}^+$$

$$(d_i, D_j) \longmapsto C(d_i, D_j) = C_{ij}$$

Le critère de décision est construit à partir des coûts C_{ij} sous forme de risque à minimiser. La procédure de décision résultante s'énonce alors :

$$\hat{\phi} : M_S \longrightarrow \Delta$$

$$m_S \longmapsto \hat{\phi}(m_S) = d_i \text{ tel que } R(d_i, \mathcal{D}) = \min_{d_j \in \Delta} R(d_j, \mathcal{D})$$

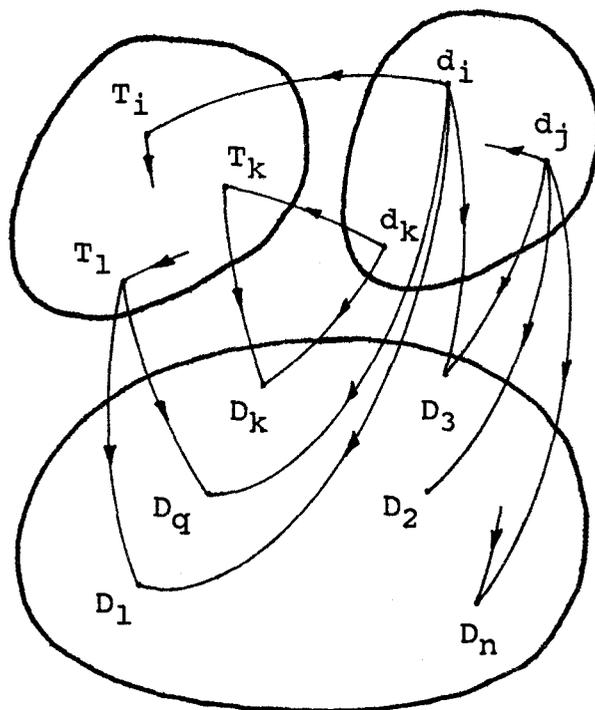
$$\text{avec } R(d_j, \mathcal{D}) = \sum_{i=1}^q C(d_i, D_j) \Pr(D_j/S=m_S)$$

Nous sommes en droit de nous interroger sur le caractère objectif ou subjectif de ces coûts C_{ij} . On peut noter que pour certaines applications, une telle approche a déjà été utilisée /HABB,81/, /WART,74/, ...

Pour une application donnée, la détermination des valeurs C_{ij} nécessite une étude spécifique. Afin d'en illustrer sa complexité, nous allons étudier les différents types d'erreurs de diagnostic qui, comme on pourra le voir, peuvent avoir ou non des conséquences sur la guérison du malade.

Pour cela, considérons les trois ensembles suivants et les différentes relations qui peuvent exister entre ces ensembles.

\mathcal{T} : Ensemble
des
traitements



$\Delta : \mathcal{P}(M_D)$

Ensemble des
propositions
diagnostiques

\mathcal{D} : Ensemble
des maladies

Pour les ensembles \mathcal{D} et Δ , nous avons vu (III.2.2) que pour tout élément de \mathcal{D} , il existait au moins une image dans Δ . Par contre pour \mathcal{D} et \mathcal{T} , ceci n'est pas forcément vrai. En effet, pour certaines maladies, il n'existe pas encore un traitement spécifique, ou un traitement conduisant à une guérison certaine du patient (exemple : cancer dépassé).

Dans ces conditions, on distinguera les quatre variétés suivantes d'erreur de diagnostic et elles seront présentées dans l'ordre des pertes croissantes qu'elles occasionnent.

a) l'erreur de diagnostic ne compromet pas la guérison du patient.

Bien que la décision thérapeutique soit basée sur un diagnostic erroné, le traitement prescrit contient dans son spectre la maladie dont le patient est atteint. Cette erreur de diagnostic n'engendre aucune conséquence grave sur la guérison du patient. Il semble alors raisonnable de penser que les pertes associées à ce type de situation ne doivent pas être plus importantes que celles associées à une situation où il y aurait une adéquation entre le traitement prescrit, la maladie à combattre et la proposition diagnostique (situation qui est bien évidemment la plus favorable).

b) l'erreur de diagnostic ne compromet pas la guérison du patient mais la rend plus tardive.

La situation est pratiquement similaire à celle exposée précédemment ; toutefois le traitement prescrit est un peu moins adapté mais il reste encore valable. Les coûts correspondant à ces situations doivent être légèrement supérieurs à ceux déjà rencontrés afin de traduire l'efficacité plus tardive du traitement choisi.

c) l'erreur de diagnostic conduit à la prescription d'un traitement inadapté pour la maladie à combattre.

Il faut distinguer deux cas de figures :

c_1) Dans le cas où il existe un traitement efficace pour lutter contre la maladie que l'on n'a pas su identifier, l'erreur de diagnostic élimine toute chance de guérison au patient. On peut alors penser que les pertes doivent être supérieures à celles du cas b).

c_2) S'il n'existe pas encore un traitement adapté à la maladie à combattre, quelle que soit la thérapie utilisée, une évolution fatale est des plus probables. L'erreur de diagnostic est "sans importance", néanmoins il nous est difficile de porter un jugement général sur les valeurs des coûts associés à ces situations vis à vis de celles déjà rencontrées.

d) L'erreur de diagnostic conduit à la prescription d'un traitement qui aggrave la maladie.

Là aussi, s'il semble raisonnable de penser que les pertes doivent être supérieures à celles associées aux situations a) et b), leur comparaison avec celles du cas c) doit aussi faire intervenir l'existence ou la non existence d'une possibilité de guérison.

d_1) aggraver la maladie alors qu'une possibilité de guérison existait doit conduire à un coût au moins supérieur au cas c_1).

d_2) aggraver une maladie pour laquelle aucun traitement n'existe peut être considéré comme moins "coûteux", de manière comparable au cas c_2).

Par ailleurs, il faut noter que le coût C_{ij} n'est pas forcément égal au coût C_{ji} ; le traitement T_i peut être efficace pour combattre la maladie D_j mais ceci n'entraîne pas obligatoirement que le spectre du traitement T_j s'étend à la maladie D_i .

REMARQUES :

1 : L'approche probabiliste développée en IV.3.2a revient à donner les pénalisations suivantes :

Δ	$d_1=D_1$	$d_i=D_i$	$d_q=D_q$	$d_{q+1}=D_1UD_2$	$d_{2q-1}=D_1U..UD_q$
D_1	0	1	1	0	0
D_2	1	1	1	0	0
D_i	1	0	1	1	0
D_q	1	1	0	1	0

On peut alors remarquer que cette approche associe le même coût à toutes les situations d'erreurs de diagnostic, ce qui est bien évidemment en contradiction avec l'analyse précédente. On en conclura donc que la procédure minimisant la probabilité d'erreur ne permet pas de prendre en compte l'efficacité thérapeutique des propositions diagnostiques, ce que l'on pouvait imaginer de façon intuitive.

2 : Ce type de modèle de décision constitue un modèle général de résolution par fonction de décision statistique. Son développement peut être illustré par l'application des concepts de la Théorie des Jeux. En effet, l'aide au diagnostic peut être considérée comme un jeu statistique, à deux joueurs et à somme nulle. L'un des joueurs est la Nature, l'autre étant le praticien. Toutefois la Nature ne peut être considérée comme un véritable "joueur" cherchant à maximiser son gain. Les procédures de choix basées sur l'approche Bayésienne considèrent alors que la Nature joue de façon aléatoire suivant un modèle constitué par les probabilités à priori $Pr(D_i)$, $i=1, \dots, q$.

3.4 VALIDITE DES DECISIONS

Les procédures de décisions exposées nécessitent l'évaluation des probabilités à postériori ainsi que celles des probabilités à priori des modalités de la variable diagnostic. Il en résulte que la validité de leurs résultats est fonction de l'estimation de ces quantités. Celles-ci pourront être déterminées de façon fiable au moyen des fréquences observées sur le noeud $\Omega_t(\underline{w})$ si ce dernier est de cardinal suffisant /CONA,72//CRON,63//SAPO,78/.

$$\hat{\Pr}(D_i/S_t=m_S) = \frac{|\{w \in \Omega_t(\underline{w}), \phi(w) = D_i \text{ et } S_t(w) = m_S\}|}{|\{w \in \Omega_t(\underline{w}), S_t(w) = m_S\}|}$$

Compte tenu que $\Omega_t(\underline{w})$ est un ensemble fini, le nombre de modalités de S_t , soit le cardinal de M_{S_t} devra être assez faible. Ainsi, S_t devra être constitué d'un petit nombre de paramètres et/ou que chacun de ses paramètres présente peu de modalités. Néanmoins, S_t devra fournir le plus d'information possible pour déterminer la forme à laquelle appartient tout individu \underline{w} .

Nous avons vu au chapitre II comment il était possible de réduire le nombre de modalités prises par les variables explicatives ; nous nous intéresserons donc au paragraphe suivant à la phase de sélection, considérant que cette opération de prétraitement des données a déjà été réalisée.

4 . SELECTION DES EXAMENS COMPLEMENTAIRES

Le processus humain de reconnaissance (plus particulièrement celui qui mène au diagnostic) est basé sur l'étude de peu de caractères très discriminants (les symptômes pantagnomoniques). La sélection automatique de variables permet d'une part de s'identifier à ce processus en éliminant les informations redondantes et inopportunes qui pourraient nuire au système de reconnaissance, et d'autre part de diminuer les calculs (donc d'en réduire les coûts) nécessaires à la procédure de décision. De plus, comme nous venons de le voir, compte tenu du fait que la population d'apprentissage Ω est un ensemble fini, diminuer l'ensemble de représentation permet de déterminer avec une bonne fiabilité les probabilités multidimensionnelles à partir des fréquences évaluées sur Ω .

Par ailleurs, dans le cadre d'une aide au diagnostic, la sélection de variables peut avoir pour conséquence de remettre en cause certains examens dont la valeur diagnostique n'est guère satisfaisante. Dans ces conditions, elle peut entraîner une évolution dans la construction du dossier médical. Certains examens risquant de disparaître, une économie au niveau des coûts physiques (pénibilité occasionnée par la mesure de tel paramètre) et financiers peut être entrevue.

Le dossier médical évoluant au cours du temps, l'ensemble des examens complémentaires mesurables à l'instant t , dans lequel une sélection est envisagée, est défini par \mathcal{S}_t .

Pour effectuer cette sélection, il existe dans la littérature deux grandes catégories de méthodes.

En premier lieu, la sélection peut être faite dans l'ensemble \mathcal{S}_t . Chaque variable de S_t est choisie parmi les p variables de \mathcal{S}_t ; on a alors : $S_t \subset \mathcal{S}_t$.

En deuxième lieu, la sélection peut s'opérer dans un ensemble \mathcal{S}'_t obtenu après transformation de l'ensemble \mathcal{S}_t . Chaque variable de S_t est alors une combinaison (linéaire ou non) des p variables de \mathcal{S}_t . Notons que les plus simples à mettre en oeuvre sont les transformations linéaires (Analyse Factorielle Discriminante /BERT,75/, /FOUC,82/).

Bien que cette deuxième approche permette de réduire la dimension de \mathcal{S}'_t , elle nécessite la connaissance (donc la mesure) de tous les paramètres de \mathcal{S}_t . Dans ce sens, ces méthodes dites "d'extraction de caractères" ne permettront pas d'entrevoir une éventuelle diminution des coûts liés à l'obtention des variables conduisant au diagnostic. De plus, elles nécessitent une structure algébrique des variables que nous ne possédons qu'en partie (les paramètres constituant le dossier médical sont aussi bien quantitatifs que qualitatifs).

Ces différentes raisons nous ont conduits à opter pour la première famille de méthode.

Néanmoins, on aurait pu envisager d'appliquer ces méthodes d'extraction de caractères une fois S_t déterminé. L'approche informationnelle développée ci-dessous démontre que cette opération ne peut que diminuer les performances obtenues avec S_t .

Rappelons tout d'abord la propriété suivante :

Pour $H =$ entropie, quels que soient D et S , et quelle que soit f une application définie sur l'ensemble des modalités de S , on a :

$$H(D/S) < H(D/f(S))$$

L'égalité n'a lieu que si f est une fonction biunivoque.

Maintenant, supposons que S désigne l'ensemble optimal des examens complémentaires, D la variable "diagnostic" à expliquer et Y , les combinaisons linéaires de S : $Y = A S$ (supposées constructibles).

Pour que $H(D/Y) = H(D/S)$, il faut nécessairement que la matrice A soit carrée inversible (f devant être une fonction biunivoque). Par conséquent, si on ne retient que les meilleures (au sens d'un certain critère) composantes de Y , il y aura forcément une perte d'information.

La détermination de S_t est bien évidemment conditionnée par la manière d'évaluer ses performances. Nous avons présenté au chapitre III les différents modes d'évaluation qui pouvaient s'appliquer à notre problème. Compte tenu des structures algébriques des variables dont nous disposons, nous avons choisi d'utiliser la théorie de l'information de Shannon pour apprécier la contribution de S_t à l'explication de la variable diagnostic (cf III.4.2.2)

4.1 MESURE DE CONTRIBUTION DE S_t A L'AFFINEMENT OU/ET LA CONFIRMATION DE D

La contribution de S_t à l'explication de la variable diagnostic D peut être quantifiée à partir de la quantité d'information commune entre ces deux variables, soit $I(D:S_t)$. Ainsi, si S_t prétend expliquer D , cette quantité doit être maximale ; en d'autres termes, il faut que l'information restant à connaître sur D soit minimale, lorsque l'on possède déjà l'information fournie par S_t .

En effet, maximiser $I(D:S_t)$ revient à minimiser $H(D/S_t)$; nous allons donc travailler en termes d'entropie conditionnelle.

