

50376  
1981  
220

50376  
1981  
220

# THÈSE

présentée à

L'UNIVERSITÉ DES SCIENCES ET TECHNIQUES DE LILLE

pour l'obtention du grade de

**DOCTEUR ES-SCIENCES PHYSIQUES**

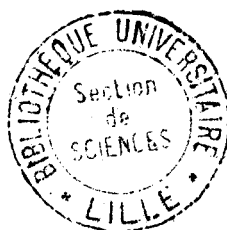
par

Jack-Gérard POSTAIRE

Docteur-Ingénieur

**OPTIMISATION DU PROCESSUS DE CLASSIFICATION  
AUTOMATIQUE PAR ANALYSE DE LA CONVEXITE  
DES FONCTIONS DE DENSITE DE PROBABILITE**

**APPLICATION A LA RECONNAISSANCE DE LA FORME  
DES ARBRES EN DENDROMETRIE.**



Soutenu le 14 septembre 1981 devant la Commission d'Examen :

MM.	P. VIDAL	Président et Rapporteur
	C. FOULARD	Rapporteur
	F. LHOTE	Rapporteur
	M. AGUILAR	Examineur
	R. TOMASSONNE	Examineur
	J.M. TOULOTTE	Examineur
	S. WEGRZYN	Examineur
	O. M'HIRIT	Examineur

## AVANT -PROPOS

Le travail de recherche présenté dans ce mémoire de thèse a débuté au Centre d'Automatique de l'Université de Lille 1, animé par Monsieur le Professeur P. VIDAL. Je le remercie très vivement pour avoir suivi, pendant de nombreuses années, le développement de mes recherches. Qu'il trouve ici l'expression de ma grande reconnaissance pour son rôle stimulant qui a permis la valorisation de ce travail et pour son attention toujours soutenue, tant sur le plan humain que sur le plan scientifique.

C'est avec mon collègue C. VASSEUR que je me suis familiarisé avec les problèmes de classification automatique et, ensemble, nous avons jeté les premières bases de cette étude. Je tiens à le remercier pour ce travail d'équipe et cette collaboration qui sont à l'origine de bien des aspects des résultats présentés dans ce mémoire.

C'est au cours d'une mission de coopération à la Faculté des Sciences de Rabat, Maroc, que ce travail a été poursuivi. Je suis particulièrement reconnaissant envers Monsieur le Professeur M. NAJIM qui, en m'accueillant dans le Laboratoire d'Electronique et d'Etude des Systèmes Automatiques, a mis généreusement à ma disposition des moyens matériels tout à fait exceptionnels.

Je sais grè à Monsieur O. M'HIRIT, Ingénieur Forestier et Professeur à l'Ecole Nationale Forestière d'Ingénieurs de Salé, Maroc, pour la confiance qu'il m'a prodiguée lorsque je lui ai proposé d'introduire, dans les problèmes de gestion des cédraies du Maroc, les méthodes de classification que j'avais mises au point. Je le remercie pour sa participation au jury de thèse.

L'enjeu de cette étude n'a pas échappé à la Direction des Eaux et Forêts et de la Conservation des Sols du Ministère de l'Agriculture et de la Réforme Agraire du Royaume du Maroc que je remercie également pour son aimable collaboration.

Que Monsieur C. FOULARD, Professeur à l'Université de Grenoble, trouve ici l'expression de ma sincère reconnaissance pour l'intérêt qu'il a porté à ce travail. Je tiens à le remercier pour ses critiques enrichissantes et pour avoir bien voulu être l'un des rapporteurs de cette thèse.

Je présente à Monsieur F. LHOTE, Professeur à l'Université de Besançon, l'expression de ma profonde gratitude pour avoir accepté de porter un jugement sur ce travail et lui suis reconnaissant d'être l'un des rapporteurs de ce mémoire.

Je remercie vivement Monsieur M. AGUILARD, Maître de Recherche au C.N.R.S./L.A.A.S. de Toulouse, pour l'intérêt qu'il a témoigné pour ce travail et pour ses incitations à approfondir certains aspects de cette étude. Je lui suis très reconnaissant d'être l'un des examinateurs de cette thèse.

Je présente à Monsieur R. TOMASSONNE, Directeur du Département de Biométrie au C.N.R.Z. de Jouy-en-Josas, l'expression de ma profonde reconnaissance pour son examen attentif de ce travail et son jugement critique. Qu'il trouve ici l'expression de mes sincères remerciements pour avoir accepté d'être examinateur dans le jury de cette thèse.

Je remercie très vivement Monsieur J.-M. TOULOTTE, Professeur à l'Université de Lille 1, pour l'intérêt tout particulier qu'il a manifesté pour ce travail et pour avoir bien voulu être l'un des examinateurs de cette thèse.

Que Monsieur S. WEGRZYN, Professeur à l'Université de Gliwice, Pologne, qui a accepté de juger le contenu de ce mémoire et de participer, comme examinateur, au jury de cette thèse, trouve ici l'expression de mes plus vifs remerciements.

Je ne saurais terminer cet avant-propos sans remercier Monsieur J. HOUZE du Centre d'Automatique de l'Université de Lille, pour l'amicale diligence avec laquelle il a participé à la confection de ce mémoire.

Mes remerciements vont enfin à tous mes amis avec lesquels j'ai eu la chance de travailler, tant au Centre d'Automatique de l'Université de Lille qu'à la Faculté des Sciences de l'Université Mohamed V de Rabat.

## SOMMAIRE

<b>CHAPITRE I</b>	<b><u>LA CLASSIFICATION AUTOMATIQUE</u></b>	<b>I.1.</b>
I.1.	LA CLASSIFICATION: UNE DEMARCHE SCIENTIFIQUE FONDAMENTALE	
I.1.1.	La classification et les sciences de la nature	I.1.
I.1.2.	Critères et difficultés de la classification naturelle	I.2.
I.2.	L'AUTOMATISATION DU PROCESSUS DE CLASSIFICATION	I.3.
I.3.	CLASSIFICATION ET THEORIE DE LA DECISION	I.5.
I.3.1.	Notion de taux d'erreur	I.5.
I.3.2.	Classification optimale	I.5.
I.4.	L'ANALYSE DES MELANGES	I.7.
I.4.1.	Analyse des mélanges monovariabes	I.7.
I.4.2.	Techniques d'apprentissage bayésien	I.7.
I.4.3.	Méthode d'estimation par maximum de vraisemblance	I.8.
I.4.4.	Approches diverses	I.8.
I.4.5.	Limitation des méthodes d'analyse des mélanges	I.9.
I.5.	CONVEXITE ET OPTIMISATION DU PROCESSUS DE CLASSIFICATION	I.10.

## **PARTIE I**      **CONVEXITE DES FONCTIONS DE DENSITE**

<b>CHAPITRE II</b>	<b><u>ANALYSE DE LA CONVEXITE DES FONCTIONS MULTIVARIABLES</u></b>	<b>II.1.</b>
II.1.	INTRODUCTION	II.1.
II.2.	FONCTIONS CONVEXES	II.2.
II.2.1.	Domaines convexes	II.2.
II.2.2.	Fonctions convexes	II.3.
II.2.3.	Fonctions localement convexes	II.4.
II.2.4.	Continuité des fonctions localement convexes	II.6.
II.3.	DOMAINES D'OBSERVATION DES FONCTIONS CONVEXES	II.7.
II.3.1.	Définition	II.7.
II.3.2.	Propriété	II.7.

II.4.	VALEUR MOYENNE DES FONCTIONS LOCALEMENT CONVEXES	II.10.
II.4.1.	Propriété fondamentale	II.10.
II.4.2.	Extremum local de la valeur moyenne	II.15.
II.5.	CONCLUSION	II.16.

**CHAPITRE III** DETERMINATION DE LA CONVEXITE LOCALE DES FONCTIONS DE DENSITE A PARTIR DES OBSERVATIONS

III.1.	INTRODUCTION	III.1.
III.2.	ESTIMATION DE LA VALEUR MOYENNE DES FONCTIONS DE DENSITE	III.2.
III.2.1.	Principe de base	III.2.
III.2.2.	Propriétés asymptotiques de l'estimateur	III.3.
III.3.	TEST DE CONVEXITE	III.4.
III.4.	ALGORITHME RAPIDE D'ESTIMATION DES FONCTIONS DE DENSITE	III.7.
III.4.1.	Algorithme d'estimation conventionnel	III.7.
III.4.2.	Détermination des hypercubes non vides	III.9.
III.4.3.	Estimation rapide des fonctions de densité	III.13.
III.5.	COMPARAISON DES ALGORITHMES D'ESTIMATION RAPIDE ET CONVENTIONNEL	III.16.
III.5.1.	Analyse des algorithmes	III.16.
III.5.2.	Performances des algorithmes	III.18.
III.6.	ALGORITHME DE TEST DE LA CONVEXITE LOCALE DES FONCTIONS DE DENSITE	III.23.
III.7.	CONCLUSION	III.25.

**PARTIE II** OPTIMISATION DU PROCESSUS DE CLASSIFICATION

**CHAPITRE IV** DETERMINATION DES PARAMETRES D'UNE DISTRIBUTION NORMALE PAR ANALYSE DE LA CONVEXITE DE SA FONCTION DE DENSITE

IV.1.	INTRODUCTION	IV.1.
IV.2.	LES FONCTIONS DE DENSITE DE PROBABILITE NORMALES	IV.1.
IV.3.	CONVEXITE DES FONCTIONS DE DENSITE NORMALES	IV.3.
IV.3.1.	Diagonalisation de la matrice de covariance	IV.3.
IV.3.2.	Condition de convexité	IV.4.
IV.3.3.	Domaine de concavité	IV.6.

IV.4.	PROPRIETES GEOMETRIQUES DU DOMAINE DE CONCAVITE	IV.8.
IV.5.	IDENTIFICATION DES DISTRIBUTIONS NORMALES	IV.11.
IV.5.1.	Détermination du domaine de concavité	IV.11.
IV.5.2.	Détermination des paramètres de la distribution	IV.12.
IV.6.	EXEMPLES D'APPLICATION	IV.17.
IV.7.	CONCLUSION	IV.22.

CHAPITRE V	<u>ANALYSE DES MELANGES GAUSSIENS ET CLASSIFICATION</u>
------------	---

V.1.	INTRODUCTION	V.1.
V.2.	LES MELANGES GAUSSIENS	V.1.
V.3.	APPROXIMATION DES FONCTIONS DE DENSITE CONDITIONNELLES	V.4.
V.4.	DETERMINATION DES PROBABILITES A PRIORI	V.7.
V.4.1.	Approximation des probabilités a priori	V.7.
V.4.2.	Détermination des valeurs approchées des probabilités a priori	V.8.
V.5.	OPTIMISATION DU PROCESSUS DE CLASSIFICATION	V.9.
V.6.	ANALYSE DES PERFORMANCES DE LA PROCEDURE DE CLASSIFICATION	V.11.
V.6.1.	Analyse de données bidimensionnelles	V.12.
V.6.2.	Analyse de données multidimensionnelles	V.25.
V.7.	CONCLUSION	V.29.

PARTIE III	APPLICATIONS PARTICULIERES DU CONCEPT DE CONVEXITE EN CLASSIFICATION AUTOMATIQUE
------------	---

CHAPITRE VI	<u>OPTIMISATION DU PROCESSUS DE CLASSIFICATION PAR ANALYSE DE LA CONVEXITE DES DENSITES MARGINALES DE PROBABILITE</u>
-------------	---

VI.1.	INTRODUCTION	VI.1.
VI.2.	POSITION DU PROBLEME	VI.2.
VI.3.	ANALYSE DES DENSITES MARGINALES DE PROBABILITE	VI.3.
VI.3.1.	Convexité des densités marginales	VI.3.
VI.3.2.	Implantation de la procédure	VI.6.

VI.4.	DETERMINATION DES PARAMETRES D'UN MELANGE MULTIVARIABLE	VI.8.
VI.4.1.	Détermination des probabilités a priori	VI.8.
VI.4.2.	Détermination des vecteurs moyenne et des matrices de covariance	VI.12.
VI.4.3.	Classification	VI.12.
VI.5.	ANALYSE GLOBALE - ANALYSE LOCALE	VI.13.
VI.6.	PERFORMANCES DE LA PROCEDURE	VI.16.
VI.7.	CONCLUSION	VI.23.

<b>CHAPITRE VII</b>	<b>RECHERCHE DE GROUPEMENTS PAR DETECTION DES MODES</b>	VII.1.
VII.1.	INTRODUCTION	VII.1.
VII.2.	LA RECHERCHE DE GROUPEMENTS	VII.2.
VII.2.1.	Approche métrique	VII.2.
VII.2.2.	Approche statistique	VII.4.
VII.3.	CONVEXITE ET DETECTION DES MODES	VII.4.
VII.3.1.	Détection des modes par analyse de la convexité des fonctions de densité	VII.4.
VII.3.2.	Détection des modes par analyse de la convexité des densités marginales	VII.5.
VII.4.	RECHERCHE DE GROUPEMENTS RESULTATS EXPERIMENTAUX	VII.6.
VII.4.1.	Recherche de groupements par analyse de la convexité des fonctions de densité	VII.6.
VII.4.2.	Recherche de groupements par analyse de la convexité des densités marginales	VII.10.
VII.5.	CONCLUSION	VII.16.

PARTIE IV	APPLICATION DES TECHNIQUES DE RECONNAISSANCE DES FORMES ET DE CLASSIFICATION AUTOMATIQUE A LA CONSTRUCTION DES TABLES DE CURAGE FORESTIERES
-----------	---

<b>CHAPITRE VIII</b>	<b>ANALYSE MORPHOMETRIQUE DU CEDRE DU MAROC</b>	VIII.1.
VIII.1.	L'ESTIMATION DES PESSOURCES FORESTIERES	
VIII.1.1.	Objectifs des inventaires forestiers	VIII.1.
VIII.1.2.	Estimation du volume brut sur pied	VIII.2.
VIII.1.3.	Etablissement des tarifs de cubage	VIII.3.



VIII.2.	CARACTERISATION DE LA FORME DES ARBRES	VIII.3.
VIII.3.	LES DONNEES : LES CEDRES DU MAROC	VIII.8.
VIII.4.	ANALYSE DE LA FORME DES CEDRES	VIII.11.
VIII.4.1.	Polymorphisme des cèdres dans chaque massif forestier	VIII.11.
VIII.4.2.	Stabilité de la forme des cèdres	VIII.12.
VIII.5.	CONCLUSION	VIII.17.

## CHAPITRE IX

### INFLUENCE DE LA FORME DES ARBRES SUR LA PRECISION DES TABLES DE CUBAGE FORESTIERES

IX.1.

IX.1.	PRECISION DES TARIFS DE CUBAGE ET ECHANTILLONNAGE	IX.1.
IX.2.	LES TARIFS D'AMENAGEMENT	IX.2.
IX.2.1.	Construction des tarifs de cubage à une à une entrée	IX.2.
IX.2.2.	Critères de précision	IX.3.
IX.3.	TECHNIQUES D'ECHANTILLONNAGE	IX.4.
IX.3.1.	Choix du type d'échantillonnage	IX.4.
IX.3.2.	Taille de l'échantillon	IX.5.
IX.4.	INFLUENCE DE LA STABILITE DE LA FORME DES CEDRES SUR LA PRECISION DES TARIFS DE CUBAGE	IX.6.
IX.5.	INFLUENCE DU POLYMORPHISME DES CEDRES SUR LA PRECISION DES TARIFS DE CUBAGE	IX.13.
IX.6.	CONCLUSION	IX.15.

## CONCLUSION GENERALE

## LISTE DES PRINCIPALES NOTATIONS

### PARTIE I

$n$	dimension des données
$q$	taille de l'échantillon
$R_0^n$	espace de représentation des données
$R_1^n$	espace de représentation des données normalisées
$X = [x_1, \dots, x_i, \dots, x_n]^T$	point de l'espace $R_0^n$
$X' = [x'_1, \dots, x'_i, \dots, x'_n]^T$	point de l'espace $R_1^n$
$X_{i,j} = [x_{1,j}, \dots, x_{i,j}, \dots, x_{n,j}]^T$	observation
$D_r(X)$	boule unité centrée en $X$
$S_r(X)$	sphère unité centrée en $X$
$D(X, \alpha)$	domaine d'observation centré en $X$
$V\{D(X, \alpha)\}$	volume du domaine d'observation centré en $X$
$f(X)$	fonction de densité de probabilité
$\rho\{D(X, \alpha)\}$	valeur moyenne de $f(X)$ sur $D(X, \alpha)$
$\hat{\rho}\{D(X, \alpha)\}$	valeur estimée de $\rho\{D(X, \alpha)\}$
$H[f(X)]$	hessien de $f(X)$

### PARTIE II

$\bar{Y} = [\bar{x}_1, \dots, \bar{x}_i, \dots, \bar{x}_n]^T$	vecteur-moyenne
$\Sigma = [\sigma_{ij}]$	matrice de covariance
$\lambda_i, i=1, 2, \dots, n$	valeurs propres de $\Sigma^{-1}$
$D$	domaine de concavité d'une fonction de densité normale
$D^\circ$	domaine modal d'une fonction de densité normale
$\hat{P}$	modèle hyperellipsoïdal du domaine modal $D^\circ$
$C_k$	kème classe, $k = 1, 2, \dots, K$
$\bar{x}_k$	vecteur-moyenne de la classe $C_k$
$\Sigma_k$	matrice de covariance de la classe $C_k$
$p(X C_k)$	fonction de densité conditionnelle de la classe $C_k$
$P(C_k)$	probabilité a priori de la classe $C_k$

$P(C_k X)$	probabilité a posteriori de la classe $C_k$
$\hat{X}_k$	valeur approchée du vecteur-moyenne de la classe $C_k$
$\hat{\Sigma}_k$	valeur approchée de la matrice de covariance de la classe $C_k$
$\hat{P}(C_k)$	valeur approchée de la probabilité a priori de la classe $C_k$
$D_k$	domaine de concavité de la classe $C_k$
$D_k^o$	domaine modal du mélange associé à la classe $C_k$
$\hat{D}_k$	modèle hyperellipsoïdal du domaine modal $D_k^o$
$\alpha_n = 1/N_n$	pas de discrétisation de l'espace normalisé $R_1^n$

### PARTIE III

$v_{i,k}$	écart-type de la distribution du ième caractère de la classe $C_k$
$f_i(x_i)$	densité marginale de la fonction de densité $f(X)$
$p_i(x_i C_k)$	densité marginale conditionnelle de la fonction de densité conditionnelle $p(X C_k)$
$d_{i,k}$	segment de concavité de la fonction de densité $p_i(x_i C_k)$
$\hat{d}_{i,k}$	segment de concavité de la fonction de densité $f_i(x_i)$
$\mathcal{D}_k$	domaine caractéristique de la classe $C_k$
$\hat{\mathcal{D}}_k$	domaine approché du domaine caractéristique $\mathcal{D}_k$
$\hat{a}_k^{(i)}$	kème fonction de décision à la ième itération

### PARTIE IV

$D_S$	diamètre à la souche (en mètres)
$D_{130}$	diamètre à hauteur d'homme (1,30m) (en mètres)
$D_{H/2}$	diamètre à mi-hauteur (en mètres)
$H$	hauteur totale (en mètres)
$\delta$	coefficient de décroissance
$\epsilon$	coefficient d'empatement
$\theta$	angle de défilement (en grades)
$f$	coefficient de forme
$F_{i,j}$	prototype de forme. $i$ : indice de forme $j$ : indice de massif forestier

$\bar{F}_i$	forme-type
$\bar{e}$	moyenne des erreurs de cubage sur un échantillon de test
$\sigma_e$	écart-type de la distribution des erreurs de cubage sur un échantillon de test
$T_C$	tarif de cubage général pour l'ensemble forestier C
$T_A$	tarif de cubage spécifique obtenu par échantillonnage simple, propre à l'ensemble forestier A
$T_B$	tarif de cubage spécifique obtenu par échantillonnage simple, propre à l'ensemble forestier B
$T'_A$	tarif de cubage spécifique obtenu par échantillonnage stratifié, propre à l'ensemble forestier A
$T'_B$	tarif de cubage spécifique obtenu par échantillonnage stratifié, propre à l'ensemble forestier B

## C H A P I T R E I

## LA CLASSIFICATION AUTOMATIQUE

I.1. LA CLASSIFICATION : UNE DEMARCHE SCIENTIFIQUE FONDAMENTALEI.1.1. LA CLASSIFICATION ET LES SCIENCES DE LA NATURE

Jusqu'à une époque relativement récente, la classification, considérée comme méthode scientifique, a été étroitement liée aux sciences biologiques. C'est du 18ème siècle, après l'ère des collections, des monographies et de l'accumulation des observations que datent les premiers essais de classification scientifique du règne animal et du règne végétal. Devant la multitude des êtres observés, les grands zoologistes (Cuvier) et botanistes (Linné, les Jussieu) tentent d'enfermer l'ordre de la nature dans un système permettant de classer tous les êtres vivants. Les techniques de classification en vigueur à cette époque procèdent par dichotomies successives. Un groupement est divisé en deux sous-groupes: l'un constitué des individus qui possèdent tel caractère, l'autre constitué des individus qui ne le possèdent pas. De cette démarche dichotomique est né le grand principe de subordination des caractères que l'on attribue généralement à Linné. En effet, certains caractères apparaissent comme hiérarchiquement élevés car il faut toujours commencer par les rechercher pour situer un individu dans un groupe d'espèces. Par contre, d'autres caractères ne doivent être considérés que pour achever le classement de l'individu à l'intérieur du groupe d'espèces auquel il appartient.

La grande critique faite à cette procédure, encore utilisée aujourd'hui, réside dans le choix des critères et de leur hiérarchie qui, sinon arbitraire, peut paraître artificiel. En effet, en prenant pour critère un seul caractère à chaque

niveau de classement, consciemment ou non, le naturaliste en fait le choix selon une certaine finalité. Celle-ci est en général la facilité et la rapidité de la détermination de l'espèce de tout individu inconnu /33/.

Cette hiérarchie des caractères, reflet des connaissances scientifiques, mais également d'habitudes, d'intuitions sinon de conceptions philosophiques a bien sûr évolué depuis son introduction par Linné. Après les tentatives d'organisation fondées sur l'observation de la morphologie des individus, la classification s'adapte aux nouvelles acquisitions de la génétique et débouche sur un réseau de plus en plus complexe entre les espèces. Au 19ème siècle, le transformisme vient donner un sens nouveau à la classification naturelle qui, comme le fait remarquer Darwin, s'identifie à l'arbre généalogique des formes vivantes.

#### I.1.2. CRITERES ET DIFFICULTES DE LA CLASSIFICATION NATURELLE

L'histoire montre comment la classification a longtemps été au service des sciences de la nature, ses principes dépendant ainsi étroitement des acquis de celles-ci. Le chercheur doit pourtant se garder d'orienter une classification en fonction de ce qu'il pense être l'organisation sous-jacente des individus étudiés. La classification doit être considérée comme un processus de connaissance objectif. Démarche exploratoire, elle doit permettre, sur la seule base des faits observables, de découvrir l'ordre du réel caché derrière la prolifération des détails.

Ouvrant une voie plus objective, Ray /31/, dès la fin du 17ème siècle, s'attache à définir la notion de type par appréciation des similitudes entre individus. L'idée est reprise cinquante ans plus tard par Buffon: "Le seul moyen de faire une méthode instructive et naturelle, c'est de mettre ensemble les choses qui se ressemblent et de séparer celles qui diffèrent les unes des autres." /5/. Le grand naturaliste français n'exploita pas cette idée et c'est à Adanson /1/ que revient le mérite d'avoir le premier proposé une alternative à la classification linnéenne. Désirant s'affranchir des postulats de hiérarchisation qu'il considère comme subjectifs, il tente de regrouper les individus d'après leurs seuls degrés de ressemblance en utilisant des

caractères aussi nombreux que possible. En 1757, il écrit: "Je me contenterai de rapprocher les objets suivant le plus grand nombre de degrés de leurs rapports et de leurs ressemblances." Postulant que tous les caractères observés ont la même importance, il élabore différents classements artificiels par dichotomies successives, chacun d'eux résultant d'une hiérarchie des caractères particulière et totalement arbitraire. De la confrontation des différents résultats devait apparaître automatiquement le classement final reposant sur la totalité des caractères observés. Dans l'esprit Adanson, ce n'était que dans une seconde phase, une fois révélés les groupements d'individus, que les résultats pouvaient être exploités pour suggérer des idées ou des interprétations.

On peut se demander pourquoi les idées émises par Buffon et la tentative empirique de Adanson restèrent sans lendemain. En fait, la voie ouverte par ces deux naturalistes nécessite le traitement de vastes ensembles de données, car chaque individu de l'échantillon étudié est représenté par des caractères dont le nombre peut être important. S'il est possible de quantifier ces caractères, les individus peuvent être représentés par des points dans un espace euclidien multidimensionnel, chaque coordonnée étant associée à un caractère. Mais l'homme reste prisonnier de son univers tri-dimensionnel et toute synthèse de ces données, par le seul jeu des facultés mentales, est impossible ou très fragmentaire.

## I.2. L'AUTOMATISATION DU PROCESSUS DE CLASSIFICATION

Avant l'apparition de l'ordinateur, les techniques permettant de découvrir les propriétés structurales d'un ensemble d'individus faisaient surtout appel aux capacités du système visuel humain. En effet, l'observateur humain n'a aucun problème pour discerner des groupements de points, pourvu qu'ils soient disposés dans un espace à deux ou trois dimensions. Cependant, les vecteurs de caractères extraits des individus que l'on étudie sont en général représentés par plus de trois composantes, ce qui rend impossible l'examen visuel.

Au début du 20ème siècle apparaissent, sous l'impulsion des psychologues /34//6/, des techniques qui visent finalement à transformer l'espace de représentation de dimension

élevée en une image bidimensionnelle. Les procédures classiques d'analyse en composantes principales /23/ et d'analyse factorielle /8/, comme le "multidimensional scaling" /19/, conduisent essentiellement à des représentations de faible dimension préservant aussi bien que possible la structure des données. Elles ne sont pas à proprement parler des méthodes de classification automatique.

En fait, c'est des années trente que datent les premières études de base relatives aux problèmes de classification, tels que nous les avons évoqués. Mais les progrès théoriques réalisés notamment par Mahalanobis (1930) /24/ qui propose différents types de distances entre groupes d'individus et par Fisher (1936) /18/ qui introduit la notion de fonction discriminante sont restés longtemps sans retombées pratiques. Il a fallu attendre l'essor des calculateurs numériques pour que soient mises en œuvre de manière efficace les méthodes d'exploration multidimensionnelle nouvellement élaborées. L'absence de moyens de calcul et de mémorisation des données a en fait longtemps barré une des voies d'exploration les plus fécondes du réel: la classification automatique.

L'objet premier de toute classification est de définir des groupes (ou classes) à partir d'un ensemble d'individus dont la structure est inconnue a priori. Son but principal est donc de condenser l'essentiel des informations multiples observées sur un lot d'individus, la description complexe et détaillée de chaque individu étant remplacée par son appartenance à une classe bien définie. Le cerveau humain étant incapable d'accomplir une synthèse multidimensionnelle, l'automatisation de la classification constitue un pas décisif dans le processus d'analyse et de compréhension des données. Si elle n'est pas une fin en soi, la classification automatique permet, dans un domaine inconnu, de découvrir dans les phénomènes étudiés des structures qui n'étaient pas visibles sur les données. Elle apparaît ainsi comme une méthode exploratoire créatrice d'hypothèses. Dans d'autres cas, elle permettra de retrouver, en les précisant, des structures que l'on soupçonnait déjà. Cette démarche est fondamentale pour de nombreux travaux scientifiques et son application n'est plus l'exclusivité des naturalistes. On assiste, depuis les années cinquante, à une infiltration croissante des techniques de classification automatique dans



les différents secteurs de l'activité scientifique, industrielle et économique, essentiellement favorisée par la diffusion des moyens informatiques.

### I.3. CLASSIFICATION ET THEORIE DE LA DECISION

#### I.3.1. NOTION DE TAUX D'ERREUR

Les différentes classes détectées par une procédure de classification automatique peuvent être définies par l'ensemble des individus qui les composent, comme c'est souvent le cas pour les techniques de recherche de groupements ("cluster analysis"). Mais elles peuvent également être définies de manière plus conceptuelle, par exemple sous la forme de quelques paramètres définissant la loi de distribution des individus au sein de chaque classe. Dans ce cas on doit faire appel à une procédure de classement\* pour ranger tous les individus dans les différentes classes mises en évidence. En général, ce problème est résolu à l'aide de règles de décision dont la qualité peut être appréciée par la probabilité moyenne d'erreur de classement, désignée sous le nom de taux d'erreur. Il semble donc naturel de rechercher des règles de classement qui minimisent ce taux d'erreur.

La notion de probabilité d'erreur apparaît en 1928 dans le cadre de la statistique mathématique /27/. Wald (1939) /35/ généralise ensuite cette notion en introduisant le concept de fonction de coût. Mais c'est à Chow (1957) /10/ que revient le mérite d'avoir appliqué la théorie bayessienne de la décision aux problèmes de classement.

#### I.3.2. CLASSIFICATION OPTIMALE

Supposons que l'on associe à chaque individu de l'échantillon étudié une observation multidimensionnelle, constituée par exemple par un ensemble de mesures effectuées sur chaque

---

\* Par opposition à "classification" nous appellerons "classement" la procédure qui consiste à classer ou ranger des individus dans un ensemble de classes préalablement définies.

individu. Si on connaît la classe à laquelle appartient un individu, l'observation qui lui est associée peut être considérée comme une réalisation particulière d'une variable aléatoire dont les propriétés statistiques sont caractéristiques de la classe considérée.

Dans ces conditions, la distribution des observations pour des individus tirés aléatoirement et indépendamment de l'ensemble des classes constituant la population étudiée est un mélange pondéré des lois de probabilité relatives à ces différentes classes. Le coefficient de pondération de chaque loi est la probabilité a priori de la classe correspondante.

La connaissance du nombre de classes et, pour chacune d'elles, de la fonction de densité de probabilité et de la probabilité a priori permet de connaître la probabilité d'erreur associée au classement d'un individu donné dans chacune des classes /17/. La règle de classement qui affecte tout individu à la classe pour laquelle la probabilité d'erreur est minimale peut être considérée comme optimale, en ce sens qu'aucune autre stratégie ne peut donner un taux d'erreur plus petit. On désigne également cette procédure de classement optimale sous les noms de "classement à taux d'erreur minimal" ou de "classement bayésien".

Dans les problèmes pratiques, les données nécessaires pour calculer les probabilités d'erreur de classement d'un individu dans les différentes classes ne sont pas, en général, toutes disponibles. En effet, la distribution des observations n'est pas connue a priori. Dans certains cas, on ne connaît ni le nombre de classes en présence, ni la fonction de densité et la probabilité a priori attachées à chacune d'elles. Il est toutefois possible d'envisager une optimisation du classement en compensant le manque de connaissances sur le mélange par les informations apportées par les individus à classer eux-mêmes. En supposant que les fonctions de densité des différentes classes appartiennent à un ensemble de fonctions représentables par quelques paramètres (fonctions normales, de Bernouilli, etc...), le problème de l'optimisation du processus de classification se trouve posé en termes d'analyse des mélanges.

#### I.4. L'ANALYSE DES MELANGES

Le champ d'application des méthodes de classification automatique se situant à la jonction de plusieurs domaines scientifiques, le problème fondamental de l'analyse des mélanges a été abordé par des méthodes fort nombreuses et plus ou moins acceptables. Avant de poursuivre cet exposé, il importe donc de donner une description brève et aussi cohérente que possible des procédures existantes pour en préciser les fondements mathématiques, les avantages et les limitations.

##### I.4.1. ANALYSE DES MELANGES MONOVARIABLES

Historiquement, le premier travail sur la question remonte à 1894, lorsque K. Pearson utilise les moments de la distribution des observations pour déterminer les paramètres d'un mélange de deux densités normales et monovariabiles /29/.

Buchanan-Wollaston et Hodgesson (1929) /4/ proposent une méthode graphique consistant à ajuster une loi normale sur chaque pic de l'histogramme expérimental représentant la distribution des données. On retrouvera la même approche beaucoup plus tard chez Bhattacharya (1967) /2/.

Doetsch (1936) /16/ utilise la transformée de Fourier pour décomposer les mélanges gaussiens monovariabiles en supposant connue la valeur de la fonction de densité en tout point.

Dans le cadre de l'analyse des mélanges unidimensionnels, il convient de citer encore Rao (1948) /30/ qui traite le cas de mélanges à deux composantes et, plus récemment, Benzecri (1972) /3/ et Cazes (1976) /7/ qui utilisent une série de déconvolutions successives.

Si performantes soient-elles pour les problèmes monovariabiles, aucune de ces méthodes n'a pu être étendue au cas multidimensionnel.

##### I.4.2. TECHNIQUES D'APPRENTISSAGE BAYESIEN

Daly (1962) /13/ a eu le mérite d'introduire les techniques d'apprentissage bayésien en classification. Cette

approche suppose que les paramètres inconnus du mélange sont des variables aléatoires. L'information apportée séquentiellement par les individus à classer permet de réduire l'incertitude et de préciser de manière itérative la valeur des paramètres. Cette approche, reprise par Hilborn et Lainotis (1968) /21/ nécessite la connaissance a priori du nombre de classes et conduit à des calculs prohibitifs. Une procédure plus réaliste, proche de la solution bayésienne, a été proposée par Makov et Smith (1977) /25/. Elle est malheureusement limitée à des problèmes à deux classes.

#### I.4.3. METHODE D'ESTIMATION PAR MAXIMUM DE VRAISEMBLANCE

Les méthodes d'estimation par maximum de vraisemblance constituent une approche totalement différente du problème de la détermination des paramètres d'un mélange. Leur première utilisation dans ce domaine est due à Hasselblad (1966) /20/. L'extension de ces méthodes au cas multidimensionnel revient à Day (1969) /14/, mais celui-ci n'aborde sur le plan pratique que le cas de classes ayant même matrice de covariance. Wolfe (1970) /36/, qui considère des mélanges de fonctions normales et de fonctions de Bernouilli, se heurte aux mêmes difficultés pratiques et se limite, au niveau des applications, au cas de matrices de covariance égales ou diagonales.

Schröder (1976) /32/ associe la méthode du maximum de vraisemblance à un processus itératif inspiré de celui employé par Diday /15/ dans la méthode des nuées dynamiques. L'algorithme, très performant, s'applique à n'importe quel mélange de lois de même type mais nécessite la connaissance a priori du nombre de composantes à rechercher.

Kazakos (1977) /22/, enfin, utilise la méthode du maximum de vraisemblance pour estimer les probabilités a priori des différentes classes en supposant connus tous les autres paramètres du mélange.

#### I.4.4. APPROCHES DIVERSES

L'idée de Pearson a été reprise par D. Cooper et P. Cooper (1964) /11/ pour calculer les paramètres inconnus d'un mélange à partir des moments de la distribution de toutes

les observations. Mais la méthode, même après son amélioration par P. Cooper (1967) /12/ ne permet de traiter que le cas de composantes qui ne diffèrent que par leur valeur moyenne.

Chien et Fu (1967) /9/, puis Young et Coraluppi (1970) /37/ proposent des approches originales basées sur des techniques d'approximation stochastique. Mais le champ d'application de ces méthodes semble limité au cas unidimensionnel. L'utilisation de ces techniques dans le cas multidimensionnel n'a été envisagée que par Mizoguchi et Shimura (1975) /26/ et uniquement pour des problèmes à deux classes, avec matrices de covariance ou vecteurs moyenne connus a priori.

Nous terminerons ce rapide survol en mentionnant la méthode approchée donnée par Patrick et Hancock (1966) /28/ qui estiment les densités de probabilité à partir d'histogrammes. Bien qu'ils présentent la théorie dans le cas multidimensionnel, certains aspects spécifiques de leur méthode, tels que l'utilisation des quartiles, sont réservés au cas unidimensionnel.

#### I.4.5. LIMITATIONS DES METHODES D'ANALYSE DES MELANGES

Bien que le problème de l'optimisation du processus de classification puisse être ramené simplement au problème de l'estimation des paramètres d'un mélange, aucune solution entièrement satisfaisante ne semble avoir été apportée à cette question. En effet, certaines des techniques connues ne sont applicables que sous des hypothèses restrictives, telles que l'égalité des matrices de covariance; d'autres ne sont utilisables que dans des cas particuliers, tels que les mélanges à deux classes ou ceux à matrices de covariance diagonales. Accepter de telles hypothèses peut conduire à imposer une structure aux données plutôt qu'à découvrir leur organisation véritable.

D'autre part, nombre des techniques mises à la disposition des analystes nécessitent des informations a priori, telles que le nombre de classes en présence ou même la valeur de certains paramètres du mélange. Dans le stade exploratoire d'un ensemble de données inconnues, il est généralement impossible de fournir, a priori, de telles informations.

Enfin, les rares procédures qui échappent à ces critiques nécessitent une telle somme de calculs qu'elles ne

sont applicables qu'à des échantillons de taille et de dimension réduites, ce qui limite considérablement leur champ d'utilisation pratique.

Or, dans le cas où l'on accepte l'hypothèse normale pour optimiser la classification d'un ensemble de données, on devrait pouvoir disposer d'un outil performant pour identifier les paramètres des mélanges gaussiens, uniquement à partir des observations disponibles et sans hypothèse restrictive ni information a priori. C'est pourquoi nous nous proposons, dans cette étude, de mettre au point une procédure permettant de déterminer tous les paramètres d'un mélange gaussien, uniquement à partir de l'information qui peut être extraite de l'échantillon disponible.

L'algorithme résultant devra être utilisable pour analyser la structure d'échantillons de taille importante, même avec des observations de forte dimension. L'unique information fournie à cet algorithme sera la liste des paramètres mesurés sur les objets à classer. En sortie, il donnera:

- le nombre de classes en présence
- le vecteur-moyenne de chaque classe
- la matrice de covariance de chaque classe
- la probabilité a priori de chaque classe.

Ces paramètres seront ensuite utilisés pour optimiser la classification des observations constituant l'échantillon soumis à l'analyse.

## I.5 CONVEXITE ET OPTIMISATION DU PROCESSUS DE CLASSIFICATION

Le principal but de ce travail est donc d'optimiser le processus de classification en n'utilisant aucune autre information que celle apportée par les observations disponibles. L'unique hypothèse de travail concerne la normalité de la distribution des individus de chaque classe. Cette hypothèse est généralement admise lorsqu'on dispose de peu d'information a priori, mais de beaucoup d'observations.

La méthode proposée pour obtenir une description complète des mélanges gaussiens repose essentiellement sur une procédure permettant de déterminer localement la convexité de toute fonction multivariable continue (Chapitre II). Cette procédure est basée sur l'analyse des variations de la valeur moyenne de la fonction calculée sur des domaines appropriés, lorsque la taille de ces domaines augmente.

Pratiquement, on s'intéresse à la convexité de la fonction de densité de la distribution des observations disponibles. Dans le contexte de ce travail cette dernière n'est pas connue explicitement. En faisant appel à des techniques d'estimation non paramétrique on montre comment la convexité locale d'une fonction de densité quelconque peut être déterminée à partir des observations (Chapitre III).

Des estimateurs simples, bâtis autour d'un noyau, sont parfaitement adaptés à ce problème. Ce type d'estimateur n'a cependant pas que des avantages. En effet, le coût en calcul devient prohibitif dès que la dimension des données dépasse quelques unités. Pour que l'utilisation pratique du test de convexité proposé ne demeure pas limitée à des données de faible dimension, un algorithme d'estimation rapide, permettant un gain de temps très appréciable, est présenté.

Dans la seconde partie de ce travail, on montre d'abord comment la détermination du domaine à l'intérieur duquel une fonction de densité normale est concave permet d'obtenir des valeurs approchées du vecteur-moyenne et de la matrice de covariance de la distribution correspondante (Chapitre IV).

Cette procédure est exploitée pour analyser les mélanges gaussiens en déterminant, à partir des observations, les domaines où la fonction de densité est concave. L'analyse des propriétés géométriques de ces domaines permet de déterminer des valeurs approchées des paramètres statistiques de chaque composante mise en évidence. La prise en compte des observations tombant dans chacun de ces domaines permet d'autre part d'approcher la probabilité a priori de chaque classe. La détermination de fonctions de décision à partir de ces valeurs approchées conduit enfin à une automatisation complète du processus de classification avec un taux d'erreur très proche de l'optimum théorique (Chapitre V).

Deux variantes de cette approche sont proposées dans la troisième partie. La première, qui est destinée à l'analyse d'observations peu nombreuses mais de dimension élevée, nécessite des hypothèses plus fortes, à savoir l'indépendance, au sein de chaque classe, des caractères observés sur les individus. (Chapitre VI).

La seconde, qui permet de s'affranchir de l'hypothèse de normalité, est en fait une méthode de recherche de groupements par détection des modes, ceux-ci étant identifiés aux domaines où la fonction de densité de la distribution des observations est concave (Chapitre VII).

Dans la dernière partie de ce mémoire, les méthodes de classification proposées sont utilisées pour analyser la forme des cèdres du Maroc (Chapitre VIII). Cette étude dendrométrique, conduite sur plusieurs cédraies du Rif et du Moyen-Atlas, fait apparaître un polymorphisme marqué chez cet arbre. En tenant compte des différences de forme ainsi mises en évidence, il est alors possible d'améliorer sensiblement la précision et la fiabilité des tables de cubage utilisées pour l'exploitation de ces cédraies (Chapitre IX). Cette application pratique fait clairement apparaître le rôle déterminant que peuvent jouer les techniques de classification automatique et de reconnaissance des formes dans la gestion des ressources forestières.



REFERENCES (chapitre I)

- /1/ ADANSON, M. (1763)  
"Famille des plantes"
- /2/ BHATTACHARYA, C.G.  
"A simple method of resolution of a distribution into gaussian components"  
Biometrics (1967)
- /3/ BENZECRI, J.P.  
"La regression"  
Publication Labo. Statist. Math. Univ. Paris 6 (1972)
- /4/ BUCHANAN-WOLLASTON, H.G. et HODGESSON, W.G.  
"A new method of treating frequency curves in fishery statistics, with some results"  
Jour. Cons. Vol. 4 p. 207-225 (1929)
- /5/ BUFFON,  
"De la manière de traiter et d'étudier l'histoire naturelle"
- /6/ BURT, C.  
"Experimental tests of general intelligence"  
Brit. Jour. Psychol. Vol. 3 p. 94-177 (1909)
- /7/ CAZES, P.  
"Décomposition d'un histogramme en composantes gaussiennes"  
Rev. de Statist. App. Vol. 24 n°1 p. 63-82 (1976)
- /8/ CAILLIEZ, F. et PAGES, J.P.  
"Introduction à l'analyse des données"  
Smash, Paris. Chapitres 11 à 13 (1976)
- /9/ CHIEN, Y.T. et FU, F.S.  
"On Bayesian learning and stochastic approximation"  
IEEE Trans. Syst. Sci. Cyb. Vol. SSC-3 p. 28-38 (1967)
- /10/ CHOW, C.K.  
"An optimum character recognition system using decision functions"  
IRE Trans. Elec. Comp. Vol. EC-6 p. 247-254 (1957)
- /11/ COOPER, D.B. et COOPER, P.W.  
"Non supervised adaptive signal detection and pattern recognition"  
Inform. & Control Vol. 7 p. 416-444 (1964)

- /12/ COOPER, P.W.  
 "Some topics on nonsupervised adaptive detection for multivariate normal distributions"  
 Computer and Information Sciences II p. 123-146  
 Academic Press, New York (1967)
- /13/ DALY, R.F.  
 "The adaptive binary-detection problem on the real line"  
 Techn. Report 2003-3 Stanford Univ. Calif. (1962)
- /14/ DAY, N.E.  
 "Estimating the components of a mixture of normal distributions"  
 Biometrika Vol. 56 p. 463-474 (1969)
- /15/ DIDAY, E.  
 "Nouvelles méthodes et nouveaux concepts en classification automatique et reconnaissance des formes"  
 Thèse d'Etat. Univ. Paris 6 (1972)
- /16/ DOETSCH, G.  
 "Zerlegung einer funktion in Gausche fehlerkurven und zeitliche zuruckverfolgung eines temperaturzustandes"  
 Math. Zeitschrift Vol. 41 p. 283-318 (1936)
- /17/ DUDA, R.O. et HART, P.E.  
 "Pattern classification and scene analysis"  
 Wiley, New York. Chapitre 2 (1973)
- /18/ FISCHER, R.A.  
 "The use of multiple measurements in taxonomic problems"  
 Ann. Eugenics Vol. 7 Part II p. 179-188 (1936)
- /19/ GREEN, P.E. et CARMONE, F.J.  
 "Multidimensional scaling and related techniques in marketing analysis"  
 Allyn & Bacon, Boston, Mass. (1970)
- /20/ HASSELBLAD, V.  
 "Estimation of parameters for a mixture of normal distributions"  
 Technometrics Vol. 8 p. 431-444 (1966)
- /21/ HILLBORN, C.G. et LAINIOTIS, D.G.  
 "Optimal unsupervised learning multicategory dependant hypotheses pattern recognition"  
 IEEE Trans. Info. Theory Vol. IT-14 p. 468-470 (1968)
- /22/ KAZAKOS, D.  
 "Recursive estimation of prior probabilities using a mixture"  
 IEEE Trans. Info. Theory Vol. IT-23 n°2 p. 203-211 (1977)

- /23/ LEBART, L. ; MORINEAU, A. et TABART, N.  
 "Techniques de la description statistique"  
 Dunod, Paris (1977)
- /24/ MAHALANOBIS, P.C.  
 "On tests and measures of group divergence"  
 Jour. Asiat. Soc. Beng. Vol. 26 p. 541-588 (1930)
- /25/ MAKOV, U.E. et SMITH, A.F.M.  
 "A quasi-Bayes unsupervised learning procedure for priors"  
 IEEE Trans. Info. Theory Vol. IT-23 n°6 p.761-764 (1977)
- /26/ MIZOGUCHI, R. et SHIMURA, M.  
 "An approach to unsupervised learning classification"  
 IEEE Trans. Comput. Vol. C-24 n°10 p. 979-983 (1975)
- /27/ NEYMAN, J. et PEARSON, E.S.  
 "On the use and interpretation of certain test criteria  
 for purposes of statistical inference"  
 Biometrika Vol. 20A p. 175-240 (1928)
- /28/ PATRICK, E. et HANCOCK, J.  
 "Nonsupervised sequential classification and recognition  
 of patterns"  
 IEEE Trans. Info. Theory Vol. IT-12 n°4 p. 362-372 (1966)
- /29/ PEARSON, K.  
 "Contribution to the mathematical theory of evolution"  
 Philos. Trans. Royal Soc. of London Vol. 185 p. 71-110 (1894)
- /30/ RAO, C.R.  
 "Utilization of multiple measurements in problems of  
 biological classification"  
 Jour. R. Statist. Soc. B. Vol. 10 n°2 p; 159-193 (1948)
- /31/ RAY (1693)  
 "Synopsis methodica"
- /32/ SCHROEDER, A.  
 "Analyse d'un mélange de distributions de probabilité  
 de même type"  
 Rev. Statist. App. Vol. 24 n°1 p. 39-62 (1976)
- /33/ SOKAL, R.R. et SNEATH, P.H.A.  
 "Principles of numerical taxonomy"  
 W.H. Freeman, San Francisco, Calif (1963)
- /34/ SPEARMAN, C.  
 "General intelligence objectively determined and measured"  
 Amer. Jour. Psychol. Vol. 15 p. 202-292 (1904)

- /35/ WALD, A.  
"Contribution to the theory of statistical estimation  
and testing of hypotheses"  
Ann. Math. Stat. Vol. 10 p. 299-326 (1939)
- /36/ WOLFE, J.H.  
"Pattern clustering by multivariate mixture analysis"  
Multiv. Behav. Res. Vol. 5 p. 329-350 (1970)
- /37/ YOUNG, T.Y. et CORALUPPI, C.  
"Stochastic estimation of a mixture of normal density  
functions using an information criterion"  
IEEE Trans. Info. Theory Vol. IT-16 n°3 p. 258-263 (1970)

PREMIERE PARTIE

CONVEXITE DES FONCTIONS DE DENSITE

## C H A P I T R E II

ANALYSE DE LA CONVEXITE DES FONCTIONS MULTIVARIABLES
---

II.1 INTRODUCTION

La théorie de la convexité n'est pas, à vrai dire, une discipline nouvelle. La plupart des résultats concernant la convexité sont connus depuis plus de cinquante ans. Pendant longtemps la convexité n'intéressa que les mathématiciens. Mais depuis l'apparition de l'ordinateur, par le biais de la théorie des jeux /8/, de la programmation linéaire /5/, de la programmation non linéaire /2//10/ et des diverses techniques d'optimisation /11/, le champ d'application de la théorie de la convexité n'a cessé de s'étendre à des domaines très variés extérieurs aux mathématiques /2//8/.

Dans ce travail, nous nous proposons d'utiliser cette théorie pour résoudre certains problèmes de classification automatique. Nous montrerons en effet, dans les seconde et troisième parties de ce mémoire, comment la détermination de la convexité des fonctions de densité de probabilité permet d'aborder l'analyse des données multidimensionnelles. Avant de nous consacrer à l'étude spécifique des fonctions de densité, nous proposons dans ce chapitre une méthode d'analyse de la convexité des fonctions multivariées /14/.

II.2. FONCTIONS CONVEXES

II.2.1. DOMAINES CONVEXES

Afin de fixer la terminologie et les notations, nous précisons quelques notions topologiques essentielles /3/ que nous utiliserons souvent par la suite.

Nous considérons des fonctions à valeurs réelles définies sur  $R^n$ . Les éléments de cet espace, appelés indifféremment "points" ou "vecteurs", sont notés:

$$X = [x_1, \dots, x_i, \dots, x_n]^T$$

Sur l'espace  $R^n$  muni d'une norme, on définit le voisinage- $\epsilon$  d'un point  $X_0$  comme le domaine ouvert (ou boule) tel que:

$$D(X_0, \epsilon) = \{ X \in R^n : \|X - X_0\| < \epsilon \}$$

La frontière de ce domaine (ou sphère) est définie par:

$$S(X_0, \epsilon) = \{ X \in R^n : \|X - X_0\| = \epsilon \}$$

où  $\|X - X_0\|$  représente la distance entre les points  $X$  et  $X_0$ .

Remarque Parmi les normes usuelles employées sur  $R^n$  on peut citer:

$$\|X\| = |x_1| + |x_2| + \dots + |x_n|$$

$$\|X\| = \max(|x_1|, |x_2|, \dots, |x_n|)$$

$$\|X\| = (x_1^2 + x_2^2 + \dots + x_n^2)^{1/2}$$

Pour chacune de ces normes on obtient dans  $R^2$  les voisinages- $\epsilon$  de l'origine représentés figure 2.1

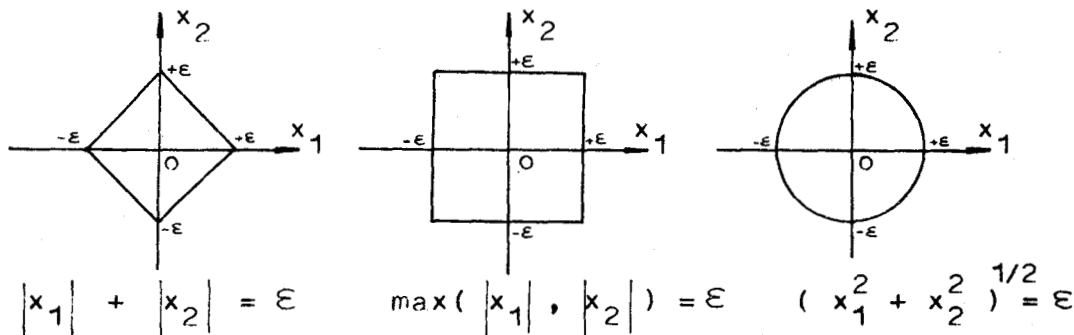


Figure 2.1.

### Points intérieurs et extérieurs

Soit  $D$  un domaine de  $R^n$ . Le point  $X$  est intérieur à  $D$  s'il existe un voisinage- $\varepsilon$  de  $X$  ne contenant que des points de  $D$ . Il est extérieur s'il existe un voisinage- $\varepsilon$  de  $X$  qui ne contient pas de points de  $D$ . Un point  $X$  qui n'est ni intérieur, ni extérieur à  $D$ , de telle sorte que tout voisinage de  $X$  contient au moins un point de  $D$  et un point qui n'appartient pas à  $D$  est appelé "point frontière".

### Domaine ouvert, domaine fermé

Le domaine  $D$  est ouvert si tout point de  $D$  est un point intérieur. Il est fermé si son complément est ouvert.

### Domaine convexe

Le but principal de ce chapitre est d'étudier les fonctions convexes. Les domaines naturels de définition de telles fonctions sont des domaines convexes /13/. On dit que  $D$  est convexe si  $x_1, x_2 \in D$  entraîne que  $Z = [\lambda x_1 + (1-\lambda)x_2] \in D$  quel que soit  $\lambda \in [0, 1]$ . D'un point de vue géométrique, un domaine convexe doit contenir le segment de droite joignant tout couple de ses points.

#### II.2.2. FONCTIONS CONVEXES

On attribue généralement à Jensen /6//7/ (1905) la mise en évidence des fonctions convexes comme objet mathématique particulier. La définition de la convexité usuellement adoptée pour les fonctions multivariées résulte d'une généralisation très naturelle des travaux de Jensen sur les fonctions monovariées /12/.

Soit une fonction  $f(X)$ , définie sur un domaine  $D$  convexe, de telle sorte que si  $x_1, x_2 \in D$ ,  $\lambda \in [0, 1]$ ,  $f(X)$  est toujours définie au point  $\lambda x_1 + (1-\lambda)x_2$ .

#### Définition 2.1.

Par définition, la fonction  $f(X)$  est convexe sur un domaine convexe  $D$  de  $R^n$  si:

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2) \quad (2.1.)$$

quels que soient  $x_1, x_2 \in D$  et  $\lambda \in [0, 1]$ .



Si la fonction  $-f(X)$  est convexe, on dit que la fonction  $f(X)$  est concave.

Remarque On note que l'inégalité (2.1.) ne fait aucune référence à la notion de norme et peut être définie sur des espaces plus généraux. Mais compte-tenu de la finalité de cette étude et pour des raisons de commodité, nous travaillerons uniquement dans des espaces normés.

Nous verrons ultérieurement que la détermination de la convexité des fonctions de densité de probabilité apparaît comme une approche intéressante des problèmes de classification automatique. Mais ces fonctions ne sont généralement connues que sous la forme d'une estimation explicite, obtenue à partir des observations disponibles. Or la définition 2.1. est mal adaptée à la détermination de la convexité d'une fonction dont on ne connaît pas l'expression analytique. En effet, cette définition fait appel à une notion globale, en ce sens que l'inégalité (2.1.) doit être vérifiée pour tout couple de points  $X_1, X_2$  appartenant au domaine sur lequel la fonction est convexe. Lorsqu'on ne dispose pas de la forme analytique de la fonction étudiée, il est difficile d'utiliser cette définition globale pour analyser sa convexité. Même avec des moyens de calcul très performants, il est peu réaliste de prendre en compte les valeurs de la fonction sur tous les segments de droite joignant tous les couples de points appartenant au domaine où elle est convexe.

Pour pallier cette difficulté, nous proposons une méthode qui consiste à tester localement (ou point par point) la convexité d'une fonction multivariable à partir de l'observation de ses valeurs sur son domaine de définition. Cette analyse fait appel à une notion de convexité en un point qu'il convient de définir avec précision.

### II.2.3. FONCTIONS LOCALEMENT CONVEXES

La définition de la convexité locale d'une fonction que nous proposons est dérivée de la notion classique de convexité des fonctions sur des domaines convexes.

Définition 2.2.

Une fonction  $f(X)$  est dite "localement convexe" au point  $X_0$  s'il existe un voisinage- $\epsilon$  de  $X_0$  convexe sur lequel la fonction  $f(X)$  est convexe.

Il est aisé de rattacher cette définition locale de la convexité à la définition classique 2.1. Nous pouvons en effet énoncer le

Théorème 2.1.

Si une fonction  $f(X)$  est convexe sur un domaine  $D$  ouvert convexe, elle est localement convexe en tout point de ce domaine.

Ce théorème permet de passer de la notion de convexité globale à la notion de convexité locale nouvellement définie. Il résulte du fait que, lorsque le domaine  $D$  est ouvert, tout point  $X$  de  $D$  est un point intérieur. Quel que soit  $X$ , il existe donc un voisinage de  $X$ , convexe, ne contenant que des points de  $D$ . Par conséquent, la fonction  $f(X)$  est convexe dans ce voisinage, c'est à dire localement convexe en  $X$ .

Pour permettre le passage de la notion locale de convexité à la notion globale, nous donnons maintenant une proposition réciproque de celle du Théorème 2.1. Pour établir cette réciproque, nous utilisons une condition classique de convexité, exprimée en fonction des dérivées de la fonction en un point /4/.

Condition nécessaire et suffisante de convexité

Une condition nécessaire et suffisante pour qu'une fonction admettant des dérivées partielles d'ordre deux continues sur un domaine convexe ouvert soit convexe sur ce domaine est que le hessien:

$$H[f(X)] = \begin{bmatrix} f_{11}(X) & \dots & f_{1n}(X) \\ \vdots & f_{ij}(X) & \vdots \\ f_{m1}(X) & \dots & f_{nn}(X) \end{bmatrix} \quad \text{où: } f_{ij}(X) = \frac{\partial^2 f(X)}{\partial x_i \partial x_j}$$

soit semi-défini positif pour tout point  $X$  de ce domaine.

Nous n'aborderons pas le problème très délicat de l'existence des dérivées du second ordre qui dépasse largement le cadre de cette étude /1//9/. Si, pour une fonction  $f(X)$  localement convexe en  $X_0$  celles-ci sont partout définies et continues, il existe un voisinage- $\varepsilon$  de  $X_0$  dans lequel le hessien  $H[f(X)]$  est positif ou nul. Le hessien en  $X_0$  est donc lui-même positif ou nul. Par conséquent, si la fonction  $f(X)$  est localement convexe en tout point d'un domaine convexe ouvert, le hessien  $H[f(X)]$  est positif ou nul en tout point de ce domaine. Il en résulte que la fonction est convexe sur ce domaine. On peut donc énoncer le:

#### Théorème 2.2.

Si une fonction  $f(X)$ , admettant des dérivées partielles d'ordre deux continues, est localement convexe en tout point d'un domaine ouvert convexe, cette fonction est convexe sur ce domaine.

Remarque Si une fonction admettant des dérivées partielles du second ordre continues est localement convexe en tout point d'un domaine ouvert non convexe, on ne peut pas conclure à la convexité de la fonction sur ce domaine. Le chapitre VII nous fournira des exemples de ce type de situation.

#### II.2.4. CONTINUITÉ DES FONCTIONS LOCALEMENT CONVEXES

Nous terminerons cette étude des fonctions localement convexes par une propriété de continuité. On sait qu'une fonction convexe et bornée supérieurement sur un domaine convexe est continue en tout point de ce domaine /4/. On déduit immédiatement de cette propriété le

#### Théorème 2.3.

Une fonction  $f(X)$  bornée supérieurement et localement convexe en  $X_0$  est continue en  $X_0$ .

Nous allons maintenant montrer comment cette nouvelle définition locale de la convexité se prête à l'analyse de la convexité des fonctions dont on ignore la forme analytique.

### II.3. DOMAINES D'OBSERVATION DES FONCTIONS CONVEXES

Notre but est de proposer une méthode pour déterminer la convexité locale d'une fonction à partir de l'observation de ses valeurs au voisinage d'un point. Pour ce faire, nous analyserons les variations de la valeur moyenne de la fonction calculée sur des domaines d'observation appropriés de dimension variable. Avant d'envisager cette procédure, il convient de définir ces domaines et d'en préciser les propriétés.

#### II.3.1. DEFINITION

Soient  $D_r$  et  $S_r$  la boule unité et la sphère unité associées à la norme définie sur  $R^n$ . Le domaine convexe  $D_r$ , symétrique par rapport à l'origine, est appelé "domaine de référence".

Sur chaque point  $X$  de  $R^n$ , on centre un domaine  $D_r(X)$  de frontière  $S_r(X)$ , obtenu par une translation de vecteur  $X$  du domaine de référence  $D_r$ .

On associe à chaque point  $X$  une famille de domaines notés  $D(X, \alpha)$ , centrés en  $X$  et homothétiques du domaine  $D_r(X)$ . L'homothétie est caractérisée par son rapport  $\alpha$ , réel positif.

Les domaines  $D(X, \alpha)$  sont appelés "domaines d'observation".

#### II.3.2. PROPRIETE

Le volume des domaines d'observation peut être aisément exprimé en fonction du volume du domaine de référence  $D_r$  et du rapport d'homothétie.

#### Théorème 2.4.

Le volume  $V\{D(X, \alpha)\}$  du domaine d'observation  $D(X, \alpha)$  est donné par l'expression:

$$V\{D(X, \alpha)\} = \alpha^n \cdot V\{D_r\} \quad (2.2.)$$

dans laquelle  $V\{D_r\}$  représente le volume du domaine de référence  $D_r$ .

Démonstration (Voir figure 2.2.)

Soit  $\Delta(X,U)$  une droite passant par  $X$  et dont la direction est donnée par le vecteur  $U$ . Le domaine  $D_r(X)$  étant convexe, cette droite coupe sa frontière  $S_r(X)$  en deux points  $P_r$  et  $P'_r$ , symétriques par rapport à  $X$ , de telle sorte que:

$$\overrightarrow{XP_r} = -\overrightarrow{XP'_r} = l(U).U$$

Le scalaire  $l(U)$ , positif ou nul, caractérise la géométrie du domaine de référence.