A l'instant t , il convient alors de prescrire l'ensemble des examens complémentaires S_t^* tel que :

$$S_t^* : \quad H(D/S_t^*) = \min_{S_t \in \mathcal{J}(\mathcal{J}_t)} (H(D/S_t))$$

Compte tenu de la monotonie de $H(D/S_t)$ par rapport à l'inclusion dans $\mathcal{J}(\mathcal{J}_t)$ ($S_t' \subseteq S_t \Rightarrow H(D/S_t) \leq H(D/S_t')$, $\forall S_t, S_t' \in \mathcal{J}(\mathcal{J}_t)$), le problème a pour solution triviale : $S_t^* = \mathcal{J}_t$.

$$\forall S_t \in \mathcal{J}(\mathcal{J}_t) : H(D/\mathcal{J}_t) \leq H(D/S_t)$$

En conséquence, minimiser l'entropie conditionnelle $H(D/S_t)$ ne peut être le seul critère pour déterminer S_t^* , et il est nécessaire d'introduire un test d'arrêt limitant le cardinal de la solution, directement (cardinal de $S_t^* < k$), ou indirectement (par exemple $\% \text{ENEPE}(D/S_t^*) < a$, $a \in [0, 1[$).

Cependant, ce choix d'examens complémentaires ne peut être dicté par les seules considérations informationnelles : un examen qui permettrait un diagnostic sûr à 100 % au prix de la vie du patient doit de toute évidence être rejeté ; il en est de même pour un examen fournissant une quantité d'information très importante mais sous un délai au cours duquel une évolution fatale est possible.

Dans ce sens, une évaluation plus complète des examens complémentaires faisant intervenir des critères relatifs à leur condition d'obtention s'impose. Elle sera utilisée au paragraphe suivant.

Que l'entropie conditionnelle $H(D/S_t)$ soit ou non le seul critère retenu, il est clair que son calcul doit tenir compte du problème posé en $t+1$ par le praticien.

Pour un problème de confirmation, la partition de D à expliquer se réduit à deux classes : $D = \{ D_i, \bar{D}_i \}$; $H(D/S_t)$ apparaît alors comme une mesure de contribution à la confirmation de D puisque l'indice permet de rechercher le sous ensemble S_t qui sépare au mieux la modalité diagnostic D_i des autres. Pour le problème d'affinement, la partition à expliquer comportera autant de classes qu'il reste de modalités D_i à discriminer.

4.2 CARACTERISATION MULTICRITERE DE S_t

D'une façon générale, à un ensemble de paramètres S_t peut être associé un vecteur " coût ", $C(S_t)$, dans le but de pouvoir établir ses performances.

$$\forall S_t \in \mathcal{P}(\mathcal{I}_t), S_t \longmapsto C(S_t) = (C_1(S_t), C_2(S_t), \dots, C_f(S_t))$$

Les coûts " C_i " peuvent s'exprimer de façon qualitative, quantitative, voire même de façon floue. Remarquons que le coût " C_i " associé à S_t n'est pas forcément la " somme " des coûts élémentaires relatifs aux composantes de S_t .

Dans le cadre de notre étude, le vecteur $C(S_t)$ est formé d'une part d'un coût qualifiant la contribution de S_t à l'affinement et/ou à la confirmation de la variable diagnostic, et d'autre part de coûts relatifs aux conditions d'obtention de S_t . De façon plus précise, on distingue :

- $C_1(S_t)$: la contribution de S_t à l'explication de D

Elle est exprimée numériquement par l'entropie conditionnelle $H(D/S_t)$ dont le calcul tient compte, comme nous venons de le voir du problème posé en $t+1$ par le praticien. Ce critère présente les deux propriétés suivantes :

$$\text{sous additif : } H(D/S_t) \leq \sum_{X_i \in S_t} H(D/X_i)$$

$$\begin{aligned} \text{monotone} & : S_t, S_t' \in \mathcal{P}(\mathcal{I}_t), S_t' \subseteq S_t \\ & \Rightarrow H(D/S_t) \leq H(D/S_t') \end{aligned}$$

- $C_2(S_t)$: pénibilité associée à la mesure de S_t .

La pénibilité associée à un examen est un critère difficilement quantifiable. C'est pourquoi, nous utiliserons des méthodes d'optimisation permettant d'exprimer les valeurs des critères sous des formes très variées : quantitatives (bien sûr ...), mais aussi qualitatives, floues, comparatives ...

- $C_3(S_t)$: coût financier des examens nécessaires à la mesure de S_t .

$$\text{sous additif : } C_3(S_t) \leq \sum_{X_i \in S_t} C_3(X_i)$$

- $C_4(S_t)$: délai d'obtention de la mesure de S_t .

sous additif : $C_4(S_t) \leq \sum_{X_i \in S_t} C_4(X_i)$

- $C_5(S_t)$: fiabilité des mesures.

·
·
·
·

Dans ces conditions, le problème de sélection de variables décisionnelles est défini par la donnée des deux ensembles suivants :

1) ensemble des décisions admissibles pour le choix de S_t .

Le dossier médical évoluant au cours du temps, cet ensemble est défini à la date t par $\mathcal{P}(S_t)$. Par ailleurs, on pourra être amené suivant le cas, à introduire pour la résolution du problème de sélection un ensemble de contraintes supplémentaires qui devront être vérifiées par la solution retenue (pénibilité inférieure à un certain seuil,...).

2) ensemble des propriétés désirées pour la décision à retenir.

Les propriétés désirées pour la sélection à retenir s'expriment par référence au vecteur $C(S_t)$ défini précédemment. On aura alors trois types de formulations possibles.

a) formulation multicritère

La solution $S_t^* \in \mathcal{P}(S_t)$ vérifie :

$C_i(S_t^*)$ optimal $i = 1, 2, \dots, f$

L'optimisation simultanée des f critères n'étant bien sur pas possible dans le cas général, cette notion d'optimalité repose en fait sur le concept de "solutions non dominées".

Ce concept consiste à dire que S_t' est non dominée ssi :

$\forall S_t, C(S_t) \geq C(S_t') \Rightarrow C(S_t) = C(S_t')$

(la comparaison entre les deux vecteurs $C(S_t)$ et $C(S_t')$ s'entendant composante à composante)

b) formulation monocritère

La solution S_t^* est définie par :

$C_i(S_t^*)$ optimal sous les contraintes $C_j(S_t^*) \in \mathcal{L}_j$, $j \neq i$

où \mathcal{L}_j représente un ensemble de modalités admissibles du coût $C_j(S_t^*)$.

c) formulation intermédiaire

On définit un problème multicritère faisant intervenir p critères parmi les "f" possibles, les "f-p" restants étant traités sous forme de contraintes.

Deux problèmes se posent alors:

Problème n 1 : Evaluer de façon "fiable" les critères ; ceci implique en particulier, pour les critères informationnels, une évaluation des probabilités à partir des fréquences observées sur un sous ensemble de Ω de cardinal suffisant.

Problème n 2 : Résoudre le problème d'optimisation.

5 . EVALUATION FIABLE DES DIFFERENTS CRITERES

En considérant l'ensemble des modalités prises par S_{t-1} sur la population $\Omega_{t-1}(w)$, le processus d'aide au diagnostic définit une structure arborescente. Compte tenu du fait que \mathcal{L} est un ensemble fini, il apparaît immédiatement que plus on avance dans l'arbre, plus le cardinal de $\Omega_t(w)$ devient faible; les fréquences observées des différentes modalités diagnostic perdent alors une signification statistique. De tels noeuds sont déclarés indivisibles bien qu'ils ne soient pas forcément terminaux ; d'où les limites d'un tel processus d'interrogation.

Il s'agit alors de développer une procédure de fusion permettant d'obtenir un noeud divisible à partir d'un noeud indivisible /ZIGH,84/.

Il est clair que ces fusions occasionnent une perte d'information ; le calcul ci-après nous permet de donner son expression :

Soit Ω_{t-1} un noeud non terminal et S_{t-1} l'ensemble des variables effectivement mesurées à l'instant t . En considérant que les variables de S_{t-1} prennent q modalités, Ω_{t-1} admet alors q successeurs notés $\Omega_t^1, \Omega_t^2, \dots, \Omega_t^q$. Le gain d'information est donné par :

$$I^+_1 = H(D/\Omega_{t-1}) - \sum_{k=1}^q (|\Omega_t^k|/|\Omega_{t-1}|) H(D/\Omega_t^k)$$

Supposons que Ω_t^i soit un noeud non terminal indivisible et Ω_t^j le noeud à fusionner avec Ω_t^i , le gain d'information, après leur fusion, se réduit à :

$$I^+_2 = H(D/\Omega_{t-1}) - \sum_{\substack{k=1 \\ k \neq i \\ k \neq j}}^q \frac{|\Omega_t^k|}{|\Omega_{t-1}|} H(D/\Omega_t^k) - \left(\frac{|\Omega_t^i| + |\Omega_t^j|}{|\Omega_{t-1}|} \right) H(D/\Omega_t^i \cup \Omega_t^j)$$

La perte d'information occasionnée par cette fusion est donc égale à :

$$\begin{aligned} \Delta H &= I^+_2 - I^+_1 \\ &= \frac{|\Omega_t^i|}{|\Omega_{t-1}|} H(D/\Omega_t^i) + \frac{|\Omega_t^j|}{|\Omega_{t-1}|} H(D/\Omega_t^j) - \left(\frac{|\Omega_t^i| + |\Omega_t^j|}{|\Omega_{t-1}|} \right) H(D/\Omega_t^i \cup \Omega_t^j) \end{aligned}$$

On peut remarquer que le même calcul est applicable à la fusion de noeuds issus de sommets différents. En effet, il suffit de remplacer le noeud Ω_{t-1} par un ancêtre commun Ω_{t-k} . La comparaison des pertes d'information liées à des fusions se situant à des niveaux différents de l'arbre conduit à prendre l'ensemble Ω comme ancêtre commun de référence.

L'approche des fusions permet d'utiliser des algorithmes de classification hiérarchique (ascendant ou descendant) /JAMB,78/ pour déterminer les fusions optimales à établir à chaque étape.

Rappelons qu'une fusion aboutit à la proposition au praticien d'une population de redémarrage et des examens complémentaires associés (cf 2.3.2). Les pertes d'informations croissantes définissent l'ordre de présentation au praticien des noeuds fusionnés candidats à constituer une population de redémarrage.

6 . RESOLUTION DU PROBLEME D'OPTIMISATION

Dans la formulation monocritère, la résolution du problème d'optimisation ne présente en général aucune difficulté, compte tenu du nombre fini de décisions possibles ; c'est pourquoi nous étudierons des algorithmes permettant une approche multicritère (avec ou sans contrainte). Ces algorithmes ont pour but de déterminer à chaque date t , le sous-ensemble S_t "optimal" de paramètres dont la connaissance pourra permettre de résoudre le problème de décision relatif au noeud $\Omega_{t+1}(w)$.

6.1 CAS MULTICRITERE

Etant donné l'ensemble des décisions admissibles pour le choix de la sélection à retenir, soit $\mathcal{P}(\mathcal{J}_t)$ pour la date t , le problème consiste à déterminer la (ou les) solution(s) non dominée(s) en prenant appui simultanément sur plusieurs critères non aisément réductibles à un seul.

Ces critères comme nous venons de le voir peuvent prendre des formes diverses : appréciation d'un facteur non quantifié (pénibilité), valeur numérique résultat d'un calcul (mesure de la contribution à l'affinement/confirmation du diagnostic). Chacun de ces critères définit une certaine structure (relation d'ordre total ou partiel) sur l'univers de choix $\mathcal{P}(\mathcal{J}_t)$.

Si dans le cas monocritère, l'organe de décision ou décideur est censé être capable de déterminer une solution optimale (l'univers de choix étant muni d'une structure unique) il n'en est pas de même dans un problème d'optimisation multicritère. La situation n'est pas aussi simple puisque l'on a sur $\mathcal{P}(\mathcal{J}_t)$ autant de structures que de critères considérés.

L'objectif des différentes méthodes discutées ci-dessous est d'agréger ces différentes structures initiales en une structure unique /BERT,75/,/JACK,80/,/MOUR,81/,/BERN,71/.

6.1.1 FONCTION D'UTILITE

Une telle fonction associe à chaque partie de $\mathcal{P}(\mathcal{J}_t)$, S_t une valeur numérique $U(S_t)$ et permet ainsi de générer un ordre total sur $\mathcal{P}(\mathcal{J}_t)$.

Généralement, le modèle d'agrégation est linéaire. Si les performances de S_t sont évaluées à partir d'un vecteur "coût" de dimension f :

$$\forall S_t \in \mathcal{P}(S_t), \quad S_t \longmapsto C(S_t) = (C_1(S_t), C_2(S_t) \dots C_f(S_t))$$

on a alors :

$$U : \mathcal{P}(S_t) \longrightarrow R$$

$$S_t \longmapsto U(S_t) = \sum_{i=1}^f p_i C_i(S_t)$$

où p_i désigne le coefficient de pondération associé à chacune des performances de S_t .

Pour notre application, confrontée à une telle hétérogénéité entre les différents coûts, la voie qui consiste à utiliser une fonction d'utilité s'avère impraticable. Soulignons les principaux obstacles suivants :

a. une connaissance imparfaite des préférences du médecin, principalement pour ce qui relève de la manière dont peut s'opérer la compensation entre un gain sur une performance et les pertes sur les autres ; ceci d'autant plus que les performances ne sont pas exprimées dans une même unité (temps, fiabilité, coût financier ...).

b. une appréciation plus ou moins précise pour évaluer les performances de chaque sous ensemble explicatif (jugements subjectifs ou évaluations grossières (fiabilité d'une réponse à un questionnaire, pénibilité de tel examen ...)).

c. une méthode fondée sur une fonction d'utilité exige l'hypothèse de l'indépendance des contributions de chaque critère à l'utilité globale. Cette hypothèse est souvent peu conforme à la réalité.

d. un manque de maniabilité de la fonction d'utilité globale pour pouvoir s'adapter aux impératifs du moment (l'état d'urgence à prendre une décision thérapeutique, la disponibilité des laboratoires d'analyses, ou des appareils de mesures, le budget du service ...).

6.1.2 HIERARCHISATION DES CRITERES

Si la hiérarchisation des critères est possible, on résoud le problème de choix multicritère de la façon suivante :

On recherche les solutions non dominées relativement au critère jugé comme le plus important par le décideur. Une fois celles-ci obtenues, on considère le critère qui apparaît en seconde place dans la hiérarchie. On obtient un nouvel ensemble de solutions issu de celui obtenu précédemment. On recommence

la procédure jusqu'à épuisement des critères ; ces derniers étant pris dans leur ordre d'importance.

Signalons que cette méthode lexicographique peut conduire à une extrémité regrettable : celle d'avoir écarté une solution jugée dominée par rapport à un critère et arrivant en tête pour tous les autres. Nous ne retiendrons donc pas cet algorithme pour notre application.

6.1.3 RELATION DE SURCLASSEMENT

La difficulté à donner une fonction d'utilité globale pour la résolution du problème de sélection de variables multicritère nous a conduits à étudier des méthodes basées sur la notion de surclassement. Elles sont généralement utilisées lorsque, dans l'analyse des conséquences d'une décision (ici, le choix d'un sous ensemble explicatif), on rencontre les difficultés suivantes :

- certaines conséquences sont non additives soit pour des raisons d'hétérogénéité (délai d'obtention, coût financier, fiabilité ... dans notre application) soit parce qu'elles sont dépendantes.

- certaines conséquences sont qualitatives (on peut les décrire mais non les mesurer, par exemple, la pénibilité engendrée par la mesure de S).

Par ailleurs, ces méthodes s'appliquent parfaitement lorsque l'univers de choix est fini.

Elles consistent à définir une relation binaire \mathcal{R} de surclassement. L'ensemble des décisions admissibles pour le choix de S étant $\mathcal{P}(\mathcal{S}_t)$ pour la date t, on doit bâtir \mathcal{R} tel que :

$$\mathcal{R} : \mathcal{P}(\mathcal{S}_t) \longrightarrow (0,1)$$

$$(S_t, S_t') \longmapsto \mathcal{R}(S_t, S_t') = 1 \text{ si } S_t \text{ préféré à } S_t'$$

$$0 \text{ si } S_t \text{ non préféré à } S_t'$$

Les règles définissant \mathcal{R} tiennent compte des vecteurs $C(S_t)$ et $C(S_t')$ ainsi que d'une éventuelle pondération appliquée à leurs composantes. Suivant le cas, on utilise des règles plus ou moins complexes (par exemple, dans les méthodes ELECTRE, des tests de concordance et de non discordance appliqués sur les composantes de $C(S_t)$ et $C(S_t')$ /ROY,70/, /GUIG,77/ ou bien sur les rangs associés à ces

critères). Pour tous les couples de sous ensembles (S_t, S_t') on obtient alors une des quatre situations suivantes :

$S_t \mathcal{R} S_t'$	$S_t' \mathcal{R} S_t$	Conclusion
0	0	S_t, S_t' incomparables
1	0	S_t domine S_t'
0	1	S_t dominé par S_t'
1	1	S_t équivalente à S_t'

A partir de cette relation de surclassement, on peut :

a) soit chercher à séparer dans $\mathcal{P}(\mathcal{S}_t)$ les meilleurs sous ensembles des moins bons ; cela conduit à chercher une dichotomie sur $\mathcal{P}(\mathcal{S}_t)$.

b) soit chercher à établir un préordre total, c'est à dire un ordre total tolérant des exaequo, sur $\mathcal{P}(\mathcal{S}_t)$; cela permet d'approcher plusieurs problèmes, choix d'un seul sous ensemble, de plusieurs, ...

Pour le problème a), on élimine les circuits du graphe de la relation \mathcal{R} en réduisant tous les sommets d'un circuit, que l'on considère comme équivalents, en un seul point. Puis on cherche le noyau \mathcal{N} du graphe sans circuit obtenu tel que l'on obtienne :

$$\forall S_t' \in \mathcal{P}(\mathcal{S}_t) \setminus \mathcal{N} \Rightarrow \exists S_t \in \mathcal{N} \text{ tel que } S_t \mathcal{R} S_t'$$

$$\forall S_t' \in \mathcal{N}, \exists S_t \in \mathcal{N} \text{ tel que } S_t \mathcal{R} S_t'$$

Pour le problème b), lorsque le graphe de \mathcal{R} est déterminé, sa décomposition hiérarchique fournit deux préordres (ascendant, descendant), permettant d'ordonner $\mathcal{P}(\mathcal{S}_t)$ (préordre médian).

Pour notre application, on se limitera à la solution donnée par l'approche a) ; si l'ensemble des solutions non dominées comprend plus d'un élément, deux approches sont à considérer.

i) le choix final appartient au praticien

ii) réitération de la procédure avec la donnée d'une nouvelle relation de surclassement imposant des conditions moins strictes.

6.2 SIMPLIFICATIONS

Il est clair qu'une méthode basée sur la notion de surclassement peut conduire à des temps de calcul importants si la recherche des solutions non dominées s'effectue sur un univers de choix de taille importante (ne serait ce que dans la construction de la relation de surclassement) ; ceci risque d'être le cas dans notre application puisque nous travaillons sur l'ensemble des parties de \mathcal{S}_t , soit $\mathcal{P}(\mathcal{S}_t)$.

Néanmoins, d'un point de vue opérationnel, nous ne conserverons de cet ensemble que les parties de \mathcal{S}_t qui vérifient certaines contraintes. En effet, le budget d'un service hospitalier connaît des limites, et il en est de même pour les souffrances que peut endurer un prématuré. On sera donc toujours amené à résoudre un problème d'optimisation faisant intervenir "i" critères avec $i = 1, 2, \dots, f$ et "j" contraintes avec $f-i \leq j < f$, (l'inégalité stricte provient du fait que nous n'imposerons pas de contraintes particulières au critère mesurant la contribution de S_t à l'affinement ou/et la confirmation de la variable diagnostic). Dans ces conditions, la relation de surclassement s'opère sur :

$$\mathcal{R} : \mathcal{P}(\mathcal{S}_t) \cap \text{contraintes} \longrightarrow (0,1)$$

Toutefois ce nouvel univers de choix peut rester encore important et il serait bon de définir quelques heuristiques simplificatrices pour répondre à cette éventualité. Nous proposons dans ce cas une approche hiérarchique ascendante et descendante pour le problème de sélection de variables /AYGA,86/.

6.2.1 RECHERCHE ASCENDANTE

Cet algorithme construit la chaîne des sous ensembles s^1, s^2, \dots jusqu'à s^k avec $s^{n+1} = s^n \cup \{X_i\}$: X_i est choisi parmi $\mathcal{S}_t \setminus s^n$ tel que :

- s^{n+1} surclasse tout autre sous ensemble de dimension $n+1$ obtenu à partir de s^n et,
- s^{n+1} vérifie les contraintes à respecter.

La solution s^k est obtenue quand il n'est plus possible de trouver une variable $X_i \in \mathcal{S}_t \setminus s^k$ qui n'entraîne pas la violation d'une contrainte. Dans cette procédure, s^0 est initialisé par l'ensemble vide.

6.2.2 RECHERCHE DESCENDANTE

La recherche de S^{k*} s'effectue par retraites successifs. L'algorithme construit la chaîne des sous-ensembles S^{p*} , $S^{(p-1)*}$, ..., S^{k*} avec $S^{n*} = S^{(n+1)*} \setminus X_i$: $\{X_i\}$ est choisi de façon à rendre "optimal" S^{n*} . La solution S^{k*} est obtenue dès que celle-ci vérifie les contraintes.

On peut remarquer qu'il peut exister une variable parmi celles écartées aux étapes précédentes, $X_i \in \mathcal{J}_t \setminus S^{k*}$, qui agrégée à S^{k*} n'entraîne pas la violation d'une contrainte. Nous proposons d'améliorer cette procédure en appliquant une recherche ascendante initialisée à partir de la solution déjà trouvée.

7 . CONCLUSION

Dans ce chapitre, nous avons vu que le problème d'une aide au diagnostic dans le contexte d'une médecine d'urgence peut se décomposer à chaque étape en trois sous-problèmes :

a) Définition d'un sous ensemble de la population d'apprentissage, Ω_t

b) Définition d'une procédure permettant de résoudre le problème de décision relatif à Ω_t

c) Prescription d'examens complémentaires dont la mesure permettrait d'affiner ou de confirmer la proposition diagnostique, compte tenu des contraintes liées à leurs conditions d'obtention.

Pour le problème a), nous avons fait apparaître que la construction de Ω_t prend en considération, les variables mesurées jusqu'à l'instant $t-1$, ainsi que le problème posé en $t-1$ par le praticien. Qu'éventuellement, d'autres facteurs peuvent intervenir comme par exemple la cohérence de la proposition diagnostique émise en t avec celle émise en $t-1$, la divisibilité du noeud.

Pour le problème b), nous avons développé deux approches : l'une vise à fiabiliser les propositions diagnostiques par la minimisation ou la majoration de probabilité d'erreur, tout en tenant compte de la quantité d'informations qu'apporte la décision sur la connaissance du diagnostic ; l'autre repose sur la donnée d'une fonction coût, permettant ainsi de traduire l'efficacité thérapeutique des décisions diagnostiques.

Pour le problème c), on a pu constater que le choix d'un ensemble d'examens complémentaires ne pouvait pas être dicté par de simples considérations informationnelles et qu'il était donc nécessaire d'introduire une évaluation multicritère. Différentes méthodes ont été discutées, et l'approche retenue utilise la notion de surclassement. Enfin, quelques simplifications ont été proposées pour accélérer la résolution du problème d'optimisation.

Ainsi, à tout instant, compte tenu des paramètres connus, l'approche adoptée fournit une ou des propositions de diagnostic et un processus conversationnel s'établit :

- en entrée : le médecin définit l'ensemble des examens réalisables, les coûts associés, les contraintes à respecter, ... ainsi que la démarche qu'il veut adopter (affinement, confirmation, ou les deux).

- en sortie : la machine propose la mesure d'un ensemble de paramètres "optimal", répondant aux objectifs à satisfaire.

- C O N C L U S I O N G E N E R A L E -

- CONCLUSION GENERALE -

Le travail présenté trouve son application dans le cadre d'une aide au diagnostic des troubles respiratoires, pour les nouveau-nés prématurés admis au Centre de Pathologie Néo Natale d'Arras, présentant ce symptôme.

L'approche proposée, faisant appel au modèle de la Reconnaissance des Formes est basée sur la donnée d'un échantillon représentatif de la pathologie considérée. La construction de celui-ci (sur un horizon d'observation approprié) à partir de la Banque de données élaborée au Centre pose d'emblée quelques problèmes : présence de dossiers incomplets dans la population d'apprentissage, codage des données.

Une première phase du travail a donc consisté à formuler ces deux problèmes, ce qui nous a conduits à la définition d'un individu fictif (individu le mieux reconstitué) dont les caractéristiques peuvent être utilisées pour la discrimination et le classement.

Dans une deuxième phase, nous avons développé une procédure d'aide au diagnostic basée sur des segmentations successives de la population d'apprentissage. Cette démarche peut être comparée à celles utilisant la théorie des questionnaires /PICA,72/, /TERR,76/, /ZIGH,84/. Cependant notre approche se caractérise par :

1 - Une plus grande conformité avec l'approche adoptée par le médecin lui même. En effet, on ne résoud pas en chaque noeud un problème toujours identique ; au contraire, le problème formulé en chacun des noeuds dépend de la qualité des résultats déjà obtenus en ce noeud. Il est entièrement sous le contrôle du médecin (affinement, confirmation ou les deux).

2 - Une évaluation plus réelles des éventuels examens complémentaires par le biais d'une représentation multicritère. Parmi ces critères, un seul est déterminé mathématiquement (contribution à l'affinement et/ou à la confirmation), les "valeurs" des autres étant à chaque étape sous le contrôle du médecin qui peut ainsi tenir compte des conditions réelles d'obtention (capacité réelle du nouveau-né à accepter tel examen, habileté des personnes dans le service, disponibilité du laboratoire d'analyse, ...).

3 - L'entropie de Shannon fournit des indices dont l'interprétation est aisée. Contrairement à d'autres approches qui utilisent une entropie adaptée (mais moins interprétable) pour traiter les noeuds d'effectif faible, nous utilisons l'entropie de Shannon dans tous nos traitements, que ce soit pour évaluer la contribution d'un ensemble de paramètres à l'affinement et/ou la confirmation de la proposition diagnostique, ou pour déterminer les fusions de noeuds à opérer par des méthodes de classification hiérarchique.

4 - Notre procédure prend en compte les différentes possibilités d'incohérence qui peuvent être observées lors de son déroulement et propose au médecin les populations de redémarrage les mieux adaptées à une nouvelle tentative.

L'approche de segmentation retenue peut être comparée à celle mise en oeuvre par les Systèmes Experts. Dans les deux cas, la procédure parcourt une arborescence :

- dans notre étude, un noeud de l'arborescence représente un sous-ensemble de la population d'apprentissage, caractérisé par une valeur connue d'un certain nombre de variables.

- dans l'approche Systèmes Experts, un noeud représente un sous-ensemble d'une population générale (dont la population d'apprentissage peut être considérée comme un échantillon) caractérisé par le fait que ses éléments vérifient un certain nombre de prémisses.