De la même manière, si  $\mathcal{Z}$  est le rapport d'homothétie définissant le domaine  $D(X,\mathcal{Z})$ ,  $\Delta(X,U)$  coupe la frontière  $S(X,\mathcal{Z})$  de  $D(X,\mathcal{Z})$  en deux points  $P_{\mathcal{Z}}$  et  $P'_{\mathcal{Z}}$  tels que:

$$\overrightarrow{XP_{\mathcal{Z}}} = -\overrightarrow{XP'_{\mathcal{Z}}} = \mathcal{Z}.l(U).U \quad (2.3.)$$

Soient deux éléments de surface  $d\sigma_r$  et  $d\sigma_{\mathcal{Z}}$  de  $S_r(X)$  et  $S(X,\mathcal{Z})$ , entourant respectivement  $P_r$  et  $P_{\mathcal{Z}}$  et vus de  $X$  sous le même angle solide  $d\Omega(X,U)$  de telle sorte que:

$$d\Omega(X,U) = \frac{(U \times N).d\sigma_r}{[l(U)]^{n-1}} = \frac{(U \times N).d\sigma_{\mathcal{Z}}}{[\mathcal{Z}.l(U)]^{n-1}} \quad (2.4.)$$

Dans cette expression, la quantité  $U \times N$  représente le produit scalaire du vecteur  $U$  par le vecteur unitaire  $N$  normal à  $S_r(X)$  en  $P_r$ .

De (2.4.) on déduit immédiatement:

$$d\sigma_{\mathcal{Z}} = \mathcal{Z}^{n-1}.d\sigma_r \quad (2.5.)$$

Une variation  $d\mathcal{Z}$  de  $\mathcal{Z}$  engendre une variation  $d\overrightarrow{XP_{\mathcal{Z}}}$  de  $\overrightarrow{XP_{\mathcal{Z}}}$ . Il en résulte un déplacement élémentaire de  $d\overrightarrow{\omega_{\mathcal{Z}}}$  qui engendre un élément de volume  $d^2\omega_{\mathcal{Z}}$  d'expression:

$$d^2\omega_{\mathcal{Z}} = (N \times d\overrightarrow{XP_{\mathcal{Z}}}).d\overrightarrow{\omega_{\mathcal{Z}}}$$

De (2.3.) on déduit:

$$d\overrightarrow{XP_{\mathcal{Z}}} = l(U).U.d\mathcal{Z}$$

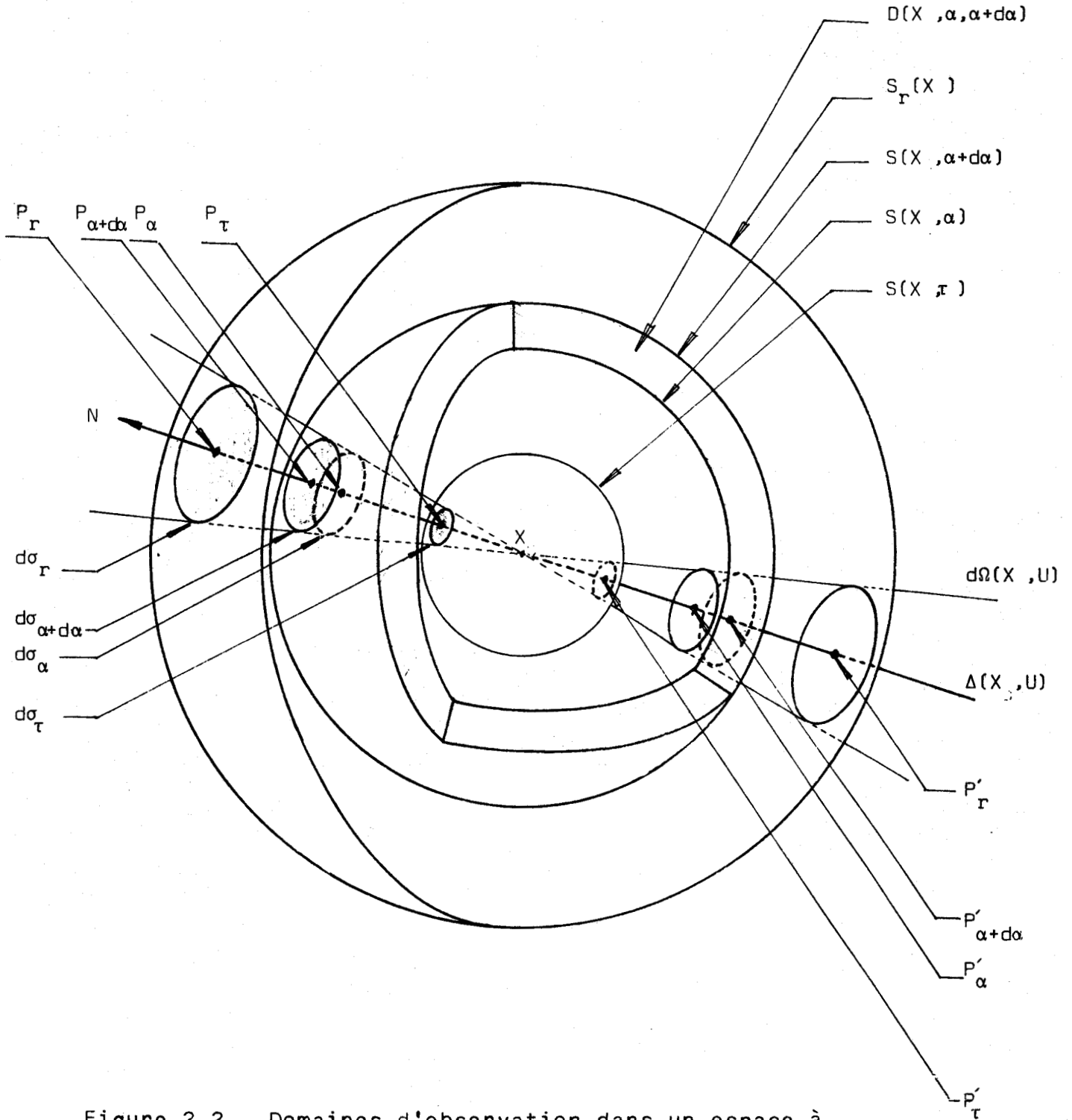


Figure 2.2. Domaines d'observation dans un espace à trois dimensions associés à la norme euclidienne.

(Figure représentée dans le cas  $0 < \alpha < \alpha+da < 1$ )



Tenant compte de (2.5.), il vient:

$$d^2\omega_{\bar{z}} = 1(U) \cdot (UxN) \cdot d\sigma_r \cdot \bar{z}^{n-1} \cdot d\bar{z} \quad (2.6.)$$

La double intégration de (2.6.) sur  $S_r(X)$  d'une part et pour  $\bar{z}$  variant de 0 à  $\alpha$  d'autre part, conduit à l'expression de  $V\{D(X, \alpha)\}$  suivante:

$$V\{D(X, \alpha)\} = \int_0^\alpha \int_{S_r(X)} 1(U) \cdot (UxN) \cdot d\sigma_r \cdot \bar{z}^{n-1} d\bar{z} \quad (2.7.)$$

$$= \frac{\alpha^n}{n} \int_{S_r(X)} 1(U) \cdot (UxN) \cdot d\sigma_r \quad (2.8.)$$

En évaluant (2.8.) pour  $\alpha = 1$ , on obtient:

$$V\{D_r(X)\} = V\{D_r\} = \frac{1}{n} \int_{S_r} 1(U) \cdot (UxN) \cdot d\sigma_r \quad (2.9.)$$

A partir de (2.8.) et de (2.9.) on obtient finalement l'équation (2.2.).

≠≠

## II.4. VALEUR MOYENNE DES FONCTIONS LOCALEMENT CONVEXES

### II.4.1. PROPRIETE FONDAMENTALE

La valeur moyenne  $\rho\{D(X_0, \alpha)\}$  d'une fonction  $f(X)$  sur un domaine d'observation  $D(X_0, \alpha)$  centré en  $X_0$  est donnée par l'expression:

$$\rho\{D(X_0, \alpha)\} = \frac{\int_{D(X_0, \alpha)} f(Z) dZ}{\int_{D(X_0, \alpha)} dZ} \quad (2.10.)$$

Pour un point  $X_0$  donné, cette quantité peut être considérée comme une fonction de  $\alpha$ , c'est à dire de la "taille" des domaines d'observation centrés en  $X_0$ . Nous allons maintenant démontrer la propriété fondamentale suivante, qui lie le sens de variation de cette fonction de  $\alpha$  à la convexité

locale de la fonction  $f(X)$  en  $X_0$ .

Théorème 2.5.

La valeur moyenne  $\rho\{D(X_0, \alpha)\}$  de la fonction  $f(X)$  est une fonction monotone croissante de  $\alpha$ ,  $\alpha > 0$ , au voisinage de zéro si la fonction  $f(X)$  est localement convexe en  $X_0$ . Cette valeur moyenne est une fonction monotone décroissante de  $\alpha$  au voisinage de zéro si la fonction  $f(X)$  est localement concave en  $X_0$ .

Démonstration (Voir figure 2.3.)

Nous allons établir la démonstration de cette propriété dans le cas où  $f(X)$  est une fonction localement convexe en  $X_0$ .

Soit  $d\rho$  la variation de  $\rho\{D(X_0, \alpha)\}$  résultant d'un accroissement  $d\alpha$  de  $\alpha$ .

$$d\rho = \rho\{D(X_0, \alpha + d\alpha)\} - \rho\{D(X_0, \alpha)\}$$

D'après la définition (2.10) de la valeur moyenne de  $f(X)$ , on peut écrire:

$$d\rho = \frac{\int_{D(X_0, \alpha + d\alpha)} f(Z) dZ}{\int_{D(X_0, \alpha + d\alpha)} dZ} - \frac{\int_{D(X_0, \alpha)} f(Z) dZ}{\int_{D(X_0, \alpha)} dZ} \quad (2.11.)$$

Soit  $D(X_0, \alpha, \alpha + d\alpha)$  le domaine compris entre les domaines  $D(X_0, \alpha)$  et  $D(X_0, \alpha + d\alpha)$ , de telle sorte que:

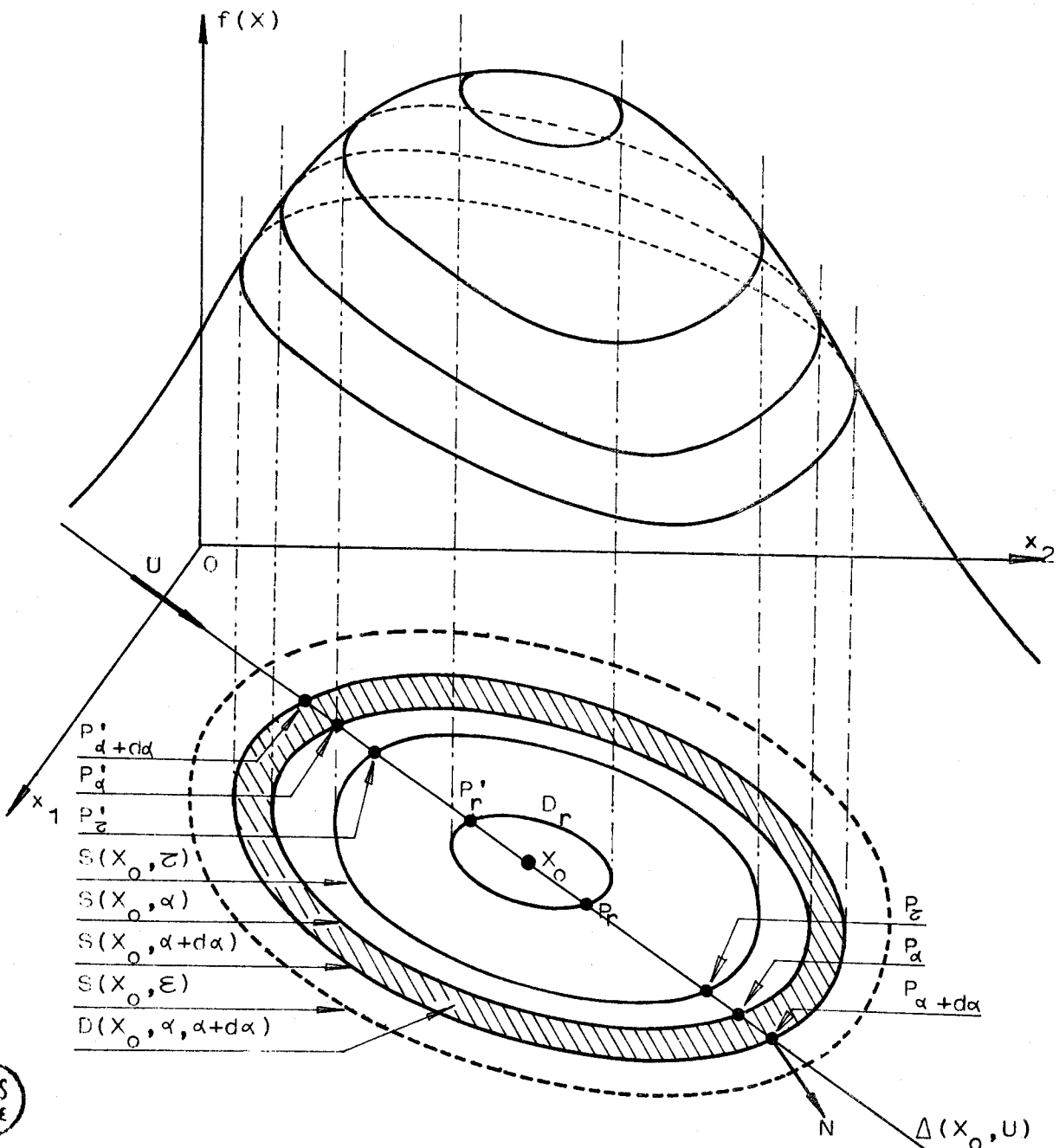
$$\int_{D(X_0, \alpha + d\alpha)} f(Z) dZ = \int_{D(X_0, \alpha)} f(Z) dZ + \int_{D(X_0, \alpha, \alpha + d\alpha)} f(Z) dZ$$

et:

$$\int_{D(X_0, \alpha + d\alpha)} dZ = \int_{D(X_0, \alpha)} dZ + \int_{D(X_0, \alpha, \alpha + d\alpha)} dZ$$

Dans ces conditions, (2.11.) peut s'écrire:





BUS  
LILLE

Figure 2.3. Domaines d'observation d'une fonction bidimensionnelle localement concave en  $X_0$ .  
 $( 1 < z < \alpha < \alpha + d\alpha < \varepsilon )$

$$d\rho = \frac{\int_{D(X_0, \alpha, \alpha+d\alpha)} dZ}{\int_{D(X_0, \alpha+d\alpha)} dZ} \left[ \frac{\int_{D(X_0, \alpha, \alpha+d\alpha)} f(Z) dZ}{\int_{D(X_0, \alpha, \alpha+d\alpha)} dZ} - \frac{\int_{D(X_0, \alpha)} f(Z) dZ}{\int_{D(X_0, \alpha)} dZ} \right]$$

ou encore:

$$d\rho = \frac{V\{D(X_0, \alpha, \alpha+d\alpha)\}}{V\{D(X_0, \alpha+d\alpha)\}} \left[ \rho\{D(X_0, \alpha, \alpha+d\alpha)\} - \rho\{D(X_0, \alpha)\} \right]$$

Le signe de  $d\rho$  est donc le même que celui de la différence  $d\rho'$  telle que:

$$d\rho' = \rho\{D(X_0, \alpha, \alpha+d\alpha)\} - \rho\{D(X_0, \alpha)\}$$

Nous allons maintenant utiliser les propriétés de symétrie des domaines d'observation et les propriétés de convexité de la fonction  $f(X)$  pour déterminer le signe de  $d\rho'$ .

Tout d'abord, l'évaluation de l'intégrale:

$$\int_{D(X_0, \alpha)} f(Z) dZ$$

s'effectue en utilisant la procédure de la démonstration du théorème 2.4. Par analogie avec (2.7.), il vient:

$$\int_{D(X_0, \alpha)} f(Z) dZ = \int_0^\alpha \left\{ \int_{S_r(X_0)} f(P_z) \cdot l(U) \cdot (U \times N) d\sigma_r \right\} z^{n-1} dz$$

En faisant intervenir le point  $P_z'$  symétrique de  $P_z$  par rapport à  $X_0$ , on peut encore écrire:

$$\int_{D(X_0, \alpha)} f(Z) dZ = \frac{1}{2} \int_0^\alpha \left\{ \int_{S_r(X_0)} [f(P_z) + f(P_z')] \cdot l(U) \cdot (U \times N) \cdot d\sigma_r \right\} z^{n-1} dz \quad (2.12)$$

D'autre part, compte-tenu des propriétés de symétrie des domaines d'observation, si  $P_\alpha$  et  $P_\alpha'$  sont les points d'intersection de  $\Delta(X_0, U)$  avec  $S(X_0, \alpha)$ , il est toujours possible de trouver pour chaque valeur de  $z$ ,  $0 \leq z \leq \alpha$ , un réel  $\lambda$  positif appartenant à l'intervalle  $[0, 1]$  tel que:

$$\begin{cases} P_z = \lambda P_\alpha + (1-\lambda) P_{\alpha'} \\ P_z' = (1-\lambda) P_\alpha + \lambda P_{\alpha'} \end{cases}$$

Enfin, la fonction  $f(X)$  étant supposée localement convexe en  $X_0$ , il existe un domaine  $D(X_0, \varepsilon)$  à l'intérieur duquel la fonction  $f(X)$  est convexe. En supposant  $\alpha < \varepsilon$ , la définition classique de la convexité permet d'écrire:

$$\begin{cases} f(P_z) = f(\lambda P_\alpha + (1-\lambda) P_{\alpha'}) \leq \lambda f(P_\alpha) + (1-\lambda) f(P_{\alpha'}) \\ f(P_z') = f((1-\lambda) P_\alpha + \lambda P_{\alpha'}) \leq (1-\lambda) f(P_\alpha) + \lambda f(P_{\alpha'}) \end{cases} \quad (2.13.)$$

Par addition des deux inégalités (2.13.), on obtient:

$$f(P_z) + f(P_z') \leq f(P_\alpha) + f(P_{\alpha'})$$

Cette inégalité reportée dans (2.12) conduit à:

$$\int_{D(X_0, \alpha)} f(Z) dZ \leq \frac{I(\alpha)}{2} \int_0^\alpha z^{n-1} dz$$

avec:

$$I(\alpha) = \int_{S_r(X_0)} [f(P_\alpha) + f(P_{\alpha'})] \cdot 1(U) \cdot (U \times N) \cdot d\sigma_r \quad (2.14.)$$

En utilisant le théorème 2.4., on aboutit à l'inégalité:

$$\rho \{D(X_0, \alpha)\} \leq \frac{I(\alpha)}{2 \cdot n \cdot V\{D_r\}} \quad (2.15.)$$

De manière analogue, quel que soit  $z$  compris entre  $\alpha$  et  $\alpha + d\alpha$ , il existe un réel positif  $\lambda$ ,  $\lambda \in [0, 1]$ , tel que:

$$\begin{cases} P_\alpha = \lambda P_z + (1-\lambda) P_z' \\ P_{\alpha'} = (1-\lambda) P_z' + \lambda P_z \end{cases}$$

Si  $\alpha + d\alpha < \varepsilon$ , le domaine d'observation  $D(X_0, \alpha + d\alpha)$  est intérieur au domaine  $D(X_0, \varepsilon)$  à l'intérieur duquel la fonction  $f(X)$  est localement convexe. Dans ces conditions, il vient:

$$\begin{cases} f(P_\alpha) = f(\lambda P_z + (1-\lambda)P_z') \leq \lambda f(P_z) + (1-\lambda)f(P_z') \\ f(P_\alpha') = f((1-\lambda)P_z + \lambda P_z') \leq (1-\lambda)f(P_z) + \lambda f(P_z') \end{cases}$$

Ces deux relations conduisent finalement à l'inégalité:

$$\rho\{D(X_0, \alpha, \alpha+d\alpha)\} \geq \frac{I(\alpha)}{2.n.V\{D_r\}} \quad (2.16.)$$

De (2.15.) et (2.16.) on déduit immédiatement:

$$d\rho \geq 0$$

Par conséquent,  $d\rho$  est positif ou nul pour un accroissement  $d\alpha$  de  $\alpha$ .  $\rho\{D(X_0, \alpha)\}$  est donc une fonction croissante de  $\alpha$  tant que le domaine d'observation  $D(X_0, \alpha+d\alpha)$  reste intérieur au voisinage  $D(X_0, \varepsilon)$  de  $X_0$  à l'intérieur duquel la fonction  $f(X)$  est localement convexe.  $\varepsilon$  étant inconnu, mais strictement positif, on en déduit que  $\rho\{D(X_0, \alpha)\}$  est une fonction de  $\alpha$  localement croissante au voisinage de zéro.

Lorsque la fonction  $f(X)$  est localement concave en  $X_0$ , un raisonnement en tout point analogue conduit à  $d\rho \leq 0$ , démontrant ainsi la seconde proposition du théorème 2.5.

≠

#### II.4.2. EXTREMUM LOCAL DE LA VALEUR MOYENNE

Nous avons vu qu'une fonction bornée, localement convexe en  $X_0$ , est continue en  $X_0$ . Ainsi, si une fonction bornée  $f(X)$  est localement convexe en  $X_0$ , la valeur moyenne de cette fonction sur le domaine d'observation  $D(X_0, \alpha)$  tend vers  $f(X_0)$  lorsque le volume de ce domaine tend vers zéro.

D'autre part, le sens de variation de la fonction  $\rho\{D(X_0, \alpha)\}$  au voisinage de zéro étant lié au sens de la convexité locale de la fonction en  $X_0$ , on peut énoncer le théorème suivant:

Théorème 2.6.

Si une fonction  $f(X)$  est localement convexe en  $X_0$ ,  $f(X_0)$  est un minimum local de la valeur moyenne  $\rho\{D(X_0, \alpha)\}$  au voisinage de  $\alpha = 0$ ,  $\alpha > 0$ .  
 Si  $f(X)$  est localement concave en  $X_0$ ,  $f(X_0)$  est alors un maximum local de  $\rho\{D(X_0, \alpha)\}$  au voisinage de zéro.

C'est cette forme légèrement modifiée du théorème 2.5. que nous utiliserons pour tester le sens de la convexité en un point d'une fonction localement concave ou convexe en ce point.

II.5. CONCLUSION

Nous avons introduit, dans ce chapitre, la notion de convexité locale d'une fonction en un point. Nous avons ensuite démontré que le sens de cette convexité est relié au sens de variation de la valeur moyenne de la fonction calculée sur des domaines d'observation de taille croissante centrés sur le point considéré.

Il convient de remarquer que nous n'avons pas abordé le problème de l'existence de cette convexité locale. En effet, il existe des fonctions qui, en certains points, ne sont ni localement convexes, ni localement concaves. Nous reviendrons sur ce problème dans la seconde partie de ce travail. Pour le moment nous nous contentons de proposer une méthode permettant de tester le sens de la convexité d'une fonction en un point où nous savons que la fonction est soit localement convexe, soit localement concave.

Dans la seconde partie de ce mémoire, nous appliquerons ce test de convexité à l'analyse de la convexité des fonctions de densité de probabilité rencontrées dans les problèmes de classification automatique. Il s'avère donc nécessaire, avant toute chose, d'adapter les résultats présentés dans ce chapitre aux exigences particulières de ce type de problème.

REFERENCES (chapitre II)

- /1/ ALEXANDROFF, A.D.  
 "Almost everywhere existence of the second differential  
 of a convex function and some properties of convex  
 surfaces connected with it"  
 Uch. Zap. Leningrad Gos. Univ. Ann. Ser. Math.  
 Vol. 6 p. 3-35 (1939)
- /2/ BRACKEN, J. et Mc CORMICK, G.P.  
 "Selected applications of nonlinear programming"  
 Wiley, New York (1968)
- /3/ BUCK, R.C.  
 "Advanced calculus"  
 Mc Graw-Hill, New York (1965)
- /4/ EGGLESTON, H.G.  
 "Convexity"  
 Cambridge Univ. Press, Camb. Tract in Math. and Math. Phys.  
 Chapitre 3 p. 45-58 (1963)
- /5/ HADLEY, G.  
 "Linear programming"  
 Addison-Wesley, Reading, Mass. (1962)
- /6/ JENSEN, J.L.W.V.  
 "Om konvekse funktioner og uligheder mellem middelværdier"  
 Nyt. Tidsskr. Math. Vol. 16B p. 46-69 (1905)
- /7/ JENSEN, J.L.W.V.  
 "Sur les fonctions convexes et les inégalités entre les  
 valeurs moyennes"  
 Acta. Math. Vol. 30 p. 175-193 (1906)
- /8/ KARLIN, S.  
 "Mathematical methods and theory of games, programming  
 and economics"  
 Addison-Wesley, Reading, Mass. (1959)
- /9/ LORCH, E.R.  
 "Differentiable inequalities and the theory of convex bodies"  
 Trans. Amer. Math. Soc. Vol. 71 p. 243-266 (1951)
- /10/ MANGASARIAN, O.L.  
 "Nonlinear programming"  
 Mc Graw-Hill, New York (1969)
- /11/ PSHENICHNY, B.M. et DANILIN, Y.M.  
 "Numerical methods in extremal problems"  
 Mir Publishers, Moscou (1978)

- /12/ ROBERTS, A.W. et VARBERG, D.E.  
"Convex functions"  
Academic Press, New York Chapitre 4 p. 88-120 (1973)
- /13/ VALENTINE, F.  
"Convex sets"  
Mc Graw-Hill, New York (1964)
- /14/ VASSEUR, C. et POSTAIRE, J.G.  
"Convexité des fonctions de densité. Application à la  
détection des modes en reconnaissance des formes"  
RAIRO Automatique Vol. 13 n°2 p. 171-188 (1979)

## C H A P I T R E III

DETERMINATION DE LA CONVEXITE LOCALE DES FONCTIONS DE DENSITE A PARTIR DES OBSERVATIONS
--

III. INTRODUCTION

Les résultats présentés au chapitre précédent permettent de déterminer le sens de la convexité locale d'une fonction lorsque cette convexité est définie.

Dans le cadre de cette étude sur la classification automatique, nous appliquerons ces résultats à l'analyse de la convexité des fonctions de densité de probabilité. Dans la pratique, ces fonctions ne sont en général connues que par l'intermédiaire des observations, réalisations particulières d'une variable aléatoire ayant pour densité de probabilité la fonction étudiée.

L'objet de ce chapitre est de mettre en place sur ordinateur numérique une procédure simple, efficace et rapide permettant, à partir d'un ensemble d'observations (ou échantillon), de tester le sens de la convexité locale de la fonction de densité sous-jacente /11/.



III.2. ESTIMATION DE LA VALEUR MOYENNE DES FONCTIONS DE DENSITEII.2.1. PRINCIPE DE BASE

Lorsque la fonction  $f(X)$  est une densité de probabilité, le numérateur de l'expression (2.10.) donnant la valeur moyenne  $\rho\{D(X_0, \alpha)\}$  de  $f(X)$  sur  $D(X_0, \alpha)$  peut être interprété comme la probabilité  $P$  pour qu'une observation  $X$  soit située à l'intérieur du domaine  $D(X_0, \alpha)$ :

$$P = \int_{D(X_0, \alpha)} f(Z) dZ$$

Supposons que l'on dispose d'un échantillon  $\{X_1, \dots, X_j, \dots, X_n\}$  de  $n$  observations indépendantes et identiquement distribuées suivant une fonction de répartition à laquelle correspond la densité de probabilité  $f(X)$ . La probabilité  $P_k$  pour que  $k$  des  $n$  observations de l'échantillon tombent dans  $D(X_0, \alpha)$  est alors donnée en fonction de  $P$  par la loi binomiale:

$$P_k = C_k^n P^k (1 - P)^{n-k}$$

On en déduit l'espérance mathématique de  $k$  :

$$E\{k\} = n.P$$

qui indique que  $k/n$  est un estimateur non biaisé de  $P$ .

Dans ces conditions, on peut prendre:

$$\hat{\rho}\{D(X_0, \alpha)\} = \frac{k/n}{V\{D(X_0, \alpha)\}} \quad (3.1.)$$

pour estimateur de la valeur moyenne de  $f(X)$  sur  $D(X_0, \alpha)$ .

Nous avons vu au chapitre précédent que pour tester le sens de la convexité locale d'une fonction en un point  $X_0$ , il fallait analyser les variations de sa valeur moyenne sur des domaines de taille tendant asymptotiquement vers zéro. Il importe donc d'étudier le comportement asymptotique de l'estimateur  $\hat{\rho}\{D(X_0, \alpha)\}$  lorsque  $\alpha$  tend vers zéro.

## III.2.2. PROPRIETES ASYMPTOTIQUES DE L'ESTIMATEUR

Nous devons nous assurer que l'estimateur  $\hat{p}\{D(X_0, \alpha)\}$  converge, comme la valeur moyenne  $p\{D(X_0, \alpha)\}$  elle-même, vers  $f(X_0)$  lorsque  $\alpha$  tend vers zéro. (cf. théorème 2.6.) Cependant, lorsqu'on ne dispose que d'un nombre limité  $n$  d'observations, le volume du domaine  $D(X_0, \alpha)$  ne peut tendre arbitrairement vers zéro. En effet, en faisant diminuer  $\alpha$  sans précaution, le domaine  $D(X_0, \alpha)$  pourrait devenir si petit qu'il ne renfermerait aucune observation, de telle sorte que la limite de l'estimateur serait nulle. Si par contre une observation venait à coïncider avec  $X_0$ , l'estimateur divergerait vers l'infini.

Pour assurer la convergence de l'estimateur vers  $f(X_0)$ , il a été montré que le volume du domaine intervenant dans l'expression de  $\hat{p}\{D(X_0, \alpha)\}$ , ainsi que le nombre d'observations  $k$  tombant dans ce domaine, doivent être liés au nombre total  $q$  d'observations disponibles.

Plus précisément, soit  $\alpha_q$  la valeur de  $\alpha$  et  $k_q$  la valeur de  $k$  correspondant à un échantillon de  $q$  observations. On démontre /16/ que  $\hat{p}\{D(X_0, \alpha)\}$  converge en moyenne quadratique vers  $f(X_0)$  si:

$$(a) \lim_{q \rightarrow \infty} V\{D(X_0, \alpha_q)\} = 0$$

$$(b) \lim_{q \rightarrow \infty} \frac{k_q}{q} = 0$$

$$(c) \lim_{q \rightarrow \infty} \frac{k_q}{q} = 0$$

Il y a deux manières d'obtenir des domaines  $D(X_0, \alpha_q)$  satisfaisant ces conditions.

La première consiste à réduire le volume du domaine  $D(X_0, \alpha_q)$  lorsque  $q$  augmente en spécifiant  $V\{D(X_0, \alpha_q)\}$  comme une fonction de  $q$ . On peut utiliser par exemple des relations du type:

$$V\{D(X_0, \alpha_q)\} = V_0 q^{-\gamma}, \gamma \in ]0, 1[ \quad (3.2.)$$

ou encore:

$$V\{D(X_0, \alpha_q)\} = V_0 / \log q \quad (3.3.)$$

Il faut alors vérifier que les variables aléatoires  $k_q$  et  $k_q/q$  satisfont aux conditions (b) et (c).

Cette approche, initialement proposée par Rosenblatt (1956) /13/ dans le cas monovarié, puis approfondie par Parzen (1962) /10/ a été étendue au cas multivarié par Murthy (1966) /9/ et par Cacoullos (1966) /2/.

La seconde méthode consiste à se donner  $k_q$  comme une fonction de  $q$ , par exemple:

$$k_q = k_0 \sqrt{q}$$

On fait alors croître  $\alpha_q$  jusqu'à ce que le domaine  $D(X_0, \alpha_q)$  englobe  $k_q$  observations. C'est la méthode des  $k_q$  plus proches voisins /3//7/.

Ces deux méthodes convergent et chacune d'elles présente des avantages spécifiques /14//15/. Nous allons montrer que la première est mieux adaptée au problème de l'analyse de la convexité que la seconde.

### III.3. TEST DE CONVEXITE

L'application du théorème 2.6. au problème de la détermination du sens de la convexité locale en un point  $X_0$  d'une fonction de densité  $f(X)$  à partir des observations est immédiate. Nous commencerons par estimer la valeur de  $f(X_0)$ , extremum local de  $\rho\{D(X_0, \alpha_q)\}$  au voisinage de  $\alpha = 0$ . Que nous utilisions l'approche de Rosenblatt-Parzen ou la méthode des  $k_q$  plus proches voisins, la valeur estimée de cet extremum sera donnée par:

$$\begin{aligned} \hat{f}(X_0) &= \hat{\rho}\{D(X_0, \alpha_q)\} \\ &= \frac{k_q/c}{V\{D(X_0, \alpha_q)\}} \end{aligned} \quad (3.4.)$$

L'estimation de  $\rho\{D(X_0, \beta)\}$  à partir de l'équation (3.1.), pour une valeur de  $\beta$  légèrement supérieure à  $\alpha_q$ , permettra ensuite de savoir si  $f(X_0)$  est un maximum ou un minimum local de la valeur moyenne de  $f(X)$ .

Cependant, comme nous venons de le voir,  $\alpha_q$  ne pourra tendre effectivement vers zéro que si le nombre d'observa-

tions  $\alpha$  tend vers l'infini. Dans le cas où l'on ne dispose que d'un nombre limité d'observations, il faut donc rechercher la procédure d'estimation conduisant à la plus faible valeur de  $\alpha_q$  afin de se rapprocher au mieux des conditions d'application du théorème 2.6.

Lorsqu'on utilise la méthode des  $k_q$  plus proches voisins, le volume du domaine  $D(X_0, \alpha_q)$  est fonction de la densité des observations autour du point  $X_0$ . Si cette densité est élevée, le domaine  $D(X_0, \alpha_q)$  contenant les  $k_q$  plus proches voisins de  $X_0$  est de taille réduite et la valeur correspondante de  $\alpha_q$  est faible. Par contre, si les observations sont peu denses autour de  $X_0$ , le domaine  $D(X_0, \alpha_q)$  doit être relativement grand pour englober  $k_q$  observations et  $\alpha_q$  peut prendre des valeurs éloignées de zéro.

Lorsqu'on utilise l'approche de Rosenblatt-Parzen, la valeur de  $\alpha_q$  n'est pas fonction de la densité locale des observations. Pour un échantillon de taille  $n$  donnée, cette valeur, calculée à partir des relations (3.2.) ou (3.3.), ne dépend que du choix de  $V_0$ . Si  $V_0$  est trop grand, il faudra disposer d'un nombre  $n$  d'observations très important pour que  $\alpha_q$  se rapproche suffisamment de zéro. On cherchera donc à prendre  $V_0$  aussi petit que possible. On sera toutefois limité par le fait que si  $V_0$  est trop petit, on obtiendra une estimation très bruitée de la densité de probabilité.

Nous utiliserons cependant cette dernière approche qui permet de garder le contrôle de la valeur de  $\alpha_q$ . Nous verrons en effet, dans la seconde partie de ce travail, qu'il est généralement possible d'obtenir des estimations satisfaisantes des fonctions de densité avec des valeurs de  $\alpha_q$  suffisamment petites pour permettre l'utilisation du test de convexité suivant:

#### Test de convexité

En tout point  $X_0$  où on désire tester le sens de la convexité locale, supposée définie, de la fonction de densité  $f(X)$ , on estime d'abord  $f(X_0)$  par:

$$\hat{f}(X_0) = \frac{k_\alpha / \alpha}{V\{D(X_0, \alpha)\}}$$

le volume  $V\{D(X_0, \alpha)\}$  étant ajusté en fonction de la taille de l'échantillon pour assurer la convergence en moyenne quadratique de l'estimateur.

On estime ensuite la valeur moyenne de la fonction  $f(X)$  sur un domaine d'observation  $D(X_0, \beta)$  de taille légèrement supérieure à celle de  $D(X_0, \alpha)$  en utilisant le même estimateur:

$$\hat{p}\{D(X_0, \beta)\} = \frac{k_\alpha / \alpha}{V\{D(X_0, \beta)\}}$$

Si:

$$\hat{p}\{D(X_0, \beta)\} > \hat{f}(X_0)$$

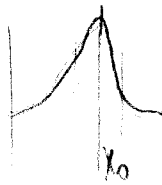
on conclut que la fonction est localement convexe en  $X_0$ . Dans le cas contraire, on conclut qu'elle est concave en  $X_0$ .

Remarque La validité du résultat de ce test n'est garantie que si la fonction  $f(X)$  est définie localement convexe ou concave à l'intérieur d'un voisinage de  $X_0$  contenant le domaine  $D(X_0, \beta)$ .

Comme  $\beta$  ne peut tendre arbitrairement vers zéro puisque  $\beta > \alpha_q$ , ce voisinage ne peut être arbitrairement petit.

Nous verrons comment tenir compte de cette remarque lors des applications de ce test de convexité.

Une des principales critiques faites à la méthode d'estimation des fonctions de densité de Rosenblatt-Parzen utilisée dans ce test est que son application demeure limitée à des problèmes de faible dimension. En effet, cette estimation locale se fait point par point. Or le nombre de points nécessaires pour décrire correctement une fonction multivariable croît exponentiellement avec la dimension de l'espace de définition. De ce fait, le temps d'exécution des algorithmes d'estimation sur ordinateur



numérique devient vite prohibitif dès que s'élève la dimension des données.

Le test de convexité proposé nécessite deux fois plus de calculs que l'estimation des fonctions de densité, puisqu'en chaque point il faut d'abord estimer la fonction elle-même, puis sa valeur moyenne sur un domaine d'observation. Il serait donc irréaliste d'envisager l'utilisation de ce test sans disposer d'une procédure permettant d'accélérer la mise en oeuvre de la méthode d'estimation de Rosenblatt-Parzen.

C'est pourquoi nous proposons maintenant un algorithme d'estimation rapide dont les performances permettront de résoudre des problèmes pratiques de forte dimension en des temps tout à fait raisonnables.

#### III.4. ALGORITHME RAPIDE D'ESTIMATION DES FONCTIONS DE DENSITE

##### III.4.1. ALGORITHME D'ESTIMATION CONVENTIONNEL

Lorsque le domaine  $D(X_0, \alpha_0)$  est un hypercube de côté  $\alpha_0$ , l'estimateur (3.4.) n'est en fait qu'un cas particulier de l'estimateur général de Rosenblatt-Parzen:

$$\hat{f}'(X_0) = \frac{1}{n} \sum_{j=1}^n \frac{1}{V\{D(X_0, \alpha_0)\}} \varphi\left(\frac{X_0 - X_j}{\alpha_0}\right) \quad (3.5.)$$

dans lequel la fonction  $\varphi(U)$  est appelée "noyau" de la fenêtre/1/. En effet, en prenant pour  $\varphi(U)$  la fonction:

$$\varphi(U) = \begin{cases} 1 & \text{si } |u_i| \leq 1/2, \quad i=1,2,\dots,n \\ 0 & \text{autrement} \end{cases}$$

avec:  $U = [u_1, \dots, u_i, \dots, u_n]^T$ , on retrouve l'estimateur (3.4.).

Ce noyau cubique est certainement l'un des plus utilisés dans la pratique /4//5//6/. Ce choix est généralement motivé par des considérations de programmation sur ordinateur numérique. En effet, le calcul des valeurs absolues est beaucoup plus rapide que le calcul des normes euclidiennes associées aux

noyaux de type sphérique ou exponentiel, par exemple. De plus, l'estimateur  $\hat{f}(X_0)$  de l'équation (3.4.) est obtenu très simplement en comptant le nombre d'observations situées dans l'hypercube de côté  $\alpha_0$  centré en  $X_0$ . La position d'une observation par rapport à cet hypercube peut être déterminée en testant ses composantes une à une. Il en résulte qu'un grand nombre d'observations situées à l'extérieur de cet hypercube peuvent être éliminées avant d'avoir testé toutes les composantes.

Cependant, même avec le noyau cubique, l'utilisation de cette technique d'estimation est limitée à des données de faible dimension. En effet, l'opération relativement simple et rapide d'estimation de la densité en un point doit être répétée pour tous les points où on désire estimer la densité. Le processus de discrétisation généralement adopté revient à estimer la densité en chaque noeud d'un réseau hypercubique. Le nombre d'estimations locales de la fonction est alors une fonction exponentielle de la dimension des observations et le temps d'exécution des programmes devient vite prohibitif /5/.

Il faut toutefois noter que dans le cas le plus défavorable, c'est à dire lorsqu'il y a au plus une observation par hypercube, le nombre d'hypercubes non vides est égal au nombre  $q$  d'observations. Ainsi, pour un échantillon de taille  $q$ , quelle que soit la dimension des données, l'estimation de la fonction de densité est nulle, sauf en un nombre limité de points au plus égal à  $q$ . Dans la pratique, le nombre d'estimations non nulles est généralement très nettement inférieur au nombre de points de discrétisation de la fonction de densité.

Tenant compte de cette remarque, nous proposons une procédure permettant de déterminer directement les hypercubes non vides, sans avoir à prendre en considération l'ensemble des hypercubes centrés en tous les points où la densité doit être estimée. Pour une taille d'échantillon donnée, le temps de calcul sera ainsi borné supérieurement puisque, quelle que soit la dimension du problème, il y aura au maximum  $q$  estimations locales non nulles à déterminer.

III.4.2. DETERMINATION DES HYPERCUBES NON VIDESa) Normalisation des données

Nous verrons ultérieurement que la méthode de classification automatique proposée dans ce mémoire n'est pas sensible à une transformation diagonale de l'espace  $R_0^n$  de représentation des observations. Dans un tel cas, il est généralement conseillé, avant d'envisager tout traitement des données, de normaliser les plages de variation de chacune des composantes des observations /8/. Par souci de simplicité et sans perte de généralité, l'espace  $R_0^n$  est normalisé de telle sorte que les valeurs extrêmes de chaque composante soient 0 et 1. (cf. figure 3.1.).

Soient:

$$x_i^M = \max_j x_{i,j} \quad j = 1, 2, \dots, n$$

et:

$$x_i^m = \min_j x_{i,j} \quad j = 1, 2, \dots, n$$

En faisant subir à l'espace  $R_0^n$  une translation de vecteur:

$$T = [t_1, \dots, t_i, \dots, t_n]^T$$

avec:

$$t_i = -x_i^m \quad i = 1, 2, \dots, n$$

suivie d'une transformation diagonale de matrice:

$$M = \begin{bmatrix} m_1 & & & \\ & \ddots & & \\ & & m_i & \\ & & & \ddots \\ & & & & m_n \end{bmatrix}$$

avec:

$$m_i = \frac{1}{x_i^M - x_i^m} \quad i = 1, 2, \dots, n$$

on obtient les nouvelles coordonnées:

$$\begin{aligned} X' &= M \cdot [X + T] \\ &= [x'_1, \dots, x'_i, \dots, x'_n]^T \end{aligned}$$

Dans ce nouvel espace, noté  $R_1^n$ , les coordonnées des observations:



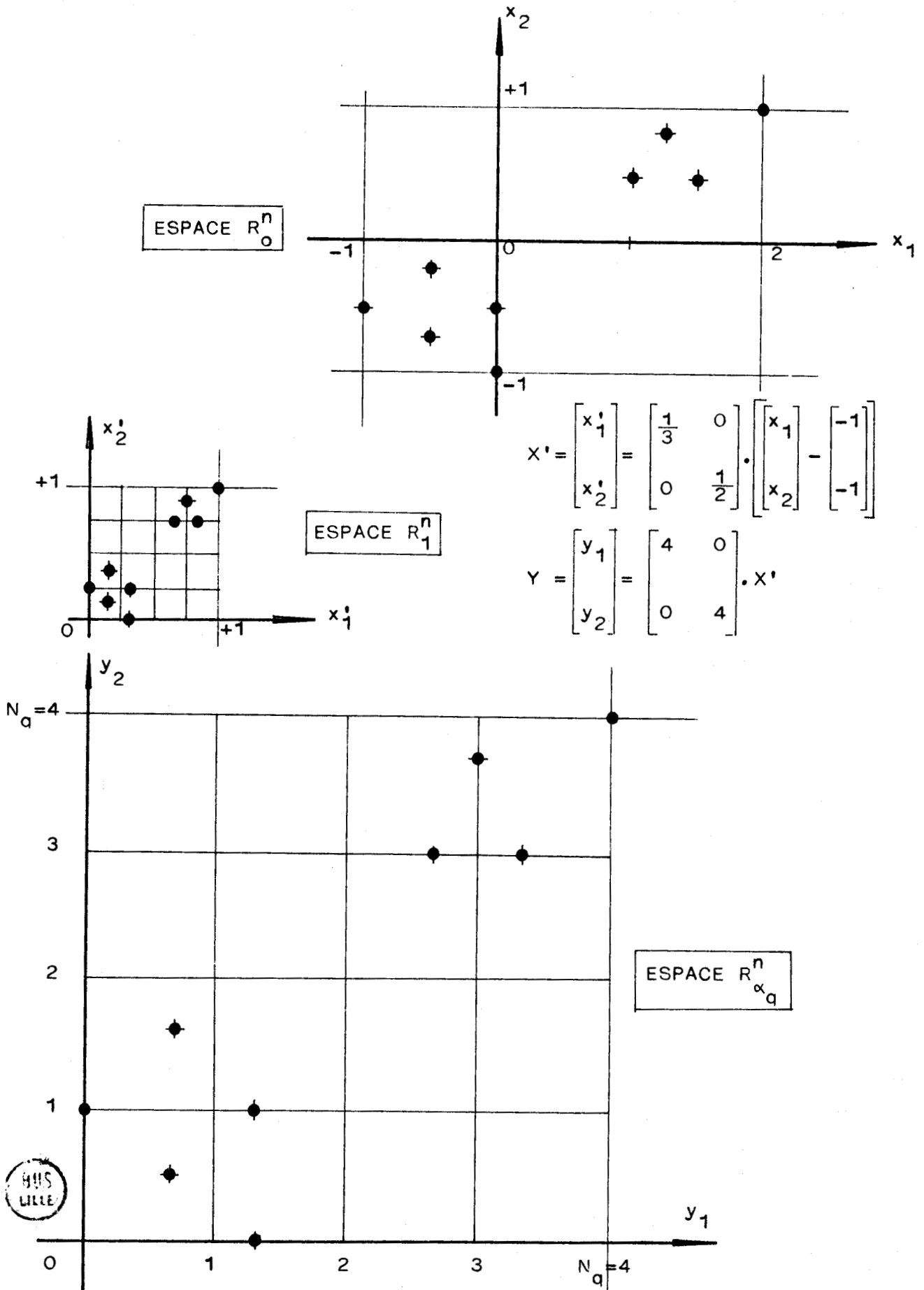


Figure 3.1. Les trois espaces de représentation des données bidimensionnelles pour  $1/\alpha_q = N_q = 4$

$$x'_j = [x'_{1,j}, \dots, x'_{i,j}, \dots, x'_{n,j}]^T, \quad j = 1, 2, \dots, q$$

sont normalisées de telle sorte que:

$$0 \leq x'_{i,j} \leq 1, \quad \forall i = 1, 2, \dots, n; \quad \forall j = 1, 2, \dots, q$$

Par souci de commodité,  $\alpha_q$  est alors choisi de la forme:

$$\alpha_q = 1 / N_q, \quad N_q : \text{entier}$$

Ainsi, en divisant chaque axe de l'espace normalisé  $R_1^n$  en  $N_q$  intervalles égaux et adjacents, on réalise une partition de cet espace en  $N_q^n$  hypercubes de côté  $\alpha_q$ . (cf. figure 3.1.).

#### b) Principe de programmation

La normalisation des données est suivie d'un changement d'échelle de l'espace  $R_1^n$ , obtenu par la transformation linéaire:

$$\begin{aligned} Y &= A.M. [X + T] \\ &= A.X' \end{aligned}$$

de matrice:

$$A = \frac{1}{\alpha_q} [I^n] = N_q [I^n]$$

où  $[I^n]$  est la matrice unité d'ordre n.

On note  $R_{\alpha_q}^n$  l'espace ainsi obtenu. Dans cette transformation élémentaire, les composantes de tout vecteur  $X'$  de l'espace  $R_1^n$  sont divisées par  $\alpha_q$  de telle sorte qu'un hypercube de côté  $\alpha_q$  de l'espace  $R_1^n$  devient un hypercube de côté unité dans l'espace  $R_{\alpha_q}^n$ . (cf. figure 3.1.)

On considère alors la partition de l'espace  $R_{\alpha_q}^n$  en hypercubes unité, transformée de la partition de l'espace  $R_1^n$  en hypercubes de côté  $\alpha_q$ . Les centres de ces hypercubes unité, dont les coordonnées sont des multiples de 0,5, forment les nœuds d'un réseau hypercubique  $\mathcal{H}_0$  de pas unité. (cf. figure 3.2.)

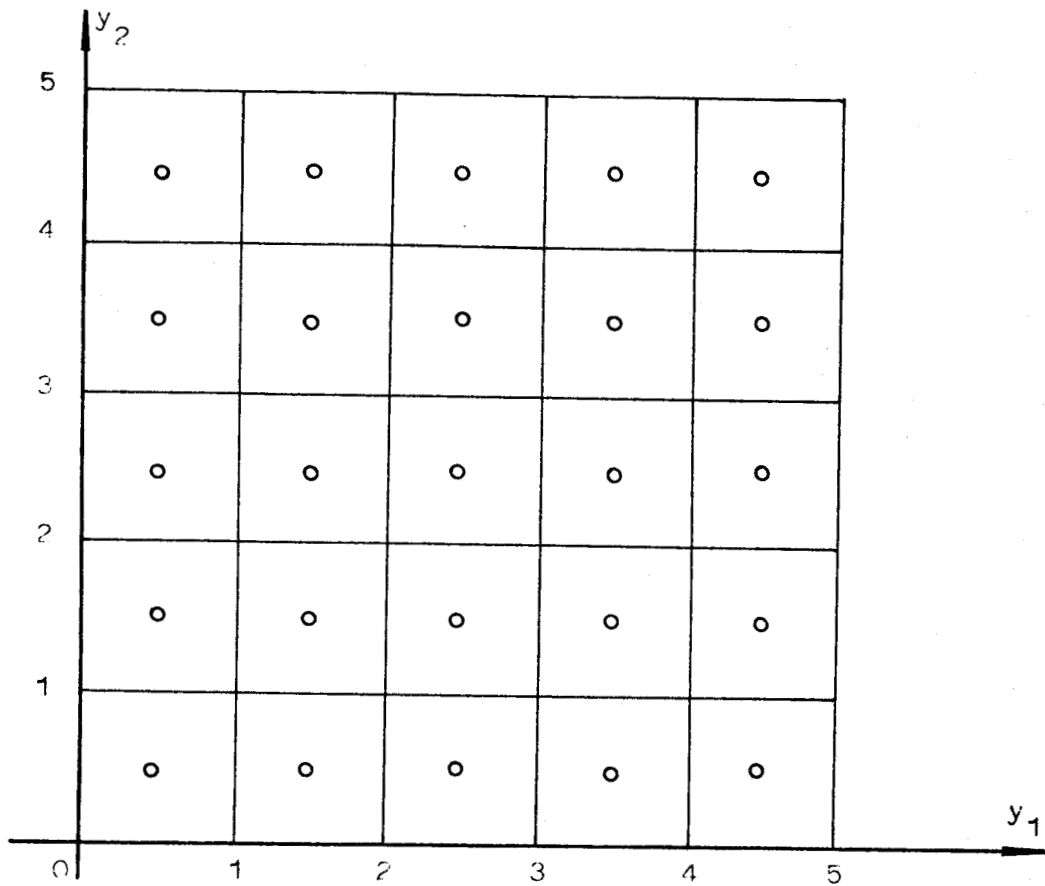


Figure 3.2. Réseau carré  $\mathcal{H}_0$  des centres des carrés de côté unité dans l'espace  $R_{\alpha_0}^2$  à deux dimensions avec  $\alpha_0 = 0,2$ .

Soit:

$$Y_j = A \cdot X_j'$$

la j<sup>ème</sup> observation dans l'espace  $R_{\alpha_0}^n$  telle que:

$$Y_j = [y_{1,j}, \dots, y_{i,j}, \dots, y_{n,j}]^T$$

Cette observation tombe dans l'hypercube unité centré au point  $H_0^j$  du réseau  $\mathcal{H}_0$  tel que:

$$H_0^j = [\text{ENT}(y_{1,j})+0,5, \dots, \text{ENT}(y_{i,j})+0,5, \dots, \text{ENT}(y_{n,j})+0,5]^T \quad (3.8.)$$

où  $\text{ENT}(y)$  est la partie entière du nombre réel positif  $y$  qui peut être obtenue directement par l'instruction "INT" du langage FORTRAN .

Il est donc aisé, sur le plan de la programmation, de dresser une liste des centres des hypercubes non vides en programmant l'équation (3.8.) et en l'appliquant à toutes les observations de l'échantillon. Toutefois, comme différentes observations peuvent définir le même hypercube non vide, il est possible que certains éléments apparaissent plusieurs fois dans cette liste brute. En fait, le nombre d'apparitions d'un même point est égal au nombre d'observations tombant à l'intérieur de l'hypercube unité centré en ce point.

En éliminant les répétitions de cette liste brute, il est donc facile de déterminer le nombre d'observations tombant dans chaque hypercube non vide. Le calcul de la valeur de l'estimateur au centre de ces hypercubes est alors immédiat. On sait, par ailleurs, que la valeur de l'estimateur au centre des hypercubes ne figurant pas dans cette liste est nulle.

#### III.4.3. ESTIMATION RAPIDE DES FONCTIONS DE DENSITE

La procédure exposée ci-dessus ne permet d'estimer la fonction de densité qu'aux noeuds du réseau hypercubique  $\mathcal{H}_0$ . Si on désire estimer la fonction en d'autres points, on est amené à modifier légèrement cette procédure en translatant le réseau  $\mathcal{H}_0$ .

Soit:

$$V = [v_1, \dots, v_i, \dots, v_n]^T$$

le vecteur définissant cette translation et soit  $\mathcal{H}(V)$  le réseau  $\mathcal{H}_0$  translaté. Les  $i$ èmes coordonnées des noeuds de ce réseau  $\mathcal{H}(V)$  sont des multiples de  $(0,5 + v_i)$ . (cf. figure 3.3.).

Comme précédemment, on peut déterminer directement l'hypercube unité centré sur un point de  $\mathcal{H}(V)$  dans lequel tombe l'observation  $Y_j$ . Le centre  $H_V^j$  de cet hypercube est défini en fonction des coordonnées de  $Y_j$  et de  $V$  par:

$$H_V^j = [ \text{ENT}(y_{1,j} - v_1) + v_1 + 0,5, \dots, \text{ENT}(y_{i,j} - v_i) + v_i + 0,5, \dots, \dots, \text{ENT}(y_{n,j} - v_n) + v_n + 0,5 ]^T$$

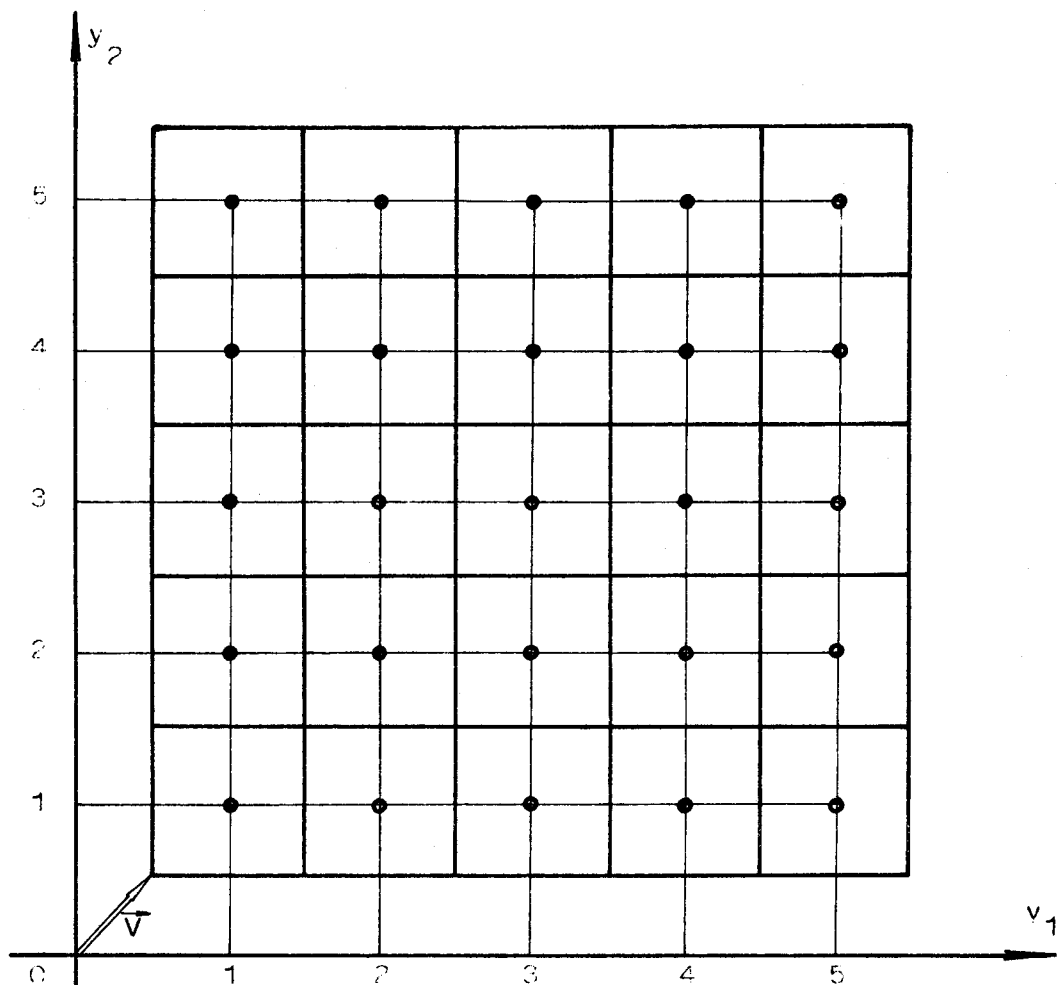


Figure 3.3. Réseau  $\mathcal{H}(v)$  obtenu par translation du réseau  $\mathcal{H}_0$  de la figure 3.2. par le vecteur :

$$v = [0,5 \ , \ 0,5]^T$$

L'utilisation de l'instruction "INT" permet d'établir la liste brute des centres  $H_V^j$ ,  $j = 1, 2, \dots, a$  des hypercubes dans lesquels tombent les observations. L'estimation de la densité au centre de ces hypercubes résulte immédiatement de l'élimination des répétitions apparaissant dans cette liste brute.

Grâce à cette procédure, il est possible d'estimer la fonction de densité en tout point d'un réseau hypercubique de pas  $1/E$ ,  $E$  entier positif inférieur à 1, couvrant l'espace  $R_{\alpha}^n$ . En effet, un tel réseau peut être obtenu en tradant  $(E^n - 1)$  fois le réseau  $\mathcal{H}_0$ . Les composantes des  $(E^n - 1)$  différents vecteurs non nuls définissant ces translations sont obtenues par tous les choix possibles de  $n$  valeurs parmi les éléments de l'ensemble des  $E$  valeurs de  $e/E$ ,  $e = 0, 1, 2, \dots, E-1$ . (cf. figure 3.4.).

La méthode proposée ne limite donc pas la finesse du pas de discrétisation de la fonction de densité qui peut être estimée aux nœuds d'un réseau aussi dense qu'on le désire.