Les modes de progression dans ces arborescences peuvent être aussi comparés :

- le chaînage avant correspondant à la procédure d'affinement (étude globale).

- le chaînage arrière correspondant à la procédure de confirmation (étude locale).

Il convient à présent de trouver les moyens nécessaires à la mise en oeuvre pratique de la démarche proposée. Celle-ci conduit à extraire des règles à caractère directement opérationnel à partir de la connaissance contenue dans une population d'apprentissage. Ces règles peuvent être complétées par d'autres s'appliquant à une population plus générale caractéristique de la pathologie considérée et formulées à partir de l'expérience du médecin.

- A N N E X E S -

- ANNEXE I -

Ces définitions sont issues de la référence /CAIL,76/.

1 . INDICE DE SIMILARITE

Sur un ensemble E, c'est une application S de ExE dans \mathbb{R}^+

$$\begin{array}{ccc} S : \text{ExE} & \longrightarrow & \mathbb{R}^+ \\ (i,j) & \longmapsto & S(i,j) \end{array}$$

telle que :

$$S(i,j) = S(j,i) \quad \text{pour tout } (i,j) \in \text{ExE}$$

$$S(i,i) = S(j,j) = S_{\max} > S(i,j) \quad \text{pour tout } \begin{array}{l} (i,j) \in \text{ExE} \\ i \neq j \end{array}$$

2 . INDICE DE DISSIMILARITE

Sur un ensemble E, c'est une application D de ExE dans \mathbb{R}^+

$$\begin{array}{ccc} D : \text{ExE} & \longrightarrow & \mathbb{R}^+ \\ (i,j) & \longmapsto & D(i,j) \end{array}$$

telle que :

$$D(i,j) = D(j,i) \quad \text{pour tout } (i,j) \in \text{ExE}$$

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

Ces deux définitions peu contraignantes permettent certaines incohérences comme par exemple :

$$\left\{ \begin{array}{l} i \neq j \\ D(i,j) = 0 \\ \exists k \in E / D(i,k) \neq D(j,k) \end{array} \right.$$

Afin d'éviter ce type d'incohérence, on adjoint deux autres propriétés :

$$(a) : D(i,j) = 0 \quad \Rightarrow \quad i = j$$

$$(b) : D(i,j) \leq D(i,k) + D(k,j) \quad \text{pour tout } (i,j,k) \in E^3$$

3 . INDICE DE DISTANCE

C'est un indice de dissimilarité qui vérifie la propriété (a) :

$$\begin{aligned}
 d : E \times E &\longrightarrow \mathbb{R}^+ \\
 (i, j) &\longmapsto d(i, j) \\
 d(i, j) &= d(j, i) \\
 d(i, j) = 0 &\iff i = j
 \end{aligned}$$

4 . ECARTS, PSEUDO-DISTANCE, QUASI-DISTANCE

C'est un indice de dissimilarité qui vérifie la propriété (b) :

$$\begin{aligned}
 e : E \times E &\longrightarrow \mathbb{R}^+ \\
 (i, j) &\longmapsto e(i, j) \\
 e(i, j) &= e(j, i) \\
 e(i, i) &= 0 \\
 e(i, j) &\leq e(i, k) + e(k, j) \quad \text{pour tout } (i, j, k) \in E^3
 \end{aligned}$$

5 . DISTANCE

C'est un indice de dissimilarité qui vérifie à la fois les propriétés (a) et (b). On parle de distance ultramétrique quand la propriété (b) est remplacée par :

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

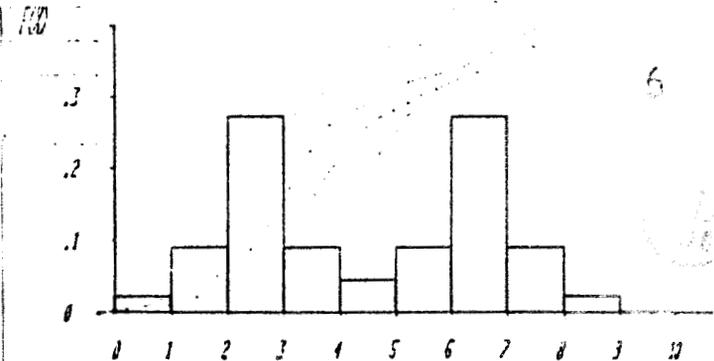
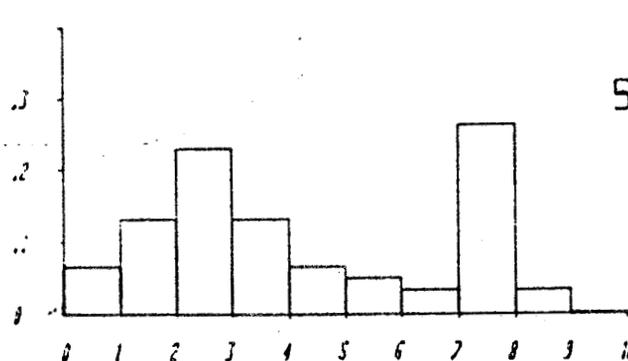
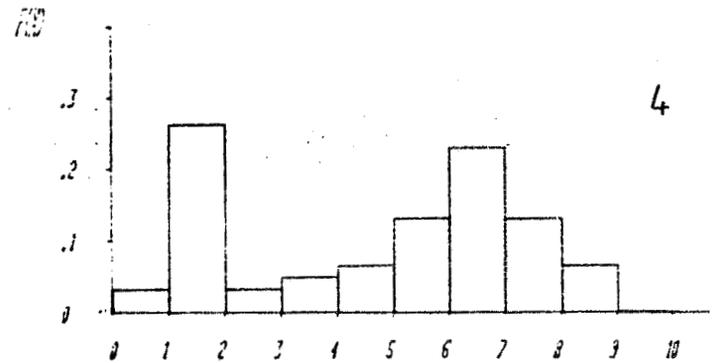
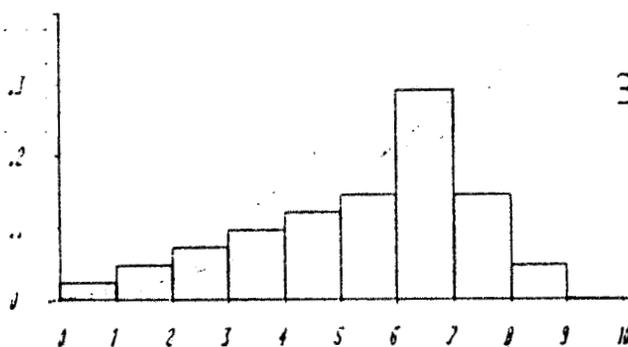
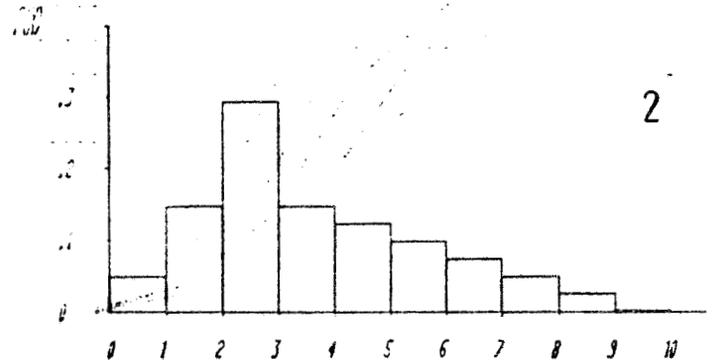
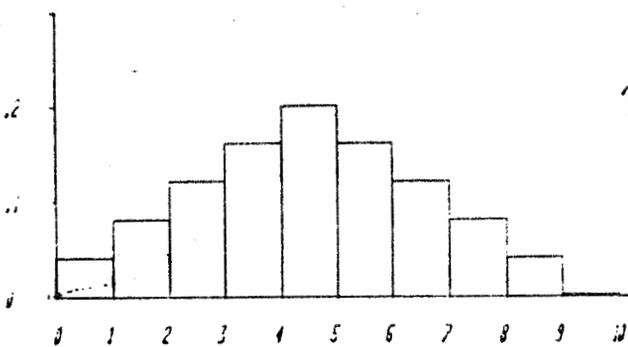
- ANNEXE II -

Les différents exemples de reconstruction donnés ci-dessous, nous permettent d'énoncer les remarques suivantes :

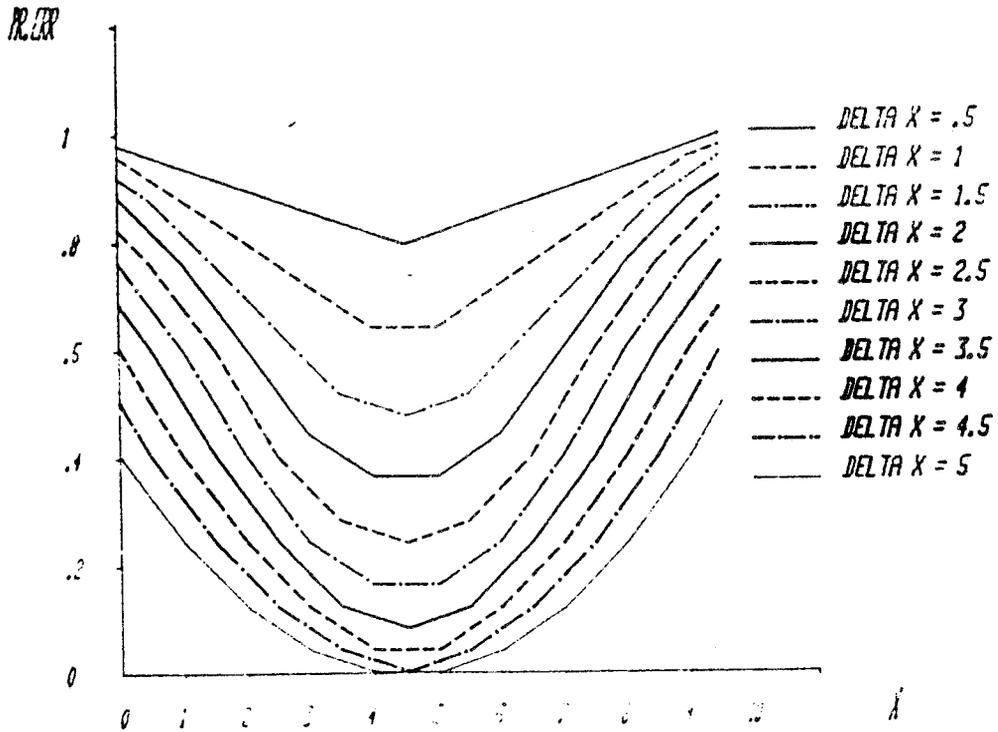
a) dans le cas où la fonction de distribution est unimodale sur l'ensemble des plus proches voisins (cf exemple 1, 2, 3), le mode est reconnu comme étant la meilleure (au sens du critère) estimation pour le paramètre à reconstituer.

b) dans le cas où la fonction de distribution est multimodale, on peut noter que le mode "isolé" ne constitue pas de façon générale l'estimation du paramètre manquant (cf exemple 4, 5, 6).

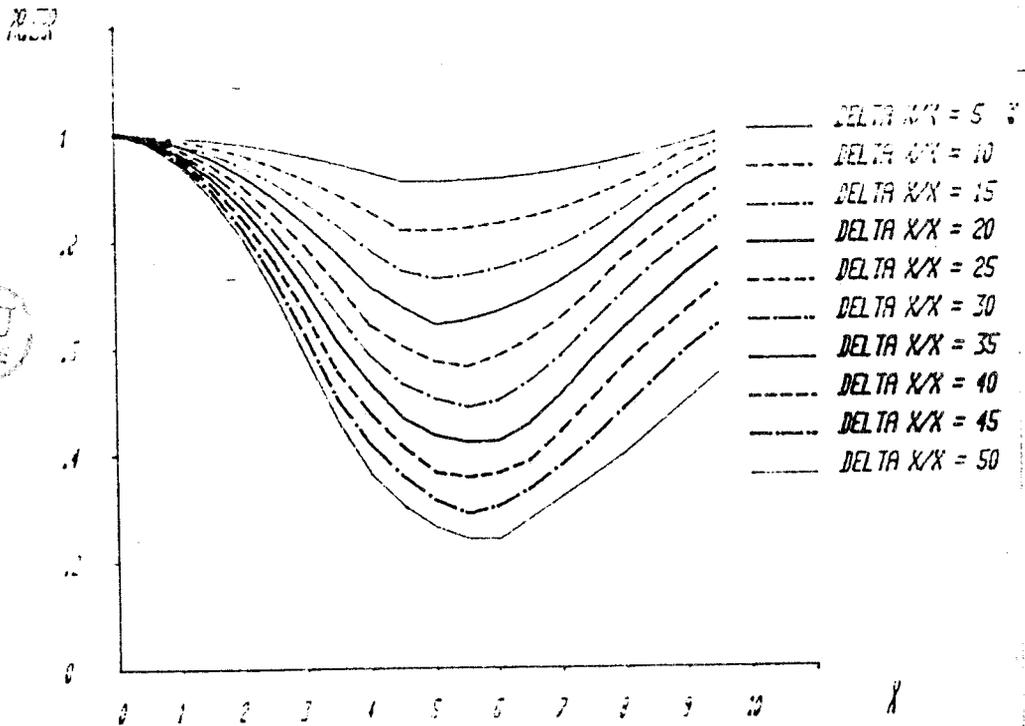
Soulignons aussi que lorsque la précision accordée à l'estimation est évaluée à partir de l'erreur relative, l'estimateur favorise les estimations de plus grande valeur. De plus, lorsque la précision devient faible, l'estimation optimale tend vers la valeur moyenne.



EXEMPLE N 1



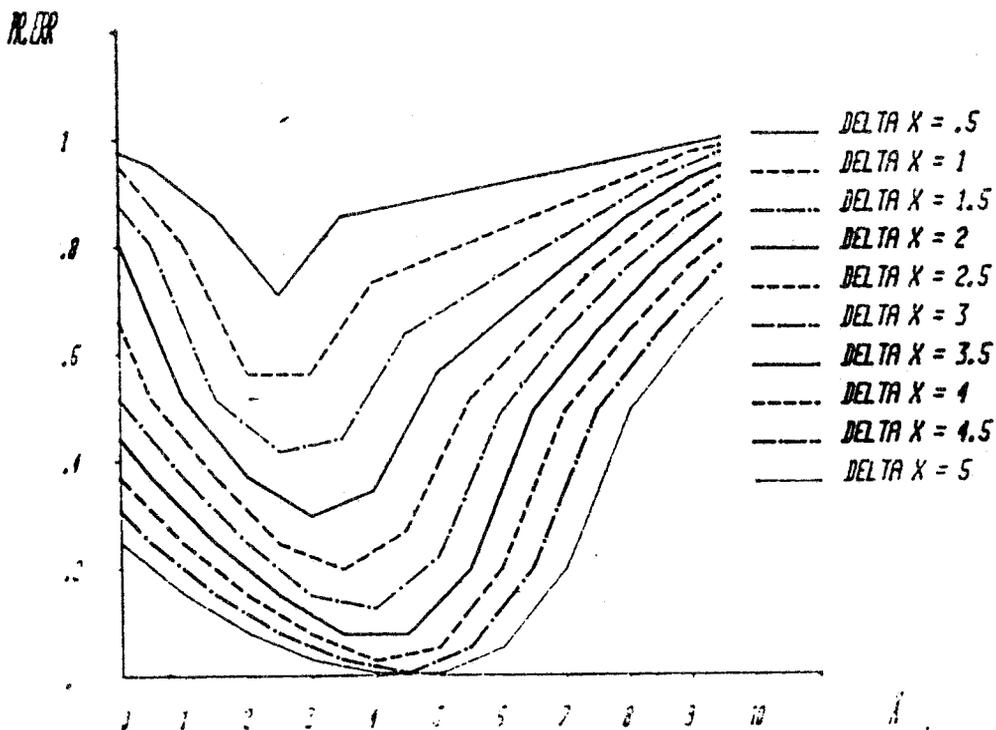
- PROP. D'ERREUR = $F(x)$ POUR DELTA X = CSTE -



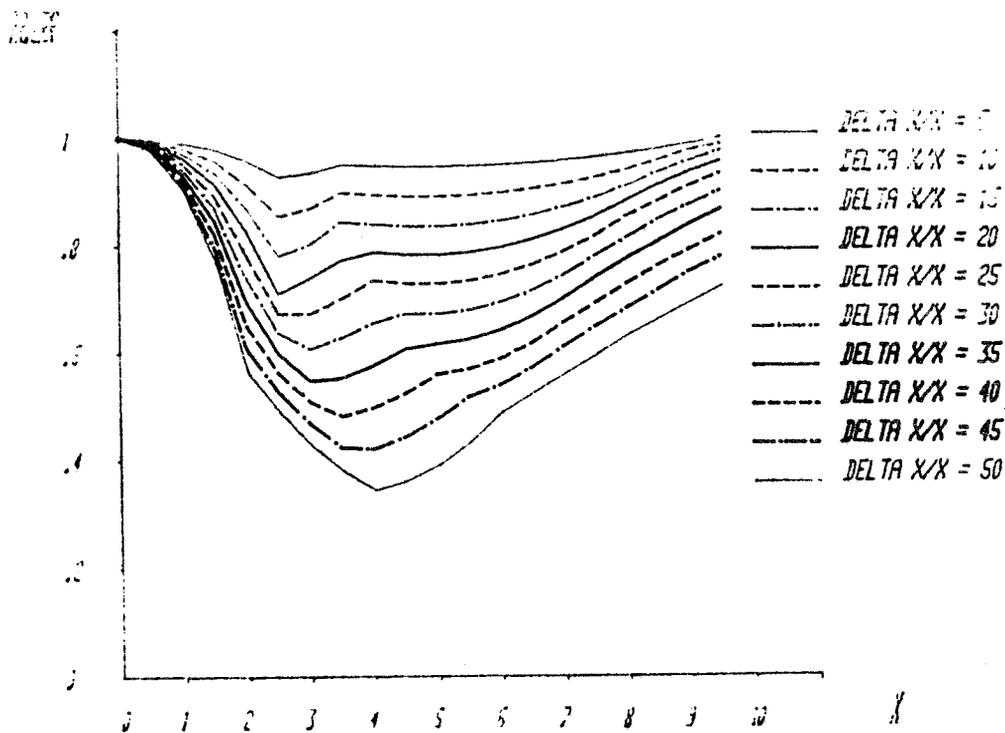
- PROP. D'ERREUR = $F(x)$ POUR DELTA X/X = CSTE -



EXEMPLE N 2



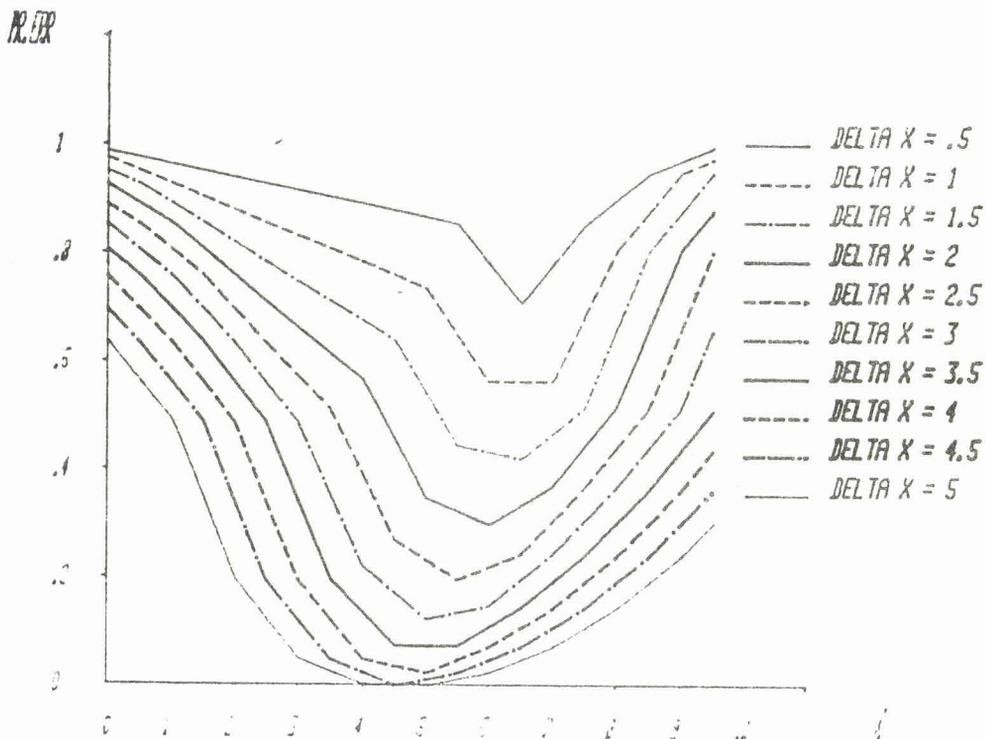
- PROB. D'ERREUR = F(X) POUR DELTA X = CSTE -



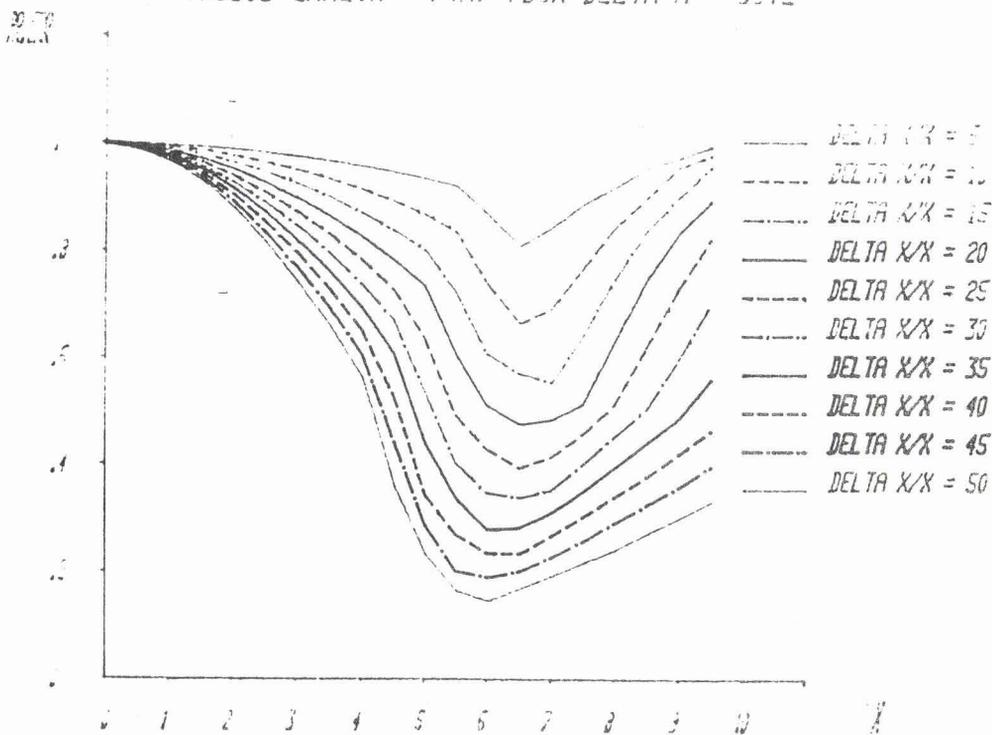
- PROB. D'ERREUR = F(X) POUR DELTA X/X = CSTE -



EXEMPLE N 3



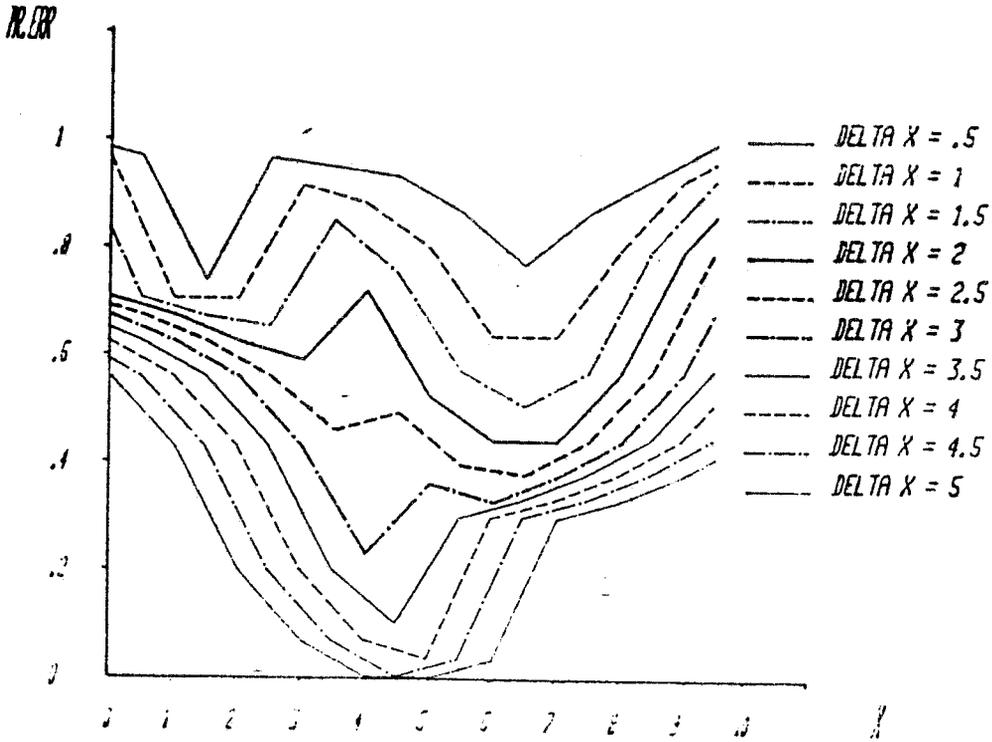
- PROB. D'ERREUR = 5(X) POUR DELTA X = CSTE -



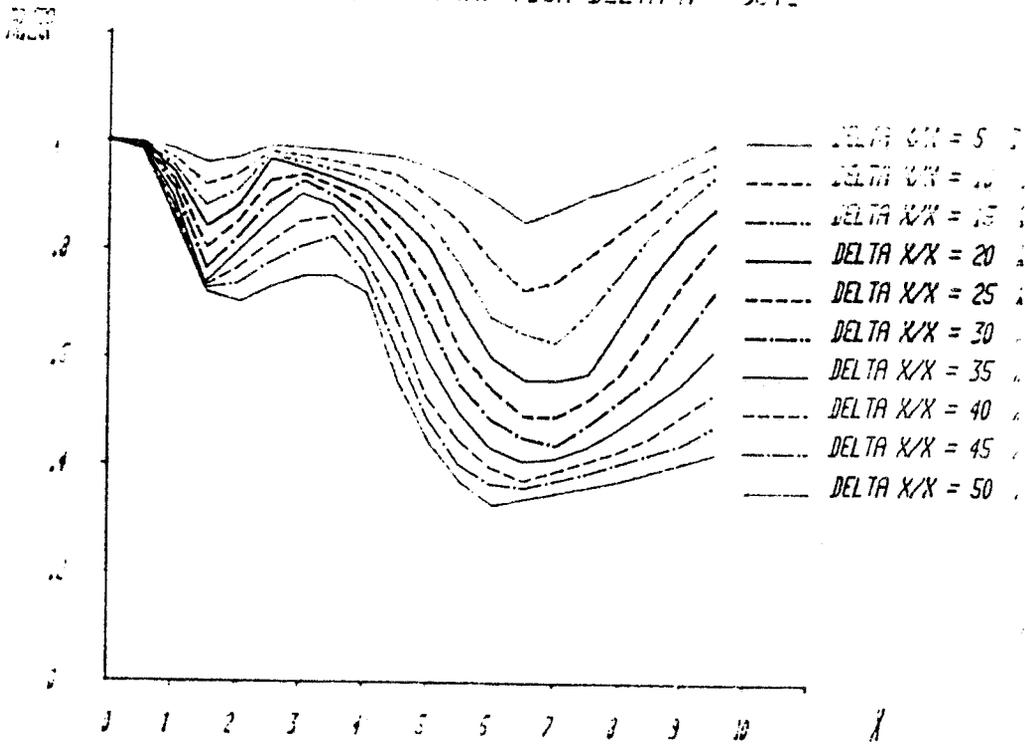
- PROB. D'ERREUR = 5(X) POUR DELTA X/X = CSTE -