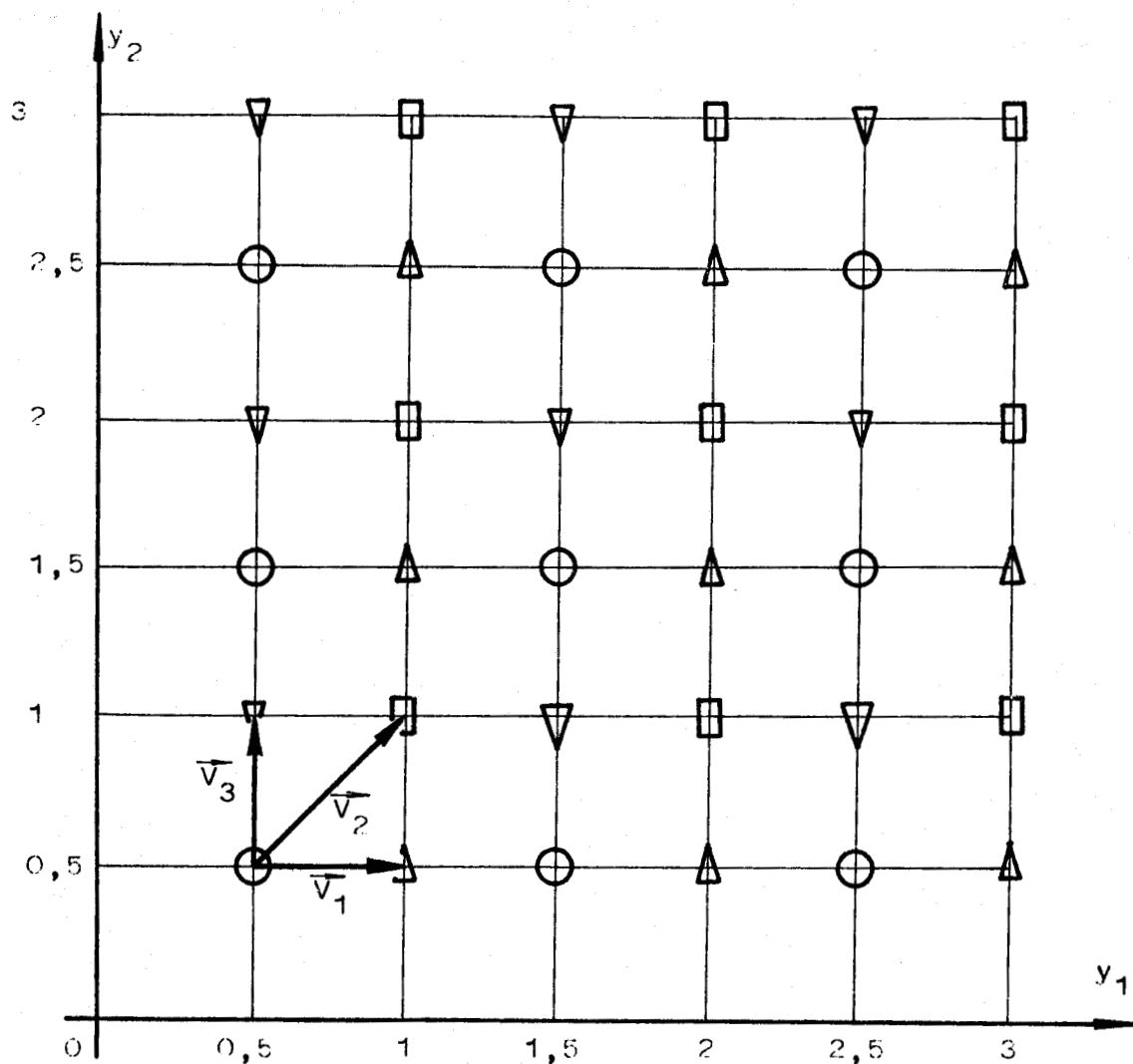


Figure 3.4. Les quatre réseaux de pas unitaire nécessaires pour construire un réseau de pas 0,5 dans l'espace  $R_{\alpha_n}^2$ . ( $\alpha_n = 1/3$ )

- Nœuds du réseau  $\mathcal{H}_0$
- △ Nœuds du réseau  $\mathcal{H}(v_1)$ ,  $v_1 = [0,5, 0]^T$
- Nœuds du réseau  $\mathcal{H}(v_2)$ ,  $v_2 = [0,5, 0,5]^T$
- ▽ Nœuds du réseau  $\mathcal{H}(v_3)$ ,  $v_3 = [0, 0,5]^T$

### III.5. COMPARAISON DES ALGORITHMES D'ESTIMATION RAPIDE ET CONVENTIONNEL

#### III.5.1. ANALYSE DES ALGORITHMES

##### a) Opération de base

La procédure proposée est fondamentalement différente de la technique d'estimation conventionnelle dans laquelle il est nécessaire d'identifier les observations tombant dans l'ensemble des hypercubes centrés en tous les points où la fonction doit être estimée. Pour comparer les performances des deux algorithmes, il est commode de définir une opération élémentaire de base, commune aux deux procédures. Soient :

$$\text{et: } Y_1 = [y_{1,1}, \dots, y_{i,1}, \dots, y_{n,1}]^T$$

$$Y_2 = [y_{1,2}, \dots, y_{i,2}, \dots, y_{n,2}]^T$$

deux points de l'espace  $R_{\alpha_0}^n$ . La norme associée aux domaines d'observation hypercubiques étant :

$$\|Y\| = \max \{|y_1|, \dots, |y_i|, \dots, |y_n|\}$$

la distance entre  $Y_1$  et  $Y_2$  est :

$$D(Y_1, Y_2) = \|Y_1 - Y_2\|$$

$$= \max_i |y_{i,1} - y_{i,2}|$$

L'opération de base consiste à tester si :

$$D(Y_1, Y_2) > d/2, \quad d \geq 0$$

Lorsque  $d$  est strictement positif, cette opération élémentaire permet de savoir si le point  $Y_2$  tombe dans l'hypercube de côté  $d$  centré en  $Y_1$ . Si  $d$  est nul, l'opération de base est simplement un test pour savoir si  $Y_1$  et  $Y_2$  sont deux points identiques. Dans les deux cas elle nécessite au plus  $n$  comparaisons entre les coordonnées des deux points. Comme nous l'avons déjà noté (cf page III.8.), ce nombre  $n$  constitue une borne supérieure du nombre de comparaisons effectivement nécessaires dans ce type d'opération.

b) Algorithme rapide

Une fois effectué le changement d'échelle de matrice  $A$  de l'équation (3.7.), l'algorithme rapide détermine l'hypercube unité non vide associé à chaque observation. Il s'agit essentiellement d'une opération de codage des  $n$  composantes de chacune des  $a$  observations. Dans l'étape suivante, les centres des hypercubes ainsi définis sont comparés les uns aux autres pour éliminer les répétitions. Cette opération nécessite au maximum:

$$a + (a-1) + (a-2) + \dots + 2 + 1 = \frac{a(a+1)}{2}$$

opérations de base avec  $d=0$ . A la fin de cette étape, on connaît le nombre d'observations tombant dans chacun des hypercubes non vides centrés aux noeuds du réseau  $\mathcal{H}_0$ . On détermine immédiatement la valeur de l'estimateur en ces points. Pour revenir à l'espace normalisé  $R_1^n$ , il suffit d'effectuer la transformation inverse de matrice  $A^{-1}$  sur les points du réseau  $\mathcal{H}_0$  où la densité estimée est différente de zéro.

Sachant qu'une multiplication et qu'une opération de codage sont beaucoup plus rapides que l'opération de base, on peut majorer le temps d'exécution des  $2nq$  multiplications associées aux transformations de matrice  $A$  et  $A^{-1}$  et des  $na$  codages par celui de  $3nq$  opérations de base. On peut donc affirmer que le temps de calcul de l'algorithme rapide ne dépasse pas celui de:

$$Q_1 = 3nq + \frac{q(a+1)}{2} = q \left( 3n + \frac{a+1}{2} \right) \quad (3.9.)$$

opérations de base.

Si on désire estimer la densité aux noeuds d'un réseau plus dense, de pas  $\alpha_q/E$ ,  $E$  entier, il faut translater  $(E^n - 1)$  fois le réseau  $\mathcal{H}_0$  dans l'espace  $R_{\alpha_q}^n$ . Le temps d'exécution de l'algorithme est alors majoré par  $E^n \cdot Q_1$  fois le temps d'exécution moyen de l'opération de base.

c) Algorithme conventionnel

Pour le même pas de discrétisation  $\alpha_q$ , l'algorithme conventionnel fait intervenir les  $(N_{\alpha_q})^n$  hypercubes



élémentaires constituant la partition de l'espace  $R_1^n$ . Il nécessite donc :

$$Q_2 = \alpha \cdot \left( \frac{1}{\alpha_\alpha} \right)^n \quad (3.10.)$$

opérations de base avec  $d = \alpha_\alpha$ . Si, comme précédemment, on désire diviser le pas de discrétisation par un entier  $E$ , l'algorithme nécessite alors  $E^n Q_2$  opérations de base.

d) Comparaison des temps d'exécution

De (3.9.) et (3.10.) on conclut que le rapport des temps d'exécution moyens des deux algorithmes est inférieur à :

$$\frac{Q_1}{Q_2} = \alpha_\alpha^n \cdot \left( 3n + \frac{\alpha + 1}{2} \right) \quad (3.11.)$$

Pour donner un ordre de grandeur, prenons un problème à 5 dimensions pour lequel on dispose d'un échantillon de 1000 observations. Si on fixe  $\alpha_\alpha = 0,1$ , on peut affirmer que l'algorithme proposé sera au minimum 200 fois plus rapide que l'algorithme conventionnel.

Cette estimation du rapport des temps d'exécution des deux algorithmes est en fait assez grossière car nous nous sommes placé dans le cas le plus défavorable pour analyser la complexité de l'algorithme rapide. En effet, le nombre d'hypercubes non vides dépend de la configuration de la distribution spatiale des observations et peut être nettement inférieur à  $\alpha$ . De plus, les opérations sur les entiers mises en jeu dans l'algorithme proposé sont plus rapides que les opérations en point flottant de l'algorithme conventionnel. Dans ces conditions, il est impossible de comparer les temps de calcul avec précision. Pour compléter cette analyse, nous présentons donc quelques simulations afin de chiffrer les temps d'exécution respectifs des deux algorithmes sous différentes conditions.

### III.5.2. PERFORMANCES DES ALGORITHMES

Les données utilisées pour comparer les performances de l'algorithme proposé et de l'algorithme conventionnel sont des

vecteurs aléatoires issus de deux distributions normales  
 $P(\bar{x}_i^n, \Sigma_i^n)$ ,  $i = 1, 2$ , où:

$$\bar{x}_1^n = [+2, +2, \dots, +2]^T$$

$$\bar{x}_2^n = \underbrace{[-2, -2, \dots, -2]}_{n \text{ fois}}^T$$

et:

$$\Sigma_1^n = \Sigma_2^n = [I^n]$$

Les temps d'exécution des deux algorithmes ont été comparés pour des données de différentes dimensions et pour des échantillons de tailles variées en gardant, pour tous les essais, la même valeur de  $\alpha_n$ . Celle-ci a été fixée à  $\alpha_n = 0,1$ . Cela revient à discrétiser la densité de probabilité en divisant en  $n_n = 10$  intervalles égaux la plage de variation normalisée de chacune des composantes des observations.

Il faut noter que ce choix arbitraire d'une valeur fixe pour  $\alpha_n$  ne permet pas d'assurer la convergence de l'estimateur lorsque la taille de l'échantillon augmente. De plus, cette procédure ne conduit pas, en général, à une estimation satisfaisante des fonctions de densité. Cependant, nous ne traitons pas, à ce niveau, le problème du choix de  $\alpha_n$  et, sans aucune information supplémentaire, aucune valeur n'est a priori meilleure que les autres. Les simulations que nous présentons n'ont d'autre but que de mettre en évidence les effets de la dimension des données et de la taille de l'échantillon sur les temps d'exécution des deux algorithmes, toutes choses restant égales par ailleurs.

Ces temps sont représentés figure 3.5. pour trois tailles d'échantillon ( $n = 100$ ,  $n = 500$  et  $n = 1000$ ) et pour  $n$  variant de 2 à 6. Comme les ordres de grandeur sont très différents, les temps de calcul ont été reportés en coordonnées logarithmiques.

La croissance exponentielle du temps d'exécution de l'algorithme conventionnel en fonction de la dimension du problème confirme les conclusions de l'analyse précédente.

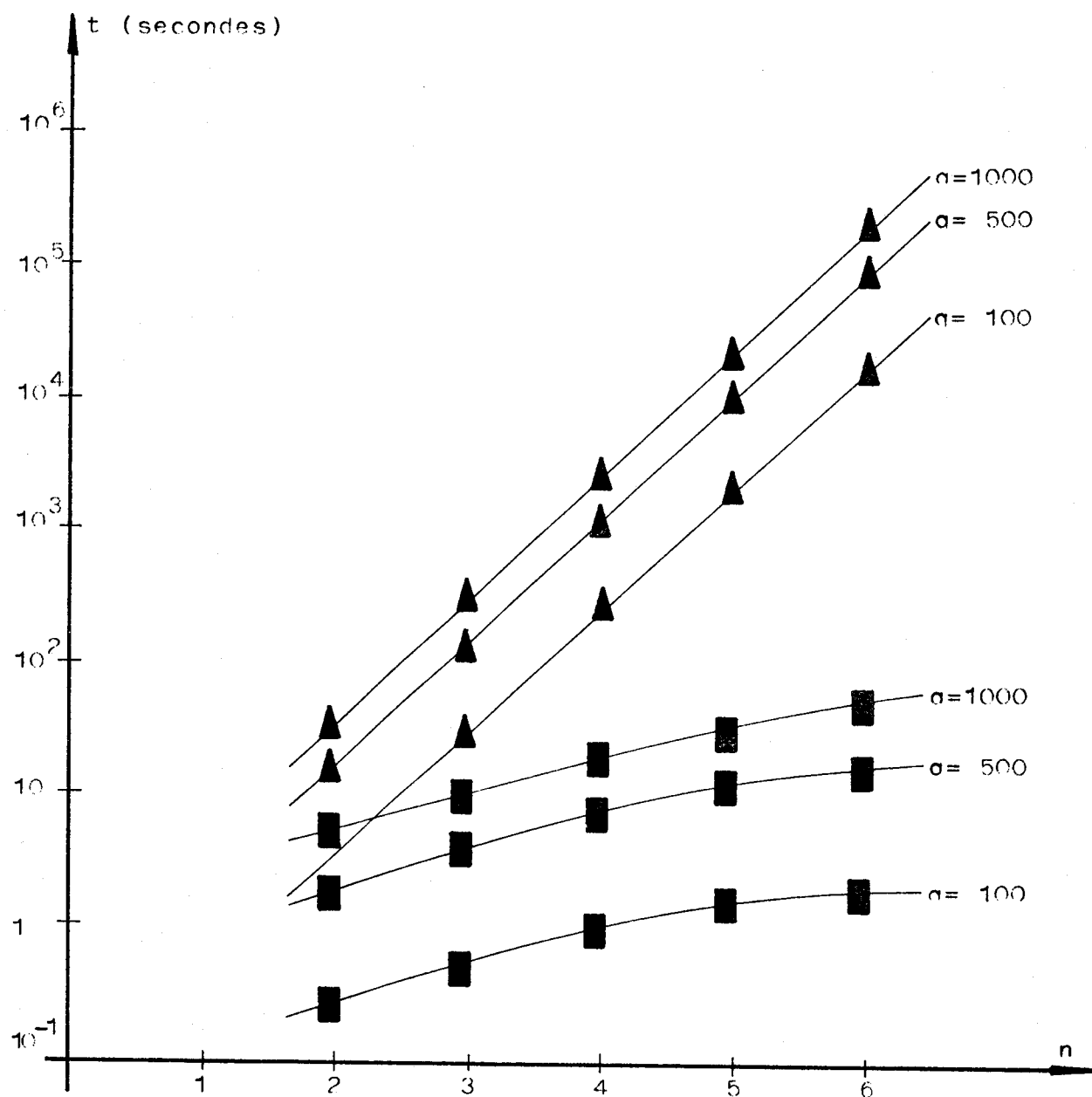


Figure 3.5. Temps d'exécution des deux algorithmes pour  $\alpha = 100$ ,  $\alpha = 500$ ,  $\alpha = 1000$  et pour  $n$  variant de 2 à 6.

- Algorithme rapide
- ▲ Algorithme conventionnel



Pour une taille d'échantillon donnée, la croissance du temps de calcul de l'algorithme rapide avec la dimension résulte du mode de génération des données artificielles. En effet, les deux distributions normales utilisées pour ces simulations ont des matrices de covariance unité, quelle que soit la dimension des données. Ainsi, pour une taille donnée de l'échantillon, la densité des observations dans chacun des deux groupements correspondant à ces deux classes tend à décroître lorsque la dimension augmente. En conséquence, le nombre d'hypercubes non vides tend à augmenter, sans toutefois pouvoir dépasser  $\alpha$ . La figure 3.6. indique l'augmentation enregistrée pour les échantillons de 100, 500 et 1000 observations utilisés pour les simulations. L'allure de cette croissance explique l'évolution du temps d'exécution de l'algorithme rapide avec la dimension. Il faut cependant noter que cette croissance est beaucoup plus lente que celle enregistrée pour l'algorithme conventionnel. On sait, de plus, qu'elle est bornée supérieurement.

Ainsi, la grande différence de temps d'exécution entre les deux algorithmes tend à augmenter avec la dimension du problème traité. Par exemple, pour  $q = 1000$ , les temps d'exécution sont respectivement de 2 et de 20 secondes pour le cas bidimensionnel, alors qu'ils sont de 30 secondes et de 5 heures et demie pour des données à 5 dimensions. On pourra rapprocher ces résultats des conclusions de l'analyse présentée au paragraphe précédent.

L'algorithme rapide proposé permet donc d'obtenir l'estimation des fonctions de densité par la méthode du noyau cubique avec un gain de temps très appréciable par rapport à la procédure conventionnelle. Cet algorithme ouvre donc un vaste champ d'application à la méthode d'estimation de Rosenblatt-Parzen qui peut désormais être utilisée pour résoudre des problèmes de dimension élevée. Nous allons maintenant appliquer cet algorithme à l'analyse de la convexité locale des fonctions de densité.

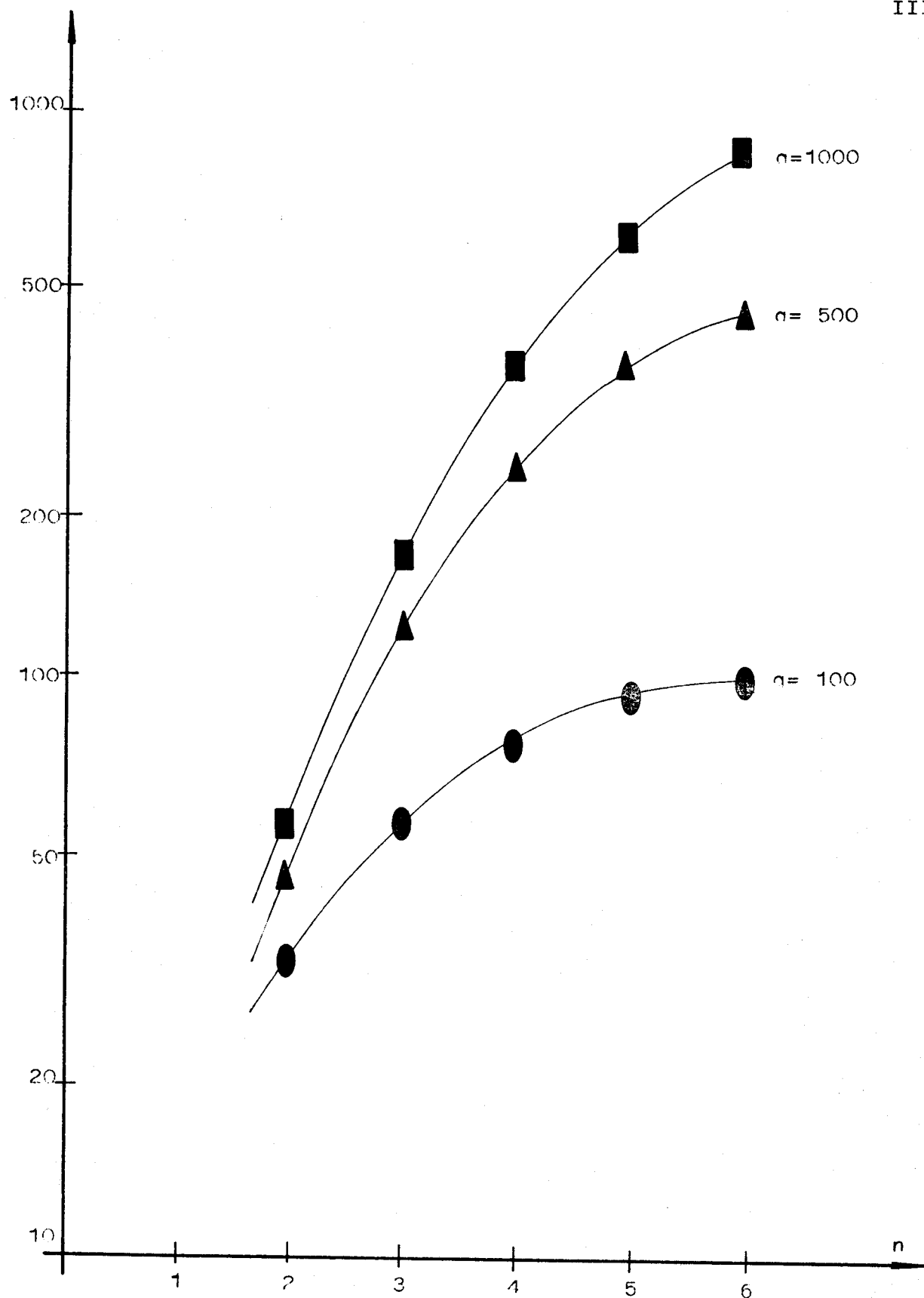


Figure 3.6. Nombre d'hypercubes non vides pour des échantillons de 100, 500 et 1000 observations en fonction de la dimension des données.

### III.6. ALGORITHME DE TEST DE LA CONVEXITE LOCALE DES FONCTIONS DE DENSITE

La mise en œuvre sur ordinateur du test de convexité présenté au paragraphe III.3. nécessite en premier lieu l'estimation de la fonction de densité soumise à l'analyse. Nous utiliserons donc l'algorithme d'estimation rapide en ajustant  $\alpha_n$  en fonction de la taille de l'échantillon et de la nature des observations. Dans tout ce qui suit, le paramètre  $\alpha_n$ , inverse d'un entier, sera pris le plus proche possible de:

$$\alpha'_n = \alpha_0 \cdot n^{-\gamma/n}, \gamma \in ]0, 1[ \quad (3.12.)$$

Ce choix permet de respecter les contraintes de programmation ainsi que l'équation (3.2.) lorsque les domaines utilisés pour l'estimation de la fonction de densité sont des hypercubes. Pratiquement, il est d'usage de fixer  $\gamma$ , généralement à la valeur  $\gamma=1/2$ , et d'ajuster  $\alpha'_n$  en jouant sur le paramètre  $\alpha_0$ . Mais le choix de  $\alpha_0$  demeure essentiellement arbitraire et nous ne disposons pas encore, à ce stade de l'étude, d'une méthode permettant d'optimiser la valeur de ce paramètre fondamental. C'est seulement lorsque nous aborderons les problèmes de classification automatique, au chapitre V, que nous pourrons proposer des critères objectifs qui serviront à l'ajustement de la taille des domaines d'observation.

Pour tester le sens de la convexité locale de la fonction de densité, on doit ensuite estimer sa valeur moyenne sur des hypercubes de côté supérieur à  $\alpha_n$  et centrés aux points où la fonction a été estimée. Pour des raisons de rapidité de calcul, ces valeurs moyennes sont estimées sur des domaines hypercubiques de côté  $3\alpha_n$ , constitués de  $3^n$  hypercubes élémentaires de côté  $\alpha_n$ . (cf. figure 3.7.) Le nombre d'observations tombant dans ces domaines est alors rapidement obtenu comme la somme du nombre d'observations tombant dans chacun des hypercubes élémentaires. /12/.

La détermination du sens de la convexité locale de la fonction de densité est alors immédiate puisque l'on dispose, en chaque point où ce sens doit être testé, des estimations de la fonction et de sa valeur moyenne sur un domaine d'observation

centré en ce point. Il faudra tenir compte, lors des applications pratiques, du fait que la validité du résultat de ce test n'est assurée que si la fonction étudiée est localement convexe (ou concave) dans un voisinage hypercubique du point de test de côté  $3\alpha_n$ .

Dans la suite de ce travail, nous nous appliquerons surtout à déterminer les domaines où une fonction est concave. De tels domaines seront matérialisés par des ensembles d'hypercubes adjacents de côté  $\alpha_n$ , centrés aux points de test, à l'intérieur desquels le résultat du test de convexité indique une fonction localement concave. Ces hypercubes seront appelés "hypercubes à convexité négative", par opposition aux "hypercubes à convexité positive" pour lesquels le test indique que la fonction est localement convexe.

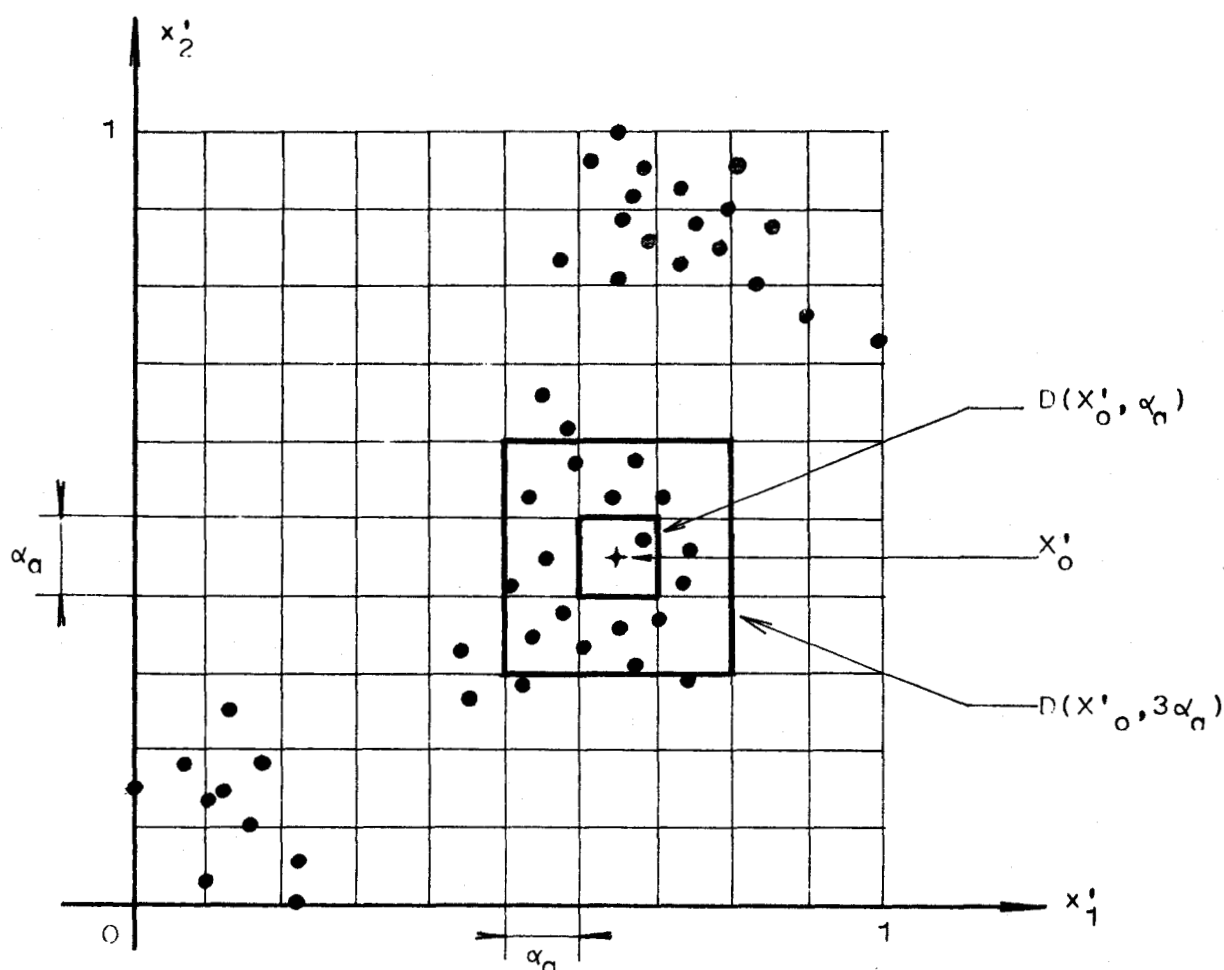


Figure 3.7. Domaines de l'espace  $R_2^2$  utilisés pour tester la convexité d'une fonction de densité en un point  $x'_0$ .

III.7. CONCLUSION

Nous avons mis au point un algorithme rapide d'estimation des fonctions de densité de probabilité par la méthode du noyau cubique de Rosenblatt-Parzen. Utilisé pour tester le sens de la convexité locale des fonctions de densité à partir des observations, cet algorithme permettra de traiter des problèmes de forte dimension avec un temps de calcul relativement réduit.

Il faut toutefois garder à l'esprit le fait que l'application de cette procédure ne permet de déterminer le sens de la convexité locale de la fonction analysée que si celle-ci est définie dans un certain voisinage des points de test. Avant d'envisager l'utilisation systématique de ce test dans les problèmes de classification automatique, il importe donc de connaître les propriétés de convexité des fonctions de densité rencontrées dans ces problèmes. C'est dans cet esprit que nous abordons, dans la seconde partie de ce travail, une étude détaillée de la convexité des fonctions de densité de probabilité normales.



REFERENCES (chapitre III)

- /1/ BANON, G.  
 "Estimation non paramétrique de densité de probabilité pour les processus de Markov"  
 Thèse d'état, Univ. P. Sabatier, Toulouse 1ère partie (1977)
- /2/ CACOULLOS, T.  
 "Estimation of a multivariate density"  
 Ann. Inst. Stat. Math. Vol. 18 p. 179-189 (1966)
- /3/ COVER, T.M. et HART, P.E.  
 "Nearest neighbor pattern classification"  
 IEEE Trans. Info. Theory Vol. IT-13 p. 21-27 (1967)
- /4/ GRABLICKI, W.  
 "Pattern recognition procedure with non parametric density estimates"  
 IEEE Trans. on Syst. Man & Cyb. Vol. SMC-8, n°11 p. 809-812 (1978)
- /5/ KITTLER, J.  
 "A locally sensitive method for cluster analysis"  
 Pattern Recog. Vol. 8 p. 23-33 (1976)
- /6/ KOONTZ, W.L.G. et FUKUNAGA, K.  
 "Asymptotic analysis of a nonparametric clustering technique"  
 IEEE Trans. on Computers Vol. C-21, n°9 p. 967-974 (1972)
- /7/ LOFTSGAARDEN, D.O. et QUESENBERNY, C.P.  
 "A nonparametric estimate of a multivariate density function"  
 Ann. Math. Stat. Vol. 36 p. 1049-1051 (1965)
- /8/ MEISEL, W.S.  
 "Computer-oriented approaches to pattern recognition"  
 Academic Press, New York (1972)
- /9/ MURTHY, V.K.  
 "Nonparametric estimation of multivariate densities with applications"  
 Multivariate Analysis Academic Press, New York p. 43-56 (1966)
- /10/ PARZEN, E.  
 "On estimation of a probability density function and mode"  
 Ann. Math. Stat. Vol. 33 p. 1065-1076 (1962)
- /11/ POSTAIRE, J.G. et VASSEUR, C.  
 "Application of the geometrical properties of convex functions to unsupervised classification"  
 Info II. 2nd Int. Conf. on Inf. Sc. & Syst. Patras Grèce (1979)

- /12/ POSTAIRE, J.G. et VASSEUR, C.  
"Analyse locale de la forme des fonctions multivariables.  
Application en classification automatique"  
Congrès AFCET-IRIA Reconnaissance des Formes et Intelligence  
Artificielle Toulouse p. 348-355 (1979)
- /13/ ROSENBLATT, M.  
"Remarks on some nonparametric estimates of a density function"  
Ann. Math. Stat. Vol. 27 p. 832-837 (1956)
- /14/ VAN RYZIN, J.  
"On strong consistency of density estimates"  
Ann. Math. Stat. Vol 40 p. 1765-1772 (1969)
- /15/ WAGNER, T.J.  
"Convergence of the nearest neighbor rule"  
IEEE Trans. on Info. Theory Vol. IT-17 p. 566-571 (1971)
- /16/ WEGMAN, E.J.  
"Non parametric probability density estimation.  
I. A summary of available methods"  
Technometrics Vol. 14 p. 533-546 (1972)

DEUXIEME PARTIE

OPTIMISATION DU PROCESSUS  
DE CLASSIFICATION

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

DETERMINATION DES PARAMETRES D'UNE DISTRIBUTION NORMALE PAR ANALYSE DE LA CONVEXITE DE SA FONCTION DE DENSITE
--

### IV.1. INTRODUCTION

Les résultats présentés dans la première partie de ce travail permettent de déterminer la convexité locale de la fonction de densité d'une distribution en n'utilisant aucune autre information que celle apportée par les observations.

Nous allons maintenant montrer comment exploiter la procédure proposée au chapitre précédent pour déterminer les paramètres de la fonction de densité d'une distribution normale d'observations multivariées. Cette étape constitue le premier pas vers l'optimisation du processus de classification par analyse des mélanges gaussiens.

### IV.2. LES FONCTIONS DE DENSITE DE PROBABILITE NORMALES

Les fonctions de densité normales multivariées sont de la forme :

$$p(x) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left\{ -1/2 (x-\bar{x})^T \Sigma^{-1} (x-\bar{x}) \right\} \quad (4.1.)$$

où:

$$X = [x_1, \dots, x_i, \dots, x_n]^T$$

$\bar{X}$  est le vecteur-moyenne de la distribution:

$$\begin{aligned} \bar{X} &= \mathbb{E}\{X\} \\ &= [\bar{x}_1, \dots, \bar{x}_i, \dots, \bar{x}_n]^T \end{aligned}$$

avec  $\bar{x}_i = \mathbb{E}\{x_i\}$ ,  $i = 1, 2, \dots, n$ ; le symbole  $\mathbb{E}\{x\}$  désignant l'espérance mathématique de la variable aléatoire  $x$ .

$\Sigma$  est la matrice de covariance:

$$\Sigma = \begin{bmatrix} \sigma_{1,1} & & & \sigma_{1,n} \\ & \ddots & & \\ & & \sigma_{i,i} & \\ & & & \ddots \\ \sigma_{n,1} & & & & \sigma_{n,n} \end{bmatrix}$$

avec:

$$\sigma_{i,j} = \mathbb{E}\{(x_i - \bar{x}_i)(x_j - \bar{x}_j)\}$$

La matrice de covariance est toujours symétrique et semi-définie positive /1/. Nous ne considérerons cependant que les cas où elle est définie positive, excluant ainsi toute situation où l'ensemble des observations appartiendrait à un sous-espace de  $R^n$ .

La fonction de densité normale  $p(X)$  est complètement déterminée par  $n + \frac{n(n+1)}{2}$  paramètres, à savoir les  $n$  éléments du vecteur moyenne  $\bar{X}$  et les  $\frac{n(n+1)}{2}$  éléments indépendants de la matrice de covariance  $\Sigma$ . Deux approches classiques permettent de déterminer ces paramètres à partir des observations. L'une est basée sur les techniques d'apprentissage bayésien /2/, l'autre est une méthode d'estimation par maximum de vraisemblance /3/. Cependant, comme nous l'avons noté au cours du premier chapitre, l'application de ces deux approches à l'analyse des mélanges gaussiens ne peut être envisagée que sous des hypothèses restrictives ou nécessite certaines informations a priori qui ne sont pas toujours disponibles.

C'est pourquoi nous proposons une nouvelle approche au problème de la détermination des paramètres d'une

distribution normale dont l'intérêt majeur réside dans ses possibilités d'application à l'analyse des mélanges gaussiens. La méthode proposée tire essentiellement parti des relations simples existant entre les paramètres recherchés et les caractéristiques géométriques du plus grand domaine à l'intérieur duquel la fonction de densité  $p(X)$  est concave.

#### IV.3. CONVEXITE DES FONCTIONS DE DENSITE NORMALES

##### IV.3.1. DIAGONALISATION DE LA MATRICE DE COVARIANCE

Pour simplifier la présentation des calculs, mais sans aucune perte de généralité, considérons une fonction de densité normale à vecteur moyenne nul:

$$\begin{aligned} p(X) &= \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} X^T \Sigma^{-1} X \right\} \\ &= A \exp \left\{ -\frac{1}{2} X^T \Sigma^{-1} X \right\} \end{aligned}$$

avec:

$$A = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}}$$

Si le vecteur-moyenne  $\bar{X}$  de la distribution considérée n'est pas nul, une simple translation de l'origine de vecteur  $\bar{X}$  permet de ramener le problème à celui d'une distribution à vecteur-moyenne nul.

Considérons la forme quadratique réelle:

$$Q = X^T \Sigma^{-1} X$$

Comme la matrice  $\Sigma^{-1}$  est réelle et symétrique, elle peut toujours être diagonalisée par une transformation orthogonale /4/. En choisissant les vecteurs propres de la matrice  $\Sigma^{-1}$  comme nouveaux vecteurs de base, le changement de base:

$$X = M Y \tag{4.2.}$$

permet d'écrire  $Q$  sous la forme:

$$Q = Y^T M^T \Sigma^{-1} M Y \tag{4.3.}$$

$$= Y^T \Lambda Y \tag{4.4.}$$

où:

$$\Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_i & \\ & & & \ddots \\ & & & & \lambda_n \end{bmatrix}$$

est une matrice diagonale.

Dans cette nouvelle base, la fonction de densité considérée prend la forme particulière:

$$p(Y) = A \exp \left\{ -\frac{1}{2} Y^T \Lambda Y \right\} \quad (4.5.)$$

qui se prête particulièrement bien à l'analyse de sa convexité.

#### IV.3.2. CONDITION DE CONVEXITE

Comme les dérivées partielles du second ordre de  $p(Y)$  sont continues quel que soit  $Y$ , on peut utiliser le théorème 2.2. La recherche du plus grand domaine  $D$  à l'intérieur duquel la fonction  $p(Y)$  est concave se ramène alors à la détermination du plus grand domaine convexe de l'espace où la forme quadratique:

$$F = Y^T \cdot H[p(Y)] \cdot Y$$

est semi-définie négative.

Rappelons que  $H[p(Y)]$  est le hessien de  $p(Y)$  (cf. chapitre II, paragraphe II.2.)

$$H[p(Y)] = \begin{bmatrix} p_{11}(Y) & & & p_{1n}(Y) \\ & \ddots & & \\ & & p_{ii}(Y) & \\ & & & \ddots \\ p_{m1}(Y) & & & p_{nn}(Y) \end{bmatrix}$$

avec:

$$p_{ij}(Y) = \frac{\partial^2 p(Y)}{\partial y_i \partial y_j}$$

Compte-tenu de (4.5.),  $H[p(Y)]$  s'écrit:

$$\begin{aligned} H[p(Y)] &= H\left[A \cdot \exp\left\{-\frac{1}{2} Y^T \Lambda Y\right\}\right] \\ &= A \cdot H\left[\exp\left\{-\frac{1}{2} Y^T \Lambda Y\right\}\right] \end{aligned}$$

La constante  $A$  étant positive, le signe de la forme quadratique  $F$  est le même que celui de la forme:

$$F' = Y^T \cdot H\left[\exp\left\{-\frac{1}{2} Y^T \Lambda Y\right\}\right] \cdot Y$$

qui peut être évaluée grâce aux relations:

$$\frac{\partial^2 \exp\left\{-\frac{1}{2} Y^T \Lambda Y\right\}}{\partial y_i^2} = \left[ \left(\frac{\partial Y^T}{\partial y_i} \Lambda Y\right)^2 - \frac{\partial Y^T}{\partial y_i} \Lambda \frac{\partial Y}{\partial y_i} \right] \exp\left\{-\frac{1}{2} Y^T \Lambda Y\right\} \quad 0/2$$

et:

$$\frac{\partial^2 \exp\left\{-\frac{1}{2} Y^T \Lambda Y\right\}}{\partial y_i \partial y_j} = \left[ \left(\frac{\partial Y^T}{\partial y_i} \Lambda Y\right) \left(\frac{\partial Y^T}{\partial y_j} \Lambda Y\right) - \frac{\partial Y^T}{\partial y_i} \Lambda \frac{\partial Y}{\partial y_j} \right] \exp\left\{-\frac{1}{2} Y^T \Lambda Y\right\}$$

Le hessien  $H\left[\exp\left\{-\frac{1}{2} Y^T \Lambda Y\right\}\right]$  prend alors la forme:

$$H\left[\exp\left\{-\frac{1}{2} Y^T \Lambda Y\right\}\right] = (H_1 + H_2) \cdot \exp\left\{-\frac{1}{2} Y^T \Lambda Y\right\}$$

avec:

$$\begin{aligned} H_1 &= \begin{bmatrix} \left(\frac{\partial Y^T}{\partial y_1} \Lambda Y\right)^2 & & \left(\frac{\partial Y^T}{\partial y_1} \Lambda Y\right) \left(\frac{\partial Y^T}{\partial y_n} \Lambda Y\right) \\ & \left(\frac{\partial Y^T}{\partial y_i} \Lambda Y\right)^2 & \\ \left(\frac{\partial Y^T}{\partial y_n} \Lambda Y\right) \left(\frac{\partial Y^T}{\partial y_1} \Lambda Y\right) & & \left(\frac{\partial Y^T}{\partial y_n} \Lambda Y\right)^2 \end{bmatrix} \\ H_2 &= \begin{bmatrix} \frac{\partial Y^T}{\partial y_1} \Lambda \frac{\partial Y}{\partial y_1} & & \frac{\partial Y^T}{\partial y_1} \Lambda \frac{\partial Y}{\partial y_n} \\ & \frac{\partial Y^T}{\partial y_i} \Lambda \frac{\partial Y}{\partial y_i} & \\ \frac{\partial Y^T}{\partial y_n} \Lambda \frac{\partial Y}{\partial y_1} & & \frac{\partial Y^T}{\partial y_n} \Lambda \frac{\partial Y}{\partial y_n} \end{bmatrix} \end{aligned}$$



En introduisant la matrice:

$$\left[ \frac{\partial Y^T}{\partial y_1}, \dots, \frac{\partial Y^T}{\partial y_i}, \dots, \frac{\partial Y^T}{\partial y_n} \right]^T = I_n$$

(  $I_n$  matrice unité d'ordre  $n$  ),  $H_1$  peut s'écrire:

$$\begin{aligned} H_1 &= (I_n \wedge Y)(I_n \wedge Y)^T \\ &= (\Lambda Y)(\Lambda Y)^T \end{aligned}$$

De manière analogue:

$$\begin{aligned} H_2 &= I_n \wedge I_n^T \\ &= \Lambda \end{aligned}$$

Ainsi:

$$H \exp\left\{-\frac{1}{2} Y^T \Lambda Y\right\} = [(\Lambda Y)(\Lambda Y)^T - \Lambda] \exp\left\{-\frac{1}{2} Y^T \Lambda Y\right\}$$

La détermination de la convexité de la fonction  $p(X)$  se ramène finalement à l'étude du signe de la forme quadratique:

$$F'' = Y^T [(\Lambda Y)(\Lambda Y)^T - \Lambda] Y$$

#### IV.3.3. DOMAINE DE CONCAVITE

Soient  $\Delta_1, \Delta_2, \dots, \Delta_i, \dots, \Delta_n$ , les mineurs principaux de la matrice:

$$H_r = (\Lambda Y)(\Lambda Y)^T - \Lambda$$

$$\Delta_1 = \lambda_1^2 y_1^2 - \lambda_1$$

$$\Delta_2 = \begin{vmatrix} \lambda_1^2 y_1^2 - \lambda_1 & \lambda_1 \lambda_2 y_1 y_2 \\ \lambda_2 \lambda_1 y_2 y_1 & \lambda_2^2 y_2^2 - \lambda_2 \end{vmatrix}$$

$$\Delta_i = \begin{vmatrix} \lambda_1^2 y_1^2 - \lambda_1 & \lambda_1 \lambda_2 y_1 y_2 & \dots & \lambda_1 \lambda_i y_1 y_i \\ \lambda_2 \lambda_1 y_2 y_1 & \lambda_2^2 y_2^2 - \lambda_2 & \dots & \lambda_2 \lambda_i y_2 y_i \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_i \lambda_1 y_i y_1 & \dots & \dots & \lambda_i^2 y_i^2 - \lambda_i \end{vmatrix}$$

$$\Delta_n = |H_r|$$

Une condition nécessaire et suffisante pour que la forme quadratique  $F''$  soit définie négative est /4/:

$$\left. \begin{array}{l} \Delta_i < 0 \text{ si } i \text{ impair} \\ \Delta_i > 0 \text{ si } i \text{ pair} \end{array} \right\} i = 1, 2, \dots, n$$

En multipliant la première colonne par  $y_1$ , la seconde par  $y_2$ , ... et la  $i$ ème par  $y_i$ ; puis en divisant la première ligne par  $y_1$ , la seconde par  $y_2$ , ... et finalement la  $i$ ème par  $y_i$ , le mineur principal  $\Delta_i$  peut s'écrire sous la forme:

$$\Delta_i = (\lambda_1 \cdot \lambda_2 \cdot \dots \cdot \lambda_n) \cdot \begin{vmatrix} \lambda_1 y_1^2 - 1 & \lambda_2 y_2^2 & & & \lambda_i y_i^2 \\ \lambda_1 y_1^2 & \lambda_2 y_2^2 - 1 & & & \lambda_i y_i^2 \\ & & \lambda_1 y_1^2 & & \lambda_i y_i^2 \\ & & & \ddots & \lambda_i y_i^2 \\ & & & & \lambda_1 y_1^2 \\ \lambda_1 y_1^2 & & & & \lambda_i y_i^2 - 1 \end{vmatrix}$$

Comme les  $\lambda_i, i = 1, 2, \dots, n$ , sont les valeurs propres de la matrice définie positive  $\Sigma^{-1}$ ,  $\Delta_i$  a le même signe que le déterminant:

$$\Delta'_i = \Delta_i / (\lambda_1 \cdot \lambda_2 \cdot \dots \cdot \lambda_n)$$

L'addition de toutes les colonnes de  $\Delta'_i$  de rang plus grand que 1 à la première colonne, suivie de la soustraction de la première ligne à toutes les lignes de rang supérieur à 1 permettent d'écrire  $\Delta'_i$  sous la forme:

$$\Delta'_i = \begin{vmatrix} \lambda_1 y_1^2 + \lambda_2 y_2^2 + \dots + \lambda_i y_i^2 - 1 & \lambda_2 y_2^2 & \lambda_3 y_3^2 & \dots & \lambda_i y_i^2 \\ & 0 & & & \\ & & -1 & & \\ & & & \ddots & \\ & & & & -1 \\ & & & & & 0 \\ & & & & & & -1 \end{vmatrix}$$

Finalement, il vient:

$$\Delta_i' = (-1)^i (1 - \lambda_1 y_1^2 - \lambda_2 y_2^2 - \dots - \lambda_n y_n^2)$$

La forme quadratique  $F''$  est donc semi-définie négative si et seulement si:

$$1 - \sum_{j=1}^i \lambda_j y_j^2 \geq 0, \quad i = 1, 2, \dots, n$$

Comme les valeurs propres  $\lambda_i$ ,  $i = 1, 2, \dots, n$  de la matrice  $\Sigma^{-1}$  sont positives, cet ensemble d'inégalités est vérifié si et seulement si:

$$Y^T \Lambda Y \leq 1$$

La fonction de densité  $p(X)$  est donc concave dans le domaine convexe  $D$  situé à l'intérieur de la surface  $\Gamma$  d'équation:

$$Y^T \Lambda Y = 1$$

Il est aisé de montrer qu'à l'extérieur de ce domaine la forme quadratique  $F''$  n'est ni semi-définie négative, ni semi-définie positive. A l'extérieur de  $D$ , la fonction  $p(X)$  n'est donc ni localement concave, ni localement convexe. Le domaine  $D$  intérieur à  $\Gamma$  est donc le plus grand domaine convexe à l'intérieur duquel la fonction de densité  $p(X)$  est concave. Il sera appelé "domaine de concavité de  $p(X)$ ".

#### IV.4. PROPRIETES GEOMETRIQUES DU DOMAINE DE CONCAVITE

La surface  $\Gamma$  d'équation:

$$Y^T \Lambda Y = 1$$

délimitant le domaine de concavité  $D$  d'une fonction de densité normale de vecteur-moyenne nul est un hyperellipsoïde centré à l'origine. Ses axes principaux ont les mêmes directions que les vecteurs propres de la matrice  $\Sigma^{-1}$ . De plus, les longueurs  $L_i$ ,  $i = 1, 2, \dots, n$  de ses axes sont liées aux éléments diago-

naux  $\lambda_i$ ,  $i = 1, 2, \dots, n$  de la matrice  $\Lambda$  par les relations:  
(cf. figure 4.1.)

$$L_i = \frac{2}{\sqrt{\lambda_i}}, \quad i = 1, 2, \dots, n \quad (4.6.)$$

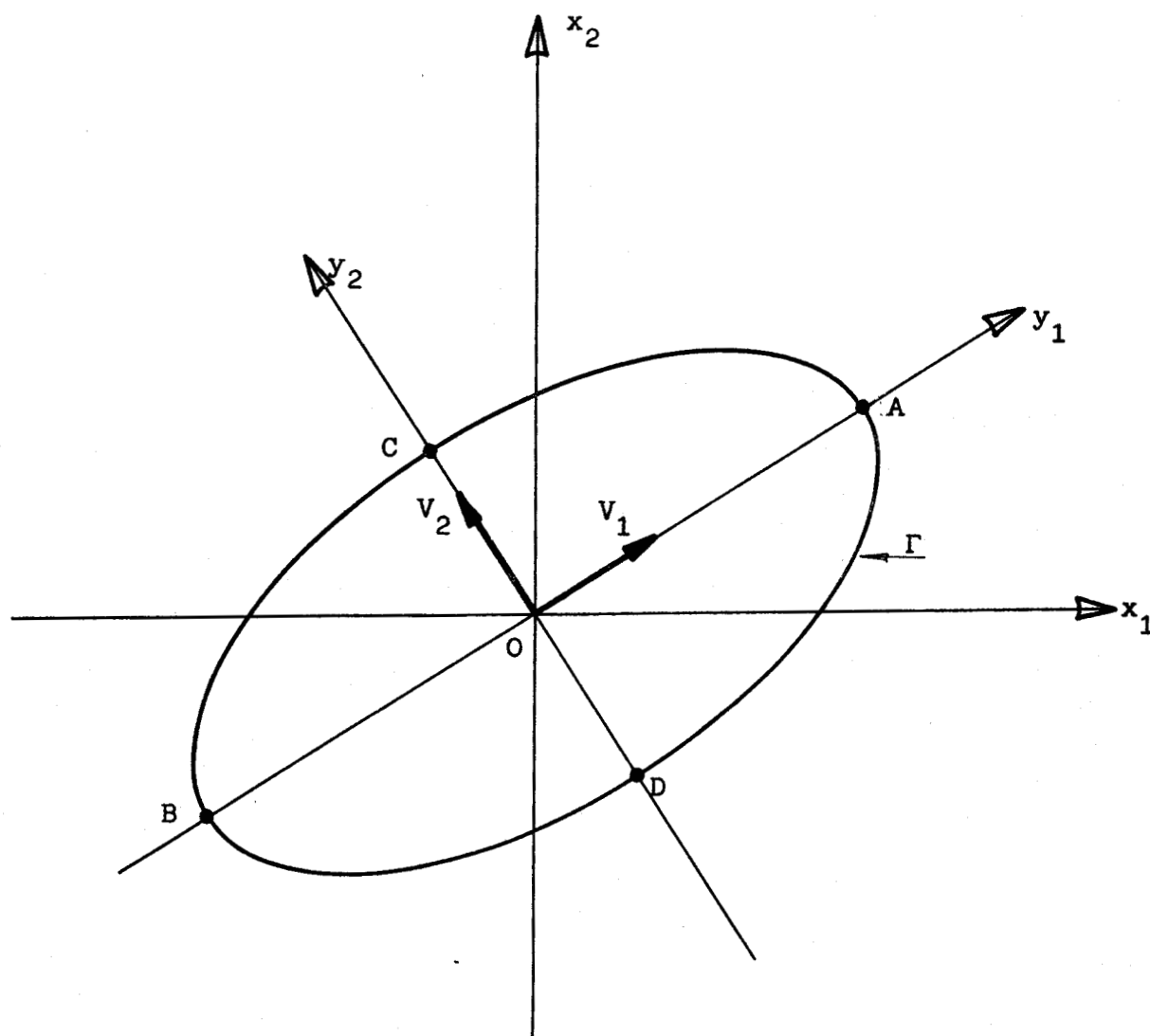


Figure 4.1. Domaine de concavité  $D$  d'une fonction de densité de probabilité normale à deux variables

$$|AB| = L_1$$

$$|CD| = L_2$$



Il est ainsi possible de déterminer la matrice de covariance  $\Sigma$  à partir des caractéristiques géométriques de l'hyperellipsoïde  $\Gamma$ . En effet, la transformation orthogonale qui diagonalise la matrice  $\Sigma^{-1}$  est définie par la matrice  $M$  telle que:

$$M = [V_1, V_2, \dots, V_i, \dots, V_n]$$

dans laquelle les vecteurs  $V_i$ ,  $i = 1, 2, \dots, n$  sont les vecteurs propres normés de  $\Sigma^{-1}$  ( cf. équation 4.2. ), c'est à dire les vecteurs unitaires des axes principaux de l'hyperellipsoïde  $\Gamma$

D'après (4.3.) et (4.4.) et compte-tenu de (4.6.)

nous avons:

$$\Sigma^{-1} = M \cdot \begin{bmatrix} \frac{4}{L_1^2} & & & \\ & \ddots & & \\ & & \frac{4}{L_i^2} & \\ & & & \ddots \\ & & & & \frac{4}{L_n^2} \end{bmatrix} \cdot M^{-1} \quad (4.7.)$$

Dans le cas d'une distribution de vecteur-moyenne  $\bar{X}$  non nul, il est évident que le centre de l'hyperellipsoïde  $\Gamma$  indique la valeur de ce vecteur-moyenne.

A ce stade, nous disposons donc d'un schéma pour déterminer les paramètres d'une distribution normale lorsque le domaine de concavité hyperellipsoïdal de la fonction de densité sous-jacente est connu. La procédure peut être décomposée en quatre pas distincts.

- Pas n°1 La détermination du centre du domaine de concavité  $D$  donne le vecteur-moyenne  $\bar{X}$  de la distribution.
- Pas n°2 La détermination des axes principaux du domaine hyperellipsoïdal  $D$  donne les vecteurs propres de l'inverse de la matrice de covariance de la distribution.

- Pas n°3 La détermination des longueurs des axes principaux de l'hyperellipsoïde  $\Gamma$  donne les valeurs propres de l'inverse de la matrice de covariance.
- Pas n°4 La connaissance de ces valeurs propres et des vecteurs propres associés permet enfin de déterminer l'inverse de la matrice de covariance ( cf équation 4.7.), puis, finalement, la matrice de covariance elle même.

L'utilisation de ce schéma pour l'identification d'une distribution à partir d'un échantillon d'observations nécessite donc, en premier lieu, la détermination du domaine de concavité  $D$  de la fonction de densité sous jacente.

#### IV.5. IDENTIFICATION DES DISTRIBUTIONS NORMALES

##### IV.5.1. DETERMINATION DU DOMAINE DE CONCAVITE

Pour déterminer, à partir d'un ensemble d'observations distribuées normalement, le domaine de concavité de la fonction de densité sous jacente, on utilise le test de convexité présenté au chapitre III.

Rappelons que les résultats de ce test, qui permet de déterminer le sens de la convexité de la fonction de densité aux nœuds d'un réseau hypercubique de pas  $\alpha_0$ , ne sont valables que si les domaines d'observation restent intérieurs à des régions où la convexité de la fonction est définie.

Dans le cas de distributions normales, on sait que la fonction de densité est concave à l'intérieur du domaine de concavité. Cette région sera donc matérialisée par un ensemble d'hypercubes à convexité négative adjacents ( cf. chapitre III, paragraphe III.6.).

On sait, par contre, qu'à l'extérieur de ce domaine, la fonction n'est ni concave, ni convexe. Ainsi, pour des points de test tels que la totalité ou une partie du domaine d'observation est située à l'extérieur du domaine de concavité, le test

de convexité indique indifféremment l'un ou l'autre des deux sens de convexité, sans qu'on puisse attribuer une signification précise au résultat.

Appliqué à tous les points du réseau hypercubique, le test fait donc apparaître des hypercubes à convexité négative dans la région où la convexité de la fonction n'est pas définie. Mais, alors qu'à l'intérieur du domaine de concavité, les hypercubes de ce type sont théoriquement adjacents, ils se trouvent mêlés, à l'extérieur de celui-ci, à des hypercubes à convexité opposée. Une simple procédure d'agrégation, portant sur les hypercubes à convexité négative, permet alors d'isoler, par chaînage, l'ensemble des hypercubes adjacents correspondant au domaine de concavité. (cf. figure 4.2.(a)). Les petits ensembles d'hypercubes à convexité négative adjacents pouvant apparaître à l'extérieur du domaine de concavité sont, de par leur taille réduite, aisément éliminés.

L'ensemble des hypercubes à convexité négative ainsi isolé constitue un domaine  $D^0$ , appelé "domaine modal", qui est en fait une matérialisation assez grossière du domaine de concavité hyperellipsoïdal  $D$  de la fonction de densité. Les imprécisions dans la détermination de ce domaine résultent d'une part de la discrétisation de l'espace (le pas du réseau hypercubique ne peut être choisi arbitrairement petit), d'autre part de la mise en défaut des conditions d'application du test de convexité pour les points de test situés près de la frontière du domaine de concavité.

Malgré ces sources d'erreur et les imprécisions qui en résultent, il est possible d'assimiler le domaine  $D^0$  à un domaine hyperellipsoïdal  $\hat{D}$ . Cette représentation mathématique du domaine modal sera considérée comme une approximation du domaine de concavité  $D$ .

#### IV.5.2. DETERMINATION DES PARAMETRES DE LA DISTRIBUTION

Soient  $M_j$ ,  $j = 1, 2, \dots, \nu$ , les centres des hypercubes adjacents à convexité négative constituant le domaine modal  $D^0$ . Ce domaine est considéré comme un corps solide constitué de particules de masse unité, situées aux points  $M_j$ ,  $j = 1, 2, \dots, \nu$ .

On choisit alors pour nouvelle origine le centre d'inertie  $\hat{O}$  du solide ainsi formé. Ce point est considéré comme le centre géométrique du domaine  $\hat{D}$ . Ses coordonnées fournissent en fait une approximation du vecteur-moyenne de la distribution.

Le calcul des produits d'inertie pour chacun des points  $M_j$ ,  $j = 1, 2, \dots, \varnothing$ , permet ensuite de déterminer le tenseur central d'inertie de ce corps [5]. La recherche des axes principaux du domaine hyperellipsoïdal  $\hat{D}$  se ramène alors à la recherche des vecteurs propres de ce tenseur d'inertie. Les vecteurs propres  $V_i$ ,  $i = 1, 2, \dots, n$  de l'inverse de la matrice de covariance peuvent ainsi être approchés par les vecteurs unitaires  $\hat{V}_i$ ,  $i = 1, 2, \dots, n$  des axes principaux d'inertie.

La longueur  $\hat{L}_i$  du domaine  $\hat{D}$  dans la  $i$ ème direction principale est alors obtenue simplement en projetant les points  $M_j$ ,  $j = 1, 2, \dots, \varnothing$  sur le  $i$ ème axe principal: (cf. figure 4.2. (b)).

$$\hat{L}_i = \text{Max}_j \overline{OM}_j \cdot \hat{V}_i - \text{Min}_j \overline{OM}_j \cdot \hat{V}_i$$

Une approximation  $\hat{\lambda}_i$  de la  $i$ ème valeur propre de la matrice  $\Sigma^{-1}$  est alors donnée par: (cf. équation 4.6.)

$$\hat{\lambda}_i = 4 / \hat{L}_i$$

La matrice:

$$\hat{\Sigma}^{-1} = \hat{O} \cdot \begin{bmatrix} 4/\hat{L}_1^2 & & & \\ & \ddots & & \\ & & 4/\hat{L}_i^2 & \\ & & & \ddots \\ & & & & 4/\hat{L}_n^2 \end{bmatrix} \cdot \hat{O}^{-1}$$

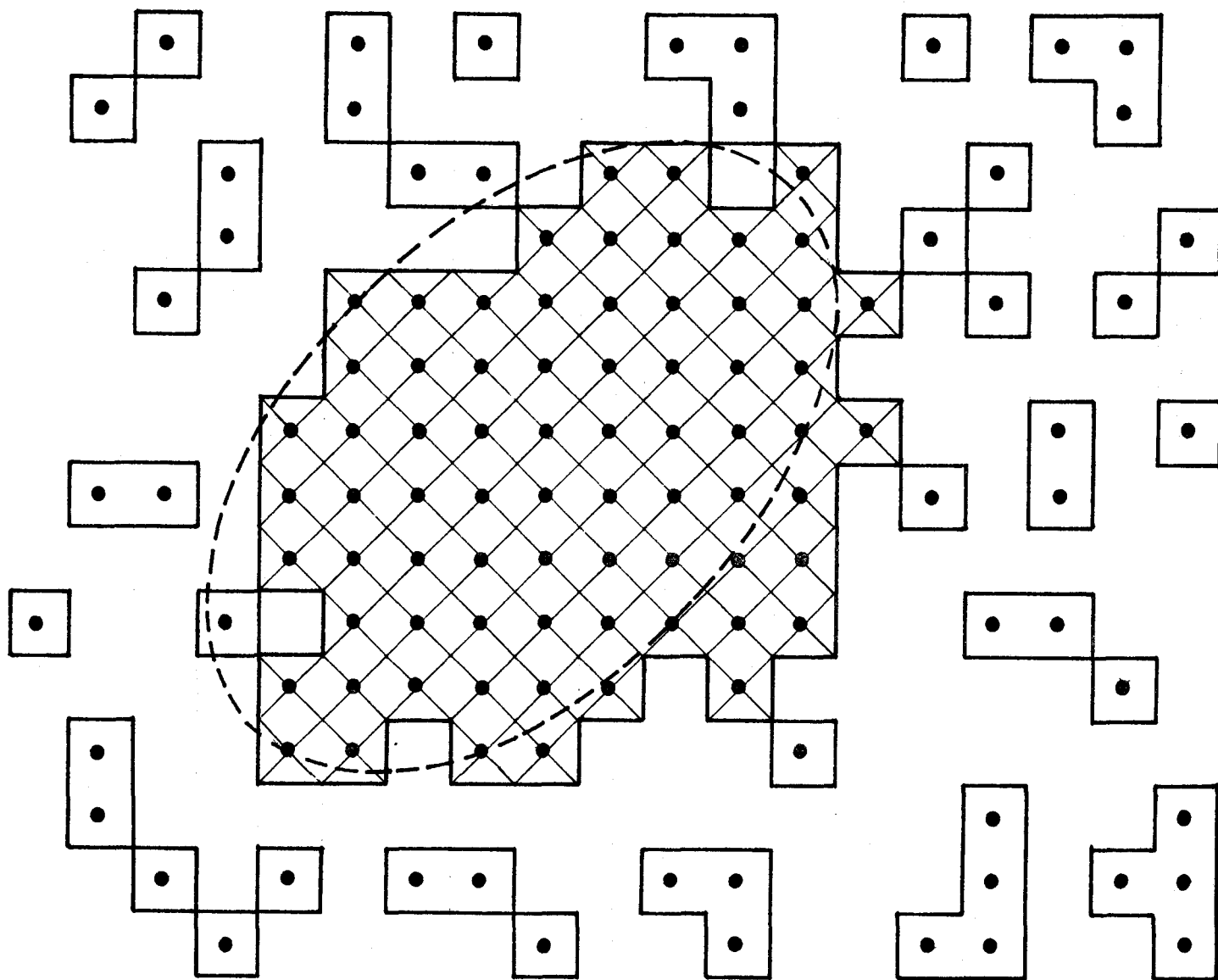
avec:

$$\hat{O} = [\hat{V}_1, \hat{V}_2, \dots, \hat{V}_n]$$

constitue une approximation de l'inverse de la matrice de covariance (cf. équation 4.7.)