EXEMPLE N 4

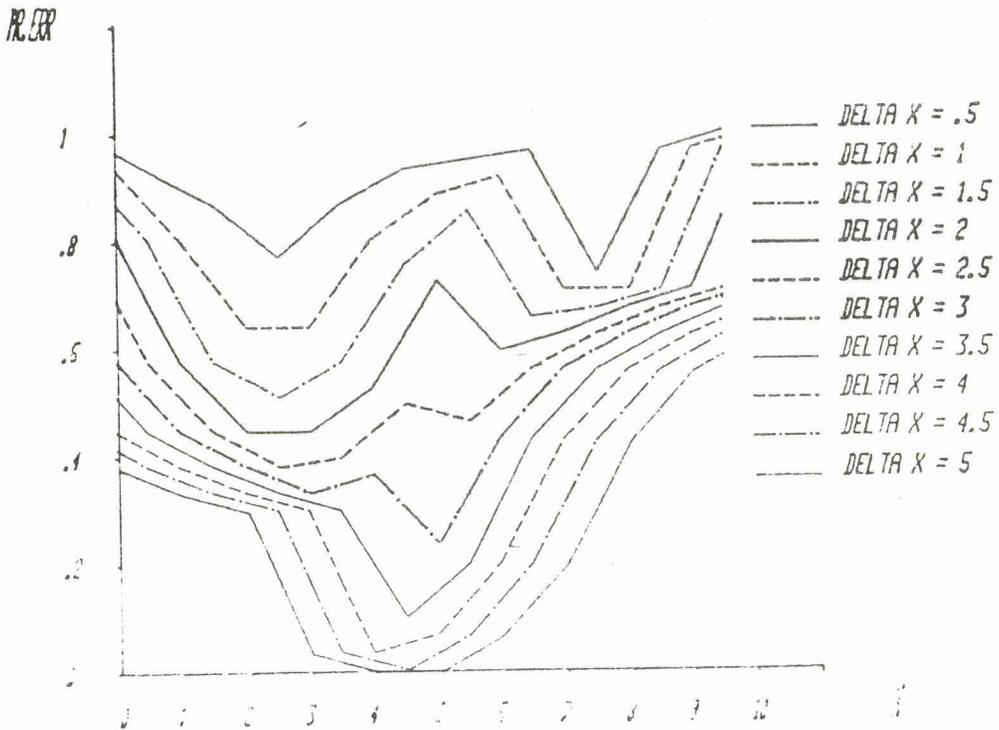


- PROB. D'ERREUR = F(X) POUR DELTA X = CSTE -

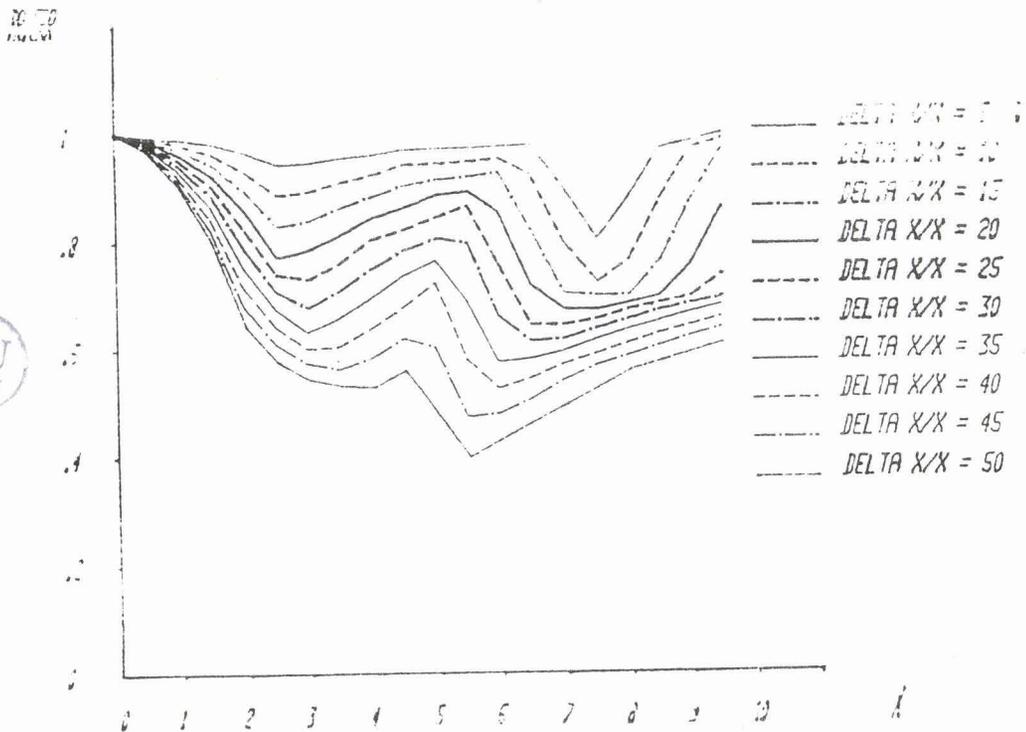


- PROB. D'ERREUR = F(X) POUR DELTA X/X = CSTE -

EXEMPLE N 5



- PROB. D'ERREUR = F(x) POUR DELTA X = CSTE -

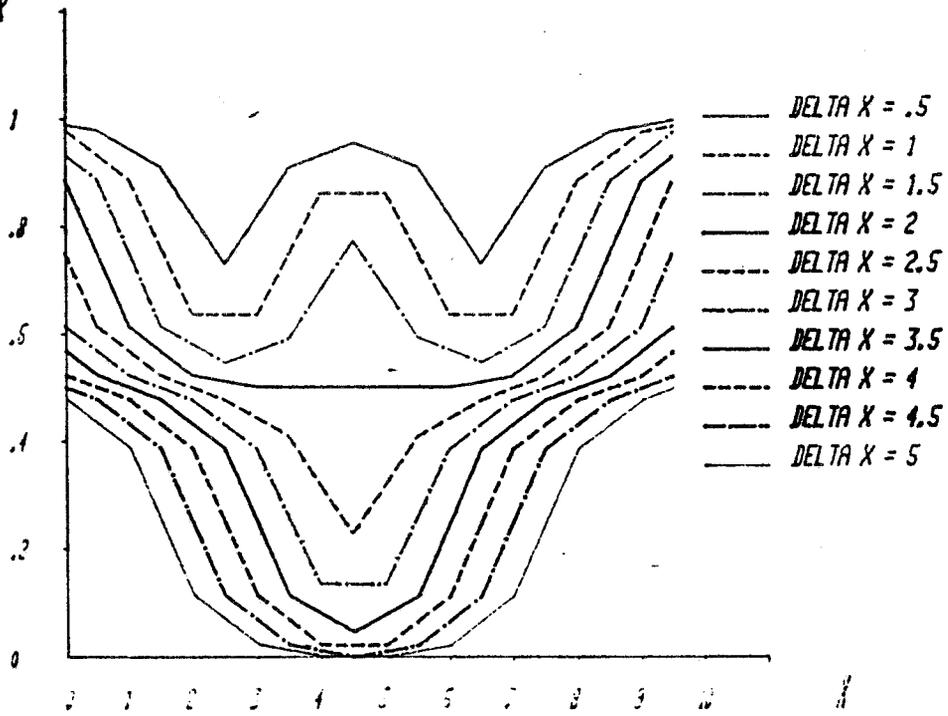


- PROB. D'ERREUR = G(x) POUR DELTA X/X = CSTE -



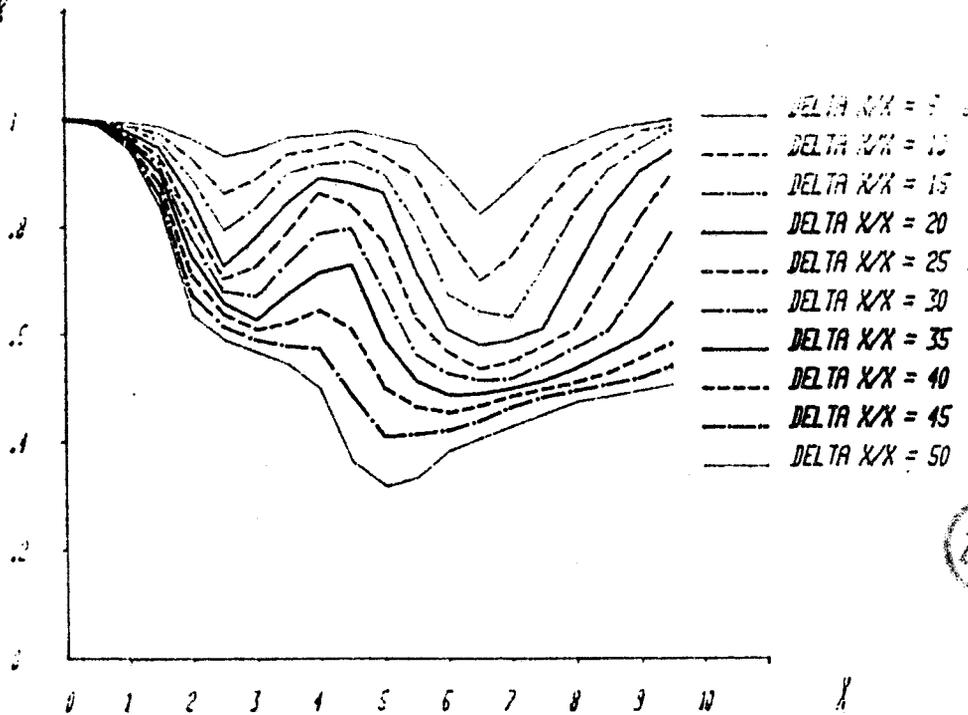
EXEMPLE N 6

PROB



- PROB. D'ERREUR = F(X) POUR DELTA X = CSTE -

PROB



- PROB. D'ERREUR = G(X) POUR DELTA X/X = CSTE -



- ANNEXE III -COMPARAISON DES PARTITIONS $P_\phi(\Omega)$ ET $P_{\hat{\phi}}(\Omega)$ (III.4.1c)

PROPOSITION : L'égalité des partitions $P_\phi(\Omega)$ et $P_{\hat{\phi}}(\Omega)$ implique forcément l'égalité des contenus sémantiques $\hat{\phi}$ si la procédure d'estimation $\hat{\phi}$ minimise la probabilité d'erreur, et ceci quel que soit l'ensemble des décisions choisi.

PREUVES :

1 . ETUDE CLASSIQUE : $\Delta = M_D$

A partir des partitions $P_\phi(\Omega)$ et $P_{\hat{\phi}}(\Omega)$, on peut construire le tableau suivant représentant l'intersection de ces deux partitions :

$P_\phi(\Omega) \backslash P_{\hat{\phi}}(\Omega)$	d_1		d_j		d_q	$\sum_j N_{ij}$
D_1	N_{11}		N_{1j}		N_{1q}	$N_{1.}$
D_i	N_{i1}		N_{ij}		N_{iq}	$N_{i.}$
D_q	N_{q1}		N_{qj}		N_{qq}	$N_{q.}$
$\sum_i N_{ij}$	$N_{.1}$		$N_{.j}$		$N_{.q}$	N

où N_{ij} désigne le nombre d'individus de Ω appartenant à la forme D_i pour lesquels on a décidé d_j ,

$N_{i.}$ le nombre d'individus de Ω appartenant à la forme D_i ,

$N_{.j}$ le nombre d'individus de Ω pour lesquels on a décidé d_j .

a) Comme sur la population d'apprentissage Ω , toutes les formes sont représentées, on a alors les q données suivantes :

$$\forall i = 1, q \quad N_{i.} > 0$$

b) L'égalité des partitions $P_\phi(\Omega)$ et $P_{\hat{\phi}}(\Omega)$ n'ayant lieu que si $P_{\hat{\phi}}(\Omega)$ possède le même nombre de classes que $P_\phi(\Omega)$, on a aussi :

$$\forall j = 1, q \quad N_{.j} > 0$$

Dire que les partitions $P_\phi(\Omega)$ et $P_{\hat{\phi}}(\Omega)$ sont équivalentes revient à dire qu'il existe une correspondance biunivoque entre ces deux partitions ; à toute classe de $P_{\hat{\phi}}(\Omega)$ (resp. $P_\phi(\Omega)$), il est possible d'associer une et une seule classe de $P_\phi(\Omega)$ (resp. $P_{\hat{\phi}}(\Omega)$). Soit en d'autres termes, que le tableau représentant l'intersection de $P_\phi(\Omega)$ et de $P_{\hat{\phi}}(\Omega)$ ne possède qu'un élément non nul par ligne et par colonne. Le nombre de solutions possibles est alors donné par :

$$|\mathcal{S}| = \binom{q}{1} + \binom{q-1}{2} + \dots + \binom{1}{1} = q!$$

Vérifions maintenant que si la procédure minimise la probabilité d'erreur, la seule solution admissible conduit à l'égalité des contenus sémantiques, c'est à dire celle définie par :

$$\mathcal{S} = \{ N_{ii} = N_{i.}, N_{ij} = 0 \quad \forall i, j = 1, q ; i \neq j \}$$