L'assimilation du domaine modal  $D^0$  à un modèle hyperellipsoïdal  $\hat{D}$  (cf. figure 4.2. (c)) permet donc de déterminer des approximations du vecteur-moyenne et de la





Carrés à convexité négative



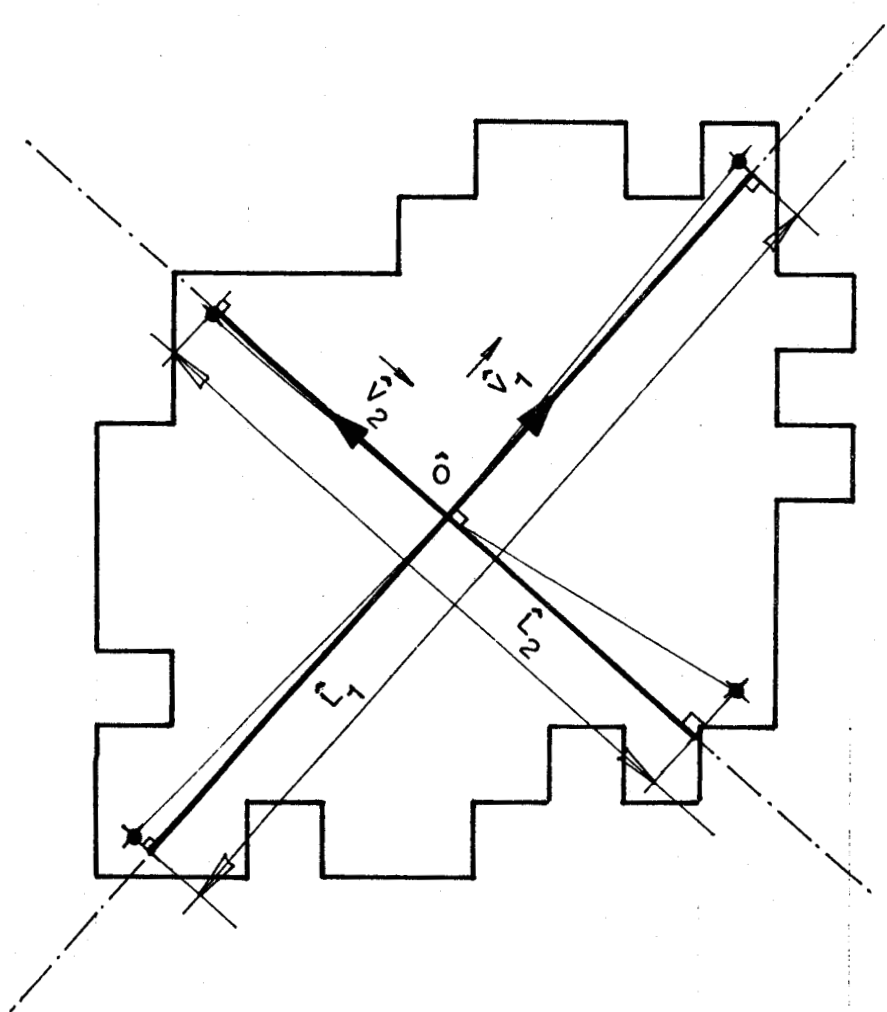
Carrés constituant le domaine modal  $D^0$



Domaine de concavité  $D$

BUS  
LILLE

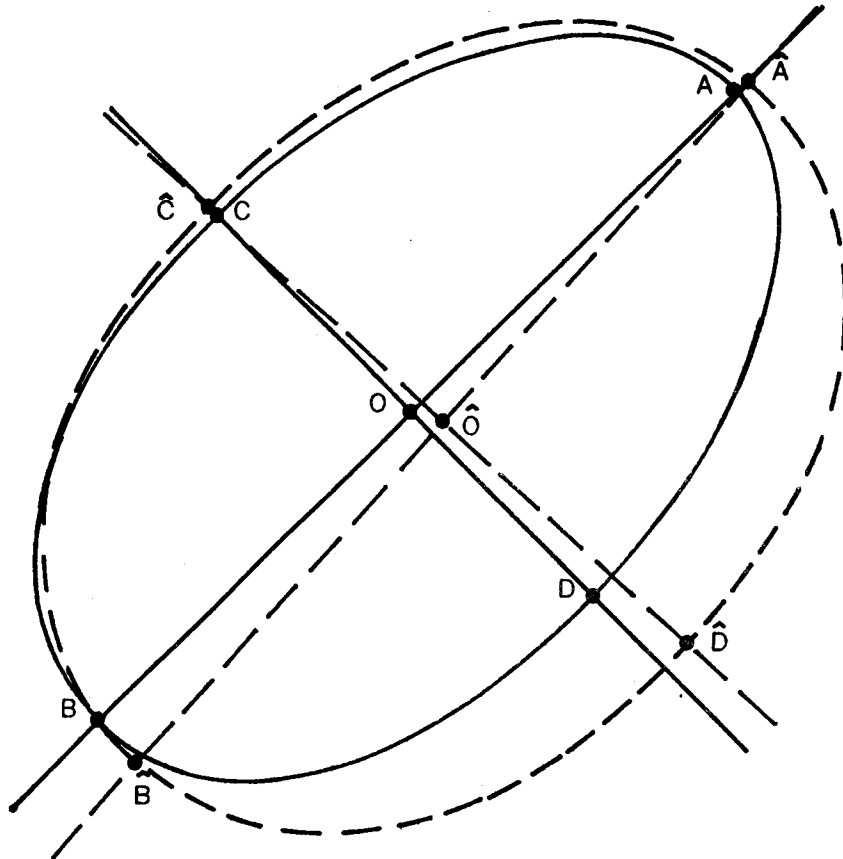
Figure 4.2. (a) Résultat du test de convexité appliqué à une distribution normale d'observations bidimensionnelles.




$\hat{O}$  Centre d'inertie du domaine modal  $D^\circ$

$$\hat{\lambda}_i = \frac{4}{\hat{L}_i^2}, \quad i = 1, 2.$$

Figure 4.2. (b) Détermination des paramètres d'une distribution normale à partir du domaine modal  $D^\circ$ .



$D$ :  Domaine de concavité réel


$\hat{D}$ :  Modèle ellipsoïdal du domaine modal  $D^\circ$

Figure 4.2. (c) Comparaison du domaine de concavité  $D$  et du modèle ellipsoïdal  $\hat{D}$  du domaine modal  $D^\circ$  pour l'exemple de la figure 4.2. (a).



matrice de covariance de la distribution considérée.

Avant d'envisager l'emploi de ce schéma pour l'analyse des mélanges gaussiens, il convient de tester cette méthode approchée sur des exemples pour analyser la précision des résultats obtenus. Il est en effet impossible d'étudier analytiquement le comportement de la méthode, même dans le cas asymptotique où le nombre d'observations disponibles tend vers l'infini, puisque la convexité de la fonction de densité normale n'est pas définie à l'extérieur du domaine de concavité. La détermination de ce domaine est donc toujours entachée d'erreur, quelle que soit la taille de l'échantillon. Afin d'apprécier et d'analyser la précision de la méthode, les résultats obtenus sont comparés à ceux de la méthode d'estimation par maximum de vraisemblance dont les propriétés asymptotiques sont parfaitement connues.

#### IV.6. EXEMPLES D'APPLICATION

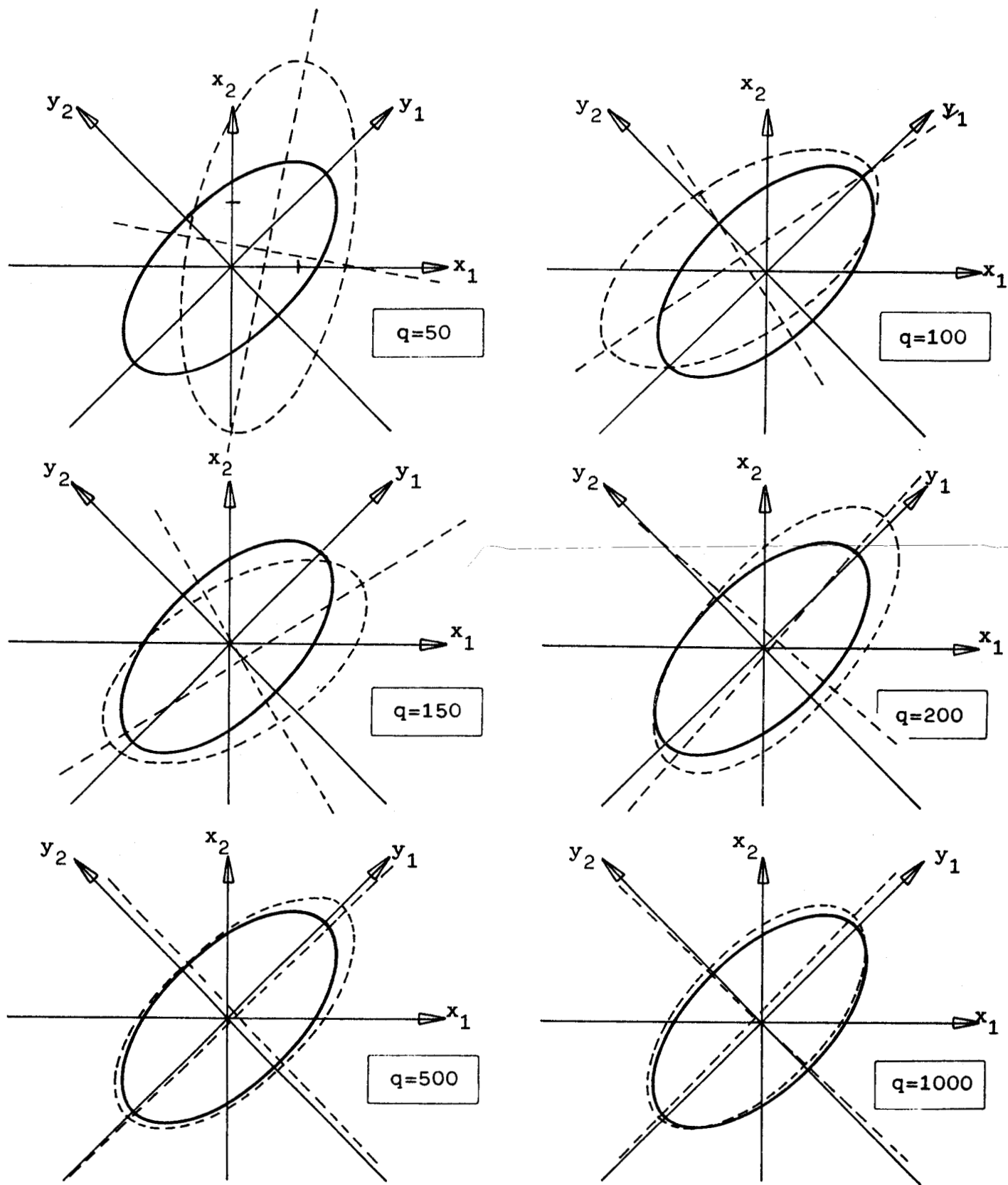
La méthode d'identification des distributions normales décrite au paragraphe précédent a été testée sur des données générées artificiellement. Les résultats présentés n'ont d'autre prétention que d'illustrer le comportement de la méthode d'approximation sur des exemples particuliers et de situer le niveau de précision de cette méthode par rapport à celui d'une méthode classique d'estimation.

Le comportement de la procédure est d'abord analysé en fonction de la taille de l'échantillon pour une distribution bidimensionnelle de vecteur-moyenne nul et de matrice de covariance:

$$\Sigma = \frac{1}{8} \begin{bmatrix} 5 & -3 \\ -3 & 5 \end{bmatrix}$$

La figure 4.3. fait apparaître, pour chaque taille d'échantillon, le domaine de concavité  $D$  de la fonction de densité sous jacente et le modèle  $\hat{D}$  du domaine modal  $D^0$ . On constate que plus le nombre d'observations augmente, plus le modèle  $\hat{D}$  se rapproche, en position, taille et orientation des axes, du domaine de concavité réel  $D$ . (Les axes de ce dernier, orientés à 45°, ont respectivement pour longueur 2 et 4.)

Note: Pour cette expérience, le paramètre de discrétisation  $\alpha_q$  a été choisi de la forme  $\alpha_q = \alpha_0 / \sqrt{q}$  avec  $\alpha_0 = 0,3$ .



- D: Domaine de concavité réel
- $\hat{D}$ : Modèle ellipsoïdal du domaine modal  $D^\circ$  déterminé à partir des observations

Figure 4.3. Modèles  $\hat{D}$  des domaines modaux  $D^\circ$  d'une distribution normale de vecteur-moyenne nul et de matrice de covariance  $\Sigma = 1/8 \begin{bmatrix} 5 & -3 \\ -3 & 5 \end{bmatrix}$  obtenus pour différentes tailles d'échantillon.

BUS  
LILLE

Les résultats obtenus sont comparés à ceux fournis par application de la méthode d'estimation classique par maximum de vraisemblance /6/. On sait en effet que les estimateurs de  $\bar{X}$  et  $\Sigma$  obtenus par cette méthode sont donnés par /3/:

$$\hat{\bar{X}}' = \frac{1}{n} \sum_{j=1}^n x_j$$

et:

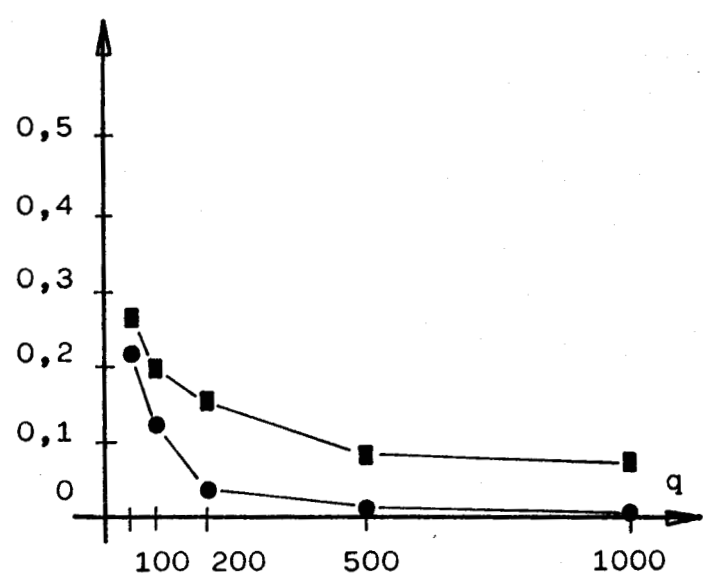
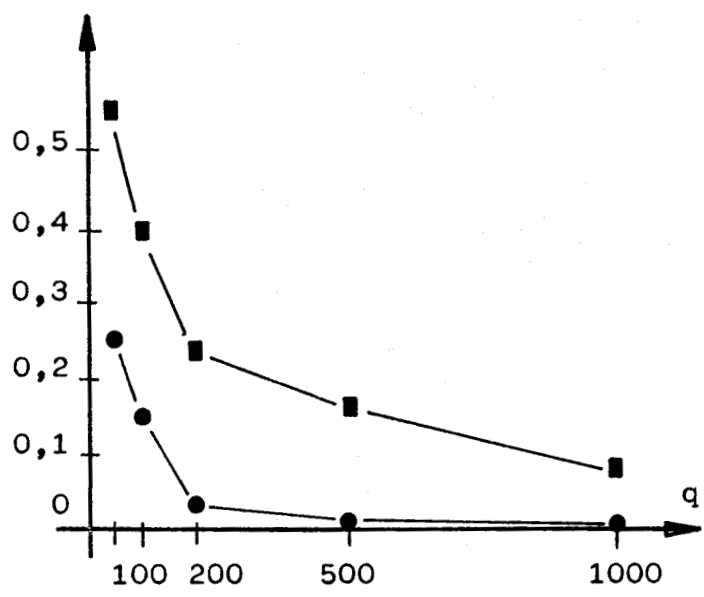
$$\hat{\Sigma}' = \frac{1}{n} \sum_{j=1}^n (x_j - \hat{\bar{X}}')(x_j - \hat{\bar{X}}')^T$$

La figure 4.4. permet de comparer les erreurs d'approximation commises sur la détermination du vecteur moyenne  $\bar{X}$  par la méthode proposée aux erreurs d'estimation résultant de l'emploi de la méthode du maximum de vraisemblance.

De manière analogue, la figure 4.5. indique les erreurs d'approximation et les erreurs d'estimation associées aux deux méthodes pour les valeurs propres  $\lambda_1$  et  $\lambda_2$  de l'inverse de la matrice de covariance.

On constate que les erreurs d'approximation résultant de la méthode proposée diminuent lorsque la taille de l'échantillon augmente, comme le font les erreurs d'estimation par maximum de vraisemblance. Rappelons cependant que cette décroissance des erreurs d'approximation ne garantit pas la convergence asymptotique des valeurs approchées vers les valeurs exactes. Dans le cas d'échantillons de taille limitée, le comportement ainsi mis en évidence permet toutefois de prévoir une amélioration de l'approximation lorsque le nombre d'observations augmente, les résultats demeurant cependant moins précis que ceux obtenus par la méthode d'estimation par maximum de vraisemblance.

La méthode proposée a été également appliquée à l'identification de distributions normales d'observations de dimension plus élevée, de vecteurs-moyenne nuls et de matrices de covariance unité. Les résultats, consignés dans le tableau 4.1., montrent que, comme pour toutes les méthodes faisant appel à des techniques non paramétriques, le nombre d'observations nécessaires pour obtenir des résultats satisfaisants est d'autant plus important que la dimension des données est élevée.



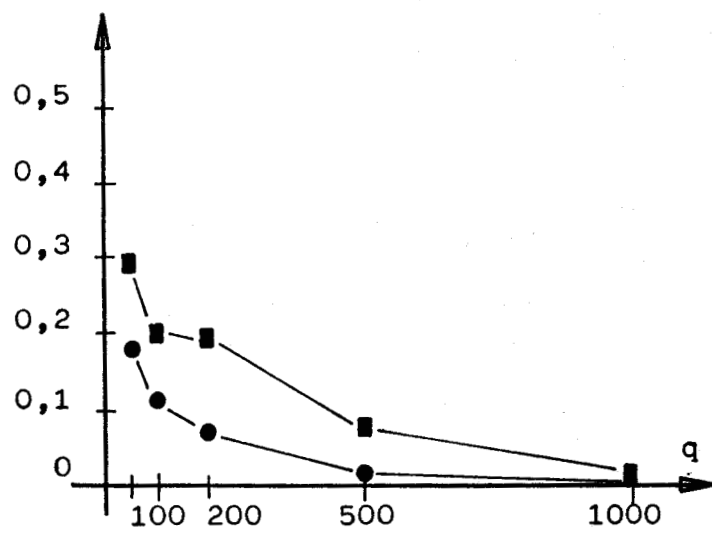
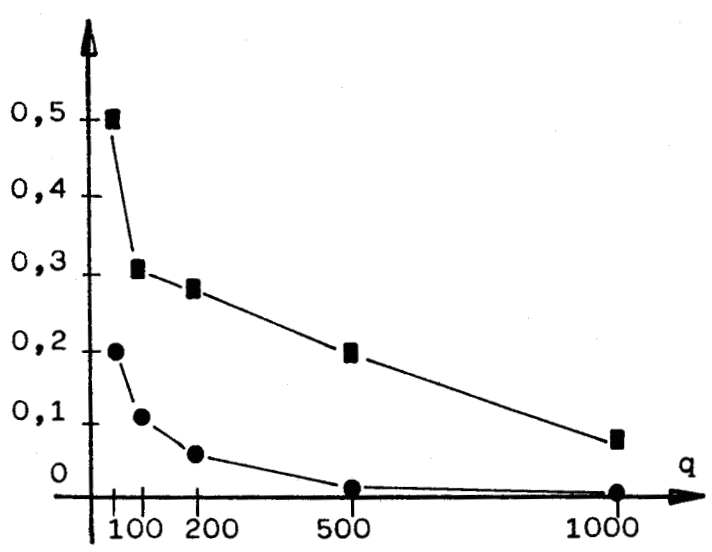
$$\blacksquare = \left| \hat{\bar{x}}_1 - \bar{x}_1 \right|$$

$$\bullet = \left| \hat{\bar{x}}'_1 - \bar{x}_1 \right|$$

$$\blacksquare = \left| \hat{\bar{x}}_2 - \bar{x}_2 \right|$$

$$\bullet = \left| \hat{\bar{x}}'_2 - \bar{x}_2 \right|$$

Figure 4.4. Comparaison des erreurs d'approximation et des erreurs d'estimation du vecteur-moyenne de la distribution de la figure 4.3.



$$\blacksquare = \frac{|\hat{\lambda}_1 - \lambda_1|}{\lambda_1}$$

$$\bullet = \frac{|\hat{\lambda}'_1 - \lambda_1|}{\lambda_1}$$

$$\blacksquare = \frac{|\hat{\lambda}_2 - \lambda_2|}{\lambda_2}$$

$$\bullet = \frac{|\hat{\lambda}'_2 - \lambda_2|}{\lambda_2}$$



Figure 4.5. Comparaison des erreurs d'approximation et des erreurs d'estimation des valeurs propres de l'inverse de la matrice de covariance de la distribution de la figure 4.3.

$n = 2$	
$n = 100$	$n = 200$
$\hat{\bar{X}} = \begin{bmatrix} 0,18 \\ 0,12 \end{bmatrix}; \hat{\Sigma} = \begin{bmatrix} 1,22 & -0,11 \\ -0,11 & 1,09 \end{bmatrix}$	$\hat{\bar{X}} = \begin{bmatrix} 0,09 \\ -0,07 \end{bmatrix}; \hat{\Sigma} = \begin{bmatrix} 1,07 & 0,09 \\ 0,09 & 1,05 \end{bmatrix}$
$n = 500$	$n = 1000$
$\hat{\bar{X}} = \begin{bmatrix} 0,05 \\ 0,05 \end{bmatrix}; \hat{\Sigma} = \begin{bmatrix} 1,04 & 0,08 \\ 0,08 & 1,06 \end{bmatrix}$	$\hat{\bar{X}} = \begin{bmatrix} 0,02 \\ 0,03 \end{bmatrix}; \hat{\Sigma} = \begin{bmatrix} 1,03 & -0,05 \\ -0,05 & 1,02 \end{bmatrix}$

$n = 3$	
$n = 100$	$n = 200$
$\hat{\bar{X}} = \begin{bmatrix} 0,15 \\ 0,23 \\ -0,02 \end{bmatrix}; \hat{\Sigma} = \begin{bmatrix} 1,33 & 0,19 & -0,01 \\ 0,19 & 1,25 & 0,09 \\ -0,01 & 0,09 & 0,86 \end{bmatrix}$	$\hat{\bar{X}} = \begin{bmatrix} 0,16 \\ 0,13 \\ 0,11 \end{bmatrix}; \hat{\Sigma} = \begin{bmatrix} 1,20 & -0,13 & 0,10 \\ -0,13 & 1,15 & 0,20 \\ 0,10 & 0,20 & 1,11 \end{bmatrix}$
$n = 500$	$n = 1000$
$\hat{\bar{X}} = \begin{bmatrix} 0,10 \\ 0,09 \\ 0,06 \end{bmatrix}; \hat{\Sigma} = \begin{bmatrix} 1,09 & -0,09 & 0,12 \\ -0,09 & 1,05 & -0,15 \\ 0,12 & -0,15 & 0,90 \end{bmatrix}$	$\hat{\bar{X}} = \begin{bmatrix} 0,03 \\ -0,06 \\ -0,02 \end{bmatrix}; \hat{\Sigma} = \begin{bmatrix} 1,05 & -0,06 & 0,08 \\ -0,06 & 1,07 & 0,02 \\ 0,08 & 0,02 & 1,01 \end{bmatrix}$

$n = 4$	
$n = 100$	
$\hat{\bar{X}} = \begin{bmatrix} 0,25 \\ -0,36 \\ 0,19 \\ -0,33 \end{bmatrix}; \hat{\Sigma} = \begin{bmatrix} 1,52 & 0,23 & -0,30 & 0,09 \\ 0,23 & 1,43 & 0,19 & -0,20 \\ -0,30 & 0,19 & 0,89 & 0,16 \\ 0,09 & -0,20 & 0,16 & 1,32 \end{bmatrix}$	
$n = 200$	
$\hat{\bar{X}} = \begin{bmatrix} 0,19 \\ 0,23 \\ -0,15 \\ 0,27 \end{bmatrix}; \hat{\Sigma} = \begin{bmatrix} 1,45 & 0,16 & 0,15 & -0,21 \\ 0,16 & 1,34 & 0,18 & -0,09 \\ 0,15 & 0,18 & 1,33 & 0,06 \\ -0,21 & -0,09 & 0,06 & 0,99 \end{bmatrix}$	

Tableau 4.1.





$n = 4$	
$n = 500$	
$\hat{\bar{X}} = \begin{bmatrix} -0,10 \\ 0,12 \\ 0,09 \\ -0,13 \end{bmatrix}$	$;\hat{\Sigma} = \begin{bmatrix} 1,12 & -0,12 & 0,01 & 0,14 \\ -0,12 & 1,18 & 0,10 & -0,09 \\ 0,01 & 0,10 & 0,95 & 0,18 \\ 0,14 & -0,09 & 0,18 & 0,89 \end{bmatrix}$
$n = 1000$	
$\hat{\bar{X}} = \begin{bmatrix} -0,08 \\ -0,09 \\ -0,10 \\ -0,11 \end{bmatrix}$	$;\hat{\Sigma} = \begin{bmatrix} 1,10 & 0,09 & 0,10 & -0,06 \\ 0,09 & 1,11 & 0,01 & 0,13 \\ 0,10 & 0,01 & 1,00 & 0,00 \\ -0,06 & 0,13 & 0,00 & 1,08 \end{bmatrix}$

Tableau 4.1. (suite)

IV.7. CONCLUSION

Nous avons proposé, dans ce chapitre, une méthode d'identification des distributions normales basée sur l'analyse de la convexité de la fonction de densité sous-jacente. La détermination, à partir des observations, du domaine à l'intérieur duquel cette fonction de densité est concave permet de calculer des valeurs approchées du vecteur moyenne et de la matrice de covariance de la distribution.

Dans le cadre restreint de l'identification d'une distribution unique, cette méthode est moins performante que la méthode classique d'estimation par maximum de vraisemblance. Nous allons maintenant montrer l'intérêt de cette nouvelle approche pour l'analyse des mélanges gaussiens et l'optimisation du processus de classification automatique pour lequel elle a été mise au point.

REFERENCES ( chapitre IV )

- /1/ CRAMER, H.  
"Mathematical methods of statistics"  
Princeton, New York Chapitre 24 p. 310-316 (1963)
- /2/ KEEHN, D.G.  
"A note on learning for gaussian properties"  
IEEE Trans. Info. Theory Vol. IT-11 p. 126-132 (1965)
- /3/ ANDERSON, T.W.  
"An introduction to multivariate statistical analysis"  
John Wiley, New York Chapitre 3 (1958)
- /4/ GANTMACHER, F.R.  
"The theory of matrices"  
Chelsea Publishing Co., New York (1960)
- /5/ CABANNES, H.  
"Mecanique"  
Dunod, Paris (1968)
- /6/ HOEL, P.G.  
"Introduction to mathematical statistics"  
John Wiley, New York (1971)

## CHAPITRE V

ANALYSE DES MELANGES GAUSSIENS ET CLASSIFICATION
---

### V.1. INTRODUCTION

Nous avons vu, dans le premier chapitre de ce mémoire, que l'optimisation en classification automatique peut être simplement ramenée au problème de l'analyse des mélanges gaussiens.

Nous allons maintenant adapter la méthode d'identification des distributions normales présentée au chapitre précédent à la détermination des paramètres statistiques des mélanges gaussiens. Le but principal de cette nouvelle approche est d'obtenir, à partir d'un échantillon fini, un modèle probabiliste de la distribution multidimensionnelle correspondante.

La détermination de ce modèle, même imparfait, permettra ensuite de résoudre le problème de l'optimisation du processus de classification des observations /9/.

### V.2. LES MELANGES GAUSSIENS

Considérons un échantillon dont les observations proviennent de  $K$  classes  $C_k$ ,  $k = 1, 2, \dots, K$ ; chacune d'elles étant caractérisée par une fonction de densité conditionnelle normale:

$$p(X|C_k) = \frac{1}{(2\pi)^{n/2} |\Sigma_k|^{1/2}} \exp\left\{-1/2(X-\bar{X}_k)^T \Sigma_k^{-1} (X-\bar{X}_k)\right\}$$

où  $\bar{X}_k$  et  $\Sigma_k$  sont respectivement le vecteur-moyenne et la matrice de covariance de la classe  $C_k$ .

Dans ces conditions, la densité de probabilité des observations pour des individus tirés aléatoirement et indépendamment de l'ensemble des classes  $C_k$ ,  $k = 1, 2, \dots, K$ , est un mélange des lois de probabilité relatives à ces différentes classes. La fonction de densité de ce mélange est de la forme:

$$f(X) = \sum_{k=1}^K p(X|C_k) \cdot P(C_k) \quad (5.1.)$$

où  $P(C_k)$  est la probabilité d'apparition a priori de la classe  $C_k$ , de telle sorte que:

$$\sum_{k=1}^K P(C_k) = 1$$

Notre but est maintenant d'utiliser les observations distribuées selon cette loi  $f(X)$  pour déterminer:

- 1) Le nombre  $K$  de classes
- 2) La probabilité a priori  $P(C_k)$  de chaque classe
- 3) Le vecteur-moyenne  $\bar{X}_k$  de chaque classe
- 4) La matrice de covariance  $\Sigma_k$  de chaque classe

Lorsque les composantes du mélange présentent un faible degré de chevauchement, il existe une correspondance bijective entre les modes du mélange et les classes en présence /4//5/. Lorsque les classes sont suffisamment séparées, on peut même affirmer que, au voisinage de chaque mode, la fonction de densité conditionnelle de la classe correspondante est fortement prépondérante sur les autres composantes. Dans ces conditions, la forme de la fonction de densité du mélange demeure très semblable, au voisinage des modes, à la forme des fonctions de densité des classes correspondantes. On peut alors s'attendre à ce que la fonction de densité du mélange soit concave dans un certain nombre de domaines très semblables aux domaines de concavité  $D_k$  des différentes composantes. Ces domaines, notés  $\underline{D}_k$ ,  $k = 1, 2, \dots, K$ ,

seront appelés "domaines de concavité du mélange". Le degré de similitude des domaines  $D_k$  et  $\underline{D}_k$  est évidemment fonction du degré de chevauchement des composantes. A titre d'exemple, la figure 5.1. illustre la similitude des domaines de concavité d'un mélange gaussien monovarié et des domaines de concavité de ses composantes.

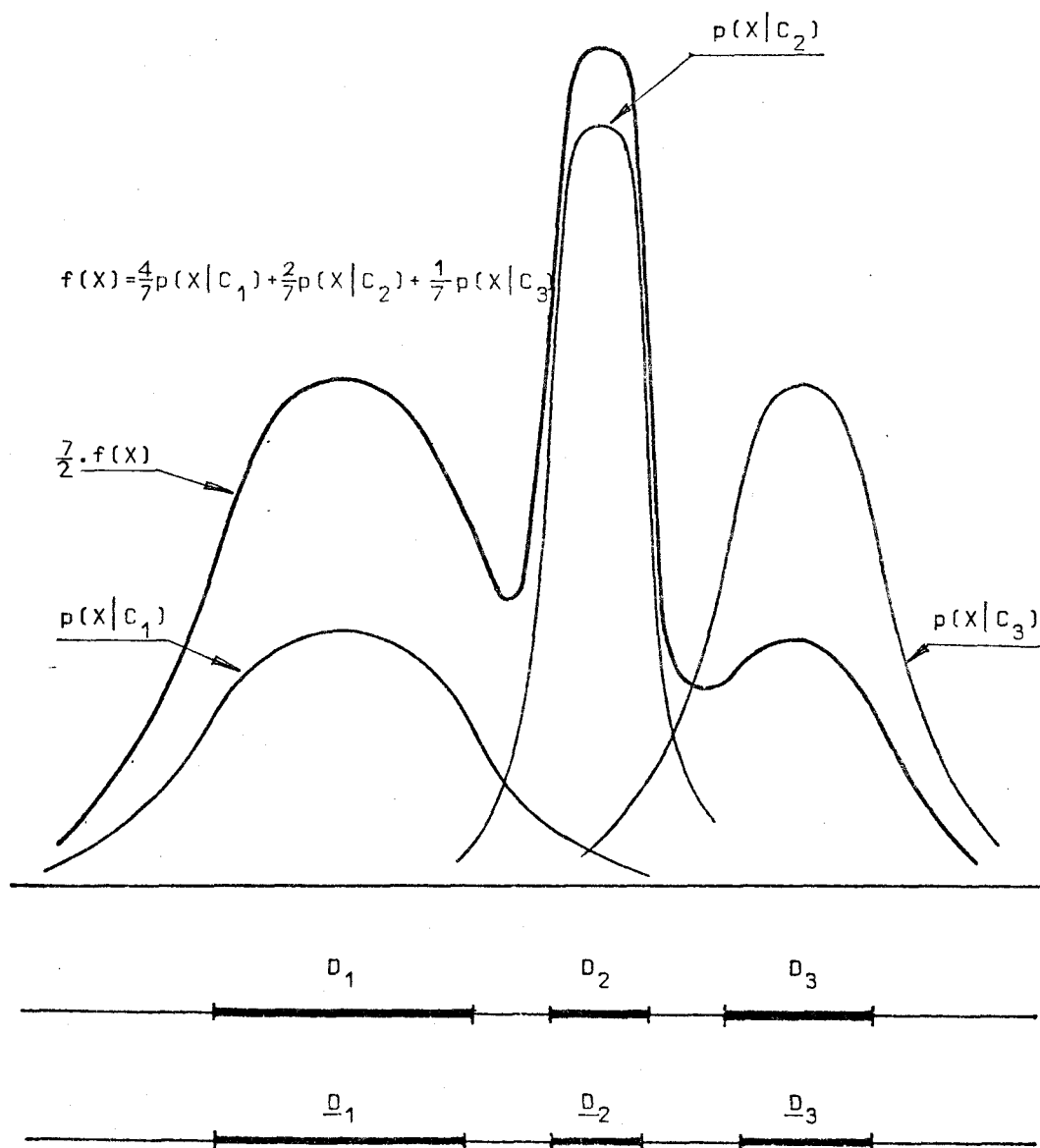


Figure 5.1. Similitude des domaines de concavité d'un mélange gaussien monovarié et des domaines de concavité de ses composantes.

$D_k, k=1,2,3$  : Domaines de concavité des composantes

$\underline{D}_k, k=1,2,3$  : Domaines de concavité du mélange

### V.3. APPROXIMATION DES FONCTIONS DE DENSITE CONDITIONNELLES

Nous ne considèrerons dans la suite que des mélanges pour lesquels, compte tenu de l'objectif final de cette étude, les domaines de concavité de la fonction de densité peuvent être assimilés à ceux des composantes. Ce n'est en effet qu'au prix de cette approximation que le problème de l'optimisation de la procédure de classification sera résolu.

Cette approximation permet d'abord de dénombrer aisément les classes présentes en comptant le nombre de domaines de concavité du mélange.

La détermination du vecteur-moyenne et de la matrice de covariance de chaque classe ainsi mise en évidence peut alors être envisagée à partir de l'analyse des propriétés géométriques du domaine de concavité du mélange correspondant, selon le schéma proposé au chapitre précédent.

A l'aide d'exemples portant sur des échantillons de structure intéressante au point de vue des applications pratiques de la méthode, nous montrerons, à la fin de ce chapitre, que cette approximation conduit à une optimisation très satisfaisante de la classification des observations.

La première étape de l'analyse d'un mélange consiste donc à étudier la convexité de sa fonction de densité. Par application du test de convexité, l'espace est d'abord partagé en hypercubes à convexité négative et en hypercubes à convexité positive. La procédure d'agrégation utilisée au chapitre précédent (paragraphe IV.5.1.) permet alors de faire apparaître les domaines modaux  $\underline{D}_k^0$ ,  $k = 1, 2, \dots, K$ , du mélange sous la forme d'ensembles d'hypercubes à convexité négative adjacents. Comme précédemment, on constate que la différence de taille entre les domaines modaux et les petits ensembles d'hypercubes à convexité négative adjacents apparaissant à l'extérieur des domaines de concavité est telle, qu'il est très aisé d'éliminer ces derniers pour ne conserver que les domaines modaux véritables (cf. figure 5.2.). La représentation de ces domaines modaux par des modèles hyperellipsoïdaux  $\hat{\underline{D}}_k$ ,  $k = 1, 2, \dots, K$ , permet finalement de déterminer des valeurs approchées du vecteur-moyenne et de la matrice de covariance de chaque composante (cf. figure 5.3.). Les seuls paramètres restant à déterminer sont donc les probabilités a priori des classes.

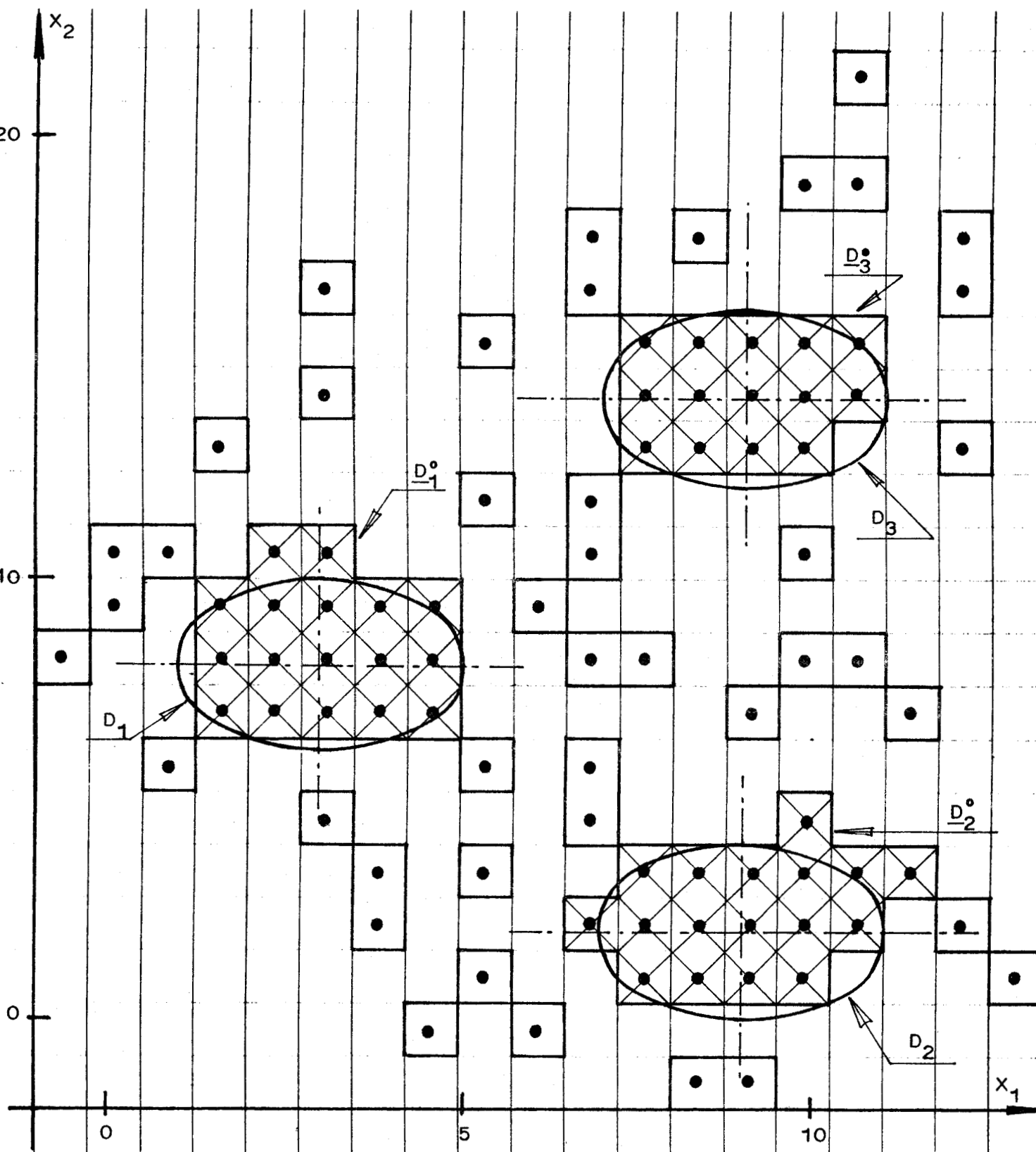


Figure 5.2. Résultat du test de convexité appliqué à un échantillon de 300 observations distribuées selon un mélange de trois lois normales équiprobables.



$D_k, k=1,2,3$  : Domaines de concavité des composantes

Carrés à convexité négative

Carrés à convexité négative constituant les domaines modaux du mélange :  $D_k^o, k=1,2,3$ .

BHS  
LILLE

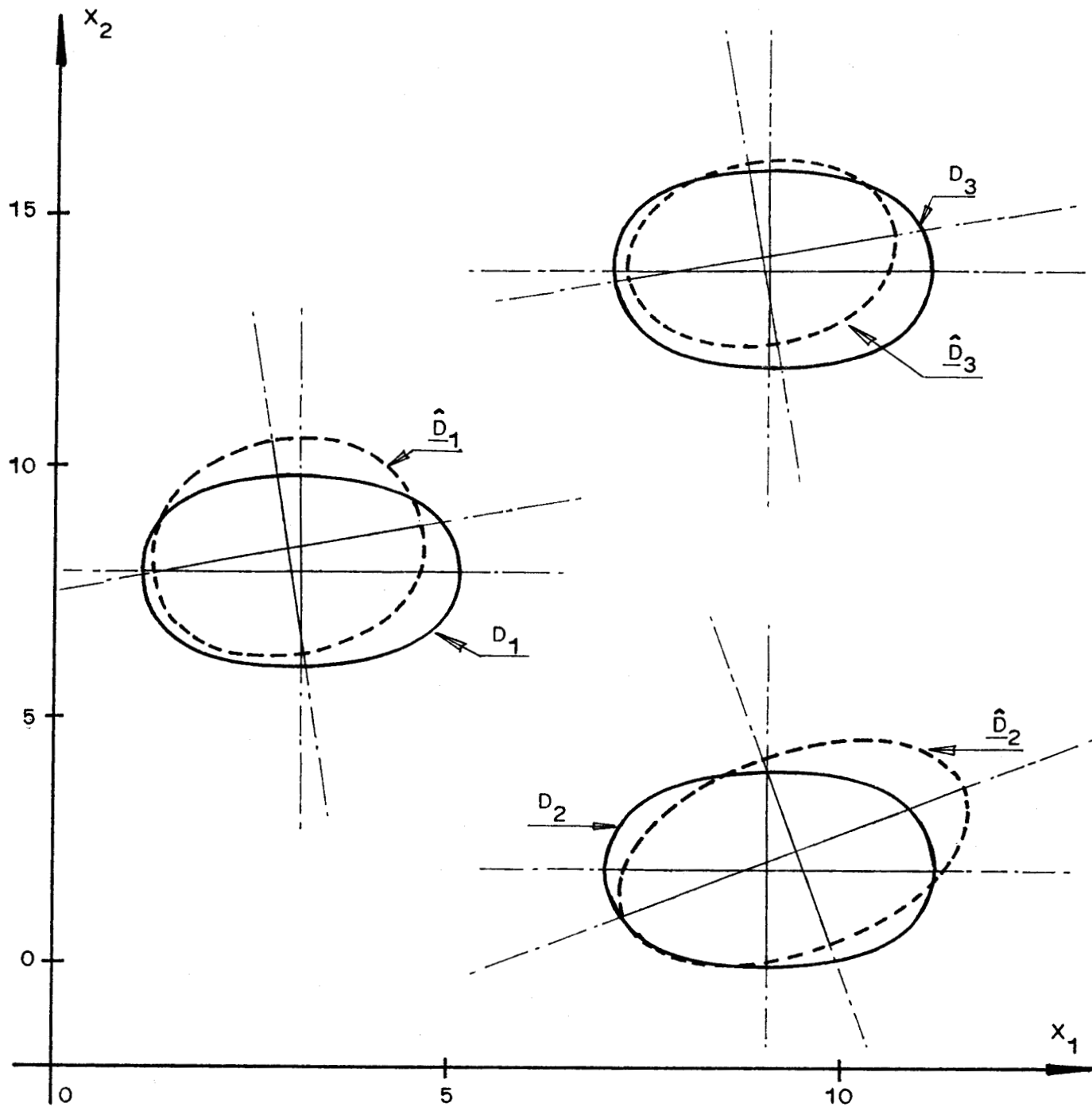


Figure 5.3. Modèles ellipsoïdaux des domaines modaux du mélange des trois lois normales de la figure 5.2.



$D_k, k=1,2,3$  : Domaines de concavité des composantes du mélange



$\hat{D}_k, k=1,2,3$  : Modèles ellipsoïdaux des domaines modaux du mélange



#### V.4. DETERMINATION DES PROBABILITES A PRIORI

##### V.4.1. APPROXIMATION DES PROBABILITES A PRIORI

Considérons la fonction de densité normale:

$$p(X) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left\{-1/2 (X-\bar{X})^T \Sigma^{-1} (X-\bar{X})\right\}$$

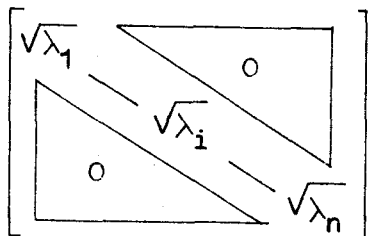
Le changement de base qui diagonalise la matrice  $\Sigma^{-1}$  (cf. chapitre IV, paragraphe 3) permet de calculer la probabilité pour qu'une observation tombe dans le domaine de concavité D de  $p(X)$  sous la forme:

$$P(Y \in D) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \int_D \exp\left\{-1/2 Y^T \Lambda Y\right\} dY$$

où D est l'intérieur de la surface définie par:

$$Y^T \Lambda Y = 1$$

Après la transformation linéaire:

$$Z = \begin{bmatrix} \sqrt{\lambda_1} & & & \\ & \sqrt{\lambda_i} & & \\ & & & \sqrt{\lambda_n} \end{bmatrix} Y$$


cette probabilité devient:

$$P(Y \in D) = \frac{\prod_{i=1}^n v_i}{(2\pi)^{n/2} |\Sigma|^{1/2}} \int_C \exp\left\{-1/2 Z^T Z\right\} dz$$

où C est le cercle de rayon unité défini par:

$$Z^T Z = 1$$

Comme:

$$|\Sigma|^{1/2} = \prod_{i=1}^n \sqrt{\lambda_i}$$

on peut énoncer le

Théorème 5.1.

La probabilité pour qu'une observation d'un échantillon normal appartienne au domaine de concavité de la fonction de densité correspondante est une constante, notée  $\beta$ , indépendante de la matrice de covariance de la distribution.

Considérons maintenant le mélange  $f(X)$  de l'équation 5.1. La probabilité pour qu'une observation d'un échantillon distribué selon la fonction  $f(X)$  provienne de la classe  $C_k$  et appartienne au domaine de concavité  $D_k$  de la fonction de densité  $p(X|C_k)$  est égale à  $\beta.P(C_k)$ .

La fonction de densité  $p(X|C_k)$  étant supposée fortement prépondérante sur les autres composantes du mélange à l'intérieur du domaine  $D_k$ , la plupart des observations appartenant à ce domaine proviennent de la classe  $C_k$ . La probabilité  $\beta.P(C_k)$  peut ainsi être approchée par la probabilité pour qu'une observation appartienne, quelle que soit la classe dont elle provient, au domaine  $D_k$ .

De plus, comme nous assimilons le domaine de concavité de chaque composante au domaine de concavité du mélange correspondant, la probabilité  $\beta.P(C_k)$  peut être approchée par la probabilité pour qu'une observation appartienne au domaine de concavité  $D_k$  du mélange. On peut donc écrire:

$$\beta.P(C_k) \approx P(X \in D_k)$$

L'importance des erreurs introduites par cette approximation dépend du degré de chevauchement des composantes du mélange analysé. Nous verrons ultérieurement que cette approximation, comme toutes les précédentes, permet d'optimiser le processus de classification de manière très satisfaisante.

V.4.2. DETERMINATION DES VALEURS APPROCHÉES DES  
PROBABILITÉS A PRIORI

Soit  $n_k$  le nombre d'observations situées à l'intérieur de l'ensemble des hypercubes à convexité négative

adjacents constituant le domaine modal  $\underline{D}_k^0$ . Le rapport  $a_k/a$  est un estimateur classique de la probabilité  $p(X \in \underline{D}_k^0) / 2/$ . Le domaine  $\underline{D}_k^0$  n'étant en fait qu'une matérialisation du domaine  $\underline{D}_k$ , on peut écrire:

$$\beta P(C_k) \approx a_k/a, \quad k = 1, 2, \dots, K$$

ou encore:

$$\frac{P(C_k)}{a_k} \approx C, \quad k = 1, 2, \dots, K$$

C étant une constante ne dépendant que du problème traité.

Des valeurs approchées  $\hat{P}(C_k)$  des probabilités a priori  $P(C_k)$  peuvent alors être obtenues en résolvant le système d'équations:

$$\begin{cases} \frac{\hat{P}(C_1)}{a_1} = \frac{\hat{P}(C_2)}{a_2} = \dots = \frac{\hat{P}(C_k)}{a_k} = \dots = \frac{\hat{P}(C_K)}{a_K} \\ \hat{P}(C_1) + \hat{P}(C_2) + \dots + \hat{P}(C_k) + \dots + \hat{P}(C_K) = 1 \end{cases}$$

A ce stade, on dispose de valeurs approchées de tous les paramètres définissant le mélange et le problème de la classification automatique des observations se trouve posé en termes probabilistes.

#### V.5. OPTIMISATION DU PROCESSUS DE CLASSIFICATION

Le contexte probabiliste dans lequel nous nous sommes placés permet d'optimiser le processus de classification automatique en minimisant le taux d'erreur. Lorsque le modèle probabiliste d'un mélange est connu, on peut en effet déterminer la probabilité pour qu'une observation  $X$  appartienne à la classe  $C_k$ . Par application du théorème de Bayes /8/, cette probabilité, appelée probabilité a posteriori de la classe  $C_k$ , est donnée par la relation:

$$P(C_k | X) = \frac{p(X | C_k) P(C_k)}{f(X)}$$

Le risque d'erreur lié au classement de l'observation  $X$  dans la classe  $C_k$  est donc:

$$R(C_k | X) = 1 - P(C_k | X)$$

La méthode la plus naturelle pour définir un classement optimal consiste à rechercher la procédure qui minimise le risque d'erreur, c'est à dire celle qui affecte une observation  $X$  à la classe  $C_k$  pour laquelle la probabilité a posteriori  $P(C_k | X)$  est maxima.

Le taux d'erreur minimum sera donc obtenu en décidant que l'observation  $X$  appartient à la classe  $C_k$  si et seulement si:

$$P(C_k | X) \geq P(C_i | X) \quad , \quad i = 1, 2, \dots, K ; i \neq k$$

La partition de l'espace de représentation  $R_0^n$  en régions  $R_k$  telles que:

$$X \in R_k \implies P(C_k | X) \geq P(C_i | X) \quad , \quad i = 1, 2, \dots, K ; i \neq k$$

permet alors de résoudre le problème du classement de manière particulièrement simple et économique au point de vue des calculs. On est alors assuré qu'aucun autre partitionnement ne peut conduire à un taux d'erreur de classement inférieur. C'est pourquoi cette procédure, connue sous le nom de classement bayésien est également appelée "classement optimal" /2/.

Le problème du classement optimal revient donc à trouver un ensemble de fonctions  $g_k(X)$ ,  $k = 1, 2, \dots, K$ , appelées fonctions de décisions, telles que l'observation  $X$  est attribuée à un individu de la classe  $C_k$  si et seulement si:

$$g_k(X) \geq g_i(X) \quad , \quad i = 1, 2, \dots, K ; i \neq k$$

Compte-tenu de ce qui précède, les fonctions de décision qui optimisent la procédure de classement peuvent être choisies de la forme /6/:

$$g_k(X) = p(X | C_k) P(C_k)$$

Les fonctions de densité normales étant de type exponentiel, on préfère, pour les mélanges gaussiens, utiliser le logarithme des fonctions  $g_k(X)$  :

$$\begin{aligned}
 \text{Log } g_k(X) &= \text{Log} [p(X|C_k) P(C_k)] \\
 &= -\frac{n}{2} \text{Log } 2\pi - \frac{1}{2} \text{Log} |\Sigma_k| \\
 &\quad - \frac{1}{2} (X-\bar{X}_k)^T \Sigma_k^{-1} (X-\bar{X}_k) + \text{Log } P(C_k)
 \end{aligned}$$

Le terme:

$$\frac{n}{2} \text{Log } 2\pi$$

ne dépend pas de  $k$  et peut être éliminé, de telle sorte que l'on utilise finalement les fonctions de décision:

$$\begin{aligned}
 g'_k(X) &= -\frac{1}{2} \text{Log} |\Sigma_k| - \frac{1}{2} (X-\bar{X}_k)^T \Sigma_k^{-1} (X-\bar{X}_k) + \text{Log } P(C_k) \quad (5.2.) \\
 &k = 1, 2, \dots, K
 \end{aligned}$$

L'optimisation du processus de classification consiste alors, dans le contexte de cette étude, à remplacer dans l'expression 5.2. les valeurs exactes du vecteur-moyenne, de la matrice de covariance et de la probabilité a priori de la classe  $C_k$  par les valeurs approchées obtenues grâce à la procédure d'analyse des mélanges proposée. Les fonctions de décision deviennent alors:

$$\begin{aligned}
 \hat{g}_k(X) &= -\frac{1}{2} \text{Log} |\hat{\Sigma}_k| - \frac{1}{2} (X-\hat{X}_k)^T \hat{\Sigma}_k^{-1} (X-\hat{X}_k) + \text{Log } \hat{P}(C_k) \\
 &k = 1, 2, \dots, K
 \end{aligned}$$

Nous présentons maintenant, sur quelques exemples, les résultats obtenus par l'application de cette procédure de classification. L'étude détaillée qui suit a pour principal but d'analyser le comportement de la méthode en fonction de différents paramètres et de préciser certains aspects pratiques de sa mise en oeuvre sur calculateur numérique. L'utilisation d'exemples artificiels permettra également de juger la qualité de l'optimisation en déterminant le taux d'erreur de classement dans différentes situations.

## V.6. ANALYSE DES PERFORMANCES DE LA PROCEDURE DE CLASSIFICATION

Avant d'illustrer par des exemples la technique utilisée pour optimiser le processus de classification, rappelons les six phases principales du traitement des données mises en oeuvre par l'algorithme proposé.

- Phase 1 Normalisation et mise à l'échelle de l'espace de représentation des données.
- Phase 2 Test de convexité de la fonction de densité sous-jacente sur la base de l'ensemble des observations disponibles
- Phase 3 Détermination des domaines modaux du mélange par agrégation des hypercubes à convexité négative adjacents.  
Dénombrement des classes en présence.
- Phase 4 Représentation des domaines modaux par des modèles hyperellipsoïdaux.  
Détermination des valeurs approchées du vecteur-moyenne, de la matrice de covariance et de la probabilité a priori de chaque classe mise en évidence.
- Phase 5 Calcul des fonctions de décision.
- Phase 6 Affectation des observations aux différentes classes.

Les observations utilisées pour les expériences de test sont des données artificielles générées par ordinateur. Parmi les nombreux avantages présentés par l'utilisation de données artificielles comme moyen de test, le plus important tient au fait que l'expérimentateur connaît la structure de ses données et en garde complètement le contrôle /1/.

#### V.6.1. ANALYSE DE DONNEES BIDIMENSIONNELLES

La qualité de l'optimisation de la classification obtenue par la procédure proposée ci-dessus est d'abord testée sur des exemples bidimensionnels qui permettent une visualisation simple des résultats.

##### Exemple 1

Les premières données utilisées sont constituées par un mélange de trois classes normales, équiprobables, dont les paramètres sont consignés dans le tableau 5.1. Pour chacune des classes, on génère un nombre identique de vecteurs aléatoires gaussiens bidimensionnels.

Nous commencerons par montrer l'influence du paramètre  $\alpha_n$ , c'est à dire du pas de discrétisation de l'espace

$R_1^n$ , sur la qualité des résultats obtenus. Nous analyserons ensuite, sur le même exemple, l'effet du nombre  $q$  d'observations disponibles.

#### Influence du pas de discretisation

L'algorithme de classification a d'abord été appliqué pour différentes valeurs de  $\alpha_q$ , à l'ensemble des 300 observations représenté sur la figure 5.4. Cet ensemble est constitué de trois lots de 100 observations distribuées selon les trois lois normales équiprobables du tableau 5.1. Pour chaque valeur de  $\alpha_q$ , les performances de la procédure de classification sont mesurées par le taux d'erreur, estimé par le rapport entre le nombre d'observations mal classées et le nombre total d'observations constituant l'échantillon analysé. Les variations de ce taux d'erreur en fonction du paramètre  $\alpha_q$  (cf. figure 5.5.) font apparaître une valeur optimale  $\alpha_q^0$  de  $\alpha_q$  pour laquelle le nombre d'observations mal classées est minimum.

La figure 5.6. indique les domaines de concavité des composantes du mélange et les modèles ellipsoïdaux des domaines modaux du mélange obtenus pour cette valeur optimale de  $\alpha_q$ . Bien que ces modèles ne soient qu'une approximation des domaines de concavité des composantes, leur utilisation pour déterminer les paramètres du mélange conduit à un taux d'erreur très proche de l'optimum théorique\* (cf. tableau 5.1.) . Mais il faut également remarquer que la qualité des résultats dépend sensiblement du choix du paramètre  $\alpha_q$  dont l'ajustement requiert donc une attention particulière.

#### Ajustement du pas de discrétisation

Nous avons vu, au chapitre III, que l'application du test de convexité nécessite un ajustement du volume des domaines utilisés pour l'estimation des fonctions de densité de probabilité en fonction de la taille de l'échantillon disponible.

---

\* L'utilisation de données générées artificiellement permet de calculer le taux d'erreur théorique minimum à partir des valeurs exactes des paramètres du mélange. Ce taux d'erreur optimal sert habituellement de référence pour juger les performances des procédures de classification automatique/7/.

	VECTEURS MOYENNES	MATRICES DE COVARIANCE	PROBABILITES A PRIORI
CLASSE 1	$\bar{X}_1 = \begin{bmatrix} 8 \\ 8 \end{bmatrix}$ $\hat{X}_1 = \begin{bmatrix} 8.09 \\ 8.52 \end{bmatrix}$	$\Sigma_1 = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$ $\hat{\Sigma}_1 = \begin{bmatrix} 4.41 & 0.32 \\ 0.32 & 5.56 \end{bmatrix}$	$P(C_1) = 0.33$ $\hat{P}(C_1) = 0.34$
CLASSE 2	$\bar{X}_2 = \begin{bmatrix} 14 \\ 14 \end{bmatrix}$ $\hat{X}_2 = \begin{bmatrix} 13.74 \\ 14.29 \end{bmatrix}$	$\Sigma_2 = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$ $\hat{\Sigma}_2 = \begin{bmatrix} 3.83 & -0.09 \\ -0.09 & 3.20 \end{bmatrix}$	$P(C_2) = 0.33$ $\hat{P}(C_2) = 0.31$
CLASSE 3	$\bar{X}_3 = \begin{bmatrix} 14 \\ 2 \end{bmatrix}$ $\hat{X}_3 = \begin{bmatrix} 14.21 \\ 2.46 \end{bmatrix}$	$\Sigma_3 = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$ $\hat{\Sigma}_3 = \begin{bmatrix} 5.67 & 0.67 \\ 0.67 & 6.67 \end{bmatrix}$	$P(C_3) = 0.33$ $\hat{P}(C_3) = 0.35$
TAUX D'ERREUR THEORIQUE MINIMUM : 3.27%    TAUX D'ERREUR DE LA CLASSIFICATION : 3.33%			

Tableau 5.1. Valeurs exactes et approchées des paramètres statistiques de la distribution de l'exemple 1



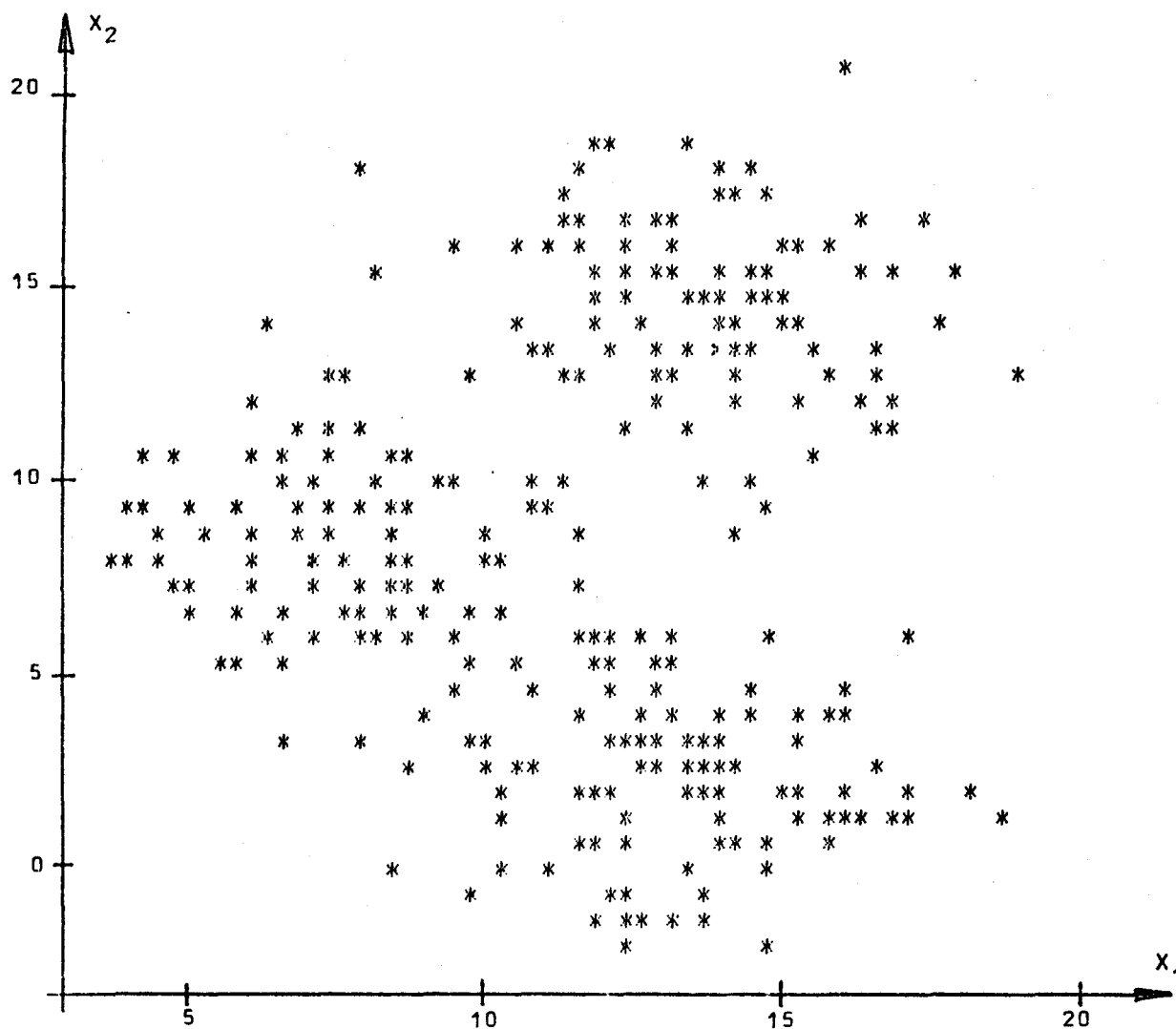


Figure 5.4. Représentation graphique des 300 observations de l'exemple 1

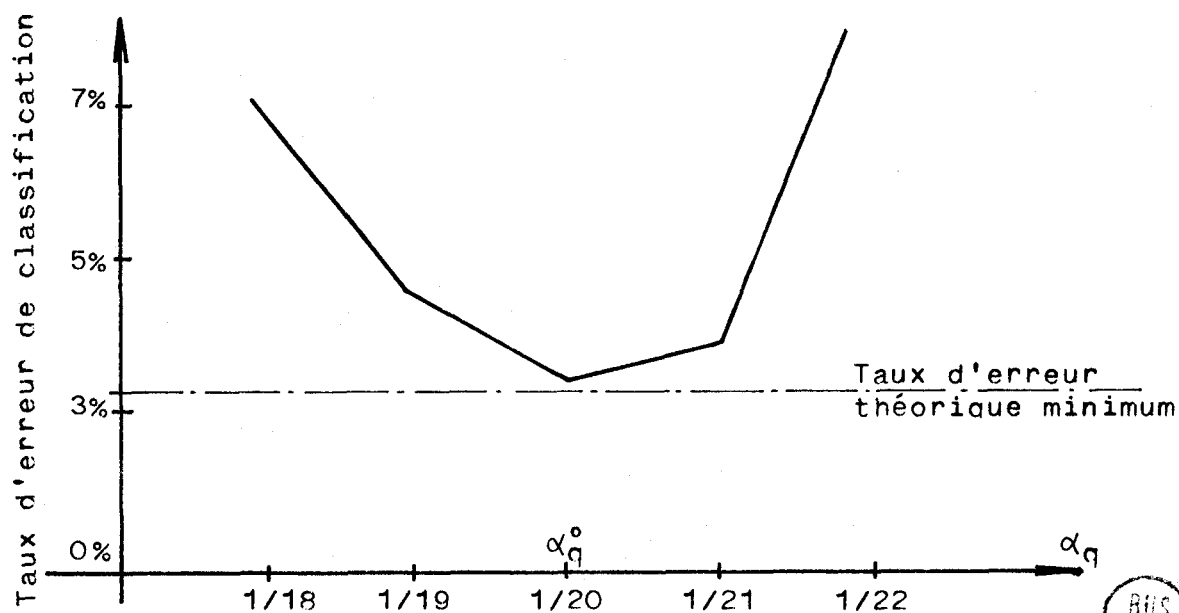


Figure 5.5. Variations du taux d'erreur de la classification en fonction du paramètre de discrétisation  $\alpha_q$  (L'échantillon analysé est celui de la figure 5.4.)

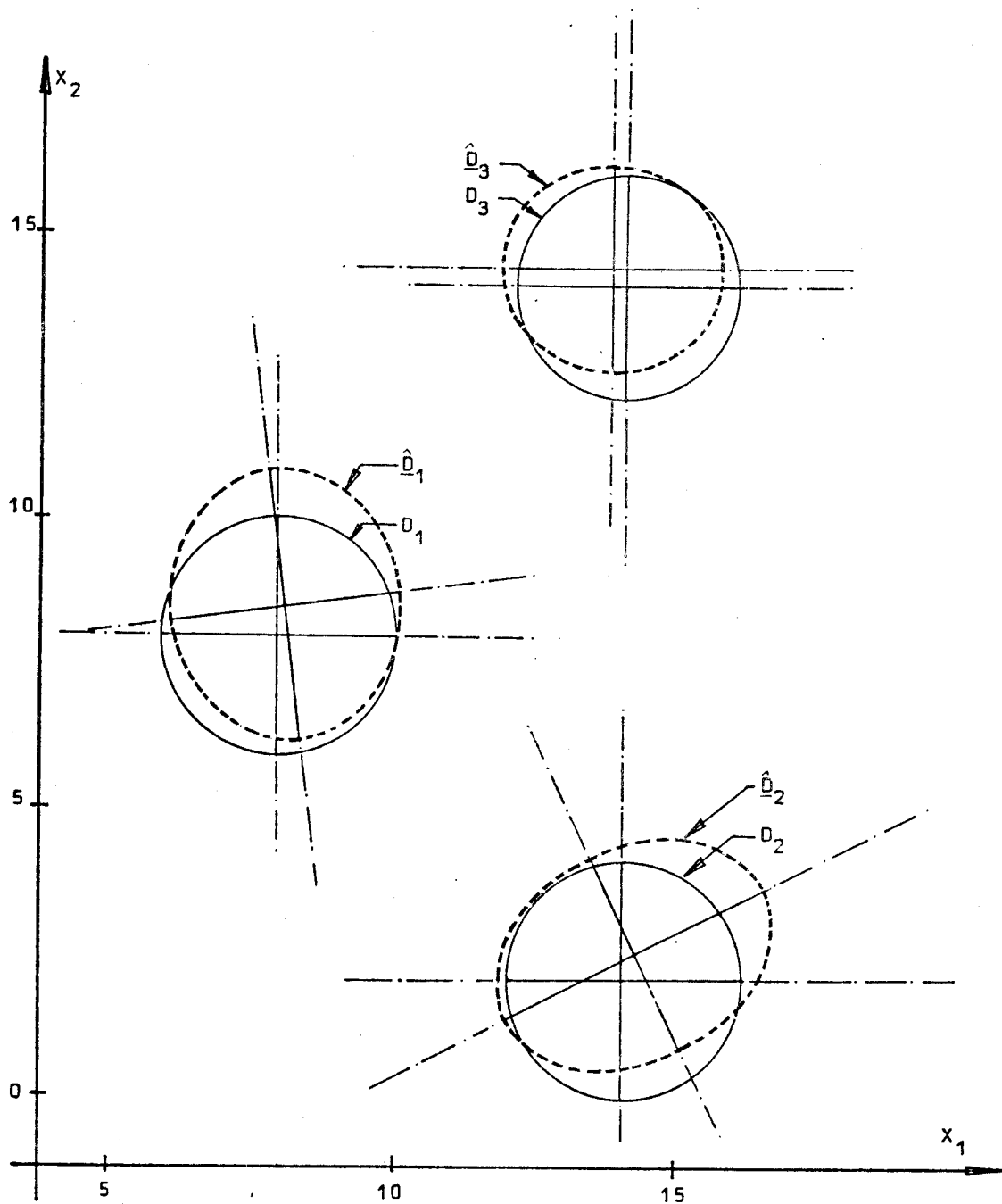


Figure 5.6. Modèles ellipsoïdaux des domaines modaux du mélange de l'exemple 1



$D_k, k=1,2,3$  : Domaines de concavité des composantes du mélange



$\hat{D}_k, k=1,2,3$  : Modèles ellipsoïdaux des domaines modaux du mélange

Nous avons choisi (cf. équation 3.12.) de prendre  $\alpha_q$  égal à l'inverse d'un entier et le plus proche possible de:

$$\alpha'_q = \alpha_0 q^{-1/2n}$$

c'est à dire tel que:

$$\frac{q^{1/2n}}{\alpha_q} = C, 1/\alpha_q \text{ entier} \quad (5.3.)$$

où  $C$  est une constante à ajuster en fonction de la nature des données.

Dans le stade d'exploration de la structure d'un échantillon inconnu de taille  $q$  et de dimension  $n$ , le choix de  $\alpha_0$  (ou, à une constante multiplicative près, de  $\alpha_q$ ) peut être guidé en faisant appel au concept de "stabilité du nombre de modes" /3/.

Dans la pratique, on ajuste en fait le nombre de divisions sur chaque axe de l'espace  $R_1^n$ , c'est à dire le paramètre  $N_q = 1/\alpha_q$ . On détermine d'abord la plus grande plage de variation de  $N_q$  pour laquelle le nombre de domaines modaux détectés dans le mélange, à l'aide du test de convexité et de la procédure d'agrégation associée, demeure constant. Le choix de ce type de paramètre de discrétisation au milieu de cette plage correspondant à la plus grande stabilité du nombre de modes a déjà été utilisé pour optimiser certains algorithmes de classification /3//10/.

Pour l'exemple étudié, le tracé du nombre de domaines modaux détectés en fonction de  $N_q = 1/\alpha_q$  montre l'intérêt de cette procédure. En effet, le milieu de la plus grande plage de stabilité du nombre de modes correspond à la valeur  $1/\alpha_q = 21$  qui est très proche de la valeur optimale  $1/\alpha_q^0 = 20$  (cf. figure 5.7.).

Il est important de préciser que, pour dénombrer les domaines modaux du mélange, il s'avère nécessaire de rejeter les ensembles d'hypercubes à convexité négative qui apparaissent uniquement du fait que la convexité locale de la fonction de densité n'est pas partout définie (cf. Chapitre IV, paragraphe IV.1.) et qui ne correspondent à aucun mode réel. La taille réduite de ces ensembles par rapport à celle des domaines modaux permet en général de les identifier sans difficulté. A chaque pas d'une procédure

itérative, on rejette l'ensemble qui contient le plus petit nombre d'observations. Il n'y a cependant rejet que si ce nombre est inférieur à une certaine fraction  $x$  du nombre d'observations contenues dans la totalité des ensembles restant à la fin de l'itération précédente. Cette procédure revient à fixer à la valeur  $x$  la limite inférieure de la probabilité a priori des classes détectables.

Pour tous les exemples artificiels présentés dans ce chapitre, seuls les modes correspondant à des classes présentant une probabilité a priori supérieure à 5% ont été retenus. Bien que l'ajustement de ce seuil ne soit pas critique (on obtient, pour tous les exemples de ce chapitre, des résultats identiques pour des valeurs de ce seuil comprises entre 5% et 21%), aucune règle précise ne permet de choisir a priori une valeur de  $x$  particulière lors de la résolution d'un problème donné: il s'agit essentiellement d'une question d'expérience et de circonstances.

Il sera toujours fait appel au concept de stabilité du nombre de modes pour ajuster le pas de discrétisation  $\alpha_q$  durant la phase d'exploration de tout ensemble de données de structure inconnue. Si, après cet ajustement, la taille de l'échantillon disponible venait à varier, par suite d'une collecte d'observations supplémentaires par exemple, l'analyste pourrait modifier le pas de discrétisation en utilisant la relation 5.3. C'est précisément cette stratégie que nous avons adoptée pour analyser l'effet de la taille de l'échantillon disponible sur la qualité de la classification de données ayant la structure du mélange décrit dans le tableau 5.1.

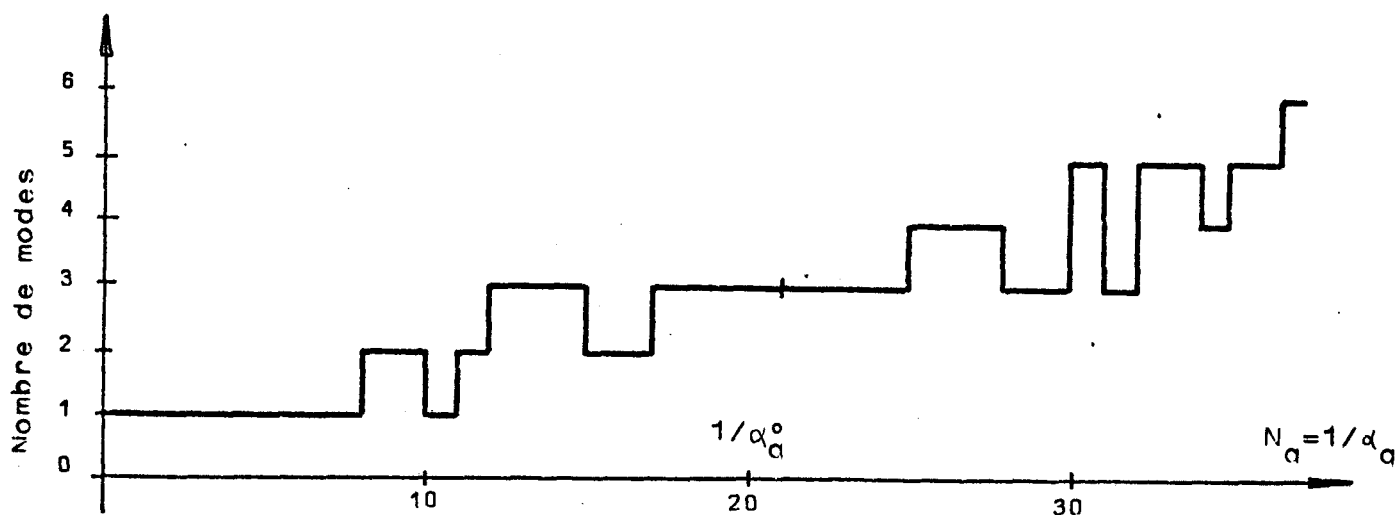


Figure 5.7. Nombre de modes détectés dans le mélange de l'exemple 1 en fonction du pas de discrétisation  $\alpha_q$  (L'échantillon analysé est celui de la figure 5.4.)

## Influence de la taille de l'échantillon

Nous avons vu au chapitre précédent que l'identification d'une distribution normale est d'autant meilleure que le nombre d'observations disponibles est important. On retrouve un phénomène analogue dans l'analyse des mélanges et le pourcentage d'observations mal classées tend à diminuer lorsque le nombre d'observations augmente (cf. figure 5.8.).

Soulignons cependant que cette décroissance du taux d'erreur ne garantit pas sa convergence vers l'optimum théorique. En effet, même avec des échantillons de taille illimitée, les domaines de concavité des mélanges resteraient toujours des approximations des domaines de concavité des composantes. Cette approche n'est donc pas consistante en ce sens qu'il n'y a pas convergence des résultats vers les valeurs exactes des paramètres. Cependant, malgré des informations a priori limitées à la seule connaissance de la forme analytique des composantes des mélanges, les taux d'erreur obtenus au cours des tests restent très proches des valeurs optimales.

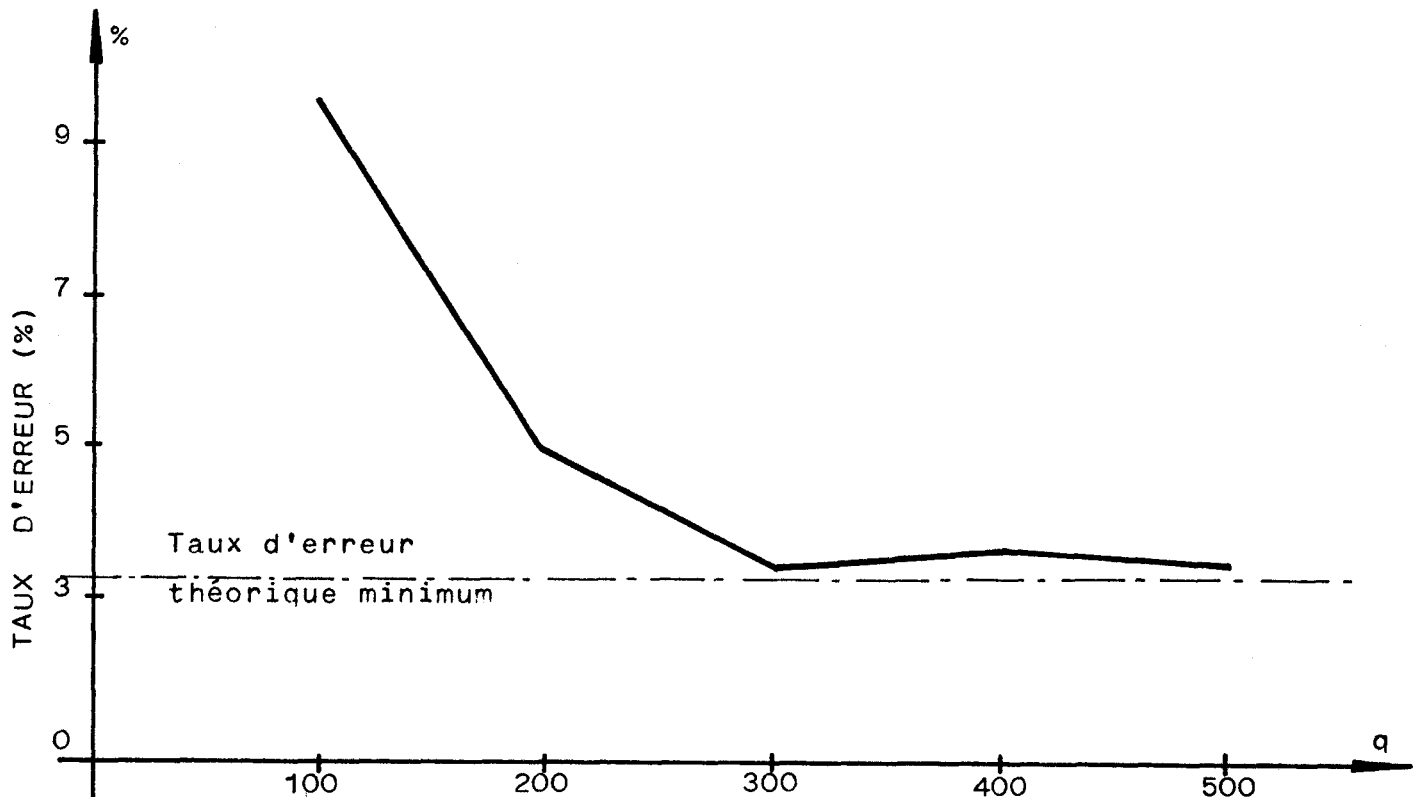


Figure 5.8. Variations du taux d'erreur en fonction de la taille de l'échantillon disponible pour l'exemple 1

Exemple 2

Après le premier exemple qui a permis de préciser les aspects pratiques de la procédure proposée et d'étudier le comportement de la méthode en fonction des différents paramètres, nous proposons un second exemple destiné à mettre en évidence les possibilités d'analyse de mélanges de composantes normales non sphériques avec des probabilités a priori différentes.

Le tableau 5.2. résume les paramètres statistiques du second mélange étudié pour lequel on dispose de l'échantillon de 300 observations représenté figure 5.9.

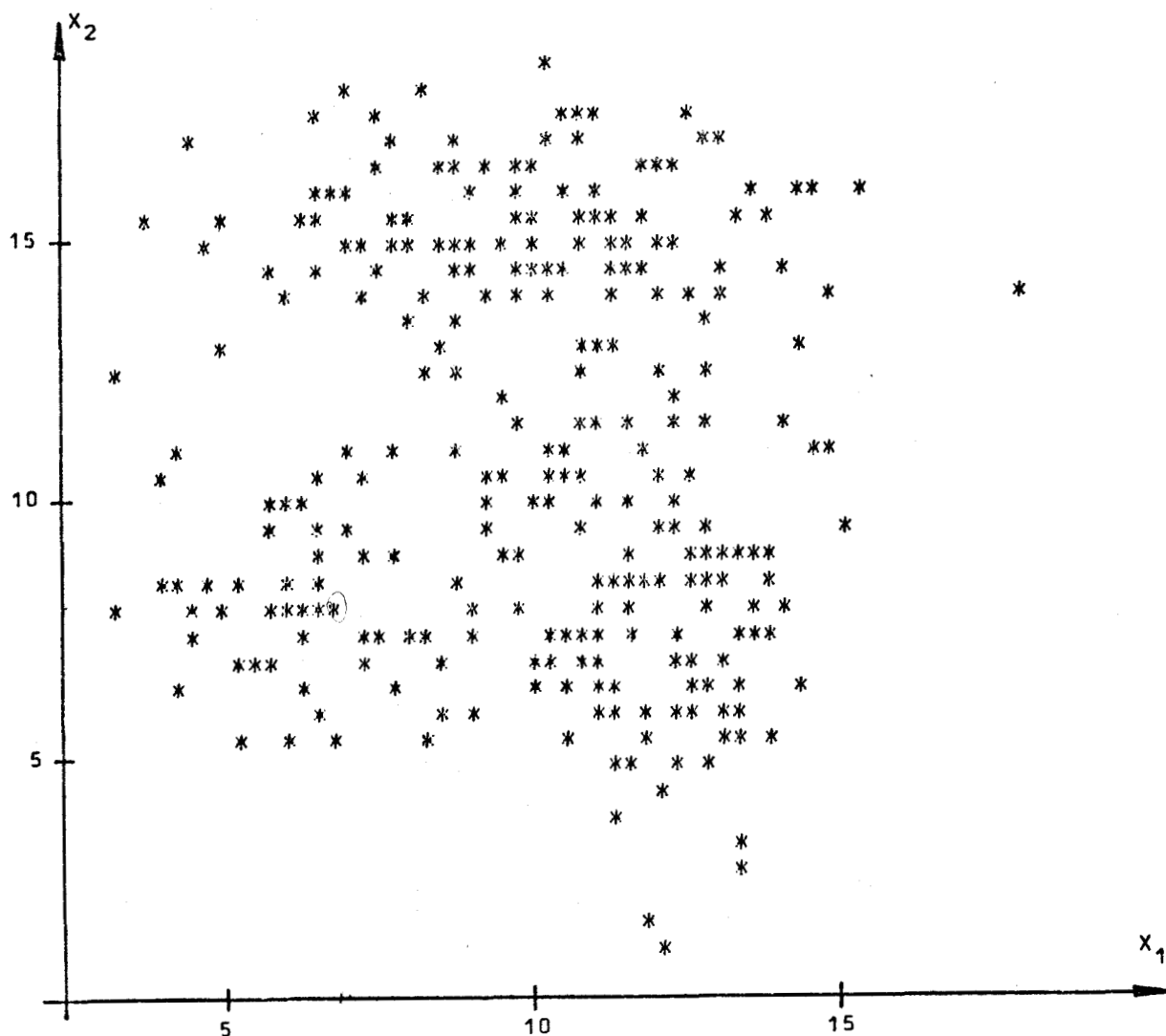


Figure 5.9. Représentation graphique des 300 observations de l'exemple 2 .

	VECTEURS MOYENNES	MATRICES DE COVARIANCE	PROBABILITES A PRIORI
CLASSE 1	$\bar{X}_1 = \begin{bmatrix} 7 \\ 8 \end{bmatrix}$ $\hat{X}_1 = \begin{bmatrix} 7,38 \\ 7,60 \end{bmatrix}$	$\Sigma_1 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ $\hat{\Sigma}_1 = \begin{bmatrix} 1,76 & -0,20 \\ -0,20 & 1,77 \end{bmatrix}$	$P(C_1) = 0,20$ $\hat{P}(C_1) = 0,21$
CLASSE 2	$\bar{X}_2 = \begin{bmatrix} 12 \\ 8 \end{bmatrix}$ $\hat{X}_2 = \begin{bmatrix} 12,19 \\ 7,37 \end{bmatrix}$	$\Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix}$ $\hat{\Sigma}_2 = \begin{bmatrix} 2,61 & -0,08 \\ -0,08 & 4,58 \end{bmatrix}$	$P(C_2) = 0,40$ $\hat{P}(C_2) = 0,40$
CLASSE 3	$\bar{X}_3 = \begin{bmatrix} 10 \\ 15 \end{bmatrix}$ $\hat{X}_3 = \begin{bmatrix} 10,14 \\ 14,49 \end{bmatrix}$	$\Sigma_3 = \begin{bmatrix} 7 & 0 \\ 0 & 2 \end{bmatrix}$ $\hat{\Sigma}_3 = \begin{bmatrix} 10,55 & 0,61 \\ 0,61 & 1,73 \end{bmatrix}$	$P(C_3) = 0,40$ $\hat{P}(C_3) = 0,39$
TAUX D'ERREUR THEORIQUE MINIMUM : 4,3%      TAUX D'ERREUR DE LA CLASSIFICATION : 5,3%			

Tableau 5.2. Valeurs exactes et approchées des paramètres statistiques de la distribution de l'exemple 2



La figure 5.10. indique les domaines de concavité des différentes composantes et les modèles ellipsoïdaux des domaines modaux du mélange obtenus après optimisation du pas de discrétisation de l'espace.

L'écart entre le taux d'erreur obtenu et la valeur optimale reste faible ici aussi puisqu'il ne correspond qu'à trois observations mal classées sur un total de 300. Notons cependant que cet écart est plus grand que dans l'exemple précédent, conséquence du chevauchement plus important des trois composantes.

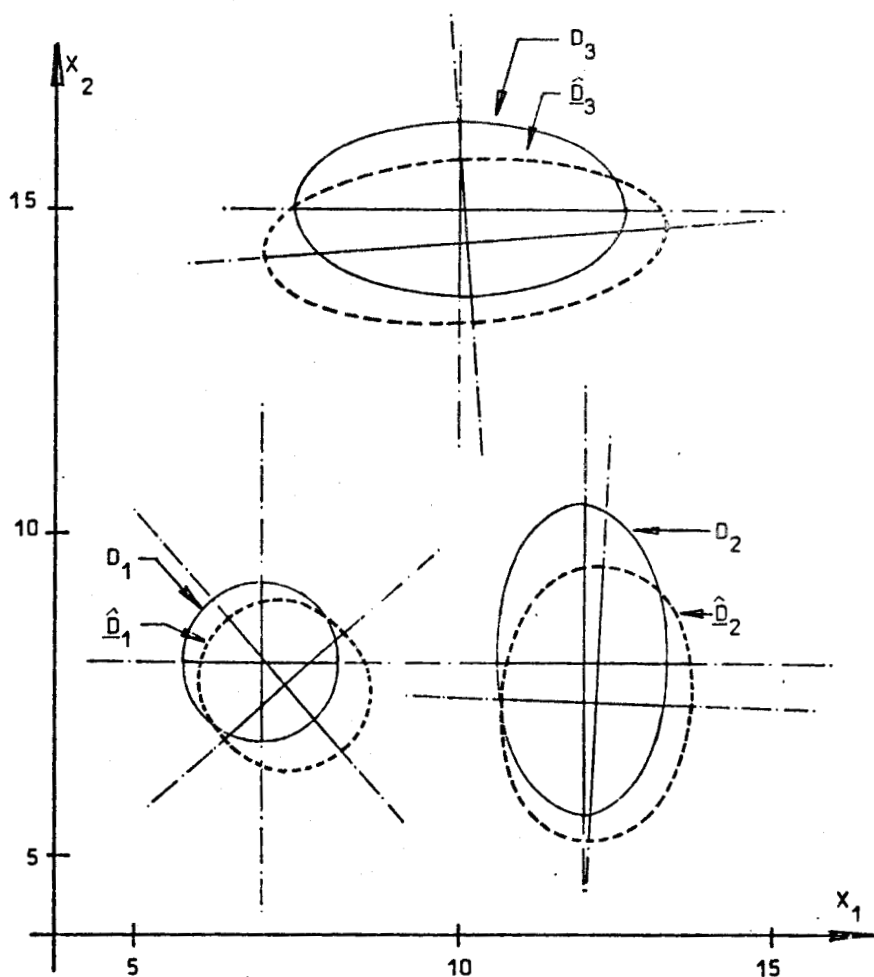




Figure 5.10. Modèles ellipsoïdaux des domaines modaux du mélange de l'exemple 2.

-   $D_k, k=1,2,3$  : Domaines de concavité des composantes du mélange
-   $\hat{D}_k, k=1,2,3$  : Modèles ellipsoïdaux des domaines modaux du mélange



Exemple 3

Le dernier exemple bidimensionnel présenté est destiné à montrer les possibilités d'analyse de mélanges de composantes ayant des matrices de covariance non diagonales (cf. tableau 5.3.). L'algorithme a été appliqué à l'échantillon constitué par les 300 observations représentées sur la figure 5.11.

Comme dans les exemples précédents, les modèles ellipsoïdaux des domaines modaux du mélange sont des approximations satisfaisantes des domaines de concavité des composantes (cf. figure 5.12.). En effet, le taux d'erreur obtenu en utilisant les paramètres statistiques déduits de ces modèles reste très proche du taux d'erreur théorique optimal (cf. tableau 5.3.).

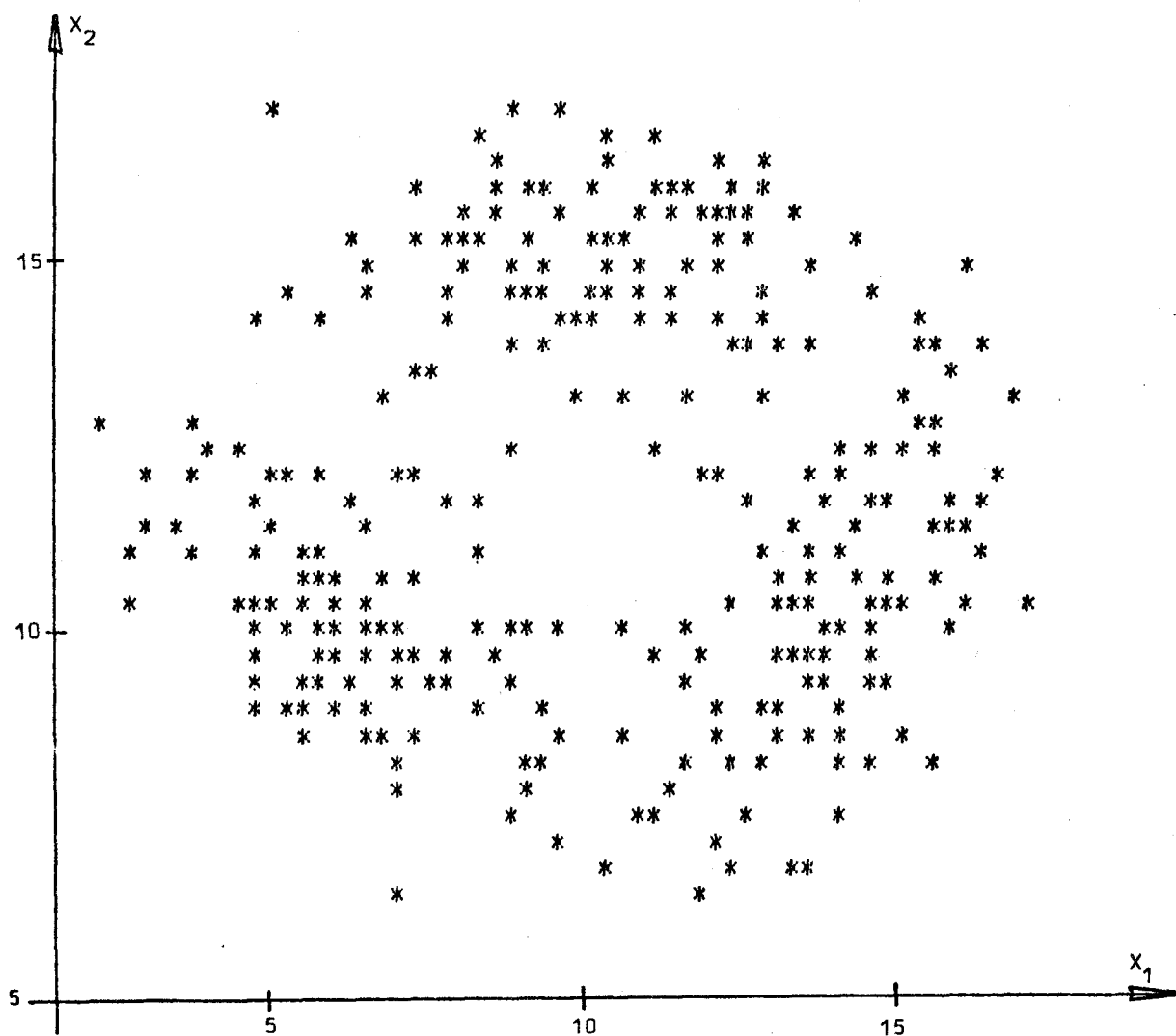


Figure 5.11. Représentation graphique des 300 observations de l'exemple 3 .

	VECTEURS MOYENNES	MATRICES DE COVARIANCE	PROBABILITES A PRIORI
CLASSE 1	$\bar{X}_1 = \begin{bmatrix} 10 \\ 15 \end{bmatrix}$ $\hat{X}_1 = \begin{bmatrix} 9,8 \\ 14,8 \end{bmatrix}$	$\Sigma_1 = \begin{bmatrix} 7 & 0 \\ 0 & 2 \end{bmatrix}$ $\hat{\Sigma}_1 = \begin{bmatrix} 8,25 & -0,69 \\ -0,69 & 2,17 \end{bmatrix}$	$P(C_1) = 0,33$ $\hat{P}(C_1) = 0,37$
CLASSE 2	$\bar{X}_2 = \begin{bmatrix} 6 \\ 10 \end{bmatrix}$ $\hat{X}_2 = \begin{bmatrix} 6,6 \\ 9,5 \end{bmatrix}$	$\Sigma_2 = \begin{bmatrix} 4,5 & 2,5 \\ 2,5 & 4,5 \end{bmatrix}$ $\hat{\Sigma}_2 = \begin{bmatrix} 4,49 & 2,74 \\ 2,74 & 5,60 \end{bmatrix}$	$P(C_2) = 0,33$ $\hat{P}(C_2) = 0,31$
CLASSE 3	$\bar{X}_3 = \begin{bmatrix} 14 \\ 10 \end{bmatrix}$ $\hat{X}_3 = \begin{bmatrix} 13,9 \\ 9,8 \end{bmatrix}$	$\Sigma_3 = \begin{bmatrix} 4,5 & -2,5 \\ -2,5 & 4,5 \end{bmatrix}$ $\hat{\Sigma}_3 = \begin{bmatrix} 4,85 & -3,46 \\ -3,46 & 6,25 \end{bmatrix}$	$P(C_3) = 0,33$ $\hat{P}(C_3) = 0,33$
TAUX D'ERREUR THEORIQUE MINIMUM : 3,9%    TAUX D'ERREUR DE LA CLASSIFICATION : 5,3%			

Tableau 5.3. Valeurs exactes et approchées des paramètres statistiques de la distribution de l'exemple 3

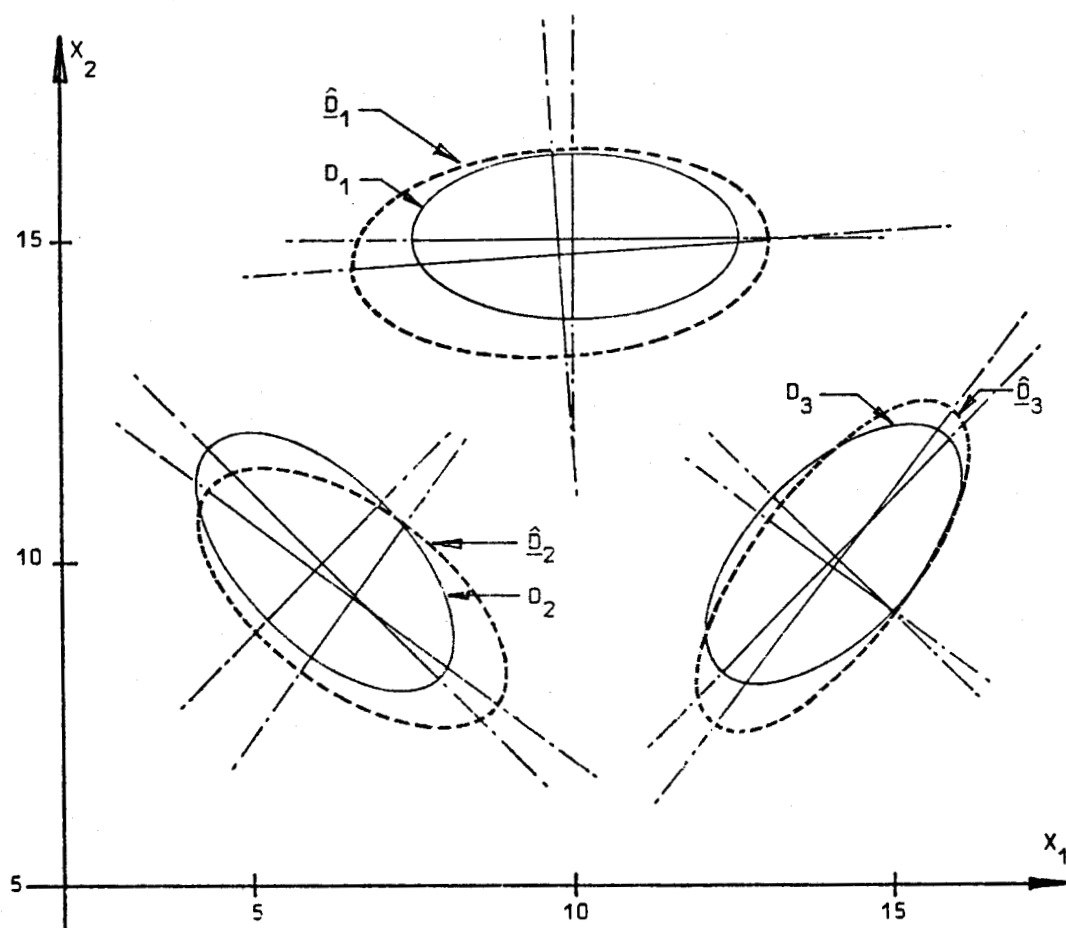




Figure 5.12. Modèles ellipsoïdaux des domaines modaux du mélange de l'exemple 3.

- 
 $D_k, k=1,2,3$  : Domaines de concavité des composantes du mélange
- 
 $\hat{D}_k, k=1,2,3$  : Modèles ellipsoïdaux des domaines modaux du mélange

#### V.6.2. ANALYSE DE DONNEES MULTIDIMENSIONNELLES

Le comportement de la procédure proposée ne se prête pas à une description analytique car il dépend de la structure des données considérées. Au cours de la série de tests sur des données bidimensionnelles, nous avons étudié l'effet de certains paramètres sur les performances de la méthode. Mais nous n'avons pas encore analysé l'effet du degré de chevauchement entre les composantes sur les résultats du processus de classification.

Afin de préciser les conséquences des approximations liées à ce chevauchement, le taux d'erreur obtenu par la méthode proposée est comparé au taux d'erreur optimal théorique dans plusieurs cas multidimensionnels. Pour cette série de tests, la dispersion des observations autour du centre de chaque classe est contrôlée par la valeur des matrices de covariance (cf. tableau 5.4.).

Les trois échantillons bidimensionnels n° 1, 2 et 3 sont représentés sur les figures 5.13., 5.4. et 5.14. L'échantillon n° 3 est le cas limite au-delà duquel, par suite d'un chevauchement trop important des composantes, il devient impossible d'identifier les trois classes constituant le mélange.

Les résultats consignés dans le tableau 5.4. font apparaître, pour les échantillons n° 1, 4 et 7, un taux d'erreur inférieur au taux théorique minimum. Ce résultat paradoxal est dû au manque de précision de l'estimation du taux d'erreur pour des échantillons de taille réduite (300 observations pour ce test).

Cette série de tests montre comment l'écart entre le taux d'erreur obtenu et sa valeur optimale, tout en restant faible compte-tenu du manque d'informations a priori sur les données, tend à augmenter avec le degré de chevauchement des composantes.

Les temps d'exécution de l'algorithme, également consignés dans le tableau 5.4., sont en accord avec l'analyse des temps de calcul de l'algorithme d'estimation rapide présenté au cours du troisième chapitre. En effet, en dehors de l'exécution du test de convexité, les procédures mises en jeu par l'algorithme de classification ne sont pas coûteuses sur le plan du temps de calcul. La taille des échantillons étant identique pour l'ensemble de ces tests, il en résulte que le temps d'exécution de la procédure est une fonction quasi linéaire de la dimension des données.

	VECTEURS MOYENNES	MATRICES DE COVARIANCE	TAUX D'ERREUR THEORIQUE MINI.	TAUX D'ERREUR DE LA CLASSIFICATION	TEMPS DE CALCUL
ECHANT. N°1	$X_1 = \begin{bmatrix} 8 \\ 8 \end{bmatrix}$ $X_2 = \begin{bmatrix} 14 \\ 14 \end{bmatrix}$ $X_3 = \begin{bmatrix} 14 \\ 2 \end{bmatrix}$	$\sum_k = 2 I_2, k=1, 2, 3$	1, 2%	0, 0%	24"
ECHANT. N°2	$X_1 = \begin{bmatrix} 8 \\ 8 \end{bmatrix}$ $X_2 = \begin{bmatrix} 14 \\ 14 \end{bmatrix}$ $X_3 = \begin{bmatrix} 14 \\ 2 \end{bmatrix}$	$\sum_k = 4 I_2, k=1, 2, 3$	3, 3%	3, 4%	27"
ECHANT. N°3	$X_1 = \begin{bmatrix} 8 \\ 8 \end{bmatrix}$ $X_2 = \begin{bmatrix} 14 \\ 14 \end{bmatrix}$ $X_3 = \begin{bmatrix} 14 \\ 2 \end{bmatrix}$	$\sum_k = 8 I_2, k=1, 2, 3$	10, 1%	11, 3%	26"
ECHANT. N°4	$X_1 = \begin{bmatrix} 8 \\ 8 \\ 8 \end{bmatrix}$ $X_2 = \begin{bmatrix} 14 \\ 14 \\ 14 \end{bmatrix}$ $X_3 = \begin{bmatrix} 14 \\ 14 \\ 2 \end{bmatrix}$	$\sum_k = 2 I_3, k=1, 2, 3$	1, 0%	0, 0%	32"
ECHANT. N°5	$X_1 = \begin{bmatrix} 8 \\ 8 \\ 8 \end{bmatrix}$ $X_2 = \begin{bmatrix} 14 \\ 14 \\ 14 \end{bmatrix}$ $X_3 = \begin{bmatrix} 14 \\ 14 \\ 2 \end{bmatrix}$	$\sum_k = 4 I_3, k=1, 2, 3$	1, 7%	2, 3%	35"
ECHANT. N°6	$X_1 = \begin{bmatrix} 8 \\ 8 \\ 8 \end{bmatrix}$ $X_2 = \begin{bmatrix} 14 \\ 14 \\ 14 \end{bmatrix}$ $X_3 = \begin{bmatrix} 14 \\ 14 \\ 2 \end{bmatrix}$	$\sum_k = 8 I_3, k=1, 2, 3$	6, 1%	7, 7%	34"
ECHANT. N°7	$X_1 = \begin{bmatrix} 8 \\ 8 \\ 8 \end{bmatrix}$ $X_2 = \begin{bmatrix} 14 \\ 14 \\ 14 \end{bmatrix}$ $X_3 = \begin{bmatrix} 14 \\ 14 \\ 2 \end{bmatrix}$	$\sum_k = 2 I_4, k=1, 2, 3$	1, 0%	0, 0%	43"
ECHANT. N°8	$X_1 = \begin{bmatrix} 8 \\ 8 \\ 8 \end{bmatrix}$ $X_2 = \begin{bmatrix} 14 \\ 14 \\ 14 \end{bmatrix}$ $X_3 = \begin{bmatrix} 14 \\ 14 \\ 2 \end{bmatrix}$	$\sum_k = 4 I_4, k=1, 2, 3$	1, 4%	1, 7%	43"
ECHANT. N°9	$X_1 = \begin{bmatrix} 8 \\ 8 \\ 8 \end{bmatrix}$ $X_2 = \begin{bmatrix} 14 \\ 14 \\ 14 \end{bmatrix}$ $X_3 = \begin{bmatrix} 14 \\ 14 \\ 2 \end{bmatrix}$	$\sum_k = 8 I_4, k=1, 2, 3$	4, 5%	6, 3%	45"

Tableau 5.4. Paramètres statistiques des distributions multidimensionnelles et performances de la procédure de classification

$I_n$  : Matrice unité d'ordre n



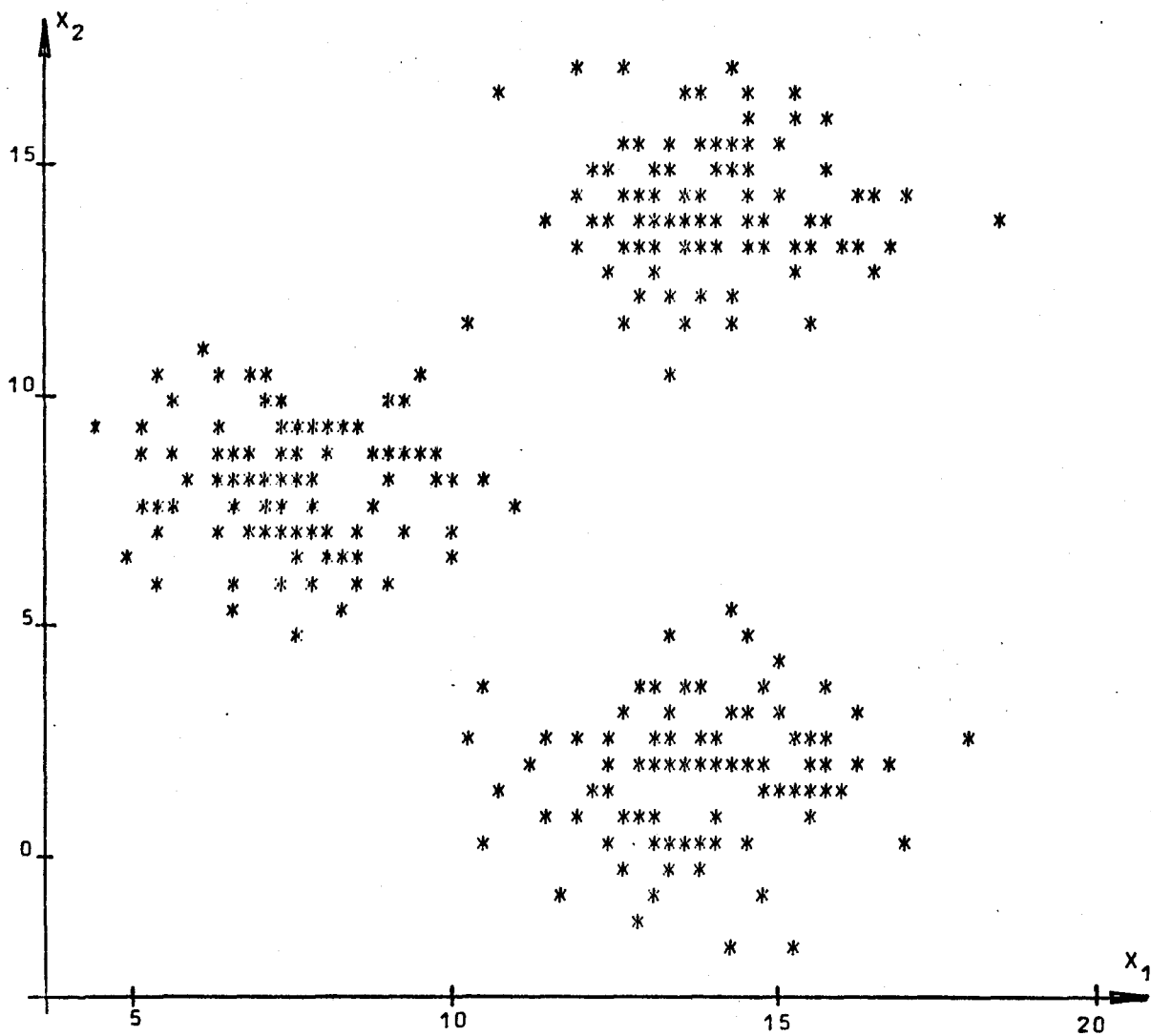


Figure 5.13. Représentation graphique des 300 observations de l'échantillon n° 1.



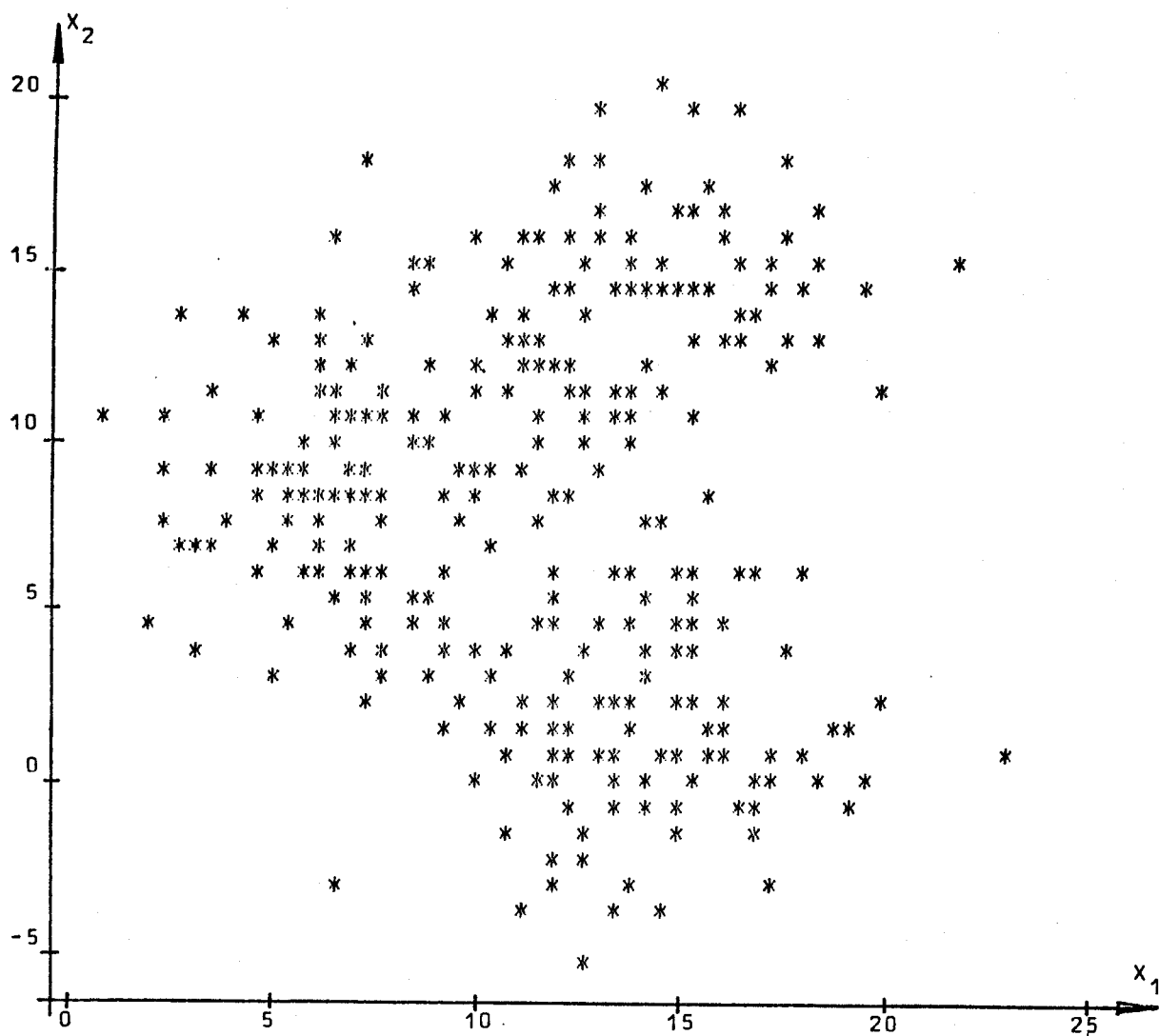


Figure 5.14. Représentation graphique des 300 observations de l'échantillon n° 3.

### V.7. CONCLUSION

La méthode décrite dans ce chapitre ramène le problème de l'optimisation du processus de classification automatique à celui de l'analyse des mélanges. Sous l'unique hypothèse que les observations de chaque classe sont distribuées normalement, nous avons proposé une méthode qui permet de déterminer le nombre de classes présentes et donne, pour chacune d'elles, des valeurs approchées du vecteur-moyenne, de la matrice de covariance et de la probabilité a priori.

La procédure utilise les relations simples qui existent entre les paramètres d'une distribution normale et les caractéristiques géométriques du domaine à l'intérieur duquel

elle est concave. Le test de convexité présenté dans la première partie de ce mémoire permet, à partir des observations, d'approcher les domaines de concavité des composantes par des modèles hyperellipsoïdaux qui sont utilisés pour déterminer les paramètres statistiques des mélanges.

L'expérience montre que le taux d'erreur de la classification basée sur les valeurs approchées des paramètres ainsi obtenus reste proche, compte-tenu du manque total d'informations a priori sur les mélanges, du taux optimal.

L'utilisation de l'algorithme rapide d'estimation présenté dans la première partie de ce travail permet d'obtenir ces résultats en un temps très court. Il est ainsi possible de multiplier les traitements et d'ajuster finement le pas de discrétisation de l'espace de représentation.

Notons que les résultats obtenus, qui sont en fait le reflet des propriétés de convexité de la fonction de densité sous-jacente, ne sont pas sensibles à des transformations linéaires des coordonnées, telles que les changements d'échelle portant sur les caractères mesurés sur les individus.

Cependant, la méthode proposée ne présente pas que des avantages. Il convient de souligner que cette approche est surtout intéressante pour des échantillons de taille importante et de dimension relativement réduite. En effet, comme pour toute procédure faisant appel à des techniques d'estimation non paramétriques, le nombre d'observations nécessaire pour obtenir des résultats satisfaisants croît avec la dimension des données /11/. Pour pallier la dégradation des performances lorsque l'on ne dispose que d'un faible nombre d'observations, nous proposons, dans la troisième partie de ce mémoire, une méthode dérivée de l'approche générale, mais adaptée à des échantillons de taille réduite et de dimension élevée.



REFERENCES (chapitre V)

- /1/ BATCHELOR, B.G.  
"Artificial data for pattern recognition"  
Proc. 2nd Int. Jt. Conf. on Pattern Recognition, Copenhagen (1974)
- /2/ DUDA, R.O. et HART, P.E.  
"Pattern Classification and Scene Analysis"  
J. Wiley New York Chapitre 2 p.10-39 (1973)
- /3/ EIGEN, D.J.; FROMM, F.R. et NORTHOUSE, R.A.  
"Cluster analysis based on dimensional information with applications to feature selection and classification"  
IEEE Trans. on Syst., Man & Cybernetics  
Vol SMC-4, n° 3, p. 284-294 (1974)
- /4/ HENRICHON, E.G. et FU, K.S.  
"On mode estimation in pattern recognition"  
Proc. 7th Symp. Adaptive Processes UCLA (1968)
- /5/ KITTLER, J.  
"A locally sensitive method for cluster analysis"  
Pattern Recognition Vol. 8 p. 23-33 (1976)
- /6/ MARILL, T. et GREEN, D.M.  
"Statistical recognition fonctions and the design of patterns recognizers"  
IRE Trans. Elec. Comp. Vol. EC-9 p. 472-477 (1960)
- /7/ MICLET, L.; GRENIER, Y. et KALIZEWSKI, J.M.  
"Reconnaissance des formes"  
Publication de l'Ecole Nat. Sup. des Telecommunications  
Paris (1978)
- /8/ PAPOULIS, A.  
"Probability, Random Variables and Stochastic Processes"  
McGraw-Hill New York (1965)
- /9/ POSTAIRE, J.G. et VASSEUR, C.  
"An approximate solution to normal mixture identification with application to unsupervised pattern classification"  
IEEE Trans. on Pattern Analysis & Machine Intelligence  
Vol. PAMI-3, n° 2, p. 163-179 (1981)
- /10/ TOMEK, I.  
"A modification of a clustering method"  
IEEE Trans. on Syst., Man & Cybernetics  
Vol. SMC-5, n° 3, p. 394-396 (1975)
- /11/ TOUSSAINT, G.T.  
"Bibliography on estimation of misclassification"  
IEEE Trans. on Info. Theory Vol. IT-20 (1974)

## TROISIEME PARTIE

APPLICATIONS PARTICULIERES DU CONCEPT DE  
CONVEXITE EN CLASSIFICATION AUTOMATIQUE

## C H A P I T R E VI

OPTIMISATION DU PROCESSUS DE CLASSIFICATION PAR ANALYSE DES DENSITES MARGINALES DE PROBABILITE
---

### VI.1. INTRODUCTION

La méthode d'optimisation du processus de classification présentée dans la seconde partie de ce mémoire nécessite un nombre d'observations disponibles d'autant plus important que la dimension des données est élevée. Dans la pratique, l'analyste peut se trouver confronté à des problèmes de forte dimension pour lesquels la taille réduite des échantillons disponibles ne permet d'obtenir ni une estimation correcte de la fonction de densité sous-jacente, ni, a fortiori, une information précise sur sa convexité.

Pour apporter une solution au problème de l'optimisation de la classification automatique dans de tels cas, nous proposons une variante de la méthode générale qui nécessite des hypothèses plus fortes, mais permet, en contre-partie, d'analyser des échantillons de taille réduite et de dimension élevée.

Il est en effet possible, dans certaines conditions, de déterminer les différents paramètres d'un mélange à partir de l'analyse des densités marginales de probabilité à la place de la

fonction de densité elle-même./6/. Ces densités sont des fonctions monovariabiles qui peuvent être estimées avec un nombre restreint d'observations, si bien que l'analyse de leur convexité ne pose pas de problème particulier, même lorsqu'on ne dispose que d'un petit échantillon./7/.

## VI.2. POSITION DU PROBLEME

On suppose, comme au chapitre V, que les données ont la structure probabiliste d'un mélange gaussien, c'est à dire que les observations proviennent de plusieurs classes, la distribution de ces observations au sein de chacune d'elles obéissant à une loi normale. On se propose d'utiliser les observations disponibles pour déterminer:

- 1) Le nombre  $K$  de classes
- 2) La probabilité a priori  $P(C_k)$  de chaque classe
- 3) Le vecteur-moyenne  $\bar{X}_k$  de chaque classe
- 4) La matrice de covariance  $\Sigma_k$  de chaque classe

afin de réaliser un classement de type bayésien pour minimiser le taux d'erreur.

Le problème est donc le même que celui traité dans le chapitre V. Afin de le résoudre lorsqu'on dispose d'un nombre trop limité d'observations pour analyser la convexité de la fonction de densité du mélange, on introduit une hypothèse supplémentaire, à savoir l'indépendance statistique des caractères des individus de chaque classe. Dans ces conditions, les matrices de covariance des différentes classes sont diagonales et peuvent s'écrire sous la forme:

$$\Sigma_k = \begin{bmatrix} v_{1,k}^2 & & & \\ & \triangle & & \\ & & v_{i,k}^2 & \\ & & & \triangle & \\ & & & & v_{n,k}^2 \end{bmatrix}$$

où  $v_{i,k}^2$  est la variance du  $i$ ème caractère pour la classe  $C_k$ .