Pour cela, nous allons montrer la propriété suivante :

$$N_{ii} > N_{ij} \quad \forall j = 1, q ; i = 1, q ; j \neq i$$

On note par n_{ij} , le nombre d'individus de Ω appartenant à la forme D_i , pour lesquels on a décidé d_j et dont la modalité prise par le paramètre S qui les caractérise est m_S .

$$n_{ij} = |\{ w \in \Omega \text{ tel que } \phi(w) = D_i, S: S(w) = m_S, \hat{\phi}(m_S) = d_j \}|$$

Si la procédure de décision $\hat{\phi}$ minimise la probabilité d'erreur (cf IV.3.2), la décision d_i est prise si et seulement si la condition suivante est vérifiée :

$$\hat{\phi}(m_S) = d_i \iff n_{ii} / \sum_{j=1}^q n_{ji} = \text{Max}_k (n_{ki} / \sum_{j=1}^q n_{ji})$$

d'où

$$N_{ii} = \sum_{d_i} n_{ii} = \text{Max}_k N_{ki}$$

Par conséquent, seuls les éléments diagonaux sont susceptibles de prendre la valeur N_{ii} . L'égalité des partitions entraîne bien l'égalité des contenus sémantiques. On aura donc:

$$d(P_{\hat{\phi}}(\Omega), P_{\hat{\phi}^*}(\Omega)) = 0 \Leftrightarrow N_{ii} = N_i \quad \forall i = 1, q$$

$$N_{ij} = 0 \quad \forall i, j = 1, q ; i \neq j$$

Il est facile de montrer que ceci reste vrai pour toute étude locale $\Delta = \Delta_i = \{d_i, d_i\}$, $i=1, q$ puisque cette étude apparaît comme un cas particulier d'une étude classique où il n'y aurait que deux formes à reconnaître.

On se propose maintenant d'étendre ces résultats aux ensembles de décision $\Delta^* = \Delta \cup \{x\}$ et $\Delta = \mathcal{P}(M_D)$, et ceci toujours dans le cas où la procédure $\hat{\phi}$ minimise la probabilité d'erreur.

2 . ETUDE $\Delta^* = \{d_1, \dots, d_q, x\}$

On peut de la même façon que précédemment construire le tableau croisé des partitions $P_{\hat{\phi}}(\Omega)$ et $P_{\hat{\phi}^*}(\Omega)$.

$P_{\hat{\phi}}(\Omega) \backslash P_{\hat{\phi}^*}(\Omega)$	d_1		d_j		d_q	x	$\sum_j N_{ij}$
D_1	N_{11}		N_{1j}		N_{1q}	N_{1q+1}	$N_{1.}$
D_i	N_{i1}		N_{ij}		N_{iq}	N_{iq+1}	$N_{i.}$
D_q	N_{q1}		N_{qj}		N_{qq}	N_{qq+1}	$N_{q.}$
$\sum_i N_{ij}$	$N_{.1}$		$N_{.j}$		$N_{.q}$	$N_{.q+1}$	N

Le nombre de solutions conduisant à l'égalité des partitions $P_{\hat{\phi}}(\Omega)$ et $P_{\hat{\phi}^*}(\Omega)$ est égal à $(q+1)!$.

La procédure minimisant la probabilité d'erreur est donnée par /HABB,79/ :

$$\hat{\phi} : S \longrightarrow \Delta^*$$

$$m_S \longmapsto \hat{\phi}(m_S) = \begin{cases} d_i & \text{si } \begin{cases} \Pr(d_i) = \text{Max} (\Pr(d_j)) \\ d_j \\ \Pr(d_i) \geq a_i, a_i \in [0,1] \end{cases} \\ x & \text{si non} \end{cases}$$

avec $\Pr(d_j) = \Pr(D_j/S=m_S)$

On distingue alors les deux cas suivants :

$$\text{a) } \hat{\phi}(m_S)=d_i \iff \begin{cases} n_{ii} = \text{Max}_k (n_{ik}) \\ n_{ii}/\sum_{j=1}^q n_{ij} \geq a_i \end{cases} \iff \begin{cases} n_{ii} = \text{Max}_k (n_{ik}) \\ n_{ii} \geq a_i/(1-a_i) \sum_{\substack{j=1 \\ j \neq i}}^q n_{ji} \\ a_i \neq 1 \end{cases}$$

Pour les q premières colonnes, on en déduit les propriétés suivantes :

$$i = 1, q \quad \left\{ \begin{array}{l} N_{ii} \geq a_i/(1-a_i) \sum_{\substack{j=1 \\ j \neq i}}^q N_{ji} \\ N_{ii} = \text{Max}_k (N_{ki}) \end{array} \right.$$

Le nombre de solutions admissibles se réduit à $q+1$. Il reste à montrer que sur la colonne associée à la décision "x", on ne peut pas avoir $N_{iq+1} = N_i$ avec $N_{jq+1} = 0, \forall j=1, q, j \neq i$ et ceci quel que soit i .

$$\text{b) } \hat{\phi}(m_S) = x \iff \begin{cases} n_{ii} = \text{Max}_k (n_{ik}) & (1) \\ n_{ii} < a_i/(1-a_i) \sum_{\substack{j=1 \\ j \neq i}}^q n_{ji} & (2) \end{cases}$$

Pour que $N_{jq+1} = 0 \forall j \neq i$, il faut nécessairement que $n_{ji}=0$, et que $n_{ii} < 0$ d'après (2). Ce qui n'est pas possible.

On en conclura donc que pour $\Delta^* = \Delta \cup \{x\}$, l'égalité des partitions $P_{\hat{\phi}}(\Omega)$ et $P_{\hat{\phi}^*}(\Omega)$ conduit aussi à l'égalité des contenus sémantiques. Notons que la comparaison des partitions permet aussi de mesurer l'efficacité de discrimination puisque l'égalité des partitions ne peut avoir lieu que si la classe x de la partition $P_{\hat{\phi}^*}(\Omega)$ est vide.

3 . ETUDE GLOBALE : $\Delta = \mathcal{P}(M_D)$

De la même façon, on peut montrer que l'égalité des partitions entraîne forcément l'égalité des contenus sémantiques si la procédure d'estimation $\hat{\phi}$ est celle donnée au IV.3.2.a. En effet, la procédure ne retenant que les décisions les plus fiables, seuls les éléments N_{ii} du tableau sont susceptibles de prendre la valeur N_i . Par ailleurs, il est facile de montrer que lorsqu'une des classes de $\hat{\mathcal{P}}(\mathcal{Q})$ associée à une décision composée n'est pas vide, il y a plus d'un élément non nul sur la colonne associée à cette décision.

- ANNEXE IV -THEORIE DE L'INFORMATION

La Théorie de l'Information de Shannon a pour objet la définition et la mesure des quantités d'information, sans référence au contenu sémantique correspondant. Elle s'applique aussi bien à des variables qualitatives que quantitatives, et propose des indices sensibles à n'importe quel type de relation entre variable (linéaire, non linéaire, ...) /SHAN,48/, /ASH,65/. De plus, l'interprétation de ses indices est aisée /MILL,63/.

Soit \mathcal{S} , un ensemble de variables aléatoires, X et Y deux variables de \mathcal{S} qui prennent respectivement leurs modalités m_X , m_Y dans les ensembles finis M_X , M_Y , la quantité d'information commune aux deux variables X et Y est définie par :

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

où $H(X)$ est l'entropie de X, quantité mesurant son indétermination à priori,

$$H(X) = - \sum_{m_X \in M_X} \Pr(X = m_X) \log_2 \Pr(X = m_X)$$

avec $\Pr(X = m_X)$: probabilité à priori pour X prenne la modalité m_X

et $H(X,Y)$ est l'entropie conjointe de X, Y :

$$H(X,Y) = - \sum_{m_X \in M_X} \sum_{m_Y \in M_Y} \Pr(X = m_X, Y = m_Y) \log_2 \Pr(X = m_X, Y = m_Y)$$

$I(X,Y)$ peut aussi s'exprimer de la façon suivante :

$$I(X,Y) = H(Y) - H(Y/X) = H(X) - H(X/Y)$$

où $H(Y/X)$ est l'entropie conditionnelle de Y sachant X :

$$H(Y/X) = - \sum_{m_X \in M_X} \Pr(X=m_X) \sum_{m_Y \in M_Y} \Pr(Y=m_Y/X=m_X) \log_2 \Pr(Y=m_Y/X=m_X)$$

avec $\Pr(Y = m_Y/X = m_X)$: probabilité conditionnelle de $Y = m_Y$ sachant que $X = m_X$.

L'entropie résiduelle $H(Y/X)$ mesure la quantité d'information qu'il reste à connaître sur Y lorsque l'on possède déjà l'information fournie par X .

Ces indices peuvent être définis de la même manière pour des variables vectorielles, c'est à dire pour des parties ordonnées de \mathcal{S} . Les propriétés de monotonie et de sous-additivité suivantes sont vérifiées par l'entropie.

- Monotonie :

$$\forall A, B \in \mathcal{P}(\mathcal{S}) \quad \text{tel que } A \subseteq B \quad \Rightarrow H(A) \leq H(B)$$

- Sous-Additivité :

$$\forall A, B \in \mathcal{P}(\mathcal{S}) \quad H(A \cup B) \leq H(A) + H(B)$$

Dans le cadre d'une application au problème d'explication d'une variable (le diagnostic D) par d'autres (résultats d'exams : \mathcal{S}), on peut définir les indices suivants /TORO,82/ /STAR,84/ :

- Pourcentage Non Explicable : %NE

$$\%NE = H(D/\mathcal{S}) / H(D)$$

Cet indice donne une idée de la faisabilité du problème posé.

- Pourcentage Encore Non expliqué par S de la Partie Explicable ($S \in \mathcal{P}(\mathcal{S})$) : %ENEPE

$$\%ENEPE = \begin{cases} (H(D/S) - H(D/\mathcal{S})) / I(D:\mathcal{S}) & \text{si } I(D:\mathcal{S}) > 0 \\ 0 & \text{si } I(D:\mathcal{S}) = 0 \end{cases}$$

Cet indice permet d'apprécier la qualité du vecteur explicatif S .

- B I B L I O G R A P H I E -

- /ARQU,79/ : P.Y ARQUES
"Décisions et traitement du signal"
Masson, 1979
- /ASH,65/ : R. ASH
"Information Theory"
John Willey & Sons, 1965
- /AUVE,86/ : B. AUVERT, P. AEGERTER, E. BENILLOUCHE,
P. BOUTIN, G. DESVE, M.F LANDRE, D. BOS
"Tropicaid : un system expert sur ordinateur
portatif pour l'aide à la décision médicale dans
les pays en voie de développement"
Les systèmes experts et leurs applications
6^{eme} Journées Internationales d'Avignon, Avril 1986
- /AYGA,86/ : P. AYGALINC, M. STAROSWIECKI, O. DUBOIS
"Selection de variables explicatives pour
l'affinement et la confirmation des diagnostics
dans une aide à la décision médicale"
IASTED International Symposium and Exhibition
Toulouse, 18/20 Juin 1986
- /AYGA,86'/ : P. AYGALINC, M. STAROSWIECKI, O. DUBOIS
"Selection of explicative variables for aided
diagnosis"
Second Catalan International Symposium on Statistics
Barcelone 18/19 septembre 1986
- /BERN,71/ : G. BERNARD, M.L BESSON
"Douze méthodes d'analyse multicritère"
RIRO, V.3, 1971
- /BERT,75/ : P. BERTIER, P.M BOUROCHE
"Analyse des données multidimensionnelles"
Presse Universitaire de France, 1975
- /BENZ,80/ : J.P BENZECRI
"L'analyse de données"
DUNOD, 1980.
- /BONN,81/ : A. BONNET
"Applications de l'intelligence artificielle"
RAIRO, Informatique computer vol 15, n 4, p 325-341,
1981
- /BRON,84/ : T.A BROWN, J. KOPLONOWITZ
"The weighted nearest neibghbour. Rule for class
dependant simple size"
Pattern Recognition Vol 13, n 2, 1984.
- /CAIL,76/ : F.CAILLEZ, J.P PAGES
"Introduction à l'analyse des données"
Société de mathématiques appliquées et de Sciences
Humaines, 1976.