La fonction de densité de la distribution des observations est donc de la forme:

$$f(X) = \sum_{k=1}^K p(X|C_k) P(C_k) \quad (6.1.)$$

avec:

$$\begin{aligned} p(X|C_k) &= \frac{1}{(2\pi)^{n/2} |\Sigma_k|^{1/2}} \exp\left\{-1/2(X-\bar{X}_k)^T \Sigma_k^{-1} (X-\bar{X}_k)\right\} \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi} v_{i,k}} \exp\left\{-1/2\left(\frac{x_i - \bar{x}_{i,k}}{v_{i,k}}\right)^2\right\} \end{aligned} \quad (6.2.)$$

### VI.3. ANALYSE DES DENSITES MARGINALES DE PROBABILITE

#### VI.3.1. CONVEXITE DES DENSITES MARGINALES

La technique d'analyse des mélanges gaussiens présentée dans ce chapitre repose sur l'étude de la convexité des densités marginales de probabilité de la fonction de densité  $f(X)$ .

Compte tenu de l'équation (6.2.) et de la définition des densités marginales/4/, les densités marginales conditionnelles  $p_i(x_i|C_k)$  des fonctions de densité conditionnelles  $p(X|C_k)$  sont de la forme:

$$p_i(x_i|C_k) = \frac{1}{\sqrt{2\pi} v_{i,k}} \exp\left\{-1/2\left(\frac{x_i - \bar{x}_{i,k}}{v_{i,k}}\right)^2\right\} \quad \begin{matrix} i=1,2,\dots,n \\ k=1,2,\dots,K \end{matrix} \quad (6.3)$$

La forme particulièrement simple des densités marginales  $f_i(x_i)$  de la fonction de densité  $f(X)$ :

$$f_i(x_i) = \sum_{k=1}^K p_i(x_i|C_k) P(C_k) \quad (6.4.)$$

obtenue sous l'hypothèse d'indépendance des caractères est exploitée pour identifier les différents paramètres du mélange  $f(X)$ .

Chaque fonction de densité monovariante  $p_i(x_i|C_k)$  peut être décomposée en un segment concave compris entre deux segments convexes (cf. figure 6.1.). On sait que le segment concave est centré sur la valeur moyenne  $\bar{x}_{i,k}$  de la distribution de densité  $p_i(x_i|C_k)$  et que sa longueur est égale au double de l'écart type  $v_{i,k}$ . Ce segment, noté  $d_{i,k}$ , tel que:

$$d_{i,k} = [\bar{x}_{i,k} - v_{i,k} ; \bar{x}_{i,k} + v_{i,k}]$$

est appelé "segment de concavité" de la fonction  $p_i(x_i|C_k)$ .

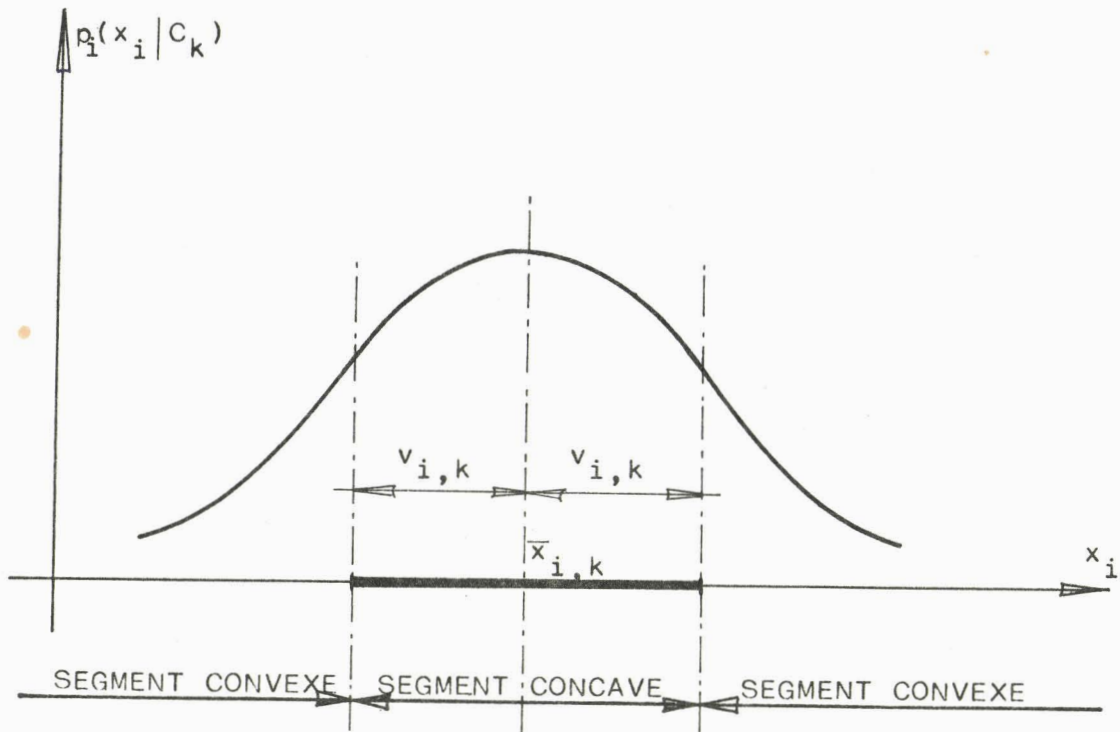


Figure 6.1. Propriétés de convexité des fonctions de densité normales monovariabiles

On suppose temporairement que les composantes des mélanges monovariabiles  $f_i(x_i), i=1,2,\dots,n$ , présentent un faible degré de chevauchement de telle sorte qu'il existe une bijection entre les modes de ces mélanges et leurs composantes. Lorsque les composantes sont suffisamment séparées, on peut affirmer que la forme des densités marginales  $f_i(x_i)$  demeure très semblable, au voisinage des modes, à la forme des composantes correspondantes. On peut donc s'attendre à ce que les densités marginales soient concaves sur un certain nombre de segments très semblables en position et en longueur aux segments de concavité de leurs composantes. Ces segments, notés  $\hat{d}_{i,k}, i=1,2,\dots,n; k=1,2,\dots,K$ , seront appelés "segments de concavité" des densités marginales et seront séparés par des segments convexes.

Le degré de similitude entre les segments de concavité des densités marginales et de leurs composantes dépend du degré de chevauchement de ces dernières. Lorsque les conditions sont favorables, la demi-longueur  $\hat{v}_{i,k}$  et la position  $\hat{x}_{i,k}$  du point milieu de chaque segment  $\hat{d}_{i,k}$  peuvent être utilisées pour approcher les valeurs de la variance  $v_{i,k}^2$  et de la moyenne  $\bar{x}_{i,k}$  de chaque composante  $p_i(x_i | C_k)$ . A titre d'exemple, la figure 6.2. illustre la similitude des segments de concavité d'un mélange et de ses composantes.

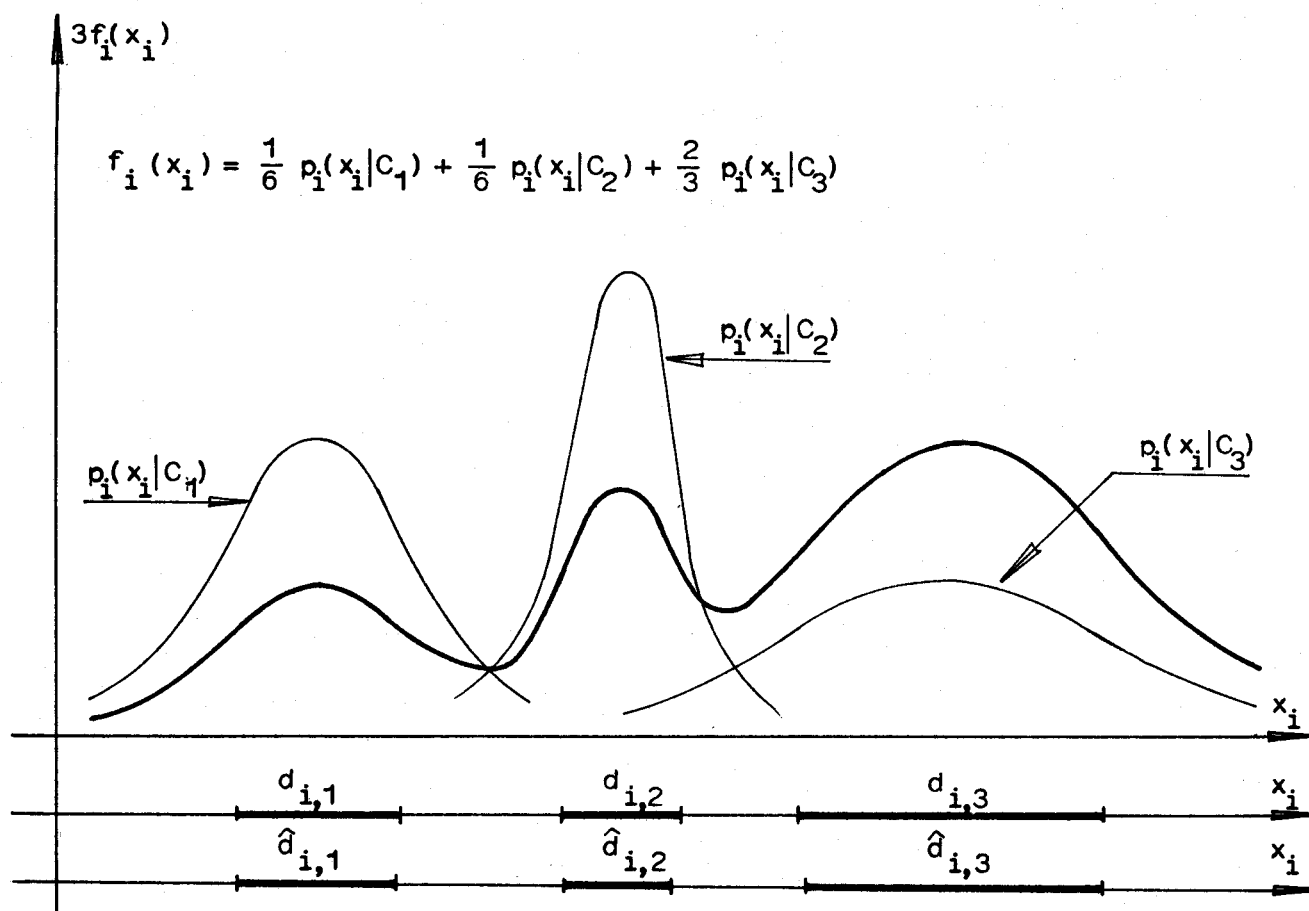


Figure 6.2.

Similitude des segments de concavité  
d'un mélange et de ses composantes

$d_{i,k}$ ,  $k=1,2,3$  : segments de concavité des composantes

$\hat{d}_{i,k}$ ,  $k=1,2,3$  : segments de concavité du mélange

Nous supposerons temporairement que toutes les composantes des densités marginales sont identifiables par analyse de la convexité de ces fonctions. Il faut cependant remarquer que, même lorsque les composantes du mélange  $f(X)$  sont nettement séparées dans l'espace  $R_0^n$ , rien ne garantit que les composantes des densités marginales le soient. (cf. figure 6.3.). Nous verrons au paragraphe VI.5. comment traiter les cas où des composantes des densités marginales sont indifférentiables alors que les classes présentent un faible degré de chevauchement dans l'espace de représentation.

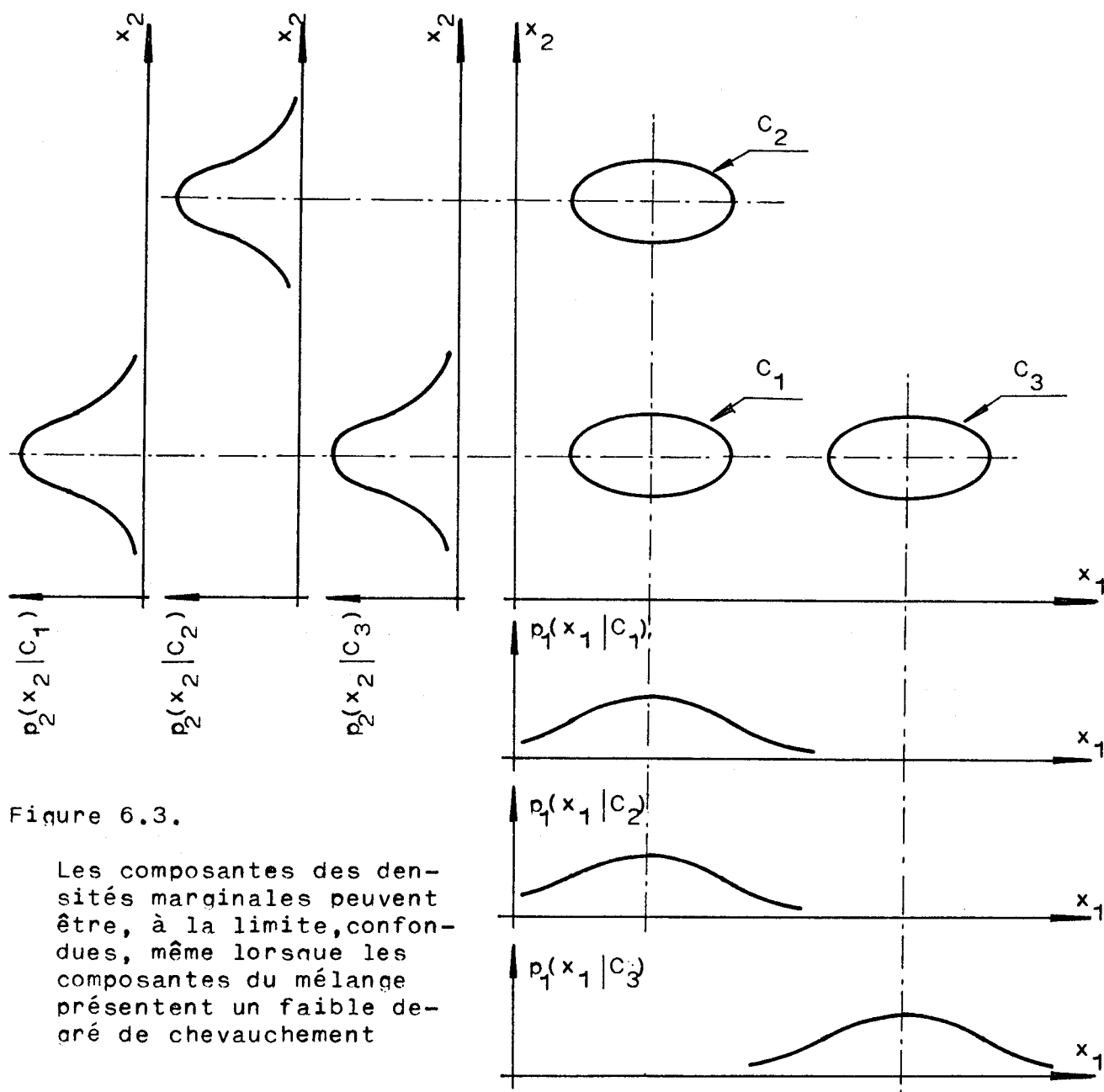


Figure 6.3.

Les composantes des densités marginales peuvent être, à la limite, confondues, même lorsque les composantes du mélange présentent un faible degré de chevauchement

### VI.3.2. IMPLANTATION DE LA PROCEDURE

Pratiquement, les densités marginales sont estimées par la méthode de Rosenblatt-Parzen /1/ à partir des observations normalisées de l'espace  $R_1^n$ . Le problème du temps de calcul étant tout à fait secondaire pour l'estimation des fonctions mono-variables, on utilise un noyau gaussien:

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} u^2\right\}$$



de telle sorte que l'estimateur  $\hat{f}_i(x_i)$  de  $f_i(x_i)$  est de la forme:

$$\hat{f}_i(x_i) = \frac{1}{a} \sum_{j=1}^a \frac{1}{h_{i,a}} \varphi\left(\frac{x_i - x_{i,j}}{h_{i,a}}\right)$$

Afin d'assurer la convergence de cet estimateur, on choisit:

$$h_{i,a} = h_i^0 / \sqrt{a}$$

où  $h_i^0$  est un paramètre à ajuster en fonction des données.

Les densités marginales sont estimées avec un pas de discrétisation  $\delta$  de la forme:  $\delta = 1/M$ ,  $M$  entier. (Rappelons que dans  $R_1^n$ , chaque densité marginale doit être estimée sur un segment de longueur unité). Dans la pratique, l'ajustement de  $\delta$  n'est pas critique mais, dans le stade d'exploration d'un ensemble de données, on a intérêt à choisir  $M$  suffisamment grand pour que des modes de faible largeur ne passent pas inaperçus. Tout expérimentateur, tant soit peu familiarisé avec les problèmes d'échantillonnage des signaux, saura ajuster  $\delta$  sur la base d'un simple examen visuel des densités marginales discrétisées.

Dans tous les exemples présentés dans ce chapitre ainsi que dans les applications pratiques de la dernière partie de ce mémoire, l'expérience a montré qu'au-delà de  $M=20$ , une discrétisation plus fine n'améliore pas sensiblement les résultats et alourdit inutilement la procédure. Cependant, tout en restant dans l'ordre de grandeur de quelques dizaines, l'expérimentateur peut, par mesure de précaution, augmenter sensiblement la valeur de  $M$  car il ne sera limité ni par l'encombrement en mémoire, ni par la durée des calculs.

La convexité de  $\hat{f}_i(x_i)$  aux points d'abscisse  $x_i^m$ ,  $m=1,2,\dots,M+1$ , est déterminée en testant la position de ces points par rapport aux segments de droite joignant les points de  $\hat{f}_i(x_i)$  d'abscisse  $x_i^{m-1}$  et  $x_i^{m+1}$ . (cf. figure 6.4.).

Pour l'ajustement des paramètres  $h_i^0$ , qui est plus délicat, on utilise à nouveau le concept de stabilité du nombre de modes détectés. Chacun de ces paramètres est ajusté au milieu de la plus grande plage de variation pour laquelle le nombre de

segments concaves détectés sur la densité marginale correspondante est constant. Les paramètres  $h_i^0$  sont ainsi ajustés différemment pour chacune des densités marginales estimées.

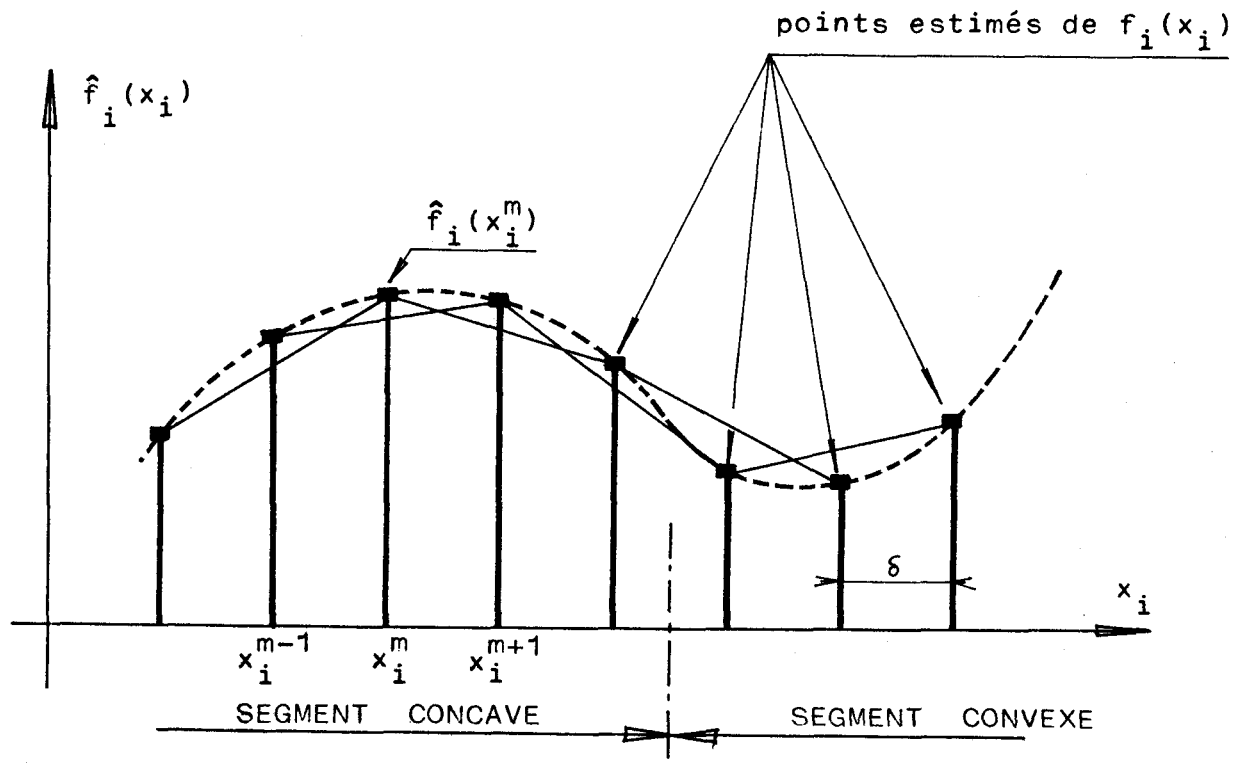


Figure 6.4. Estimation discrète des densités marginales et test de convexité

#### VI.4. DETERMINATION DES PARAMETRES D'UN MELANGE MULTIVARIABLE

Pour mener à bien l'analyse du mélange  $f(X)$  de l'équation (6.1.), il faut maintenant relier les informations obtenues par analyse des densités marginales aux paramètres statistiques définissant la fonction de densité de la distribution des observations multivariées étudiées.

##### VI.4.1. DETERMINATION DES PROBABILITES A PRIORI

Les premiers paramètres à déterminer sont les probabilités a priori de chaque classe. Considérons les  $K$  domaines hyperparallélépipédiques  $\mathcal{D}_k$ , centrés aux points  $\bar{X}_k$ , dont les côtés, parallèles aux axes, ont pour longueur  $2v_{i,k}$  dans la  $i$ ème direction,  $i=1,2,\dots,n$ . Le domaine  $\mathcal{D}_k$  est appelé "domaine caractéristique" de la classe  $C_k$ .

La probabilité pour qu'une observation de la classe  $C_k$  appartienne au domaine caractéristique  $\mathcal{D}_k$  est:

$$P(X \in \mathcal{D}_k | C_k) = \prod_{i=1}^n \int_{\bar{x}_{i,k} - v_{i,k}}^{\bar{x}_{i,k} + v_{i,k}} \frac{1}{\sqrt{2\pi} v_{i,k}} \exp\left\{-1/2 \left(\frac{x_i - \bar{x}_{i,k}}{v_{i,k}}\right)^2\right\} dx_i$$

La quantité:

$$\int_{\bar{x}_{i,k} - v_{i,k}}^{\bar{x}_{i,k} + v_{i,k}} \frac{1}{\sqrt{2\pi} v_{i,k}} \exp\left\{-1/2 \left(\frac{x_i - \bar{x}_{i,k}}{v_{i,k}}\right)^2\right\} dx_i$$

n'est autre que la probabilité pour une variable aléatoire distribuée normalement selon la loi:

$$p_i(x_i | C_k) = \frac{1}{\sqrt{2\pi} v_{i,k}} \exp\left\{-1/2 \left(\frac{x_i - \bar{x}_{i,k}}{v_{i,k}}\right)^2\right\}$$

d'appartenir au segment de concavité:

$$d_{i,k} = [\bar{x}_{i,k} - v_{i,k} ; \bar{x}_{i,k} + v_{i,k}]$$

Cette probabilité est constante et on peut écrire:

$$P(X \in \mathcal{D}_k | C_k) = \beta \quad , \quad \beta \text{ constante}$$

Considérons maintenant le mélange  $f(X)$  de l'équation (6.1.). La probabilité pour qu'une observation de l'échantillon distribué selon la fonction  $f(X)$  appartienne au domaine  $\mathcal{D}_k$  et provienne en même temps de la classe  $C_k$  est égale à  $\beta P(C_k)$ .

Dans la pratique, on s'intéresse à des mélanges  $f(X)$  pour lesquels le degré de chevauchement des composantes reste faible de telle sorte que la plupart des observations situées à l'intérieur du domaine  $\mathcal{D}_k$  proviennent de la classe  $C_k$ . La probabilité  $\beta P(C_k)$  peut donc être approchée par la probabilité  $P(X \in \mathcal{D}_k)$  pour qu'une observation, quelle que soit la classe dont elle provient, appartienne au domaine  $\mathcal{D}_k$ . Nous verrons ultérieurement que la précision obtenue par cette approximation, qui dépend du degré de chevauchement des composantes du mélange  $f(X)$ , permet une optimisation très satisfaisante du processus de classification.

Des valeurs approchées  $P'(C_k)$  des probabilités a priori  $P(C_k)$  seront obtenues en résolvant le système d'équations:

$$\begin{cases} \frac{P'(C_1)}{p(X \in \mathcal{D}_1)} = \dots = \frac{P'(C_k)}{p(X \in \mathcal{D}_k)} = \dots = \frac{P'(C_K)}{p(X \in \mathcal{D}_K)} \\ P'(C_1) + \dots + P'(C_k) + \dots + P'(C_K) = 1 \end{cases}$$

Les domaines  $\mathcal{D}_k$  introduits dans cette procédure peuvent être déterminés à partir des segments concaves mis en évidence sur les densités marginales. En effet, les caractéristiques géométriques de ces segments, étudiées au paragraphe précédent, montrent qu'ils sont des approximations des projections des domaines caractéristiques  $\mathcal{D}_k$  sur les différents axes de l'espace de représentation des données. On peut donc obtenir des domaines hyperparallélipédiques très semblables aux domaines  $\mathcal{D}_k$  par l'intermédiaire du produit euclidien de ces segments concaves. (cf. figure 6.5.).

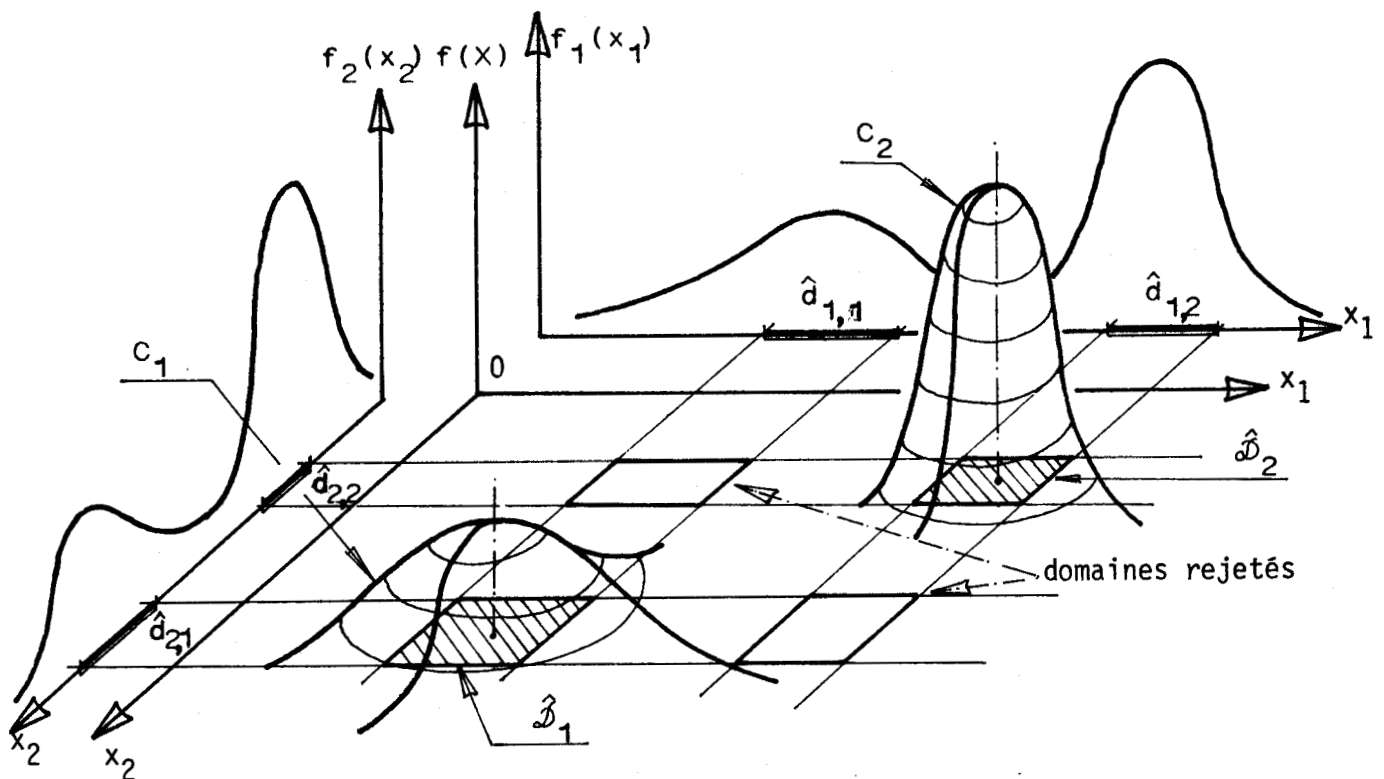


Figure 6.5. Détermination des domaines approchés ( $\hat{\mathcal{D}}_1$  et  $\hat{\mathcal{D}}_2$ ) des domaines caractéristiques d'un mélange gaussien à partir des segments concaves des densités marginales

Il est important de noter que tous les domaines définis par ce produit euclidien ne sont pas nécessairement associés à des composantes du mélange. Certains sont introduits artificiellement par la procédure et doivent être rejetés. Ils le sont de manière itérative en rejetant à chaque pas le domaine contenant le plus petit nombre d'observations. Ce rejet n'est cependant effectif que si ce nombre est inférieur à un seuil défini comme une fraction  $x$  du nombre d'observations contenues dans l'ensemble des domaines restants à la fin de l'itération précédente.

Les domaines retenus à la fin de cette procédure de rejet, notés  $\hat{D}_k$ , sont utilisés pour déterminer les valeurs approchées  $\hat{P}(C_k)$  des probabilités a priori des composantes du mélange. Soit  $\alpha_k$  le nombre d'observations situées à l'intérieur du domaine  $\hat{D}_k$ . En estimant la probabilité  $p(X \in \hat{D}_k)$  par le rapport  $\alpha_k/q$ , on obtient des valeurs approchées des probabilités a priori en résolvant le système d'équations:

$$\begin{cases} \frac{\hat{P}(C_1)}{\alpha_1} = \dots = \frac{\hat{P}(C_k)}{\alpha_k} = \dots = \frac{\hat{P}(C_K)}{\alpha_K} \\ \hat{P}(C_1) + \dots + \hat{P}(C_k) + \dots + \hat{P}(C_K) = 1 \end{cases}$$

Nous voyons donc que le seuil  $x$  de rejet des domaines définis par le produit euclidien permet d'éliminer toutes les classes générées artificiellement dont la probabilité a priori serait inférieure à la valeur de ce seuil. L'ajustement de ce seuil n'est pas critique car les domaines correspondant effectivement à des modes de la distribution contiennent généralement un nombre d'observations nettement supérieur à ceux introduits artificiellement par le produit euclidien des segments concaves. Un seuil de 10% a toujours permis, dans les exemples présentés dans ce chapitre, de rejeter les domaines indésirables, mais l'analyse des résultats montre que ceux-ci auraient été en tout point identiques pour toute valeur de ce seuil comprise entre 10% et 22% .

### VI.4.2. DETERMINATION DES VECTEURS MOYENNE ET DES MATRICES DE COVARIANCE

Les derniers paramètres restant à déterminer sont les vecteurs-moyenne et les matrices de covariance des différentes classes.

Une valeur approchée  $\hat{\bar{X}}_k$  du vecteur-moyenne  $\bar{X}_k$  est donnée par la position du centre du domaine caractéristique approché  $\hat{\mathcal{D}}_k$ . Les carrés des demi-longueurs des côtés de ce domaine fournissent des valeurs approchées des éléments de la matrice de covariance. Celle-ci est ainsi obtenue sous la forme:

$$\hat{\Sigma}_k = \begin{bmatrix} \hat{v}_{1,k}^2 & & & \\ & \hat{v}_{i,k}^2 & & \\ & & \hat{v}_{n,k}^2 & \\ & & & 0 \end{bmatrix}$$

où  $\hat{v}_{i,k}$  est la demi-longueur des côtés de  $\hat{\mathcal{D}}_k$  parallèles au  $i$ ème axe.

### VI.4.3. CLASSIFICATION

La mise en oeuvre de la procédure de classement bayésien /8/ sur la base de ces valeurs approchées des paramètres des mélanges analysés permet d'envisager une optimisation du processus de classification.

Les fonctions de décision:

$$\hat{g}_k(X) = -1/2 \text{Log } \hat{\Sigma}_k - 1/2(X - \hat{\bar{X}}_k)^T \hat{\Sigma}_k^{-1} (X - \hat{\bar{X}}_k) + \text{Log } \hat{P}(C_k)$$

$$k = 1, 2, \dots, K$$

définissent des surfaces de décision d'équation:

$$\hat{g}_k(X) - \hat{g}_{k'}(X) = 0, \quad k \neq k'$$

qui partitionnent l'espace en régions de décision  $R_k$ ,  $k=1, 2, \dots, K$ , chacune d'elles contenant exclusivement les observations assignées à l'une des classes mises en évidence.

Avant d'envisager l'application de ce schéma pour optimiser le processus de classification, il convient d'analyser

l'effet du chevauchement des composantes des densités marginales. En effet, nous avons remarqué que même lorsque les composantes multivariées sont nettement séparées dans l'espace de représentation des données, il peut arriver que les composantes des densités marginales se chevauchent au point d'être indiscernables par une analyse de convexité.

#### VI.5. ANALYSE GLOBALE - ANALYSE LOCALE

Nous avons jusqu'à présent supposé que les composantes des densités marginales présentent un faible degré de chevauchement de telle sorte qu'elles soient aisément identifiables par une analyse de convexité. Cette hypothèse est en défaut lorsqu'un segment concave mis en évidence sur une densité marginale est le résultat de la superposition de plusieurs composantes. (cf. figure 6.6.). Pour éviter une dégradation des performances dans de telles situations, nous proposons d'analyser localement la structure des données situées à l'intérieur de chacune des régions de décision déterminées par l'analyse globale de l'échantillon étudié.

Faire suivre les techniques d'analyse globale, qui conduisent parfois à des résultats peu précis, par des techniques locales destinées à améliorer la qualité des résultats n'est pas une méthodologie nouvelle en classification automatique /3//5/. Il a été montré que ces techniques locales, qui peuvent faire disparaître des classes ou en faire apparaître de nouvelles, sont en général itératives /2/.(cf. figure 6.7.).

Pour le problème particulier qui nous intéresse, l'analyse locale de l'ensemble des observations situées à l'intérieur de chacune des régions de décision déterminées à l'itération précédente, est en tout point analogue à l'analyse globale de l'échantillon complet. Il est important de noter que les paramètres  $h_i^0$ ,  $i=1,2,\dots,n$ , doivent être, pour chacune des analyses locales, identiques à ceux utilisés au cours de l'analyse globale.

Les nouveaux domaines caractéristiques approchés obtenus à la fin de chaque itération permettent de calculer de nouvelles fonctions de décision. A l'itération suivante, l'analyse séparée des observations assignées par ces fonctions à chacune des

classes mises en évidence, permet d'améliorer la détermination des domaines caractéristiques.

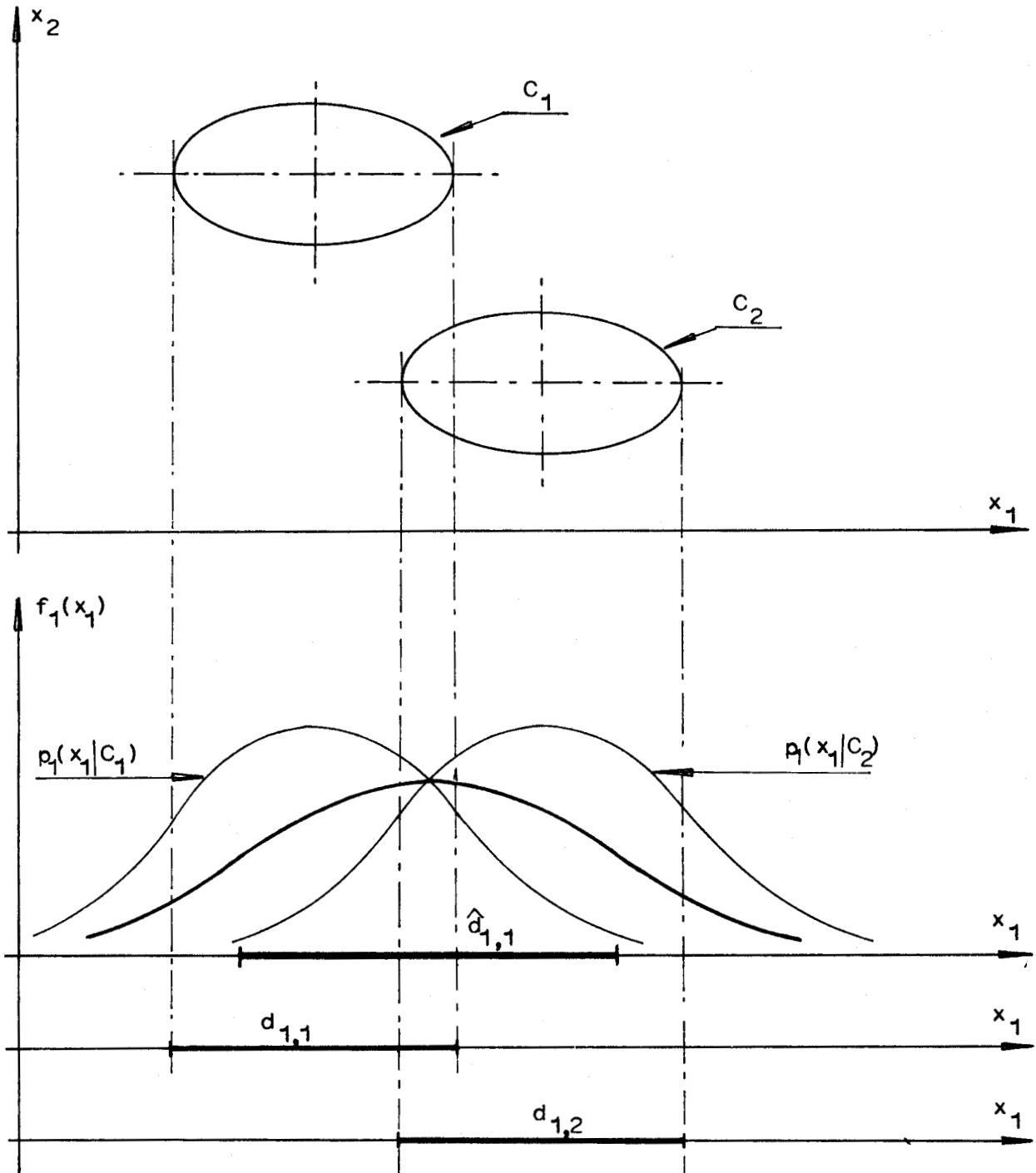


Figure 6.6. L'unique segment concave  $\hat{d}_{1,1}$  de la densité marginale  $f_1(x_1)$  ne permet pas de détecter la présence de deux composantes

$d_{1,1}$  : segment concave de  $p_1(x_1|C_1)$

$d_{1,2}$  : segment concave de  $p_2(x_1|C_2)$



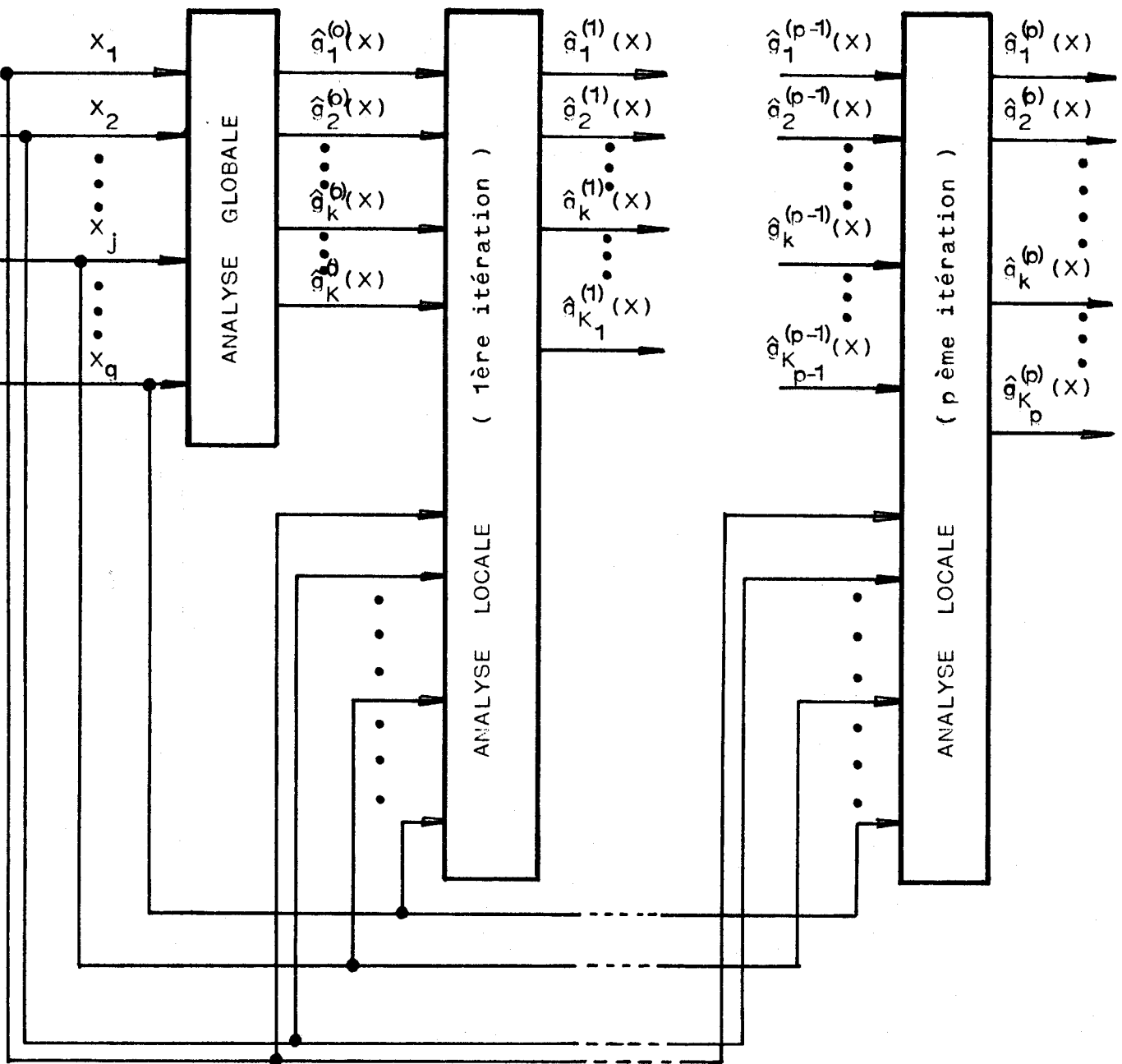


Figure 6.7. Principe de l'analyse locale itérative

$\hat{q}_k^{(i)}(x)$ ,  $k=1,2,\dots,K_i$  : ensemble des  $K_i$  fonctions de décision obtenues à la  $i$ ème itération

(Le nombre de classes mises en évidence pouvant varier à chaque itération, on peut avoir  $K_{i-1} \neq K_i$ ,  $i=1,2,\dots,p$ )

L'arrêt de cette procédure itérative est contrôlé de manière interactive en fonction du nombre d'observations reclassées différemment à chaque itération. Dans le cas idéal où il y a stabilisation de la classification, c'est à dire où le nombre d'observations reclassées différemment reste nul à partir d'un certain nombre d'itérations, l'arrêt de cette procédure ne pose aucun problème particulier. L'expérience montre d'ailleurs que ce cas est le plus fréquent. Mais rien ne permet d'affirmer qu'il y a stabilisation du résultat de la classification, les surfaces de décision pouvant osciller de part et d'autre de certaines observations. C'est pourquoi la procédure itérative n'est arrêtée que sur décision de l'opérateur, lorsque celui-ci constate qu'il y a stabilisation du résultat de la classification ou, le cas échéant, lorsqu'il constate que les variations du nombre d'observations reclassées à chaque itération tendent à se stabiliser.

A titre d'exemple, la figure 6.8. indique le résultat de l'analyse globale d'un échantillon de 300 observations. Cette analyse ne permet pas de mettre en évidence les trois classes, pourtant bien distinctes, dont sont issues les observations. Par contre, dès la première itération de l'analyse locale, on détecte les trois composantes, l'une des régions de décision déterminée au cours de l'analyse globale étant en fait bimodale.

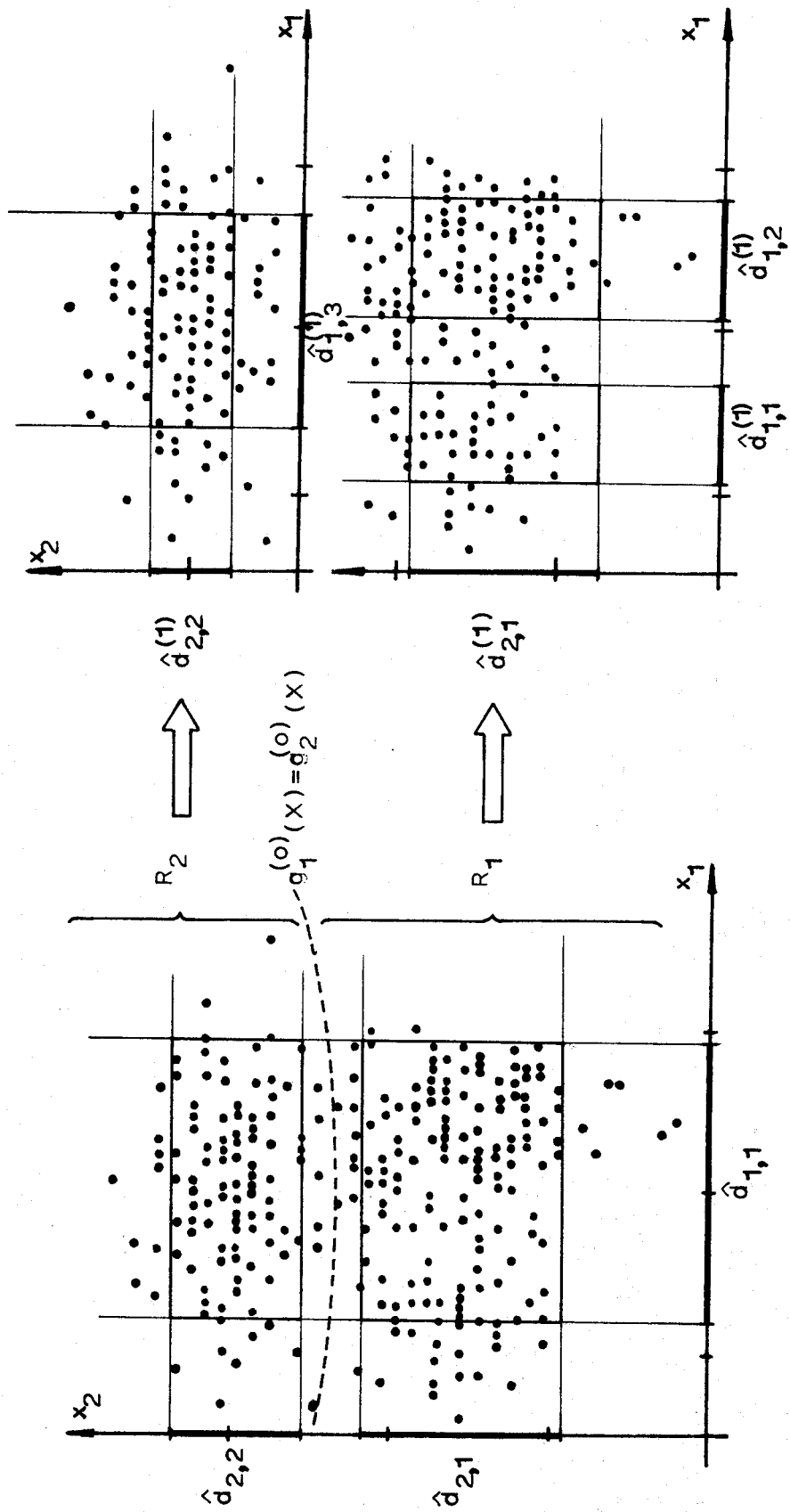
## VI.6. PERFORMANCES DE LA PROCEDURE

Afin de tester la qualité de l'optimisation obtenue sur la base des approximations proposées pour déterminer les paramètres des mélanges étudiés, la méthode de classification présentée dans ce chapitre est maintenant appliquée à des données générées artificiellement. Comme au chapitre V, nous jugerons les résultats par l'écart entre le taux d'erreur de la classification et le taux d'erreur théorique minimum calculé à partir des valeurs exactes des paramètres des mélanges.

### Exemple 1

Le comportement de la procédure est d'abord analysé sur les 300 observations bidimensionnelles de la figure 6.9., issues d'un mélange de trois composantes normales. Le tableau 6.1. résume les paramètres statistiques de ce mélange qui a déjà été rencontré au chapitre V, paragraphe V.6.1.

Figure 6.8. Analyse globale et première itération de l'analyse locale d'un échantillon tiré d'un mélange à trois composantes



ANALYSE GLOBALE

ANALYSE LOCALE ( 1ère itération )



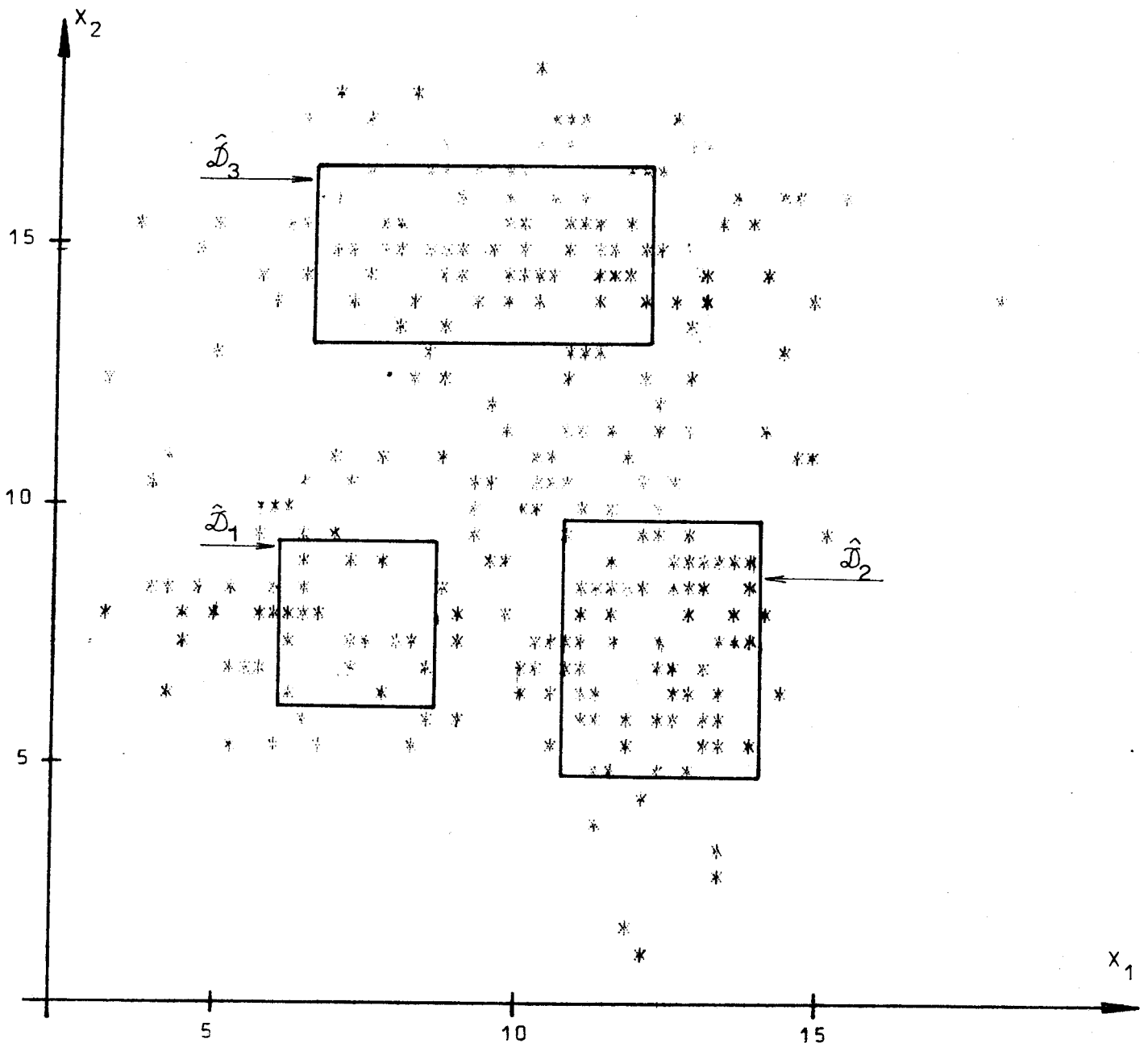


Figure 6.9. Représentation graphique des 300 observations de l'échantillon de l'exemple n°1 et domaines approchés des domaines caractéristiques obtenus à la dernière itération de l'analyse locale

Le taux d'erreur de la classification tombe à 6% au bout des quatre itérations de l'analyse locale nécessaires pour obtenir la stabilisation du résultat. Les domaines caractéristiques obtenus à la fin de l'analyse globale et de la première itération de l'analyse locale sont représentés sur la figure 6.8.

	VECTEURS MOYENNE	MATRICES DE COVARIANCE	PROBABILITES A PRIORI
CLASSE C <sub>1</sub>	$x_1 = \begin{bmatrix} 7 \\ 8 \end{bmatrix}$ $\hat{x}_1 = \begin{bmatrix} 7,41 \\ 7,72 \end{bmatrix}$	$\Sigma_1 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ $\hat{\Sigma}_1 = \begin{bmatrix} 1,69 & 0 \\ 0 & 1,78 \end{bmatrix}$	$P(C_1) = 0,20$ $\hat{P}(C_1) = 0,22$
CLASSE C <sub>2</sub>	$x_2 = \begin{bmatrix} 12 \\ 8 \end{bmatrix}$ $\hat{x}_2 = \begin{bmatrix} 12,42 \\ 7,29 \end{bmatrix}$	$\Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix}$ $\hat{\Sigma}_2 = \begin{bmatrix} 2,54 & 0 \\ 0 & 4,51 \end{bmatrix}$	$P(C_2) = 0,40$ $\hat{P}(C_2) = 0,36$
CLASSE C <sub>3</sub>	$x_3 = \begin{bmatrix} 10 \\ 15 \end{bmatrix}$ $\hat{x}_3 = \begin{bmatrix} 9,38 \\ 14,80 \end{bmatrix}$	$\Sigma_3 = \begin{bmatrix} 7 & 0 \\ 0 & 2 \end{bmatrix}$ $\hat{\Sigma}_3 = \begin{bmatrix} 7,60 & 0 \\ 0 & 2,14 \end{bmatrix}$	$P(C_3) = 0,40$ $\hat{P}(C_3) = 0,42$
TAUX D'ERREUR THEORIQUE MINIMUM : 4,3%    TAUX D'ERREUR DE LA CLASSIFICATION : 6,0%			

Tableau 6.1. Valeurs exactes et approchées des paramètres statistiques de la distribution de l'exemple 1



La figure 6.9. indique les domaines caractéristiques correspondant à la dernière itération.

Rappelons que la méthode d'optimisation basée sur l'analyse de la convexité de la fonction de densité du mélange avait permis d'aboutir à un taux d'erreur de 5,3%. Les pertes de performance entre les deux procédures ne correspondent en fait qu'à deux observations mal classées sur un total de 300.

### Exemple 2

Afin de tester les performances de la procédure lorsque le nombre d'observations disponibles est faible compte-tenu de leur dimension, on utilise maintenant un échantillon de 99 observations tirées d'un mélange de trois composantes normales équiprobables.

Les vecteurs-moyenne des composantes sont fixes et la dispersion des observations autour du centre de chaque classe est contrôlée par la valeur des éléments diagonaux des matrices de covariance.(cf. tableau 6.2.).

On peut ainsi analyser les performances de la méthode lorsque la dimension des données varie tout en jouant sur le degré de chevauchement des différentes composantes.

Dans cet exemple, le nombre d'individus par classe (33) est trop faible pour envisager l'analyse de la convexité des fonctions de densité des mélanges. Cependant, l'analyse de la convexité des densités marginales permet d'obtenir un taux d'erreur proche du taux d'erreur optimal. En règle générale, il y a toute-fois dégradation des performances (mesurées par l'écart entre le taux d'erreur de la classification et le taux d'erreur théorique minimum) lorsque la dimension des données augmente et quand croît le degré de chevauchement entre les différentes composantes. Cependant, même dans le cas le plus défavorable (échantillon n°9), la différence entre le taux d'erreur de la classification et le taux optimal représente moins de quatre observations mal classées.

Ce dernier exemple qui, comme le précédent, nous a permis de juger du bien-fondé des approximations nécessaires à la mise en œuvre de la méthode, est également utilisé pour chiffrer le temps de calcul de l'algorithme. Le nombre d'itérations

	VECTEURS MOYENNE	MATRICES DE COVARIANCE	TAUX D'ERREUR THEORIQUE MINI.	TAUX D'ERREUR DE LA CLASSIFICATION	NOMBRE D'ITERATIONS
ECHANT. N°1	$x_1 = \begin{bmatrix} 10 \\ 10 \\ 10 \end{bmatrix}$ $x_2 = \begin{bmatrix} 10 \\ 20 \\ 20 \end{bmatrix}$ $x_3 = \begin{bmatrix} 20 \\ 10 \\ 10 \end{bmatrix}$	$\Sigma_k = 2 I_2, k=1,2,3$	1,0%	1,0%	2
ECHANT. N°2	$x_1 = \begin{bmatrix} 10 \\ 10 \\ 10 \end{bmatrix}$ $x_2 = \begin{bmatrix} 10 \\ 20 \\ 20 \end{bmatrix}$ $x_3 = \begin{bmatrix} 20 \\ 10 \\ 10 \end{bmatrix}$	$\Sigma_k = 4 I_2, k=1,2,3$	2,1%	3,0%	2
ECHANT. N°3	$x_1 = \begin{bmatrix} 10 \\ 10 \\ 10 \end{bmatrix}$ $x_2 = \begin{bmatrix} 10 \\ 20 \\ 20 \end{bmatrix}$ $x_3 = \begin{bmatrix} 20 \\ 10 \\ 10 \end{bmatrix}$	$\Sigma_k = 8 I_2, k=1,2,3$	6,9%	9,0%	2
ECHANT. N°4	$x_1 = \begin{bmatrix} 10 \\ 10 \\ 10 \end{bmatrix}$ $x_2 = \begin{bmatrix} 10 \\ 10 \\ 20 \end{bmatrix}$ $x_3 = \begin{bmatrix} 20 \\ 10 \\ 10 \end{bmatrix}$	$\Sigma_k = 2 I_3, k=1,2,3$	1,0%	2,0%	3
ECHANT. N°5	$x_1 = \begin{bmatrix} 10 \\ 10 \\ 10 \end{bmatrix}$ $x_2 = \begin{bmatrix} 10 \\ 10 \\ 20 \end{bmatrix}$ $x_3 = \begin{bmatrix} 20 \\ 10 \\ 10 \end{bmatrix}$	$\Sigma_k = 4 I_3, k=1,2,3$	1,8%	4,0%	2
ECHANT. N°6	$x_1 = \begin{bmatrix} 10 \\ 10 \\ 10 \end{bmatrix}$ $x_2 = \begin{bmatrix} 10 \\ 10 \\ 20 \end{bmatrix}$ $x_3 = \begin{bmatrix} 20 \\ 10 \\ 10 \end{bmatrix}$	$\Sigma_k = 8 I_3, k=1,2,3$	6,7%	9,0%	3
ECHANT. N°7	$x_1 = \begin{bmatrix} 10 \\ 10 \\ 10 \\ 10 \end{bmatrix}$ $x_2 = \begin{bmatrix} 10 \\ 10 \\ 10 \\ 20 \end{bmatrix}$ $x_3 = \begin{bmatrix} 20 \\ 10 \\ 10 \\ 10 \end{bmatrix}$	$\Sigma_k = 2 I_4, k=1,2,3$	0,9%	2,0%	3
ECHANT. N°8	$x_1 = \begin{bmatrix} 10 \\ 10 \\ 10 \\ 10 \end{bmatrix}$ $x_2 = \begin{bmatrix} 10 \\ 10 \\ 10 \\ 20 \end{bmatrix}$ $x_3 = \begin{bmatrix} 20 \\ 10 \\ 10 \\ 10 \end{bmatrix}$	$\Sigma_k = 4 I_4, k=1,2,3$	1,5%	4,0%	4
ECHANT. N°9	$x_1 = \begin{bmatrix} 10 \\ 10 \\ 10 \\ 10 \end{bmatrix}$ $x_2 = \begin{bmatrix} 10 \\ 10 \\ 10 \\ 20 \end{bmatrix}$ $x_3 = \begin{bmatrix} 20 \\ 10 \\ 10 \\ 10 \end{bmatrix}$	$\Sigma_k = 8 I_4, k=1,2,3$	6,5%	10,0%	2

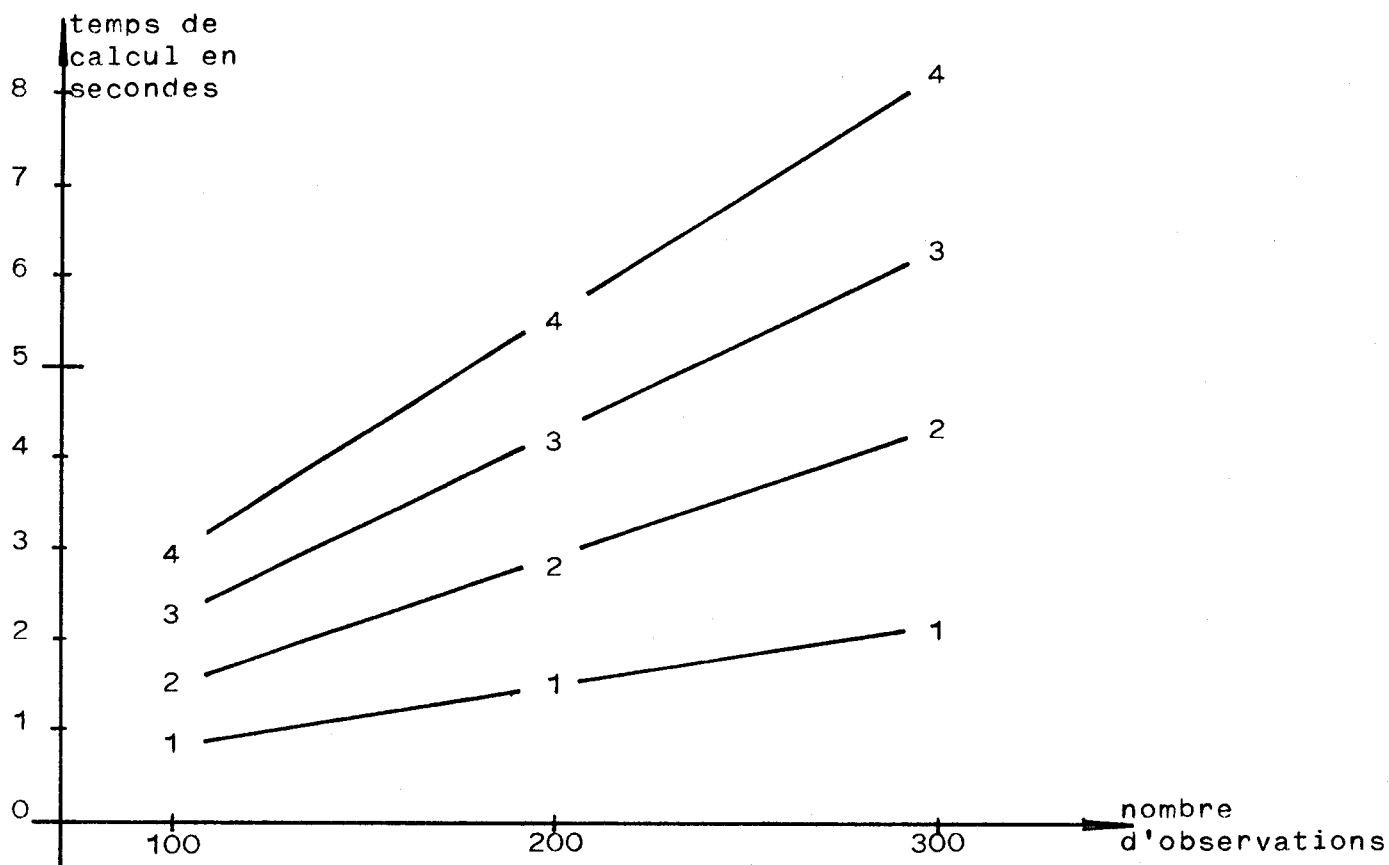
Tableau 6.2. Paramètres statistiques des distributions multidimensionnelles et performances de la procédure de classification pour l'exemple n°2

$I_n$  : Matrice unité d'ordre n



nécessaires pour obtenir la stabilisation du classement final étant différent selon les échantillons, nous n'avons représenté sur la figure 6.10. que les temps de calcul de l'analyse globale. On obtient une estimation du temps d'exécution total de la procédure de classification en attribuant le même temps de calcul à chaque analyse locale des données. Le graphique de la figure 6.10. indique que, pour une taille d'échantillon donnée, le temps de calcul est quasiment proportionnel à la dimension des données. En faisant varier le nombre d'observations soumises à l'analyse, il a été montré que, pour un problème de dimension fixe, le temps de calcul est quasiment proportionnel au nombre d'observations.