- /CAPO,85/ : R.M CAPOCELLI, I.J TANEJA
 "Generalised distance measures and error bounds"
 IEEE Proceeding, 1985
- /CHEN,76/ : CHEN
 "On information and distance measured error bounds
 and feature selection"
 Information Sciences Vol 10, pp 159-173, 1976
- /CONA,72/ : R.C CONANT
 "Detecting subsystems of complex System"
 IEEE Trans. on Systems, Man and Cybernetics.
 Vol SMC-2, n 4, pp 550-553, Septembre 1972.
- /CONA,76/ : R.C CONANT
 "Laws of Information wich govern Systems
 IEEE Trans. on Systems, Man and Cybernetics.
 Vol SMC-6, n 4, pp 240-255, Avril 1976.
- /COUV,81/ : M. COUVREUR
 "Application du traitement temps réel à la
 constitution d'une structure d'aide à la décision
 dans un service hospitalier"
 Thèse DI, UST Lille Flandres-Artois, Déc. 1981
- /COVE,67/ : T.M COVERT, P.E HART
 "Nearest neighbor pattern recognition"
 IEEE IT 13, n 1, pp 21-27, Janvier 1967
- /CRON,63/ : J.N CRONHOLM
 "A general method of obtaining exact sampling
 probabilities of the Shannon-Wiener measure of
 information H"
 Psychometrika, Vol 28, n 4, pp405-413. 1963
- /CUEN,73/ : M. CUENOD, J.L FATIO
 "Cours d'introduction aux méthodes d'identification"
 Ed. Albin Michel, 1973
- /DACU,82/ : DACHUNA-CASTELLE , DUFLO
 "Probabilités et Statistiques"
 Tome 1, Problème à temps fixe
 Masson,82
- /DAMI,84/ : P. DAMIANI H. MASSE
 "Etude de la consommation d'alcool à partir des
 distributions départementales de mortalité"
 Journées de statistique, Montpellier 21/24 Mai 1984
- /DARO,70/ : DAROCZY
 "Generalized information functions"
 Information and Control, Vol 16,1970
- /DOMB,76/ : F.T DE DOMBAL, F. GREMY
 "Decision making and medical care. Can information
 science help ?"
 North Holland Publishing Company, 1976

- /DONN,84/ : DONNA, T. ESTRIA
 "Emerge:a data driven medical decision making aid"
 Trans. in Pattern Analysis and Machine Intelligence
 Vol n 1 p87-91. 1984
- /DUDA,76/ : A. DUDANI
 "The distance weighted K nearest neighbour Rule"
 IEEE SMC-6, 1976.
- /DUFO,76/ : J. DUFOUR, G. GILLES, C. FOULARD
 "Analyse structurale et partition des systèmes
 dynamiques complexes à l'aide de la théorie de
 l'information. Etude de l'influence du nombre de
 classes"
 Compte rendu de l'académie des Sciences de Paris
 t.282, série A, pp 543-546, 8 mars 1976.
- /DUPO,86/ : D. DUPONT
 "Observation et aide à la décision, application à la
 conception d'un observateur numérique couplé à un
 système expert"
 Thèse 3^{eme} cycle, UST Lille Flandres Artois, 1986
- /EMPT,76/ : H. EMPTOZ
 "Informations utiles et Pseudoquestionnaires"
 Thèse 3^{eme} Cycle. Université de Lyon, Mars 1976
- /FACY,84/ : F. FACY, G. COVAERT, F. DAVIDSON
 "Analyse chronologique de données qualitatives en
 épidémiologie du petit enfant"
 Journées de statistique. Montpellier 21/24 Mai 1984
- /FOUC,82/ : T. FOUCART
 "Analyse factorielle : Méthodes et Programmes"
 Masson, 1982
- /FRIE,75/ : J.H FRIEDMAN, J.L BENTLEY, R. ARI FINKEL
 "An algorithm for finding nearest neighbors"
 IEEE C-24, pp 1000-1006, 1975
- /FU,68/ : K.S FU
 "Sequential methods in Pattern Recognition and
 Machine Learning"
 Academic Press, 1968
- /GREN,85/ : A. GRENIER
 "Extension de méthodes de reconnaissance statistique
 de formes pour l'applicabilité au diagnostic"
 Thèse DI, Université de technologie de Compiègne
 Mars 1985
- /GUIA,77/ : S. GUIASU
 "Information theory with applications"
 Mc-Graw Hill, London, 1977.

- /GUIG,77/ : G.L GUIGOU
"Méthodes Multidimensionnelles : Analyse de données
et choix à critères multiples"
Dunod, 1977
- /HABB,79/ : J.D.F HABBEMA, J.HERMANS
"discriminant analysis : measurement of performance
non parametric discrimination, and selection of
variables"
Cours de la commission des communautés européennes
Ed. INRIA, pp 79-124, Mars 1979
- /HABB,81/ : J.D.F HABBENA, G.J GELKE
"A Computer program for selection of variables in
diagnostic and prognostic problems"
Computer Programs in Biomedicine, pp 251-270
North-Holland Biomedical Press, Février 1981
- /JACK,80/ : E. JACQUET-LAGREZ, B. ROY
"Aide à la décision multicritère et systèmes
relationnels des préférences"
Cahier de Lamsade n 34, Juillet 1980
- /JAMB,78/ : M. JAMBU
"Classification automatique pour l'analyse des
données"
Dunod, 1978
- /KANA,70/ : L.N KANAL
"Patterns in Pattern Recognition"
IEEE Trans. on Inform. Theory IT -20 pp 697-722
1970
- /KESK,84/ : A. KESKES, Y. LECHEVALLIER
"Recherche de Facies Sismiques par une méthode de
partitionnement sous contrainte d'ordre"
Journées de statistique. Montpellier 21/24 Mai 1984
- /KLIR,69/ : G.J KLIR
"An approach to general systems theory"
Van Nostrand Reinold, New York, 1969
- /KLIR,75/ : G.J KLIR
"On the representation of Activity Arrays"
International journal of general systems, vol 2,
n 3, pp 149-168, 1975
- /KLIR,76/ : G.J KLIR
"Identification of generative structures in
Empirical data"
International journal of general systems, vol 3
pp 89-104, 1976

- /KLIR,77/ : G.J KLIR
 "On the problem of computer-aided structure identification : some experimental observations and resulting guidelines"
 International journal of Man-Machine Studies
 Vol 9, pp 593-628. 1977
- /KULI,70/ : C.A KULIKOWSKI
 "Pattern Recognition Approach to Medical Diagnosis"
 IEEE Trans. Syst. Sci. Cybern., Vol SSC-6, pp 83-89
 1970
- /LAUR,85/ : R. LAURENT
 "Modelisation et Identification des systèmes complexes"
 Thèse d'état Es-Sciences physique, UST Lille
 Flandres-Artois, n 646, Mars 1985.
- /LEBA,82/ : L. LEBART, A. MORINEAU, J.P FENELON
 "Traitements de données statistiques"
 Dunod 1982
- /LECE,83/ : C.M LECERF
 "Conception d'une structure logiciel redondante en vue d'améliorer la sureté de prise de décision en temps réel. Application à l'analyse automatique de signaux biologiques"
 Thèse DI, UST Lille Flandres-Artois, 1983.
- /LECH,76/ : Y. LECHEVALLIER
 "Classification automatique optimale sous contrainte d'ordre total"
 INRIA, Rapport de recherche n 200, 1976
- /LOPE,77/ : R. LOPEZ DE MANTARAS BADIA
 "Auto apprentissage d'une partition. Application au classement itératif des données multidimensionnelles"
 Thèse doctorat de Spécialité. Université Paul Sabatier, Toulouse. n 1998. Juin 1977.
- /MARF,83/ : C. MARFAING PALLIER
 "Contribution à l'analyse structurale des systèmes complexes par le biais de concepts en théorie de l'information"
 Thèse DI, INSA Toulouse, n 110, Déc. 1983
- /MATH,82/ : J. MATHIOT
 "Informatique et médecine"
 Instantanés Médicaux, Encyclopédie Médico-Chirurgicale, 1982-85
- /MILL,63/ : G.A MILLER "What is Information measurement"
 American Psychologist, vol 8, n 2, pp 50-51,1963

- /MINO,80/ : M. LE MINOR, A. ALPEROWITCH, J. LELLOUCH
 "Prescription séquentielle des examens
 complémentaires en médecine : Etude par simulation
 de quelques heuristiques"
 RAIRO Recherche Opérationnelle, vol n 14, pp 83-85
 1980
- /MORE,86/ : A. MOREAU, D. WILLAEYS
 "Une méthode de représentation de l'imprécis dans
 les systèmes Experts"
 Les systèmes Experts et leurs applications, 6^{ème}
 Journées Internationales d'Avignon, Avril 1986
- /MOUR,81/ : MOURMOURIS
 "Les méthodes multicritères comme outil d'aide à la
 décision dans les problèmes de transport"
 Thèse 3^{ème} Cycle. I.N.P.G , Janvier 1981
- /NAKA,77/ : J.P NAKACHE, P. LORENTE, J.P BENZECRI, J.F CHASTANG
 "Aspects pronostiques et et thérapeutiques de
 l'infactus myocardique aigu compliqué d'une
 défaillance sévère de la pompe cardiaque.
 Application des méthodes de discrimination"
 Les cahiers de l'Analyse des Données, Vol II, 1977
 n 4, pp 415-434
- /PAGE,80/ : A. PAGES, M. CONDRAN
 "Fiabilité des systèmes"
 Ed. Eyrolles, 1980
- /PASS,86/ : G. PASS
 "Consistant evaluation of uncertain reasoning
 Systems"
 Les systèmes-Experts et leurs applications, 6^{ème}
 Journées Internationales d'Avignon Avril 1986
- /PCHE,77/ : PCHENITCHNY. Y.DANILINE
 "Méthodes numériques dans les problèmes d'extrémum"
 Ed. MIR Moscou. 1977
- /PICA,72/ : C.F PICARD
 "Graphes et Questionnaires"
 Ed Gauthier Villars Paris 1972.
- /PINS,81/ : S. PINSON
 "Représentation des connaissances dans les systèmes
 Experts"
 RAIRO Informatique/ Computer Science vol 15, n 4,
 pp343-367, 1981
- /PONO,81/ : A.F PONOMARENKO
 "Information Theory and Identification"
 EUT Report 81-E-122
 Eindhoven Univ. Technol. the Netherlands.
- /POVY,75/ : L. POVY
 "Identification des processus"
 Dunod, 1975

- /RAJA,79/ : C.V RAJAGOPALAN
 "Séquentation des suites temporelles. Application à l'extraction heuristique des paramètres fondamentaux du cardiorespirogramme en pathologie néo-natale"
 Thèse DI. UST Lille Flandres-Artois, Mars 1979
- /RICH,71/ : J. RICHELET, A. RAULT, P. POULIGUEN
 "Identification des processus par la méthode du modèle"
 Gordon & Breach, 1971
- /RICH,75/ : M. RICHETIN
 "Analyse structurale des systèmes complexes en vue d'une commande hiérarchisée"
 Thèse d'état, Université Paul Sabatier Toulouse, n 674, Juillet 1975
- /RICH,80/ : M. RICHETIN, G. RIVES, M. NARANJO
 "Algorithme rapide pour la détermination des K plus proches voisins"
 RAIRO Informatique / Computer Science, Vol 14, n 4 pp 369-378, 1980
- /ROMB,82/ : M. ROMBAULT
 "Détermination des relations fonctionnelles dans un système de variables par l'analyse informationnelle des données"
 Mémoire de D.E.A, UST Lille Flandres-Artois, 1982
- /ROME,73/ : J.M ROMEDER
 "méthodes et programmes d'analyse discriminante"
 Dunod,1973
- /ROY,70/ : B. ROY
 "Décisions avec critères multiples. Problèmes et méthodes"
 7^{ème} symposium de La Haye, 1970
- /SAPO,78/ : G. SAPORTA
 "Théorie et Méthodes de la Statistique"
 Paris-Technip : 1978
- /SBAI,83/ : M.SBAI
 "Analyse structurale des systèmes complexes : Méthodes d'explication et de partition"
 Thèse 3^{ème} Cycle, UST Lille Flandres-Artois, 1983
- /SHAN,48/ : C.E SHANNON
 "A mathematical theory of communication"
 Bell system technical journal, Vol 27, 1948
- /SIMO,84/ : J.C SIMON
 "La reconnaissance des formes par algorithmes"
 Masson, 1984

- /SMOL,82/ : A. SMOLARZ
 "Développement d'une méthode de classification avec rejet, appliquée au diagnostic de fonctionnement du coeur d'une centrale nucléaire. Traitement des signaux thermiques issus d'un essai de simulation hors pile"
 Thèse DI, Université de Technologie de Compiègne, Juillet 1982
- /SPAT,73/ : A. SPATARU
 "Théorie de la transmission de l'information. Tome 2 Codes et décisions"
 Masson & Cie, 1973
- /STAR,84/ : M. STAROSWIECKI
 "Analyse structurale des systèmes complexes"
 RAIRO, Automatique : Systems analysis and Control Vol 18, n 2, 1984
- /STAR,86/ : M. STAROSWIECKI, P.AYGALINC, O. DUBOIS
 "Selection de variables explicatives et prescription d'examen complémentaires pour une aide à la décision médicale"
 Conférence Internationale IPMU, Paris 30 Juin/ 4 Juillet 1986
- /TITR,79/ : Y. TITREN
 "Analyse de données pour l'aide au diagnostic en pathologie néo-natale"
 Thèse DI, n 233, UST Lille Flandres-Artois, 1979
- /TERR,76/ : M. TERRENOIRE, D.TOUNISSOUS
 "Pseudoquestionnaires et Informations"
 Decision making and medicinal Care. North Holland Publishing Company, 1976
- /TORO,82/ : V. TORO DE CORDOBA
 "Contribution à l'analyse structurale des systèmes complexes à l'aide de l'entropie et de ses généralisations"
 Thèse 3^{ème} Cycle, UST Lille Flandres-Artois, 1982
- /VASS,82/ : C. VASSEUR
 "La notion d'événements dans les systèmes dynamiques: Détection, classification temps réel et application au concept d'une instrumentation distribuée"
 Docteur Es-Sciences Phys. UST Lille Flandres-Artois Juin 1982.
- /WART,74/ : J. WARTAK
 "Mathematical model for medical diagnosis"
 Comput. Biol.Med Pergamon Press, Vol 4, pp79-84, 1974

- /WONG,86/ : S.K.M WONG, W. ZIARKO
"On adaptative decision support system based on the
probabilistic approximated classification"
Les systèmes Experts et leurs applications, 6^{ème}
Journées Internationales d'Avignon Avril 1986
- /ZIGH,84/ : A. ZIGHED, M.TERRENOIRE, D. TOUNISSOUS
"Système interactif pour un processus
d'interrogation non arborescent"
Journées de Statistique , Montpellier 21/24 Mai 1984