Compte-tenu du nombre d'itérations nécessaires pour aboutir au classement final, les temps de calcul de l'algorithme utilisant les densités marginales sont tout à fait comparables à ceux de la méthode proposée au chapitre précédent.



$i$  —  $i$  —  $i$  : temps de calcul en fonction de la taille de l'échantillon pour des données à  $i$  dimensions

Figure 6.10. Temps de calcul de l'analyse globale pour différentes tailles d'échantillon et différentes dimensions des données



VI.7. CONCLUSION

L'analyse de la convexité des densités marginales de probabilité permet de résoudre le problème de l'optimisation du processus de classification automatique dans le cas où l'on ne dispose que d'échantillons de taille réduite, compte-tenu de la dimension des données. L'information disponible pour déterminer la structure des mélanges analysés étant alors très limitée, les résultats présentés n'ont pu être obtenus qu'au prix d'une hypothèse d'indépendance statistique des caractères pour chaque classe. En effet, sous cette hypothèse supplémentaire qui vient s'ajouter à celle de normalité, les paramètres nécessaires pour définir les mélanges sont en nombre plus faible que dans le cas général traité au chapitre V. Cet ensemble restreint de paramètres peut ainsi être déterminé à partir d'un nombre moindre d'observations.

Ainsi, sous l'hypothèse de normalité des classes en présence, l'analyse de la convexité des fonction de densité des mélanges, ou de leurs densités marginales selon les cas, permet d'effectuer une classification d'un ensemble d'observations totalement inconnu, sans aucune information a priori, avec un taux d'erreur très proche de la valeur optimale en dessous de laquelle il est théoriquement impossible de descendre. Cependant, on est en droit de se demander si cette forme analytique particulière des lois de probabilité des différentes classes est toujours bien adaptée à la structure des données. Aussi, pour les cas où l'analyste doit s'affranchir de l'hypothèse de normalité, nous proposons maintenant une méthode de détection des modes d'une distribution qui apparaît comme une application particulière des concepts et résultats présentés dans la première partie de ce mémoire.

REFERENCES (chapitre VI)

- /1/ DUDA,R.O. et HART,P.E.  
"Pattern classification and scene analysis"  
J. Wiley New York p. 88-95 (1973)
- /2/ EIGEN,D.J.;FROMM,F.R. et NORTHOUSE,R.A.  
"Cluster analysis based on dimensional information with  
applications to feature selection and classification"  
IEEE Trans. on Systems,Man & Cyber.  
Vol. SMC-4 , N° 3 , p.284-294(1974)
- /3/ MEISEL,W.S.  
"Computer oriented approach to pattern recoanition"  
Academic Press New York (1972)
- /4/ MORRISON,D.F.  
"Multivariate statistical methods"  
McGraw Hill New York p. 79-81 (1976)
- /5/ NORTHOUSE,R.A.;FROMM,F.R. et EIGEN,D.J.  
"A global-local approach to cluster analysis"  
IEEE Conf. on Decision and Control p. 550-555 (1973)
- /6/ POSTAIRE,J.G.  
"An unsupervised bayes classifier for normal patterns  
based on marginal densities analysis"  
Pattern Recognition Vol. N° p. (1981)
- /7/ POSTAIRE,J.G. et LIMOURI,M.  
"The convexity concept in cluster analysis"  
4 ème Conf. Int. on Pattern Recognition Japon (1978)
- /8/ TOU,J.T. et GONZALEZ,R.C.  
"Pattern recognition principles"  
Addisson Wesley , Readings, Mass. p.119-123 (1974)

## C H A P I T R E VII

### RECHERCHE DE GROUPEMENTS PAR DETECTION DES MODES

#### VII.1. INTRODUCTION

Nous avons mis au point des méthodes d'optimisation du processus de classification sous l'hypothèse de normalité des différentes classes dont proviennent les observations analysées. Bien que la forme générale des mélanges gaussiens permette, par le jeu des nombreux paramètres, de modéliser un nombre quasiment illimité de situations, ce modèle, comme tous les modèles, est contraignant, souvent approximatif et imparfait. Malgré ces critiques, dans de nombreux domaines l'expérience a montré que lorsqu'on travaille avec des modèles, même éloignés de la réalité, on obtient souvent de meilleurs résultats qu'en opérant sans modèle.

Néanmoins, lorsque la distribution des observations des classes en présence est très éloignée de la loi gaussienne, l'hypothèse de normalité peut, à la limite, conduire à imposer une structure aux données au lieu d'aider à découvrir cette structure. Pour les cas où l'analyste désire s'affranchir de cette hypothèse, nous adaptons maintenant les techniques d'analyse

de la convexité des fonctions de densité au problème de la détection des modes d'une distribution. Cette approche débouche sur des méthodes heuristiques qui permettent, sinon d'optimiser la classification, du moins d'établir l'existence de classes dans des lots de données qui apparaissent sous la forme de groupements d'observations.

## VII.2. LA RECHERCHE DE GROUPEMENTS

### VII.2.1. APPROCHE METRIQUE

En absence de modèle, le problème de la classification automatique peut être envisagé comme la recherche d'une partition de l'ensemble des observations telle que chaque individu de l'échantillon considéré ressemble plus aux individus intérieurs à son groupe qu'aux individus extérieurs.

Cette approche, qui s'inscrit dans le cadre des méthodes de "recherche de groupements" (en anglais: cluster analysis), a été exploitée par de nombreux auteurs. Les méthodes permettant de résoudre ce problème nécessitent généralement la définition d'une métrique sur l'espace de représentation  $R_0^n$ . En effet, la mesure la plus naturelle de la ressemblance entre deux individus est la distance séparant les deux observations correspondantes / 6/.

Certaines procédures basées sur cette notion de distance sont essentiellement intuitives / 4/, mais la plupart font appel à un critère algébrique pour mesurer la qualité des groupements effectués / 2 // 18/. Le problème consiste alors à trouver la partition qui extrêmise le critère. Certaines procédures basées sur ce principe minimisent un critère de dispersion intra-groupe de type erreur quadratique / 3 // 8 // 29 // 31 // 32/, d'autres un critère calculé à partir de la matrice de dispersion (ou de covariance) / 16 // 35 // 36 // 40/.

Faisant également référence à la notion de distance, les méthodes de recherche de groupements hiérarchiques opèrent par groupements ou séparations successifs des observations / 7 // 25 // 28 // 38/.

Enfin, la théorie des graphes permet également de formaliser l'analyse des ressemblances entre les observations

et de découvrir des groupements significatifs dans les échantillons analysés/ 1 //41/.

L'ensemble des méthodes basées sur des notions de distance et de ressemblance constitue en fait la majorité des outils disponibles pour la recherche de groupements. Mais, comme l'illustre la figure 7.1., ces méthodes, à l'exception d'une seule fondée sur l'utilisation d'un critère invariant/16/, sont sensibles à tout changement d'échelle sur les caractères. Par ailleurs, presque toutes ces méthodes supposent connu le nombre de classes a priori et les résultats dépendent du nombre de groupements demandés par l'analyste. Lorsqu'on explore un échantillon inconnu, ce choix est généralement difficile à justifier/11//39/. Notons enfin que pour certaines de ces méthodes, les résultats sont conditionnés par l'ordre dans lequel les observations sont considérées/ 3 //29/.

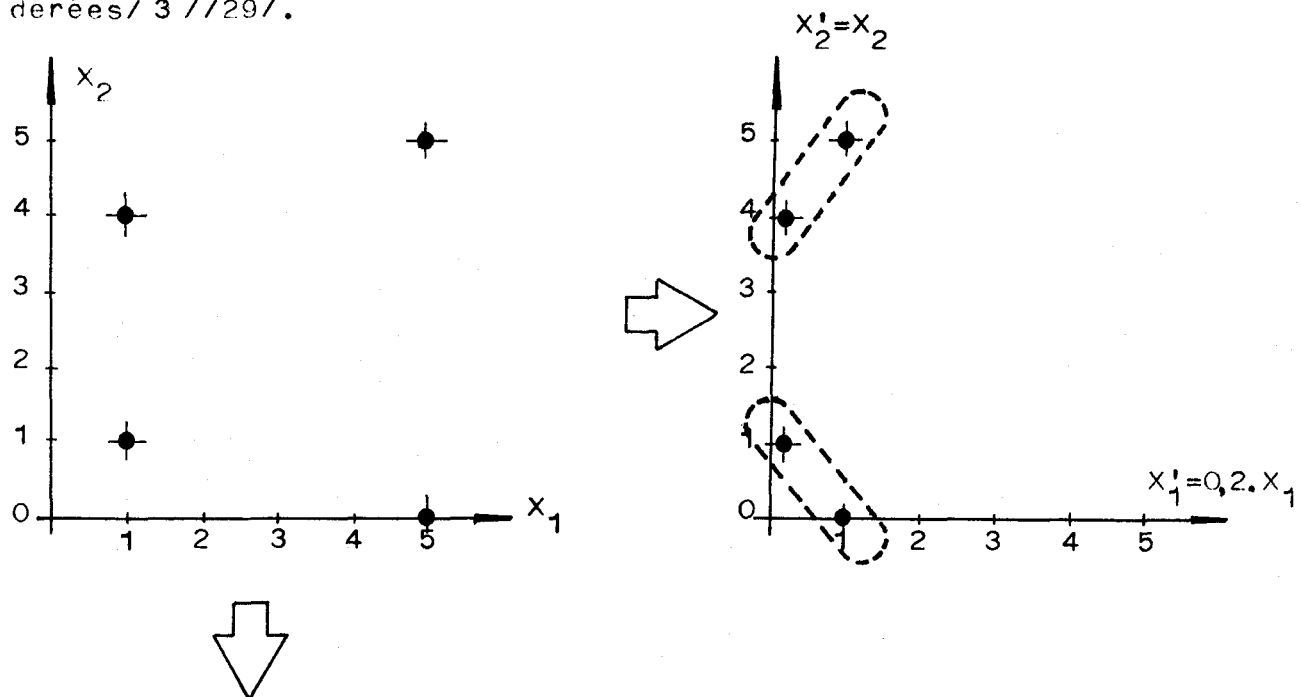


Figure 7.1. Effets des changements d'échelle sur les groupements des observations.

### VII.2.2. APPROCHE STATISTIQUE

Les inconvénients liés aux approches métriques peuvent être évités en faisant appel à des concepts statistiques plutôt qu'au concept de distance. Mais, jusqu'à présent, la recherche de groupements parmi des observations multidimensionnelles a peu été abordée d'un point de vue statistique/5//37/. Il faut même noter que certaines méthodes, bien que classées dans la catégorie des approches statistiques, ne font en fait référence à des notions de corrélation ou de variance que pour définir un critère algébrique à extrémiser. Il ne s'agit en réalité que d'une généralisation de la notion de distance/16//35//29/.

On ne peut à proprement parler de méthode de recherche des groupements statistique que lorsqu'il est fait appel à la notion de densité de probabilité/14/. Dans ce cas on suppose généralement que chaque groupement homogène d'observations correspond à un mode de la fonction de densité/19//20//30/ et le problème se trouve posé en termes de détection des modes.

Les méthodes proposées jusqu'à présent dans cette optique nécessitent en premier lieu soit une estimation de la fonction de densité à partir des observations, soit une estimation directe du gradient de cette fonction. Quelle que soit la manière dont ce gradient est obtenu, les modes sont ensuite détectés en remontant les pentes de la fonction de densité/17//21//22//23//24//30/.

Ces méthodes ont le grand avantage de tenir compte de la structure probabiliste des données, mais en contrepartie, elles présentent tous les inconvénients bien connus de l'utilisation du gradient d'une fonction. Elles sont sensibles au point de départ ainsi qu'aux irrégularités de l'estimation de la fonction de densité ou de son gradient.

### VII.3. CONVEXITE ET DETECTION DES MODES

#### VII.3.1. DETECTION DES MODES PAR ANALYSE DE LA CONVEXITE DES FONCTIONS DE DENSITE

Afin d'éviter l'utilisation du concept de gradient d'une fonction, nous proposons une nouvelle méthode de détection des modes dans laquelle ceux-ci ne sont plus considérés comme des

extrema locaux de la fonction de densité mais comme des domaines où cette fonction est localement concave /34/.

Au chapitre V, nous avons vu que cette caractérisation des modes se trouve justifiée dans le cas où la fonction de densité est un mélange de lois normales. Bien qu'en absence de modèle probabiliste aucune justification théorique ne soit envisageable, nous supposons maintenant que cette propriété s'étend aux modes des fonctions de densité multimodales, même lorsque les classes constituant l'échantillon analysé ne sont pas gaussiennes. Les résultats du test de convexité appliqué à des distributions réelles d'observations permettront de mettre en évidence l'intérêt de cette assertion dans les problèmes de recherche de groupements.

Par une procédure en tout point identique à celle présentée au chapitre V, les observations disponibles sont utilisées pour tester localement la convexité de la fonction de densité sous-jacente. L'espace de représentation des données est ainsi partagé en hypercubes à convexité négative et en hypercubes à convexité positive.

La même procédure d'agrégation des hypercubes à convexité négative adjacents permet de mettre en évidence des domaines qui, par analogie avec les mélanges gaussiens, sont appelés "domaines modaux".

A ce stade, chaque mode est caractérisé par le domaine modal correspondant, qui, du fait du caractère essentiellement local du test de convexité, peut être de forme quelconque.

### VII.3.2. DETECTION DES MODES PAR ANALYSE DE LA CONVEXITE DES DENSITES MARGINALES

Lorsque la taille de l'échantillon ne permet pas d'étudier directement la convexité de la fonction de densité sous-jacente, on peut rechercher les groupements par analyse de la convexité des densités marginales de probabilité/27/.

Nous avons vu au chapitre VI comment identifier les composantes d'un mélange gaussien par analyse de la convexité des densités marginales de sa loi de probabilité. Les densités marginales des classes gaussiennes étant des fonctions de densité normales, leur présence était détectée en identifiant les segments concaves des différentes densités marginales du mélange.

Dans le contexte de la recherche de groupements, nous supposons maintenant que les densités marginales des classes rencontrées présentent toujours un segment concave au voisinage de leur sommet. Aucun modèle probabiliste ne pouvant être accepté a priori, seule l'analyse de la convexité des densités marginales de distributions réelles d'observations permettra de justifier cette assertion/26/.

La procédure de détection des modes est alors très semblable à celle présentée au chapitre précédent, avec une première analyse globale suivie d'une analyse locale itérative. Les modes se trouvent ainsi définis par des domaines hyperparallélipédiques, définis par le produit euclidien des segments concaves mis en évidence sur les densités marginales.

#### VII.4. RECHERCHE DE GROUPEMENTS - RESULTATS EXPERIMENTAUX

Pour analyser les performances des méthodes proposées pour la détection des modes, celles-ci sont maintenant incorporées à des procédures de recherche de groupements et appliquées à des données générées artificiellement et à des données biométriques.

##### VII.4.1. RECHERCHE DE GROUPEMENTS PAR ANALYSE DE LA CONVEXITE DES FONCTIONS DE DENSITE

Pour illustrer les possibilités de détection des modes et de recherche de groupements par analyse de la convexité des densités de probabilité, on se sert de 2 échantillons bidimensionnels analogues à ceux utilisés dans /24/.

##### Exemple 1

Le premier échantillon, représenté sur la figure 7.2., est constitué de trois classes facilement identifiables à l'oeil nu, bien que deux d'entre elles ne se prêtent pas à une modélisation aisée du fait de leur structure spatiale en forme de croissant.

La procédure de détection des modes a été appliquée à cet échantillon pour différentes valeurs du paramètre de discrétisation  $\alpha_q$ . Les résultats obtenus sont représentés sur la figure 7.3. L'analyse de la stabilité du nombre de modes détectés en fonction de  $\alpha_q$  (cf. figure 7.4.) permet de retenir les

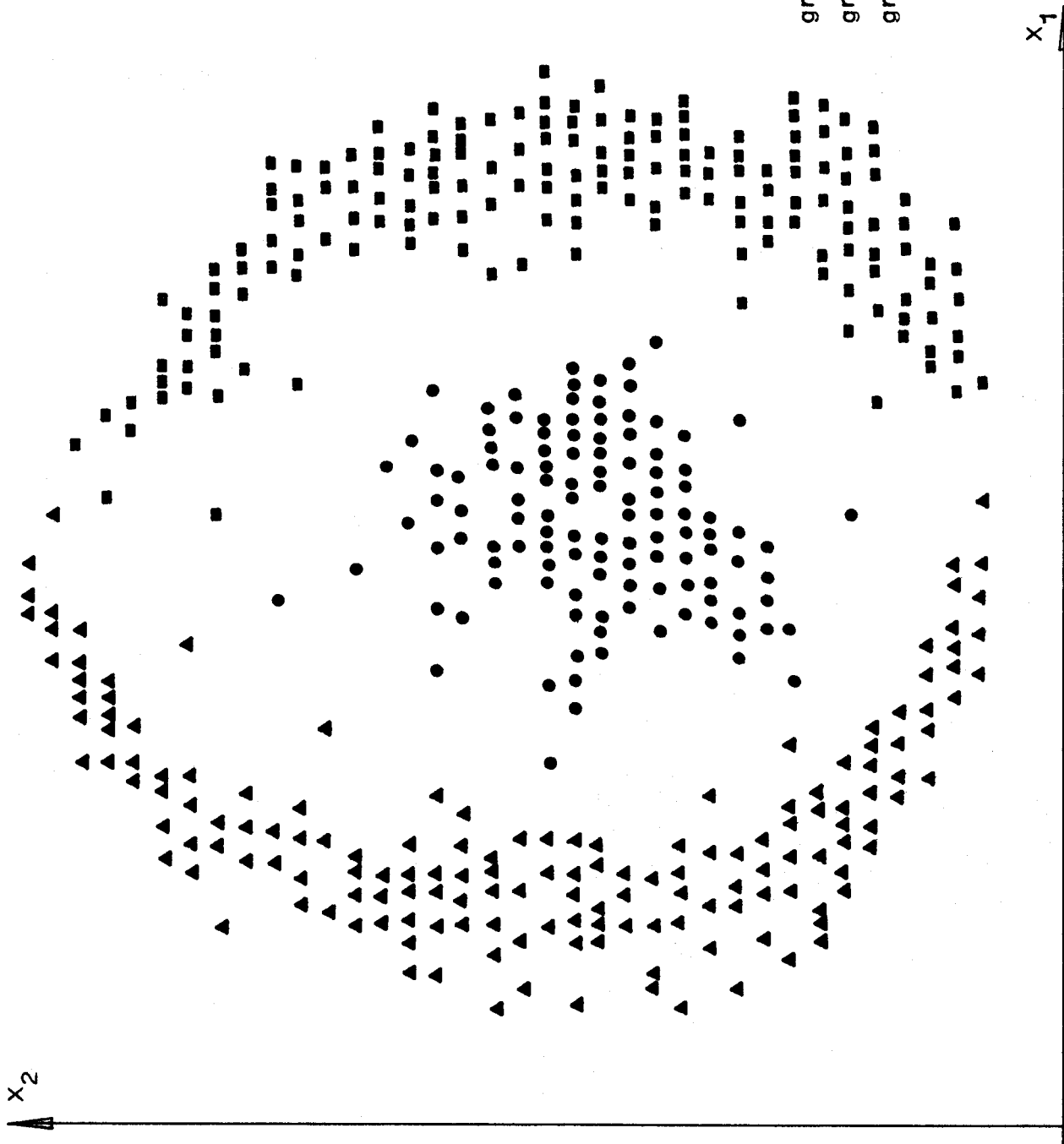


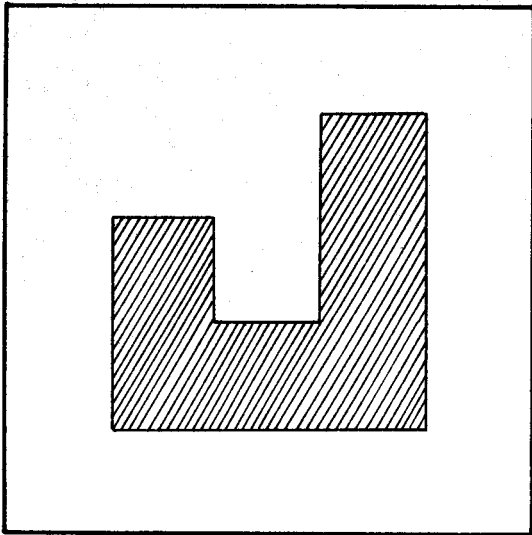
Figure 7.2.

Représentation graphique de l'échantillon de l'exemple n° 1 .

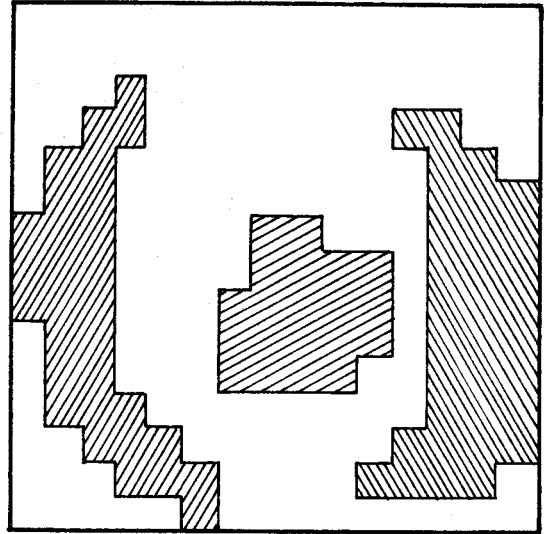
(Les groupements représentés par les trois symboles:  $\blacktriangle$ ,  $\bullet$  et  $\blacksquare$  ont été obtenus avec  $\alpha_0 = 1/20$  )

groupement n° 1 :  $\blacktriangle$   
groupement n° 2 :  $\bullet$   
groupement n° 3 :  $\blacksquare$





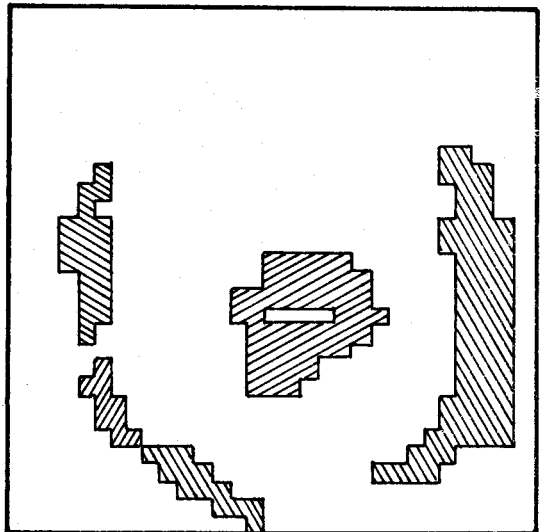
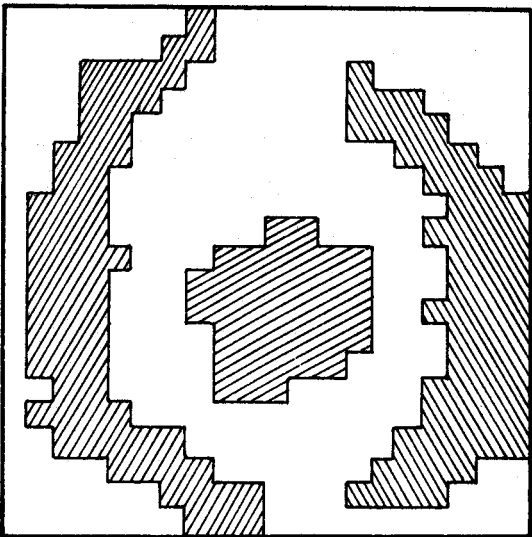
(a)



(b)

(c)

(d)



BUS  
LILLE

Figure 7.3. Domaines modaux détectés pour l'échantillon de la figure 7.2. avec:  
 (a) :  $\alpha_0 = 1/5$ ; (b) :  $\alpha_0 = 1/15$ ; (c) :  $\alpha_0 = 1/20$ ; (d) :  $\alpha_0 = 1/30$ .

domaines modaux de la figure 7.3.(c) pour regrouper les observations. Cet exemple met en évidence la stabilité de la forme des domaines modaux détectés lorsque le paramètre  $\alpha_q$  varie.

Les observations situées à l'intérieur des domaines modaux sont considérées comme des prototypes des classes correspondantes. Ainsi, lorsque tous les domaines modaux sont déterminés, la majorité des observations se trouvent assignées aux différents groupements mis en évidence. Les observations restantes, extérieures aux domaines modaux, peuvent ensuite être assignées à ces groupements par une technique classique de classement avec prototypes/33/. En affectant les observations restantes au groupement du plus proche prototype, on obtient les trois groupements représentés sur la figure 7.2.

La détermination des prototypes est théoriquement insensible aux changements d'échelle, comme le sont les propriétés de convexité des fonctions de densité. Par contre, la phase de classement des observations extérieures aux domaines modaux peut faire intervenir des notions métriques sensibles à tout changement d'échelle sur les caractères.

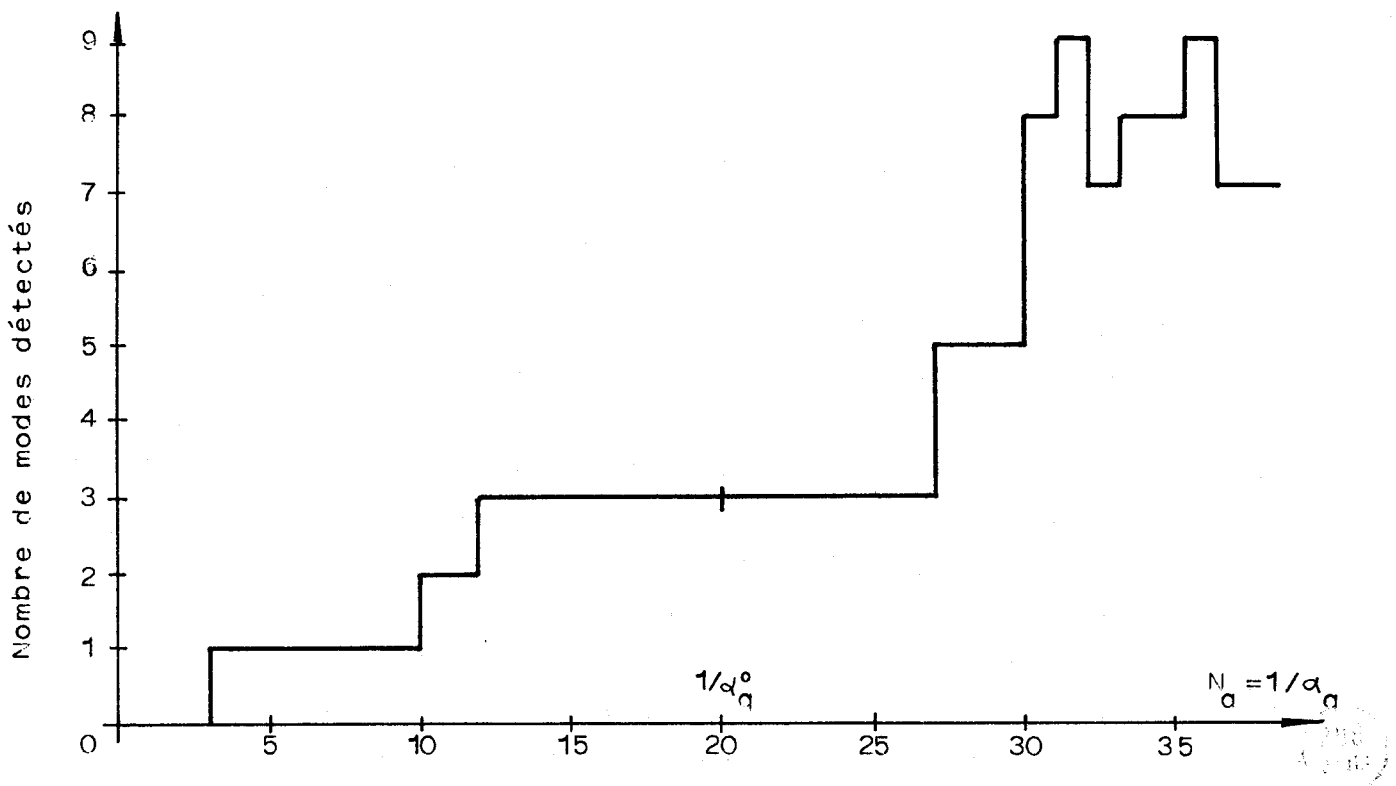


Figure 7.4. Nombre de modes détectés pour l'échantillon de la figure 7.2. en fonction du pas de discrétisation  $\alpha_q$ .

Exemple 2

L'échantillon utilisé pour le second exemple, constitué de deux groupements non sphériques, est représenté sur la figure 7.5. Les domaines modaux de la figure 7.6. ont été obtenus en ajustant la valeur du paramètre de discrétisation  $\alpha_0$  par application du concept de stabilité du nombre de modes détectés.

Pour cet exemple, comme pour le précédent, la meilleure manière de juger la qualité des résultats est certainement l'examen visuel du regroupement final des observations. Les deux groupements représentés sur la figure 7.5. concordent, à quelques exceptions près, avec les conclusions d'un examen visuel de l'échantillon.

Il est important de remarquer que les domaines modaux obtenus dans cet exemple et dans le précédent ne sont pas des domaines convexes. Ainsi, bien que d'après le test de convexité, la fonction de densité sous-jacente soit localement concave en tous les points des domaines modaux, on ne peut conclure à la concavité de la fonction sur ces domaines. (cf. chapitre II, paragraphe II.2.3.)

VII.4.2. RECHERCHE DE GROUPEMENTS PAR ANALYSE DE LA CONVEXITE DES DENSITES MARGINALES

Dans le contexte de la classification automatique, il apparaît indispensable d'associer la procédure de détection des modes par analyse de la convexité des densités marginales à une technique de recherche de groupements. On utilise pour cela les domaines obtenus par le produit euclidien des segments concaves mis en évidence sur les densités marginales et retenus pour définir les modes de la distribution. Soient  $\hat{\mathcal{D}}_k, k=1,2,\dots,K$ , les  $K$  domaines retenus. Les centres  $\hat{O}_k, k=1,2,\dots,K$ , de ces domaines hyperparallélipédiques sont considérés comme les centres des classes  $C_k, k=1,2,\dots,K$ , associés aux  $K$  modes mis en évidence. Les observations situées à l'intérieur de chacun de ces domaines sont ensuite considérées comme les prototypes de chacune des classes.

Les observations restantes sont assignées aux différentes classes en comparant leur distance au centre de

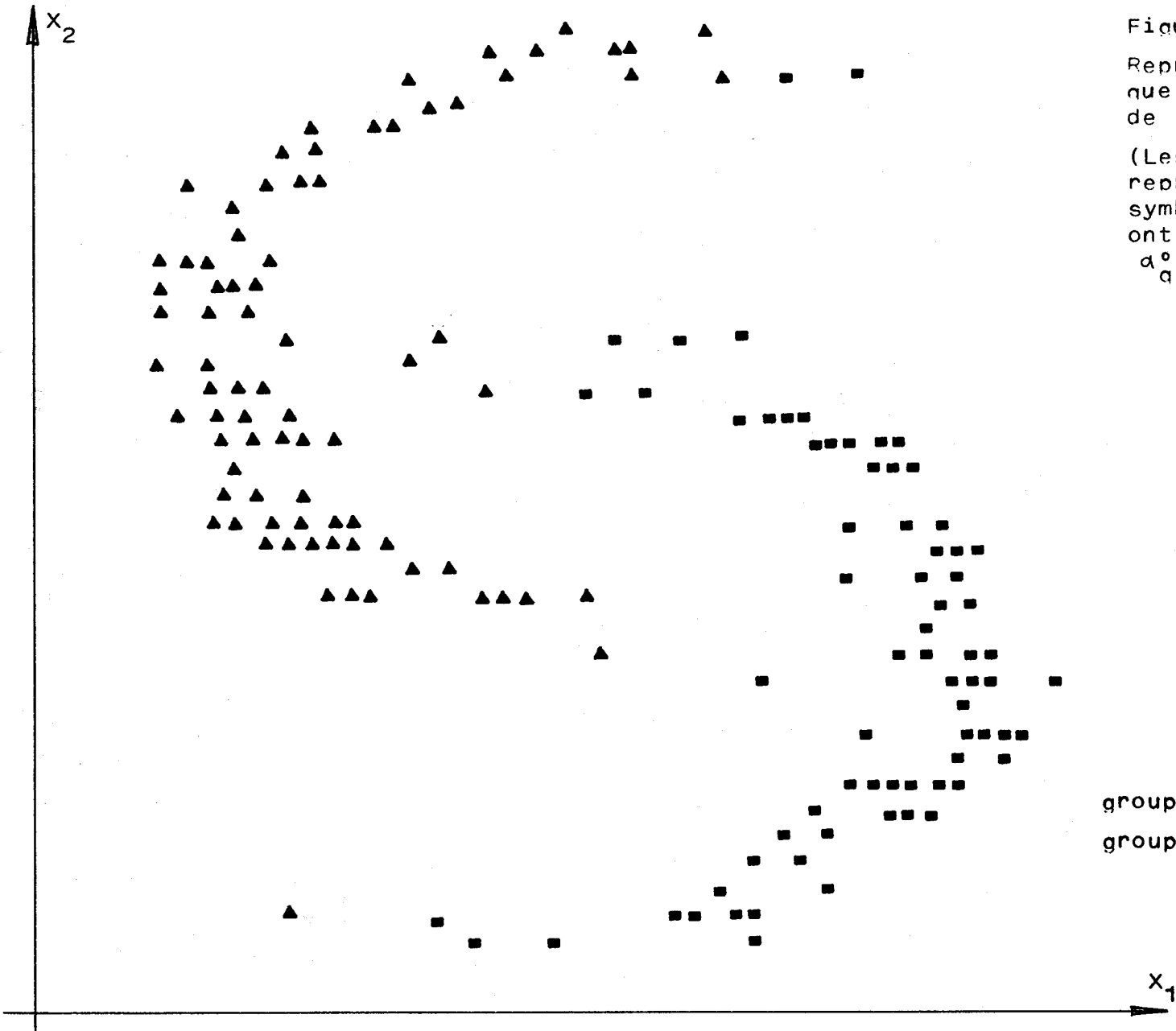


Figure 7.5.

Représentation graphique de l'échantillon de l'exemple n° 2 .

(Les groupements représentés par les 2 symboles: ▲ et ■ ont été obtenus avec  $\alpha^0 = 1/12$  )

groupement n° 1 : ▲  
groupement n° 2 : ■

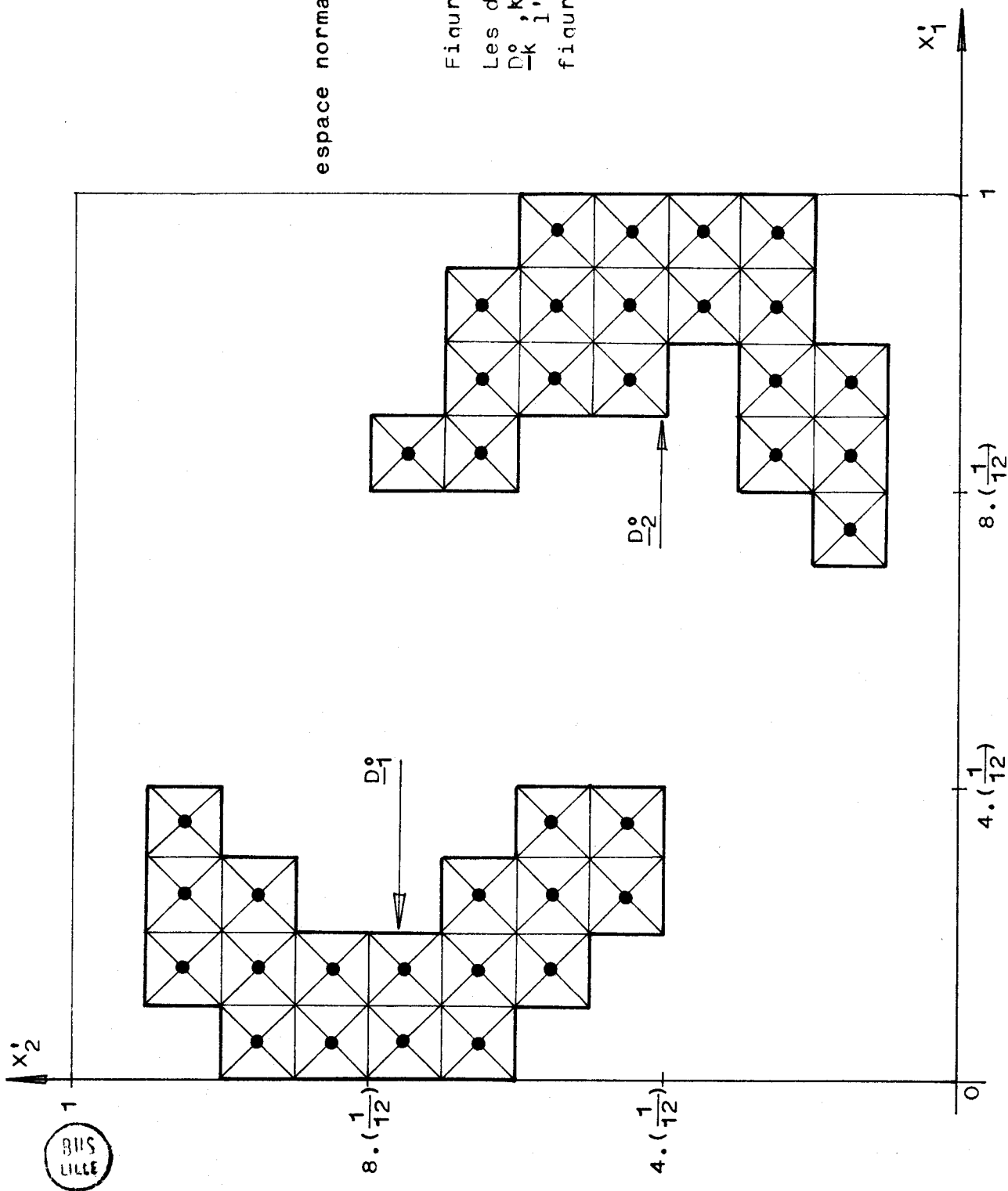


Figure 7.6.

Les deux domaines modaux  $D_k^0$ ,  $k=1,2$  détectés pour l'échantillon de la figure 7.5. avec  $\alpha_C^0=1/12$

B.I.I.S.  
LILLE

chacune des classes. Pour tenir compte de la dispersion spatiale des prototypes de chaque classe, on introduit des coefficients de pondération calculés à partir de la distance moyenne inter-prototypes des prototypes de chaque classe. On définit ainsi un ensemble de fonctions discriminantes

$$q_k(X) = \frac{d_k(X)}{\bar{D}_k}, \quad k = 1, 2, \dots, K$$

où  $d_k(X)$  est la distance du point  $X$  au centre  $\hat{O}_k$  de la classe  $C_k$  et où  $\bar{D}_k$  est la distance moyenne inter-prototypes des prototypes de la classe  $C_k$ .

L'analyse locale de l'ensemble des observations assignées à chacun des groupements mis en évidence permet ensuite d'améliorer les résultats de la classification.

Pour illustrer les possibilités de détection des modes et de recherche de groupements par analyse de la convexité des densités marginales, on utilise des données biométriques. Depuis que Fisher /13/ utilisa sa méthode d'analyse discriminante pour classer un lot d'iris provenant de trois variétés différentes sur lesquels le botaniste Anderson avait mesuré la longueur et la largeur des pétales et des sépales, cet ensemble de données quadridimensionnelles est devenu un exemple classique pour tester les performances des méthodes de classification.

Les mesures effectuées par Anderson (cf. tableau 7.1.) ne portent que sur 50 individus de chacune des variétés Versicolor, Virginica et Setosa. Ce nombre relativement faible pour des données à quatre dimensions fait de cet échantillon un exemple bien adapté à l'illustration de la recherche de groupements par analyse de la convexité des densités marginales.

Nous ne prendrons pas en compte les Iris Setosa dont la variété, nettement différenciée des deux autres /10/, ne pose aucun problème de classification.

On constate, en considérant la distribution des 100 individus restants, que l'on ne détecte deux modes que pour la densité marginale associée à la largeur des pétales.

Cela signifie que les modes associés aux deux variétés présentent un tel degré de chevauchement dans les trois autres densités marginales qu'ils demeurent indifférentiables par une analyse de convexité.

Il n'est pas étonnant, dans ces conditions, que la classification obtenue à la fin de l'analyse globale laisse subsister 7 erreurs alors qu'après quatre itérations de l'analyse locale, le raffinement de la position des modes permet d'aboutir à une classification ne comportant que 4 erreurs (cf. tableau 7.2.).

Le résultat de cette analyse peut être favorablement comparé à ceux obtenus par d'autres procédures. Ainsi la méthode des nuées dynamique /9/ laisse subsister 6 éléments mal classés. Les techniques d'analyse discriminante /15/ sont encore moins performantes puisqu'elles aboutissent à des classifications comportant au moins 7 erreurs.

IRIS VERSICOLOR								IRIS VIRGINICA							
L.S	l.S	L.P	l.P	L.S	l.S	L.P	l.P	L.S	l.S	L.P	l.P	L.S	l.S	L.P	l.P
7.0	3.2	4.7	1.4	6.4	3.2	4.5	1.5	6.3	3.3	6.0	2.5	5.8	2.7	5.1	1.9
6.9	3.1	4.9	1.5	5.5	2.3	4.0	1.3	7.1	3.0	5.9	2.1	6.3	2.9	5.6	1.8
6.5	2.8	4.6	1.5	5.7	2.8	4.5	1.3	6.5	3.0	5.8	2.2	7.6	3.0	6.6	2.1
6.3	3.3	4.7	1.6	4.9	2.4	3.3	1.0	4.9	2.5	4.5	1.7	7.3	2.9	6.3	1.8
6.6	2.9	4.6	1.3	5.2	2.7	3.9	1.4	6.7	2.5	5.8	1.8	7.2	3.6	6.1	2.5
5.0	2.0	3.5	1.0	5.9	3.0	4.2	1.5	6.5	3.2	5.1	2.0	6.4	2.7	5.3	1.9
6.0	2.2	4.0	1.0	6.1	2.9	4.7	1.4	6.8	3.0	5.5	2.1	5.7	2.5	5.0	2.0
5.6	2.9	3.6	1.3	6.7	3.1	4.4	1.4	5.8	2.8	5.1	2.4	6.4	3.2	5.3	2.3
5.6	3.0	4.5	1.5	5.8	2.7	4.1	1.0	6.5	3.0	5.5	1.8	7.7	3.8	6.7	2.2
6.2	2.2	4.5	1.5	5.6	2.5	3.9	1.1	7.7	2.6	6.9	2.3	6.0	2.2	5.0	1.5
5.9	3.2	4.8	1.8	6.1	2.8	4.0	1.3	6.9	3.2	5.7	2.3	5.6	2.8	4.9	2.0
6.3	2.5	4.9	1.5	6.1	2.8	4.7	1.2	7.7	2.8	6.7	2.0	6.3	2.7	4.9	1.8
6.4	2.9	4.3	1.3	6.6	3.0	4.4	1.4	6.7	3.3	5.7	2.1	7.2	3.2	6.0	1.8
6.8	2.8	4.8	1.4	6.7	3.0	5.0	1.7	6.2	3.8	4.8	1.8	6.1	3.0	4.9	1.8
6.0	2.9	4.5	1.5	5.7	2.6	3.5	1.0	6.4	2.8	5.6	2.1	7.2	3.0	5.8	1.6
5.5	2.4	3.8	1.1	5.5	2.4	3.7	1.0	7.4	2.8	6.1	1.9	7.9	3.8	6.4	2.0
5.8	2.7	3.9	1.2	6.0	2.7	5.1	1.6	6.4	2.8	5.6	2.2	6.3	2.8	5.1	1.5
5.4	3.0	4.5	1.5	6.0	3.4	4.5	1.6	6.1	2.6	5.6	1.4	7.7	3.0	6.1	2.3
6.7	3.1	4.7	1.5	6.3	2.3	4.4	1.3	6.3	3.4	5.6	2.4	6.4	3.1	5.5	1.8
5.6	3.0	4.1	1.3	5.5	2.5	4.0	1.3	6.0	3.0	4.8	1.8	6.9	3.1	5.4	2.1
5.5	2.6	4.4	1.2	6.1	3.0	4.6	1.4	6.7	3.1	5.6	2.4	6.9	3.1	5.1	2.3
5.8	2.6	4.0	1.2	5.0	2.3	3.3	1.0	5.8	2.7	5.1	1.9	6.8	3.2	5.9	2.3
5.6	2.7	4.2	1.3	5.7	3.0	4.2	1.2	6.7	3.3	5.7	2.5	6.7	3.0	5.2	2.3
5.7	2.9	4.2	1.3	6.2	2.9	4.3	1.3	6.3	2.5	5.0	1.9	6.5	3.0	5.2	2.0
5.1	2.5	3.0	1.1	5.7	2.8	4.1	1.3	6.2	3.4	5.4	2.3	5.9	3.0	5.1	1.8

Tableau 7.1. Données biométriques relevées sur 2 variétés d'iris  
L.S.: Longueur des sépales. L.P.: Longueur des pétales  
l.S.: Largeur des sépales. l.P.: Largeur des pétales  
Source: CHEN, C.H. "Statistical Pattern Recognition" Spartan Haygen (1973)



ANALYSE GLOBALE				ANALYSE LOCALE (dernière itération)			
$\hat{J}_1$ : Iris Versicolor Borne Inf. Borne Sup. Centre $\hat{O}_1$	L.S.	L.P.	L.P.	$\hat{J}_1$ : Iris Versicolor Borne Inf. Borne Sup. Centre $\hat{O}_1$	L.S.	L.P.	L.P.
	5,33	2,51	4,11		5,25	2,47	3,87
	6,61	3,03	5,23		7,35	3,63	4,73
	5,97	2,77	4,67		6,30	3,05	4,30
$\hat{J}_2$ : Iris Virginica Borne Inf. Borne Sup. Centre $\hat{O}_2$	L.S.	L.P.	L.P.	$\hat{J}_2$ : Iris Virginica Borne Inf. Borne Sup. Centre $\hat{O}_2$	L.S.	L.P.	L.P.
	5,33	2,51	4,11		5,60	2,50	4,90
	6,61	3,03	5,23		6,75	3,15	5,90
	5,97	2,77	4,67		6,18	2,85	5,40
Iris Versicolor mal classés	L.S.	L.P.	L.P.	Iris Versicolor mal classés	L.S.	L.P.	L.P.
	7,00	3,20	4,70		5,90	3,20	4,80
	6,90	3,10	4,90		6,70	3,00	5,00
	5,90	3,20	4,80		6,00	2,70	5,10
Iris Versicolor mal classés	L.S.	L.P.	L.P.	Iris Versicolor mal classés	L.S.	L.P.	L.P.
	6,70	3,00	5,00		5,90	3,20	4,80
	6,00	2,70	5,10		6,70	3,00	5,00
	6,00	2,70	5,10		6,00	2,70	5,10
Iris Virginica mal classés	L.S.	L.P.	L.P.	Iris Virginica mal classés	L.S.	L.P.	L.P.
	4,90	2,50	4,50		4,90	2,50	4,50
Iris Virginica mal classés	L.S.	L.P.	L.P.	Iris Virginica mal classés	L.S.	L.P.	L.P.
	6,30	2,80	5,10		4,90	2,50	4,50

Tableau 7.2.

Caractéristiques des domaines modaux et résultats de la classification de deux variétés d'iris.

L.S. : Longueur des sépales    L.S. : Largeur des sépales  
L.P. : Longueur des pétales    L.P. : Largeur des pétales (en cm.)



VII.5. CONCLUSION

Les deux méthodes de recherche de groupements proposées dans ce chapitre mettent en évidence l'intérêt de considérer les modes des distributions en termes de convexité locale plutôt que comme des extrema des fonctions de densité.

L'analyse de la convexité des fonctions de densité, telle qu'elle a été présentée dans la seconde partie de ce travail, permet de détecter des modes de forme et de taille très variées mais nécessite un nombre d'observations disponibles d'autant plus important que la dimension des données est élevée.

Par contre, l'analyse de la convexité des densités marginales permet de détecter les modes, même lorsqu'on ne dispose que d'un petit échantillon compte-tenu de la dimension du problème. Cette méthode de recherche de groupements est à rapprocher d'une procédure très voisine utilisant des histogrammes pour estimer les densités marginales /12/. Toutefois, pour de petits échantillons, la forme des histogrammes est très sensible à la position des intervalles, ce qui rend difficile la détection des modes des densités marginales. Les auteurs de la méthode, conscients de ce problème, la réservaient à l'analyse de grands échantillons.

L'utilisation d'une technique d'estimation plus élaborée et du concept de convexité permet d'utiliser l'information contenue dans les densités marginales pour analyser des échantillons de forte dimension, quelle que soit leur taille. Toutefois, comme nous disposons d'une méthode plus performante pour les grands échantillons, nous réservons cette approche aux échantillons de taille réduite.

A ce stade de l'étude, l'analyste dispose d'un ensemble de procédures qui, en exploitant le concept de convexité, permettent soit d'optimiser le processus de classification sous différentes hypothèses, soit de rechercher des groupements dans un lot de données.

Ces procédures, confiées à un ordinateur, ne constituent en fait que le premier aspect de la classification automatique. L'utilisation de ces techniques dans des cas concrets et l'interprétation ou l'exploitation des résultats obtenus constituent, pour le chercheur, un second aspect tout aussi important que le premier.

Pour illustrer ce second volet de la classification automatique, nous nous proposons maintenant d'appliquer les techniques mises au point précédemment au problème de l'estimation des ressources forestières.

REFERENCES (chapitre VII)

- /1/ AUGUSTSON, J.G. et MINKER, J.  
"An analysis of some graph theoretical cluster techniques"  
J. ACM Vol. 17 p.571-588 (1970)
- /2/ BALL, G.H.  
"Data analysis in the social sciences: what about the details"  
Proc. FJCC p. 533-560 (Spartan Books, Washington) (1965)
- /3/ BALL, G.H. et HALL, D.J.  
"A clustering technique for summarizing multivariate data"  
Behavioural Sc. Vol. 12 , p. 153-155 (1967)
- /4/ BATCHELOR, B.G. et WILKINS, B.R.  
"Methods for location of clusters of patterns to initialize  
a learning machine"  
Electronics Letters Vol. 5, n° 20, p. 481-483 (1969)
- /5/ BAYNE, C.K. , BEAUCHAMP, J.J. , BEGOVICH, C.L. et KANE, V.E.  
"Monte Carlo comparison of selected clustering procedures"  
Pattern Recon. Vol. 12 , p. 51-62 (1980)
- /6/ BOURTON, M.  
"Contribution à l'analyse, le traitement et la reconnaissance  
des formes ponctuées. Application à la classification de  
couches géologiques"  
Thèse d'état , Université de Lille 1 (1975)
- /7/ CAILLEZ, F. et PAGES, J.P.  
"Introduction à l'analyse des données"  
SMASH Paris p. 544-580 (1976)
- /8/ DIDAY, E.  
"Une nouvelle méthode en classification automatique et  
reconnaissance des formes: la méthode des nuées dynamiques"  
Revue de Stat. Appl. Vol. 19 , n° 2 , p. 20-33 (1971)
- /9/ DIDAY, E.  
"The dynamic cluster method in non hierarchical clustering"  
Int. J. of Computer & Inf. Sc. Vol.2 , n°1 , p. 61-87 (1973)
- /10/ DUDA, R.O. et HART, P.E.  
"Pattern classification and scene analysis"  
J. Wiley, New York p. 217-225 (1973)
- /11/ DUBES, R. et JAIN, A.K.  
"Validity studies in clustering methodologies"  
Pattern Recon. Vol. 11 p. 235-254 (1979)

- /12/ EIGEN, D.J. , FROMM, F.R. et NORTHOUSE, R.A.  
"Cluster analysis based on dimensional information with applications to feature selection and classification"  
IEEE Trans. on Syst., Man & Cyb. Vol. SMC-4, n°3, p.284-294 (1974)
- /13/ FISHER, R.A.  
"The use of multiple measurements in taxonomic problems"  
Ann. Eugenics. Vol. 7, Part II, p. 179-188 (1936)
- /14/ FLEISS, J.L. et ZUBIN, J.  
"On the method and theory of clustering"  
Multivariate behavioral research p. 235-250 (1969)
- /15/ FREEMAN,  
"Experiments in discrimination and classification"  
Pattern Recog. Vol. 1, n° 3, (1969)
- /16/ FRIEDMAN, H.P. et ROBIN, J.  
"On some invariant criteria for grouping data"  
J. Amer. Stat. Ass. Vol. 62 p. 1159-1178 (1967)
- /17/ FUKUNAGA, K. et HOSTETLER, L.D.  
"The estimation of the gradient of a density function with applications in pattern recognition"  
IEEE Trans. on Inf. Theory Vol IT-21, n°1, p.32-40 (1975)
- /18/ FUKUNAGA, K. et KOONTZ, W.L.G.  
"A criterion and an algorithm for grouping data"  
IEEE Trans. on Computers Vol. C-19 p.917-923 (1970)
- /19/ GITMAN, I.  
"An algorithm for nonsupervised pattern classification"  
IEEE Trans. on Syst., Man & Cyb. Vol.SMC-3 p.66-74 (1973)
- /20/ GITMAN, I. et LEVINE, M.D.  
"An algorithm for detecting unimodal fuzzy sets and its applications as a clustering technique"  
IEEE Trans on Computers Vol. C-19 p.583-593 (1970)
- /21/ JOHNSTON, B. , BAILEY, T. et DUBES, R.  
"A variation on a nonparametric clustering method"  
IEEE Trans on Pattern Anal. & Mach. Intel.  
Vol. PAMI-1, n°4 , p. 400-408 (1979)
- /22/ KATZ, J.O. et ROHLF, F.J.  
"Function point cluster analysis"  
Systematic Zoology Vol. 22 p. 295-301 (1973)
- /23/ KITTNER, J.  
"A locally sensitive method for cluster analysis"  
Pattern Recog. Vol. 8 p. 23-33 (1976)

- /24/ KOONTZ,W.L.G., NARANDRA,P.M. et FUKUNAGA,K.  
 "A graph theoretic approach to nonparametric cluster analysis"  
 IEEE Trans. on Computers Vol.C-25,n°9, p. 936-944 (1976)
- /25/ LANE,G.N. et WILLIAMS,W.T.  
 "A general theory of classificatory sorting strategies  
 1 - Hierarchical systems"  
 Computer J. Vol. 9 p. 973-980 (1967)
- /26/ LIMOURI,M.  
 "Classification automatique non supervisée par analyse des  
 densités marginales de probabilité"  
 Diplôme d'Etudes Supérieures, Univ. de Rabat, Maroc (1979)
- /27/ LIMOURI,M. et POSTAIRE,J.G.  
 "Utilisation des histogrammes en classification automatique"  
 Congrès AFCET-IRIA Reconnaissance des Formes et Intelligence  
 Artificielle Toulouse p. 373-380 (1979)
- /28/ LUKASOVA,A.  
 "Hierarchical agglomerative clustering procedure"  
 Pattern Recon. Vol.11 p. 365-381 (1979)
- /29/ MAC QUEEN,J.  
 "Some methods for classification and analysis of  
 multivariate observations"  
 Proc. 5th Berkeley Symp. on Math. Stat. and Prob. (1967)  
 Vol. I , p. 281-297 Lecam & Neyman,eds. Univ. of Calif. Press
- /30/ MIZOGUCHI,R. et SHIMURA,M.  
 "Non parametric learning without a teacher based on  
 mode estimation"  
 IEEE Trans. on Computers Vol. C-25 p. 1109-1117 (1976)
- /31/ SFRETYEN,G.S.  
 "Pattern recognition by an adaptive process of sample  
 set construction"  
 IRE Trans. on Info. Theory Vol. IT-8 p. 82-91 (1962)
- /32/ THORNDIKE,R.L.  
 "Who belong in the family?"  
 Psychometrica Vol. 18 p. 267-276 (1953)
- /33/ TOU,J.T. et GONZALEZ,R.C.  
 "Pattern recognition principles"  
 Addison-Wesley p. 75-86 (1974)
- /34/ VASSEUR,C. et POSTAIRE,J.G.  
 "A convexity testing method for cluster analysis"  
 IEEE Trans. on Syst.,Man & Cyb. Vol.SMC-10 p. 145-149 (1980)

- /35/ VOGEL, M.A. et WONG, A.K.C.  
"PFS clustering method "  
IEEE Trans. on Pattern Analysis & Machine Intel.  
Vol. PAMI-1 , n° 3 p. 237-245 (1979)
- /36/ WILKS, S.S.  
"Multidimensional statistical scatter"  
Contrib. to Proba. & Stat. I. Olkin, ed. Stanford, Calif. (1960)
- /37/ WISHART, D.  
"Estimating the modes of a multivariate sample density"  
Proc. Joint Meeting of the Classification Soc. &  
British Pattern Recon. Ass. London (1978)
- /38/ WISHART, D.  
"Mode analysis: a generalization of nearest-neighbor  
which reduces chaining effects"  
Numerical Taxinomy p. 282-308  
A.J. Cole, ed. Academic Press New York (1969)
- /39/ WOLFE, J.H.  
"Pattern clustering for multivariate mixture analysis"  
Multiv. Behavioral Research Vol. 5 p. 329-350 (1970)
- /40/ YAU, S.S. et CHANG, S.C.  
"A direct method for cluster analysis"  
Pattern Recognition Vol. 7 , p. 215-224 (1975)
- /41/ ZAHN, C.T.  
"Graph theoretical methods for detecting and describing  
gestalt clusters"  
IEEE Trans. on Computers Vol. C-20 p. 68-86 (1971)

## QUATRIEME PARTIE

APPLICATION DES TECHNIQUES DE RECONNAISSANCE DES  
FORMES ET DE CLASSIFICATION AUTOMATIQUE A LA  
CONSTRUCTION DES TABLES DE CUBAGE FORESTIERES



## C H A P I T R E VIII

### ANALYSE MORPHOMETRIQUE DU CEDRE DU MAROC

#### VIII.1. L'ESTIMATION DES RESSOURCES FORESTIERES

##### VIII.1.1. OBJECTIFS DES INVENTAIRES FORESTIERS

Face à l'évolution rapide de la consommation de bois constatée ces dernières années, on se trouve actuellement dans une phase de transition entre la période où les ressources forestières étaient considérées comme inépuisables et un avenir immédiat où l'utilisation du bois se ressentira nécessairement de la pénurie qui ne cesse de s'aggraver.

L'exploitation des ressources forestières en vue de répondre aux besoins des divers secteurs économiques et industriels doit être abordée aujourd'hui avec un souci d'efficience maximum. Il s'agit de planifier une exploitation optimisée des forêts qui garantisse une satisfaction rationnelle des besoins tout en préservant le patrimoine forestier. Une évaluation aussi précise que possible des ressources présentes et à venir au niveau

de la parcelle, du massif forestier et même du pays devient, dès lors, un outil de travail indispensable. Aussi est-il nécessaire de procéder à des inventaires très complets afin de connaître les volumes de bois disponibles en quantité (volume à l'hectare) et en qualité (bois d'oeuvre, bois d'industrie, bois de trituration).

Lorsqu'il s'agit d'établir l'inventaire de forêts à pénétration difficile, telles les forêts tropicales, les inventaires forestiers se font sur la base de techniques de photo-interprétation et de cartographie, appuyées éventuellement par des reconnaissances au sol.

Pour les forêts plus accessibles, on peut obtenir des inventaires beaucoup plus fiables et précis grâce à des sondages directs sur le terrain. Ceux-ci permettent d'acquérir les données nécessaires à l'estimation du volume de bois disponible.

#### VIII.1.2. ESTIMATION DU VOLUME BRUT SUR PIED

Le but premier de tout inventaire forestier est l'estimation du volume brut sur pied de chaque essence inventoriée. On appelle volume brut d'une essence une estimation du volume total constitué par toutes les tiges des arbres d'une dimension minimale choisie préalablement.

L'estimation du volume brut moyen à l'hectare des différentes espèces inventoriées se fait sur un nombre d'arbres réduit constituant une "unité de sondage".

La mesure directe du volume des arbres sur pied est une opération très longue et onéreuse qui ne peut être envisagée pour tous les arbres de l'unité de sondage.

Sur le terrain, on se contente d'effectuer des mesures rapides et simples qui sont consignées sur des "feuilles de pointage". Le volume brut de chaque élément de l'unité de sondage est ensuite déterminé à l'aide d'un "tarif de cubage". Il s'agit d'un tableau à une ou deux entrées donnant le volume de chaque arbre en fonction d'un ou de deux paramètres facilement accessibles sur le terrain. Il est d'usage d'utiliser le diamètre  $D$  à hauteur d'homme dans le cas où on se contente d'un seul paramètre. Pour constituer les tableaux à deux entrées, on adjoint généralement la hauteur  $H$  de l'arbre.

### VIII.1.3. ETABLISSEMENT DES TARIFS DE CUBAGE

Les tarifs de cubage sont calculés à l'aide d'échantillons distincts de ceux utilisés pour les sondages.

Les tarifs à une ou deux entrées sont généralement établis par régression progressive multiple pondérée / 8 /, donnant le volume sous la forme d'un polynôme en  $D$  ou d'un polynôme en  $D$  et  $H$ . On obtient ainsi pour chaque essence un tarif moyen unique dont l'application à différents lots permet d'estimer les volumes avec une précision plus ou moins bonne.

Le manque de précision des tarifs de cubage est un des plus vieux problèmes qui se soient posés aux ingénieurs forestiers. Cette imprécision résulte en partie du fait que les méthodes utilisées pour établir et employer les tarifs de cubages supposent implicitement que la forme des arbres considérés est unique et constante. Or, nous verrons dans ce chapitre que les tiges d'une même espèce peuvent présenter de grandes différences de forme, même à l'intérieur d'une zone géographique restreinte.

Pour améliorer la précision et la fiabilité des tarifs de cubage, nous proposons une nouvelle approche basée sur une étude préalable de la forme des tiges pour chaque essence. Nous verrons dans ce chapitre comment l'application des méthodes de classification automatique présentées dans ce mémoire permet de mettre en évidence trois types de forme chez le Cèdre du Maroc.

Au chapitre suivant, nous tiendrons compte de ce polymorphisme pour améliorer la représentativité des échantillons choisis pour établir les tarifs de cubage. Les tarifs ainsi obtenus, plus précis et plus fiables que les tarifs conventionnels, permettront une meilleure gestion des forêts de cèdres du Maroc.

### VIII.2. CARACTERISATION DE LA FORME DES ARBRES

L'ensemble des mécanismes et des facteurs responsables de la forme des arbres est extrêmement complexe. On sait que le contexte bioclimatique, l'environnement de l'arbre, sa position sociale au sein du peuplement et ses possibilités de nutrition

influencent la forme des tiges/3/. Les variations de forme peuvent également être liées aux contraintes mécaniques subies par l'arbre (poids des branches, vent, neige, ...), à des pratiques culturales (élagage) et même des variations d'ordre génétique ne sont pas exclues/ 7 /.

Pour tenir compte de tous ces phénomènes, les forestiers ont tenté de caractériser la forme des arbres par des grandeurs calculées à partir des mesures dendrométriques.

Le paramètre le plus utilisé est sans doute le "coefficient de forme"  $f$  qui s'exprime par la formule

$$f = \frac{V}{G.H} \quad (8.1.)$$

où  $V$  est le volume réel de la tige,  $G$  la surface terrière de l'arbre à 1,30 m et  $H$  sa hauteur totale. Ce type de coefficient permet d'avoir une estimation assez précise de la forme des tiges/10/.

D'autres auteurs préfèrent caractériser la forme des arbres par une relation entre les diamètres à deux niveaux différents/ 4 // 9 /.

Cependant, en se limitant à enfermer les formes dans un seul paramètre, aussi représentatif soit-il, il paraît difficile d'envisager une étude de leur variabilité. En effet, une forme n'est pas un rapport unique, c'est un ensemble de rapports dont seule une analyse multidimensionnelle permet de cerner les variations/ 2 /.

Une approche dans ce sens, faisant appel à un modèle allométrique multidimensionnel, a été proposée pour étudier la forme des tiges/7/. Cependant, l'auteur faisant son étude sur un peuplement de peupliers très homogène ne pouvait dégager qu'une forme unique moyenne, caractéristique du peuplement considéré.

L'une des premières approches mises au point pour apprécier la forme des arbres consiste à classer les tiges par "type dendrométrique"/10/.

Cette méthode revient à assimiler la tige de chaque arbre à un solide de révolution constitué de portions de cylindre, de paraboloïde, de cône et de néloïde se raccordant tangentielllement.

Cette approche, de mise en œuvre très lourde, nous a suggéré d'assimiler les tiges à des assemblages de trois troncs de cône de révolution (cf. figure 8.1.). Le premier représente l'empatement de l'arbre, c'est à dire la partie située entre le sol et une hauteur de 1,30 m, la "hauteur d'homme" qui joue un rôle privilégié chez les forestiers. Le second permet de décrire la tige entre 1,30 m et la moitié de la hauteur totale  $H$  de l'arbre. Le troisième, enfin, représente le haut de l'arbre, entre la mi-hauteur et le sommet.

Ces trois troncs de cône permettent de schématiser directement la forme des tiges à partir des mesures utilisées couramment en foresterie, à savoir:

- Le diamètre à la souche  $D_S$
- Le diamètre à hauteur d'homme  $D_{130}$
- Le diamètre à mi-hauteur  $D_{H/2}$
- La hauteur totale  $H$

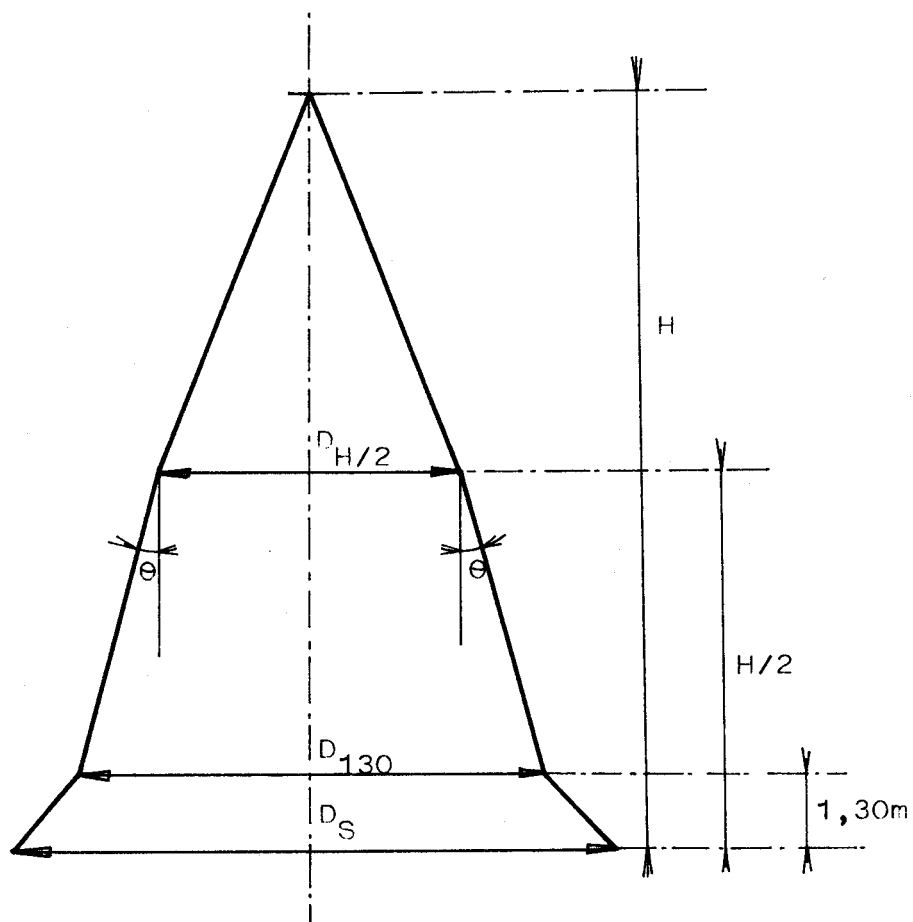


Figure 8.1. Schématisation de la forme d'une tige par assemblage de trois troncs de cône de révolution.

Afin d'éliminer le facteur taille de cette étude sur la forme des tiges et dans le but d'assurer la reconstruction de la forme schématique proposée à partir des paramètres qui la définissent, nous avons retenu, pour caractériser chaque tige, les trois paramètres suivants:

$$\delta = \frac{D_{H/2}}{D_{130}}$$

$$\varepsilon = \frac{D_{130}}{D_S}$$

$$\theta = \text{Arctg} \left( \frac{D_{130} - D_{H/2}}{H - 2,6} \right) \quad (\text{exprimé en grades})$$

Le paramètre  $\delta$ , appelé "coefficient de décroissance"\* est le rapport entre le diamètre au milieu de la tige et celui à hauteur d'homme.

Le paramètre  $\varepsilon$  traduit l'empattement à la base de l'arbre. Nous l'avons appelé "coefficient d'empattement".

L'angle  $\theta$ , que nous nommerons "angle de défilement", caractérise la "décroissance métrique sur le diamètre"\* depuis la hauteur d'homme jusqu'au milieu de l'arbre, c'est à dire la diminution moyenne du diamètre par mètre de hauteur.

Note: Le facteur taille étant absent de ces trois paramètres, la reconstruction de la forme schématique d'une tige en grandeur réelle nécessite la connaissance d'au moins une des dimensions de l'arbre.

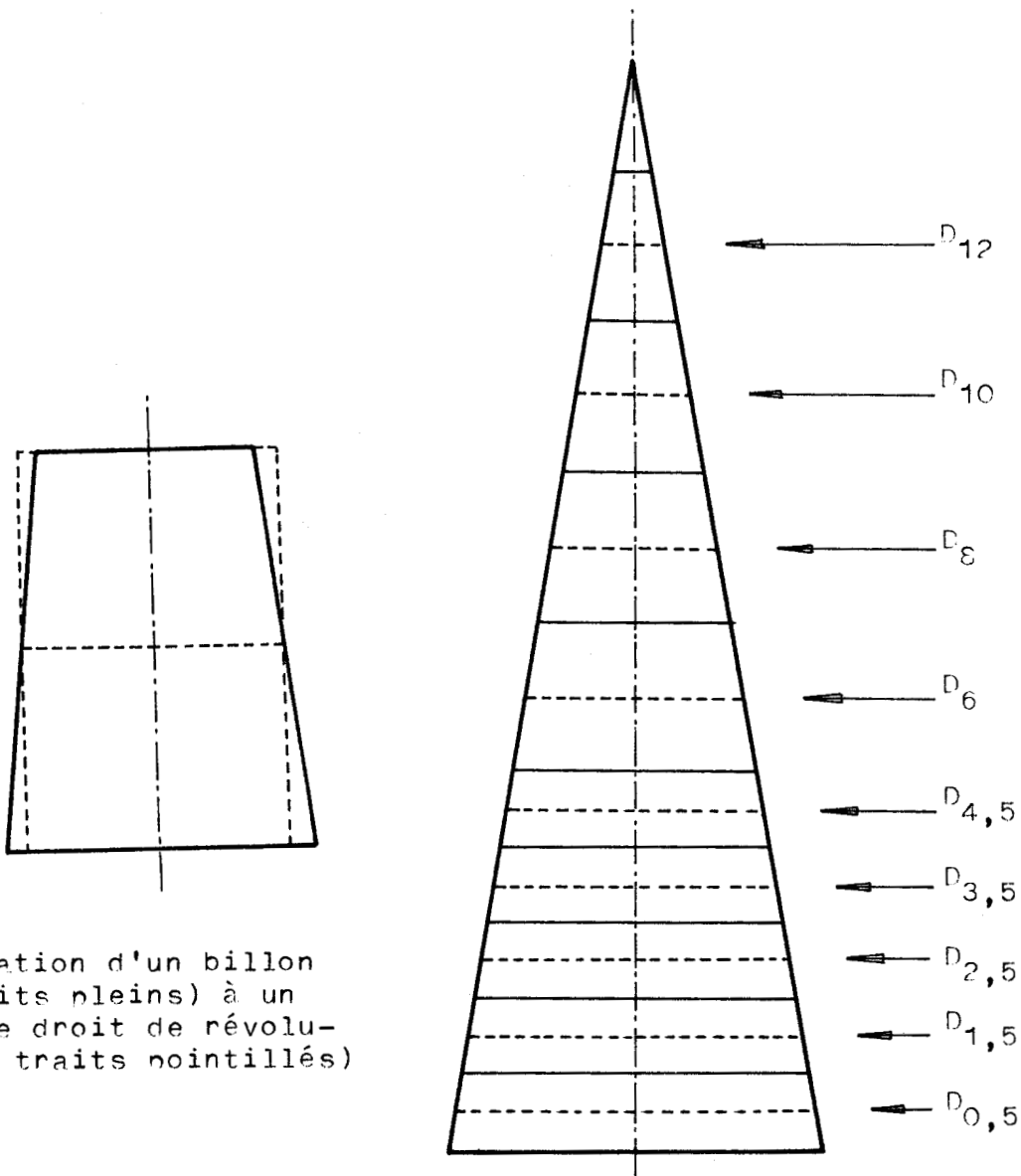
Pour nuancer cette schématisation de la forme qui peut paraître un peu grossière, nous adjoindrons à ces trois paramètres le coefficient de forme  $f$  défini par l'équation 8.1., c'est à dire:

$$f = \frac{4 V}{\pi \cdot D_{130}^2 \cdot H}$$

La détermination de ce coefficient nécessite la connaissance du volume réel des tiges. C'est pourquoi cette analyse de la forme des tiges est de préférence effectuée sur des arbres abattus, leur cubage précis étant plus facile que celui des arbres debout.

\* Norme N.F. B 53017

Les cèdres sont cubés en les décomposant en billons. Le volume de ces billons est calculé en les assimilant à des cylindres de même hauteur et de section droite égale à la section médiane du billon (méthode de Huber)<sup>19</sup>. Comme la variation du diamètre est plus forte vers le bas de la tige que vers son sommet, on utilise des billons de 1 mètre de hauteur dans la partie inférieure de l'arbre et des billons de 2 mètres de hauteur dans sa partie supérieure. Chaque cèdre a ainsi été cubé en mesurant son diamètre à des hauteurs de 0,5m., 1,5m., 2,5m., 3,5m., 4,5m., 6m., 8m., 10m., 12m., etc... jusqu'au sommet de l'arbre. (cf. figure 8.2.)



Assimilation d'un billon (en traits pleins) à un cylindre droit de révolution (en traits pointillés)

Figure 8.2. Décomposition d'une tige en billons successifs  
 $D_{0,5}$  : diamètre à la hauteur 0,5m  
 $D_{1,5}$  : diamètre à la hauteur 1,5m , etc....

VIII.3. LES DONNEES : LES CEDRES DU MAROC

Un important fichier morphométrique, relatif aux cédres du Rif, a servi de base à cette étude, menée en collaboration avec Monsieur M'hirit, Ingénieur Forestier et Professeur à l'Ecole Nationale des Ingénieurs Forestiers de Salé, Maroc. Les données, collectées par Monsieur M'hirit lui-même dans le cadre d'un travail de thèse en cours, ont été recueillies à l'occasion de coupes concernant 1775 cèdres provenant de six grands massifs forestiers du Rif occidental et du Rif central (cf. tableau 8.1.)

N° d'ordre affecté au massif forestier	Nom du massif forestier	Superficie (Ha)	Altitude moyenne	Bioclimat	Substrat	Taille de l'échantillon disponible
1	Tizirène	1327	1700m	Rif occidental humide et frais	Schistes et Grès	194
2	Bab Chiker	740	1700m			191
3	Ghommara	2320	1650m	Rif central humide et froid	Grès Quartzite	154
4	Tidiqhine	1850	2000m			432
5	Beni Khennous	1200	2000m			356
6	Jbel Lerz	2400	1850m			171

Tableau 8.1. Caractéristiques des cédraies du Rif



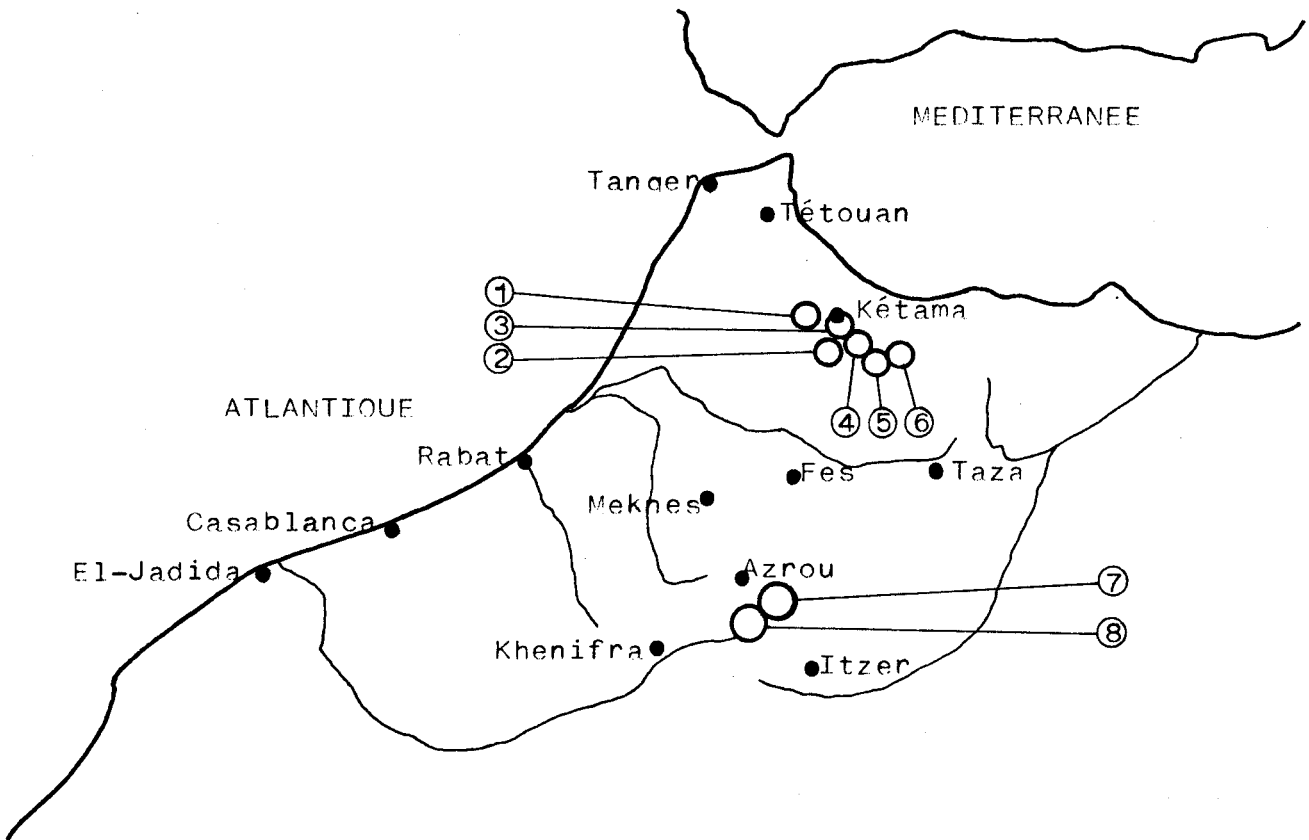


Figure 8.3. Carte des massifs forestiers étudiés  
Echelle: 1/5.000.000

① — ○ : Localisation du massif forestier n° j

Sur ces 1775 arbres répertoriés, seulement 1498 individus ont été retenus pour cette étude, répartis entre les six massifs forestiers comme l'indique la dernière colonne du tableau 8.1. En effet, il est d'usage, lors de l'établissement des tarifs de cubage, de rejeter les arbres aberrants: cassés, fourchus, à troncs jumelés, etc.... Compte tenu de la finalité de cette étude, nous avons respecté les méthodes de travail des forestiers.

Un second ensemble de données, relatives aux cèdres du Moyen Atlas, a été aimablement mis à notre disposition par la Direction des Eaux et Forêts et de la Conservation des Sols

du Ministère de l'Agriculture et de la Réforme Agraire du Royaume du Maroc. Ces données ont été recueillies, à l'occasion de travaux d'aménagement, par la Société Maroc-Développement/5/. Les mesures dendrométriques, effectuées cette fois-ci sur des arbres sur pied, concernent 627 individus répartis sur deux massifs forestiers comme l'indique le tableau 8.2. Le premier, qui englobe les forêts de Bekrit, Ajdir et une partie de celles d'Itzer et de Senoual, est situé sur une assise de roches carbonatées du secondaire que nous désignerons, sans plus de précision, sous le nom de "calcaire". Le second, qui englobe la forêt de Kerrouchen et une partie des forêts d'Itzer et de Senoual, est caractérisé par un substrat non carbonaté en grande partie constitué de "grès de Talaghine". Nous désignerons chacun de ces deux massifs par le nom de la forêt prédominante, à savoir Ajdir pour le premier et Kerrouchen pour le second/6/.

Après rejet des arbres aberrants, il reste 584 arbres utilisables dans ces deux massifs, portant ainsi à 2082 le nombre total de cèdres disponibles pour cette étude.

La carte de la figure 8.3. indique l'emplacement géographique de ces deux massifs du Moyen Atlas.

N° d'ordre affecté au massif forestier	Nom du massif forestier	Superficie (Ha)	Altitude moyenne	Bio-climat	Substrat	Taille de l'échantillon disponible
7	Ajdir	38088	1800 2000	humide et froid	Calcaire	377
8	Kerrouchen	10840	1700 1900	humide et frais	Grès de Talaghine	207

Tableau 8.2. Caractéristiques des cédraies du Moyen Atlas

VIII.4. ANALYSE DE LA FORME DES CEDRESVIII.4.1. POLYMORPHISME DES CEDRES DANS CHAQUE MASSIF FORESTIER

Les 8 massifs forestiers présentés au paragraphe précédent étant bien individualisés sur le plan écologique, la forme des cèdres a été analysée par massif.

Chaque tige, caractérisée par son coefficient de décroissance  $\delta$ , son coefficient de forme  $f$ , son angle de défilement  $\theta$ , et son coefficient d'empatement  $\varepsilon$ , est représentée par un point  $X$  tel que:

$$X = [\delta, f, \theta, \varepsilon]^T \quad (\theta \text{ en grades})$$

dans un espace  $R_0^4$  à quatre dimensions.

L'application des méthodes de classification automatique présentées dans ce mémoire à ces observations quadridimensionnelles doit permettre d'établir, si elles existent, la présence de différentes classes au sein des échantillons étudiés, chaque classe correspondant à un type de forme particulier.

Le nombre d'individus disponibles dans les échantillons provenant des massifs n° 4, 5 et 7 est suffisant, compte tenu de la dimension du problème, pour appliquer la méthode de classification du chapitre V. Par contre, le nombre limité de cèdres provenant des autres massifs ne permet pas de tester directement la convexité de la fonction de densité sous-jacente dans l'espace à quatre dimensions. Nous ferons donc appel, pour les massifs n° 1, 2, 3, 6 et 8, à la méthode décrite au chapitre VI et réservée aux échantillons de taille réduite.

Nous supposons donc implicitement que les observations suivent des lois normales et que, pour certains échantillons, les paramètres définissant les formes des tiges sont statistiquement indépendants. Plus les données se rapprocheront de ces modèles idéaux, meilleur sera le résultat de la classification. Pour cette étude, nous avons estimé, comme nous l'avons déjà mentionné, qu'il est préférable de travailler avec un modèle même très approximatif, plutôt que sans modèle. C'est pourquoi

nous avons fait appel aux méthodes de classification automatique basées sur l'analyse des mélanges gaussiens plutôt qu'aux méthodes de recherche de groupements présentées au chapitre VII .

Le premier résultat, fondamental, est la mise en évidence d'un polymorphisme marqué chez les cèdres étudiés. Trois types de forme ont ainsi été découverts dans chacun des 8 massifs forestiers considérés. Le vecteur moyenne, la matrice de covariance et la probabilité a priori de la classe correspondante à chaque type de forme sont consignés dans le tableau 8.3.

Présenté sous cette forme brute, ce résultat reste cependant d'un intérêt pratique limité. Il paraît en effet difficile, au niveau de la gestion des cédraies, de faire intervenir les 24 types de forme définis par les 24 vecteurs-moyenne du tableau 8.3. Il est cependant possible de comparer et de synthétiser les résultats obtenus dans les 8 cédraies en analysant les fluctuations des formes mises en évidence d'un massif forestier à l'autre.

#### VIII.4.2. STABILITE DE LA FORME DES CEDRES

Les vecteurs-moyenne des 3 classes mises en évidence pour chaque massif peuvent être considérés comme des prototypes des trois types de formes présents dans chaque massif.

Notons  $F_{i,j}$  ces prototypes,  $i$  étant l'indice de forme ( $i=1,2,3$ ),  $j$  étant l'indice du massif forestier ( $j=1,2,\dots,8$ ). La morphologie des arbres du jème massif est ainsi caractérisée par les trois points  $F_{i,j}$ ,  $i=1,2,3$  dont les coordonnées sont celles des trois points  $\hat{X}_{i,j}$ ,  $i=1,2,3$  .

La stabilité de la ième forme au sein de l'ensemble des massifs considérés peut être analysée en étudiant la répartition spatiale des points  $F_{i,j}$ ,  $j=1,2,\dots,8$  . Nous ferons appel, pour cette analyse, à l'algorithme de recherche de groupements du MAXIMIN / 1 /. Rappelons que cet algorithme consiste d'abord à rechercher les deux points les plus éloignés ( au sens de la distance euclidienne par exemple). Ceux-ci constituent les centres de deux premiers groupements. On affecte ensuite chacun des points

MASSIF FORESTIER N° 1		MASSIF FORESTIER N° 3	
CLASSE 1	$\hat{X}_1 = \begin{bmatrix} 0,525 \\ 0,340 \\ 1,110 \\ 0,851 \end{bmatrix}$	$\hat{\Sigma}_1 \cdot 10^4 = \begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & 9 & 0 & 0 \\ 0 & 0 & 306 & 0 \\ 0 & 0 & 0 & 35 \end{bmatrix}$	$\hat{P}_1 = 0,42$
CLASSE 2	$\hat{X}_2 = \begin{bmatrix} 0,619 \\ 0,415 \\ 0,818 \\ 0,841 \end{bmatrix}$	$\hat{\Sigma}_2 \cdot 10^4 = \begin{bmatrix} 12 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 \\ 0 & 0 & 624 & 0 \\ 0 & 0 & 0 & 51 \end{bmatrix}$	$\hat{P}_2 = 0,46$
CLASSE 3	$\hat{X}_3 = \begin{bmatrix} 0,780 \\ 0,542 \\ 0,730 \\ 0,923 \end{bmatrix}$	$\hat{\Sigma}_3 \cdot 10^4 = \begin{bmatrix} 21 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 225 & 0 \\ 0 & 0 & 0 & 31 \end{bmatrix}$	$\hat{P}_3 = 0,12$
MASSIF FORESTIER N° 2			
CLASSE 1	$\hat{X}_1 = \begin{bmatrix} 0,530 \\ 0,361 \\ 1,272 \\ 0,909 \end{bmatrix}$	$\hat{\Sigma}_1 \cdot 10^4 = \begin{bmatrix} 25 & 0 & 0 & 0 \\ 0 & 23 & 0 & 0 \\ 0 & 0 & 908 & 0 \\ 0 & 0 & 0 & 49 \end{bmatrix}$	$\hat{P}_1 = 0,31$
CLASSE 2	$\hat{X}_2 = \begin{bmatrix} 0,649 \\ 0,419 \\ 0,814 \\ 0,913 \end{bmatrix}$	$\hat{\Sigma}_2 \cdot 10^4 = \begin{bmatrix} 13 & 0 & 0 & 0 \\ 0 & 24 & 0 & 0 \\ 0 & 0 & 306 & 0 \\ 0 & 0 & 0 & 29 \end{bmatrix}$	$\hat{P}_2 = 0,46$
CLASSE 3	$\hat{X}_3 = \begin{bmatrix} 0,760 \\ 0,495 \\ 0,579 \\ 0,911 \end{bmatrix}$	$\hat{\Sigma}_3 \cdot 10^4 = \begin{bmatrix} 12 & 0 & 0 & 0 \\ 0 & 17 & 0 & 0 \\ 0 & 0 & 626 & 0 \\ 0 & 0 & 0 & 33 \end{bmatrix}$	$\hat{P}_3 = 0,23$
CLASSE 1	$\hat{X}_1 = \begin{bmatrix} 0,602 \\ 0,480 \\ 0,801 \\ 0,722 \end{bmatrix}$	$\hat{\Sigma}_1 \cdot 10^4 = \begin{bmatrix} 9 & 0 & 0 & 0 \\ 0 & 16 & 0 & 0 \\ 0 & 0 & 310 & 0 \\ 0 & 0 & 0 & 61 \end{bmatrix}$	$\hat{P}_1 = 0,33$
CLASSE 2	$\hat{X}_2 = \begin{bmatrix} 0,697 \\ 0,491 \\ 0,498 \\ 0,739 \end{bmatrix}$	$\hat{\Sigma}_2 \cdot 10^4 = \begin{bmatrix} 25 & 0 & 0 & 0 \\ 0 & 26 & 0 & 0 \\ 0 & 0 & 397 & 0 \\ 0 & 0 & 0 & 21 \end{bmatrix}$	$\hat{P}_2 = 0,26$
CLASSE 3	$\hat{X}_3 = \begin{bmatrix} 0,772 \\ 0,559 \\ 0,350 \\ 0,761 \end{bmatrix}$	$\hat{\Sigma}_3 \cdot 10^4 = \begin{bmatrix} 9 & 0 & 0 & 0 \\ 0 & 15 & 0 & 0 \\ 0 & 0 & 107 & 0 \\ 0 & 0 & 0 & 42 \end{bmatrix}$	$\hat{P}_3 = 0,41$
MASSIF FORESTIER N° 4			
CLASSE 1	$\hat{X}_1 = \begin{bmatrix} 0,626 \\ 0,469 \\ 0,653 \\ 0,852 \end{bmatrix}$	$\hat{\Sigma}_1 \cdot 10^4 = \begin{bmatrix} 17 & -2 & 16 & 1 \\ -2 & 16 & 9 & -3 \\ 16 & 9 & 221 & -1 \\ 1 & -3 & -1 & 46 \end{bmatrix}$	$\hat{P}_1 = 0,31$
CLASSE 2	$\hat{X}_2 = \begin{bmatrix} 0,704 \\ 0,493 \\ 0,552 \\ 0,869 \end{bmatrix}$	$\hat{\Sigma}_2 \cdot 10^4 = \begin{bmatrix} 7 & 4 & 11 & 0 \\ 4 & 10 & 19 & 7 \\ 11 & 19 & 153 & -9 \\ 0 & 7 & -9 & 66 \end{bmatrix}$	$\hat{P}_2 = 0,24$
CLASSE 3	$\hat{X}_3 = \begin{bmatrix} 0,751 \\ 0,578 \\ 0,340 \\ 0,788 \end{bmatrix}$	$\hat{\Sigma}_3 \cdot 10^4 = \begin{bmatrix} 16 & 5 & -9 & -2 \\ 5 & 17 & -21 & 5 \\ -9 & -21 & 149 & 7 \\ -2 & 5 & 7 & 71 \end{bmatrix}$	$\hat{P}_3 = 0,45$

Tableau F.3. (a) Résultat de la classification des cédres provenant des massifs forestiers n°1,2,3 et 4



MASSIF FORESTIER N° 5		MASSIF FORESTIER N° 7	
CLASSE 1	$\hat{X}_1 = \begin{bmatrix} 0,619 \\ 0,429 \\ 0,751 \\ 0,880 \end{bmatrix}$ $\hat{\Sigma}_1 \cdot 10^4 = \begin{bmatrix} 16 & 4 & 8 & -2 \\ 4 & 9 & 37 & 1 \\ 8 & 37 & 407 & -7 \\ -2 & 1 & -7 & 47 \end{bmatrix}$ $\hat{P}_1 = 0,32$	$\hat{X}_1 = \begin{bmatrix} 0,581 \\ 0,401 \\ 0,702 \\ 0,890 \end{bmatrix}$ $\hat{\Sigma}_1 \cdot 10^4 = \begin{bmatrix} 10 & 5 & 25 & 4 \\ 5 & 25 & 44 & -1 \\ 25 & 44 & 450 & 3 \\ 4 & -1 & 3 & 36 \end{bmatrix}$ $\hat{P}_1 = 0,35$	
CLASSE 2	$\hat{X}_2 = \begin{bmatrix} 0,696 \\ 0,482 \\ 0,602 \\ 0,838 \end{bmatrix}$ $\hat{\Sigma}_2 \cdot 10^4 = \begin{bmatrix} 7 & 3 & 6 & -2 \\ 3 & 16 & 10 & 4 \\ 6 & 10 & 98 & 2 \\ -2 & 4 & 2 & 51 \end{bmatrix}$ $\hat{P}_2 = 0,24$	$\hat{X}_2 = \begin{bmatrix} 0,685 \\ 0,472 \\ 0,560 \\ 0,885 \end{bmatrix}$ $\hat{\Sigma}_2 \cdot 10^4 = \begin{bmatrix} 7 & 5 & 16 & -12 \\ 5 & 17 & 10 & 7 \\ 16 & 10 & 321 & 14 \\ -12 & 7 & 14 & 55 \end{bmatrix}$ $\hat{P}_2 = 0,36$	
CLASSE 3	$\hat{X}_3 = \begin{bmatrix} 0,772 \\ 0,580 \\ 0,396 \\ 0,817 \end{bmatrix}$ $\hat{\Sigma}_3 \cdot 10^4 = \begin{bmatrix} 17 & 6 & 12 & 5 \\ 6 & 16 & 51 & -8 \\ 12 & 51 & 222 & 8 \\ 5 & -8 & 8 & 43 \end{bmatrix}$ $\hat{P}_3 = 0,44$	$\hat{X}_3 = \begin{bmatrix} 0,771 \\ 0,561 \\ 0,432 \\ 0,869 \end{bmatrix}$ $\hat{\Sigma}_3 \cdot 10^4 = \begin{bmatrix} 15 & 10 & 24 & -8 \\ 10 & 22 & 18 & 10 \\ 24 & 18 & 511 & -11 \\ -8 & 10 & -11 & 42 \end{bmatrix}$ $\hat{P}_3 = 0,29$	
MASSIF FORESTIER N° 6		MASSIF FORESTIER N° 8	
CLASSE 1	$\hat{X}_1 = \begin{bmatrix} 0,689 \\ 0,458 \\ 0,622 \\ 0,893 \end{bmatrix}$ $\hat{\Sigma}_1 \cdot 10^4 = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 21 & 0 & 0 \\ 0 & 0 & 219 & 0 \\ 0 & 0 & 0 & 38 \end{bmatrix}$ $\hat{P}_1 = 0,36$	$\hat{X}_1 = \begin{bmatrix} 0,640 \\ 0,451 \\ 0,649 \\ 0,863 \end{bmatrix}$ $\hat{\Sigma}_1 \cdot 10^4 = \begin{bmatrix} 25 & 0 & 0 & 0 \\ 0 & 24 & 0 & 0 \\ 0 & 0 & 610 & 0 \\ 0 & 0 & 0 & 31 \end{bmatrix}$ $\hat{P}_1 = 0,39$	
CLASSE 2	$\hat{X}_2 = \begin{bmatrix} 0,752 \\ 0,518 \\ 0,404 \\ 0,860 \end{bmatrix}$ $\hat{\Sigma}_2 \cdot 10^4 = \begin{bmatrix} 6 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 \\ 0 & 0 & 102 & 0 \\ 0 & 0 & 0 & 26 \end{bmatrix}$ $\hat{P}_2 = 0,31$	$\hat{X}_2 = \begin{bmatrix} 0,722 \\ 0,544 \\ 0,498 \\ 0,861 \end{bmatrix}$ $\hat{\Sigma}_2 \cdot 10^4 = \begin{bmatrix} 12 & 0 & 0 & 0 \\ 0 & 32 & 0 & 0 \\ 0 & 0 & 650 & 0 \\ 0 & 0 & 0 & 51 \end{bmatrix}$ $\hat{P}_2 = 0,37$	
CLASSE 3	$\hat{X}_3 = \begin{bmatrix} 0,803 \\ 0,576 \\ 0,350 \\ 0,879 \end{bmatrix}$ $\hat{\Sigma}_3 \cdot 10^4 = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 \\ 0 & 0 & 152 & 0 \\ 0 & 0 & 0 & 46 \end{bmatrix}$ $\hat{P}_3 = 0,33$	$\hat{X}_3 = \begin{bmatrix} 0,807 \\ 0,619 \\ 0,385 \\ 0,868 \end{bmatrix}$ $\hat{\Sigma}_3 \cdot 10^4 = \begin{bmatrix} 18 & 0 & 0 & 0 \\ 0 & 20 & 0 & 0 \\ 0 & 0 & 215 & 0 \\ 0 & 0 & 0 & 38 \end{bmatrix}$ $\hat{P}_3 = 0,24$	

Tableau 8.3.(b) Résultats de la classification des massifs forestiers n°5,6,7 et 8

$F_{i,j}$  restant au groupement dont le centre est le plus proche, à condition toutefois que la distance à ce centre ne dépasse pas un certain seuil, défini comme une fraction de la distance entre les deux centres les plus éloignés. Le cas échéant, si un point  $F_{i,j}$  est trop éloigné des deux centres initiaux, il constitue le centre d'un nouveau groupement et la procédure est réitérée.

La figure 8.5. représente le résultat de cette recherche de groupements sous la forme d'un graphe. Les noeuds de ce graphe représentent les points  $F_{i,j}$ ,  $j=1,2,\dots,8$ . Les longueurs des branches reliant les centres des groupements ou les points restant au centre le plus proche sont proportionnelles aux distances entre ces points dans l'espace  $R_0^4$ .

Sur la base des groupements des points  $F_{1,j}$ ,  $j=1,2,\dots,8$ , représentant le premier type de forme dans les 8 cédraies considérées, on constate qu'il apparaît deux ensembles de massifs forestiers. Le premier est constitué par les massifs n° 1 et 2 pour lesquels les formes  $F_{1,1}$  et  $F_{1,2}$  sont très voisines. Le second ensemble regroupe tous les massifs restants pour lesquels les formes  $F_{1,j}$ ,  $j=3,4,\dots,8$  sont très proches les unes des autres par rapport à l'éloignement du groupe  $\{F_{1,1}; F_{1,2}\}$  des formes du premier ensemble. (cf. figure 8.5.(a)).

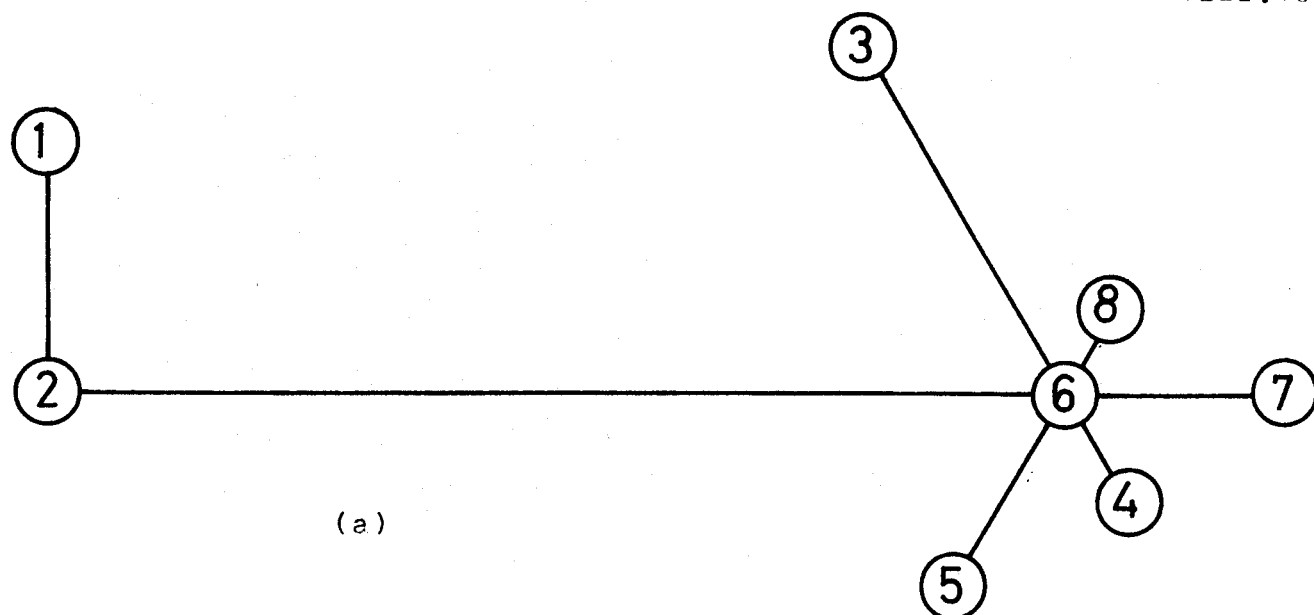
Une analyse semblable portant sur les points  $F_{2,j}$ ,  $j=1,2,\dots,8$ , représentant le second type de forme conduit aux mêmes regroupements des massifs forestiers (cf. figure 8.5.(b)).

Ce regroupement des cédraies en deux ensembles est confirmé par l'analyse de la répartition des points représentant le troisième type de forme dans les 8 massifs forestiers étudiés (cf. figure 8.5.(c)).

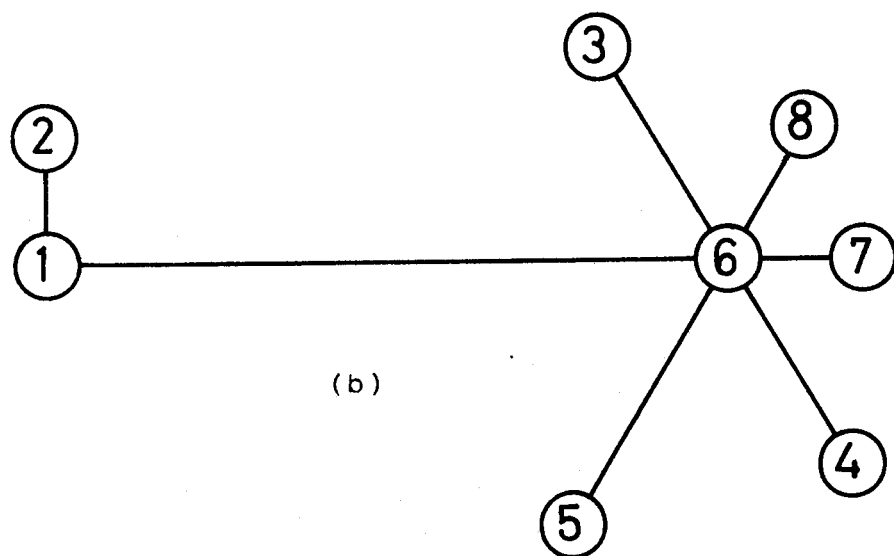
Cette analyse met en relief une certaine stabilité des trois formes mises en évidence dans le premier ensemble de massifs forestiers ( Tizirène (n°1) et Bab Chiker (2) ) d'une part, et dans le second ensemble ( Ghommara (n°3) ; Tidighine (n°4) ; Beni Khennous (n°5) ; Jbel Lerz (n°6) ; Ajdir (n°7) ; Kerrouchen (n°8) ), d'autre part.

On peut ainsi définir des formes-types pour le premier ensemble de massifs ( Ensemble A ) en calculant les barycentres  $\bar{F}_i$  des trois couples de points  $\{F_{i,1}; F_{i,2}\}$ ,  $i=1,2,3$ .

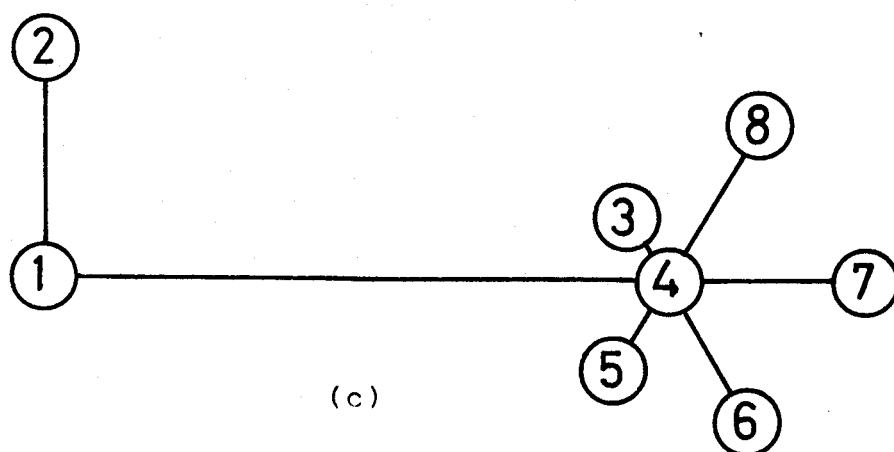




(a)



(b)



(c)

Figure 8.5. Stabilité des trois types de forme de cèdres mis en évidence dans les huit massifs forestiers étudiés.



De la même manière, le calcul des barycentres  $\bar{F}_i$  des 3 groupements de points  $\{F_{i,3} ; F_{i,4} ; F_{i,5} ; F_{i,6} ; F_{i,7} ; F_{i,8}\}$ ,  $i=1,2,3$  permet de définir trois formes-types pour le second ensemble de massifs forestiers ( Ensemble B ).

Les figures 8.6. et 8.7. représentent les formes-types ainsi mises en évidence respectivement dans les ensembles A et B .

#### VIII.5. CONCLUSION

Les fluctuations de la forme des cèdres d'un massif forestier à l'autre ne permettent pas d'envisager un traitement global pour tous les arbres répertoriés dans le Rif et le Moyen Atlas. Par contre, l'application des méthodes de classification automatique décrites dans ce mémoire permet de mettre en évidence trois types de formes d'arbres au sein de chacune des cédraies considérées.

La comparaison des résultats obtenus pour chaque cédraie montre que, sur la base de la forme des arbres, les massifs forestiers faisant l'objet de cette étude peuvent être regroupés en deux ensembles. Au sein de chacun d'eux, les arbres se répartissent en trois classes, chacune d'elle étant caractérisée par une forme-type particulière.

Cette étude de la forme des cèdres du Maroc montre que ceux-ci présentent un polymorphisme suffisamment marqué pour que lors de l'établissement des tarifs de cubage, il soit nécessaire de tenir compte des différences de forme des arbres utilisés. C'est dans ce sens que nous abordons, dans le chapitre suivant, le problème de l'échantillonnage des arbres en vue de l'amélioration de la précision et de la fiabilité des tarifs de cubage utilisés pour la gestion des cédraies du Maroc.

BIS  
LILLE

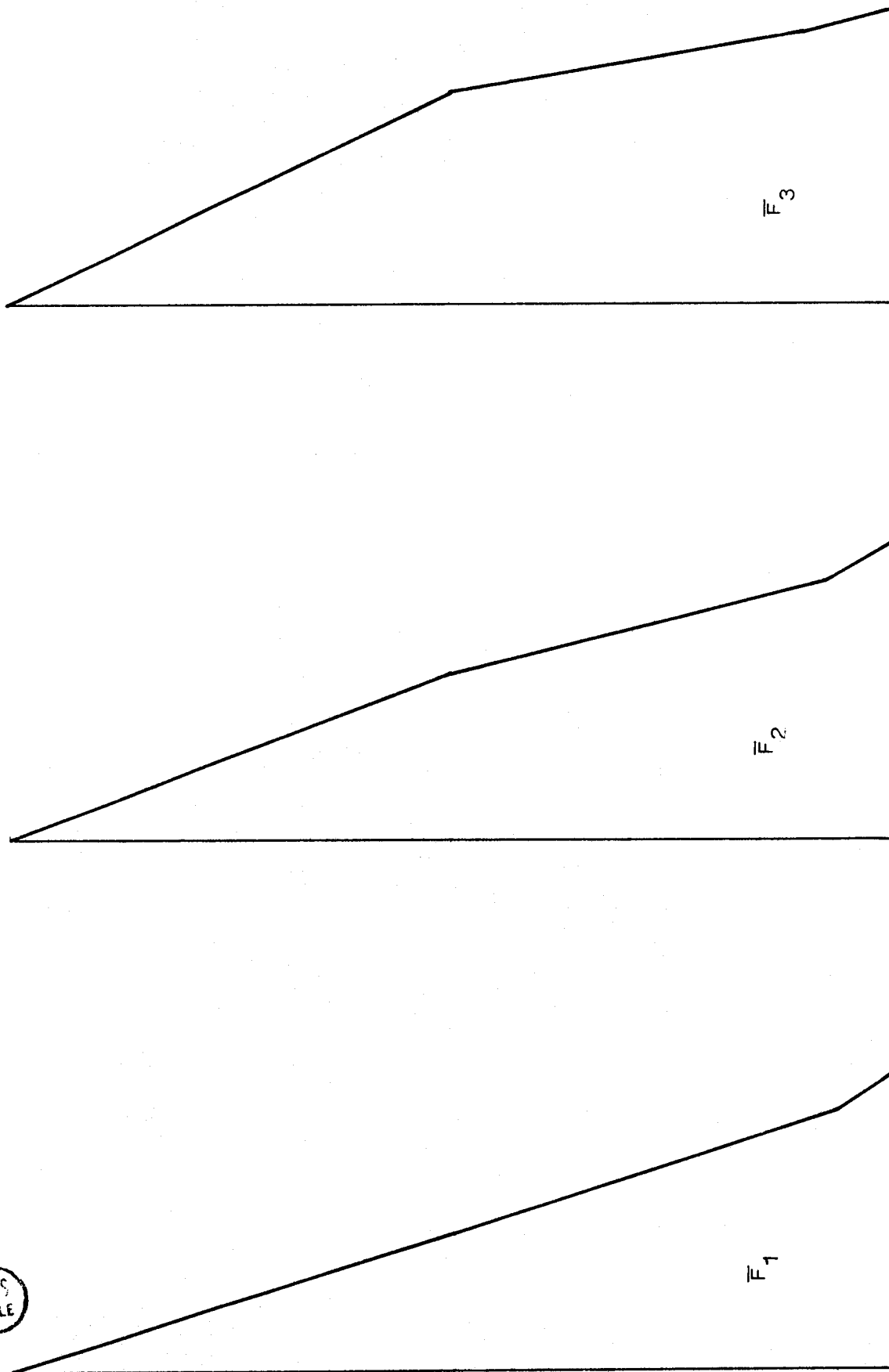


Figure 8.6. Les trois formes-types mises en évidence dans l'ensemble A des massifs forestiers ( massifs n° 1 et 2 ).

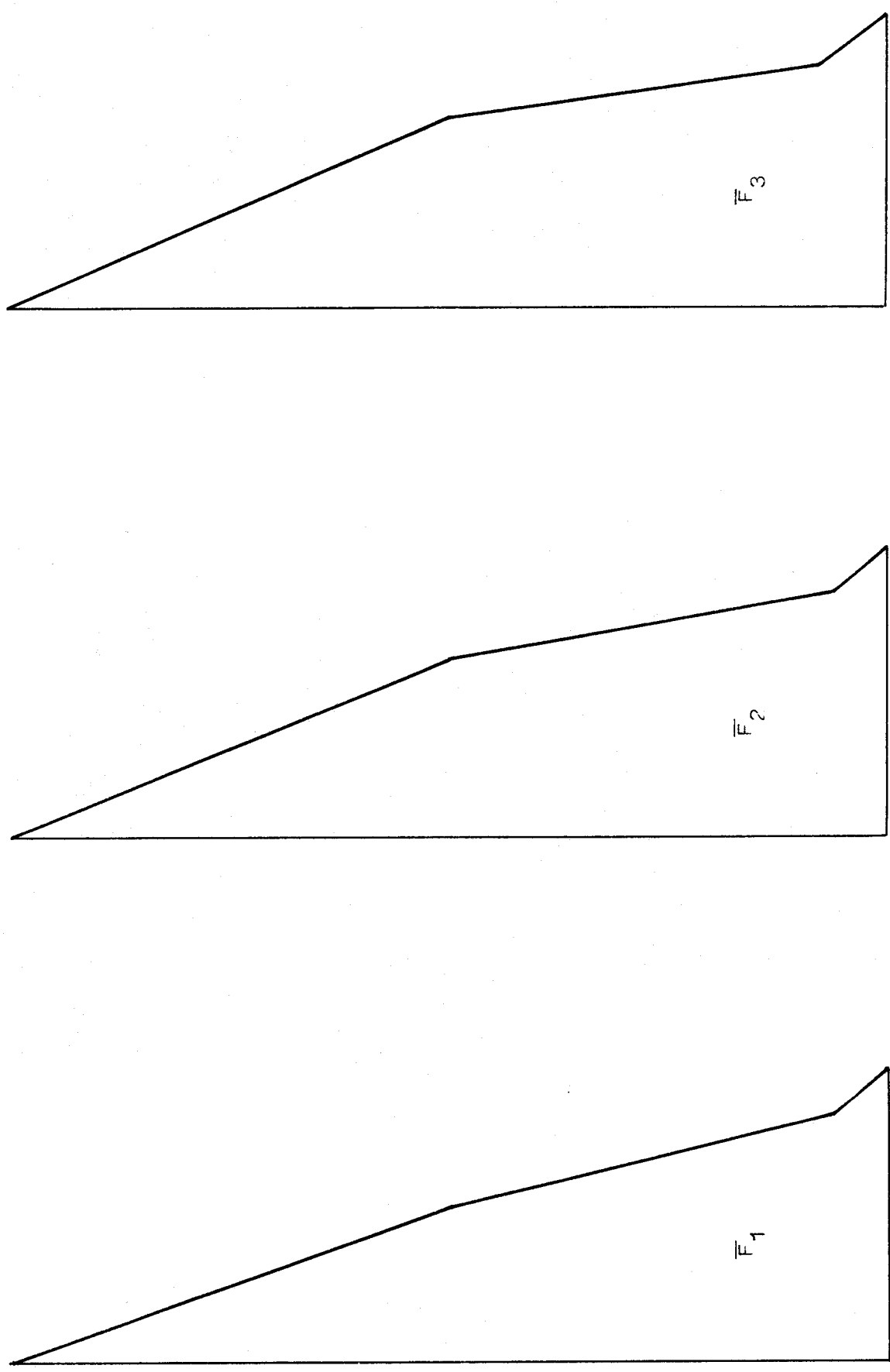


Figure 8.7. Les trois formes-types mises en évidence dans l'ensemble B des massifs forestiers (massifs n° 3, 4, 5, 6, 7 et 8).



REFERENCES (chapitre VIII)

- /1/ BATCHELOR, B.G. et WILKINS, R.R.  
 "Method for location of clusters of patterns to initialize a learning machine"  
 Electronics Letters Vol. 5, n° 20, p. 481-483 (1969)
- /2/ BENZECRI, J.P.  
 "L'allométrie: étude des variations de forme liées à la taille chez un individu, une espèce ou plusieurs espèces"  
 Les cahiers de l'analyse des données. Vol. 3, n° 3  
 Vol. 3, n° 3, p. 371-376 (1978)
- /3/ BOUCHON, J.  
 "Les tarifs de cubage"  
 E.N.G.R.E.F. Nancy (1974)
- /4/ DECOURT, N.  
 "Remarques sur la forme et la croissance des tiges"  
 Rev. Forestière Française p. 512-524 (1965)
- /5/ DIRECTION DES EAUX ET FORETS ET DE LA CONSERVATION DES SOLS  
 Division du Domaine Forestier. Royaume du Maroc  
 "Etude des travaux d'aménagement dans les cédraies de: Bekrit, Senoual, Ajdir, Kerrouchen et Itzer."  
 Rapport technique, Vol. 2 (1978)
- /6/ DIRECTION DES EAUX ET FORETS ET DE LA CONSERVATION DES SOLS  
 Division du domaine forestier. Royaume du Maroc  
 "Note sur les problèmes rencontrés lors de l'établissement de tarifs de cubage pour le cèdre"  
 Note technique N° 6704 EF/DF/A (1979)
- /7/ MENDIBOURE, P.  
 "Equation de forme: méthode-premiers résultats"  
 Ass. Forêt-Cellulose Rapport annuel p. 38-60 (1971)
- /8/ MENDIBOURE, P. et PRONIER, G.  
 "Calcul de tarifs de cubage par régression pondérée"  
 Ass. Forêt-Cellulose Rapport annuel p. 11-29 (1971)
- /9/ NASLUND, M.  
 "Stem form studies of pine in northern Sweden"  
 Swedish Univ. of Agric. Sc. Rapport n°8 (1980)
- /10/ PARDE, J.  
 "Dendrométrie"  
 E.N.E.F. Nancy (1961)

## C H A P I T R E I X

INFLUENCE DE LA FORME DES ARBRES SUR LA PRECISION DES TABLES DE CUBAGE FORESTIERES
---

### IX.1. PRECISION DES TARIFS DE CUBAGE ET ECHANTILLONNAGE

La construction des tables ou tarifs de cubage, qui permettent de passer des grandeurs mesurées sur le terrain, facilement accessibles, à l'estimation du volume d'un peuplement forestier, est certainement l'un des premiers problèmes auquel les forestiers ont été confrontés.

Pour des raisons évidentes, il est en effet exclu d'envisager la mesure directe du volume de chaque tige d'un peuplement. Pour des forêts à pénétration aisée, telles les forêts européennes, la solution généralement adoptée consiste à prélever un échantillon sur lequel il s'agit de mettre en évidence les relations statistiques qui lient le volume des tiges aux grandeurs mesurées sur les arbres. Appliquées aux arbres de tout un peuplement, ces relations doivent permettre d'estimer le volume sur pied avec précision.

Il importe donc de porter une grande attention à la manière de conduire l'échantillonnage. Il faut que les arbres retenus, formant ce que nous appellerons "l'échantillon représentatif" du peuplement, constituent un reflet aussi fidèle que possible des caractéristiques dendrométriques des arbres de la forêt à gérer. En effet, il est vain de faire appel à des outils mathématiques sophistiqués et de mettre en œuvre de puissants moyens de calcul pour construire des tarifs de cubage si les échantillons utilisés ne sont pas représentatifs des peuplements étudiés. C'est à cette question fondamentale du choix des arbres-échantillons que ce dernier chapitre est consacré. Nous allons montrer, en insistant sur l'aspect méthodologique de la démarche, que les différences de forme mises précédemment en évidence chez le cèdre du Maroc sont suffisamment importantes pour qu'il soit nécessaire de tenir compte de la morphologie des arbres lors de la construction des tarifs de cubage.

L'intérêt pratique de la méthode d'échantillonnage proposée, qui fait largement appel aux techniques de classification automatique présentées dans ce mémoire, est analysé en construisant des "tarifs d'aménagement" à une entrée, utilisés pour cuber des arbres en fonction de leur diamètre à hauteur d'homme:  $D_{130}$ . La précision de ces tarifs qui, nous le verrons, dépend de la méthode d'échantillonnage utilisée, constitue un critère objectif pour apprécier l'apport des techniques de reconnaissance des formes dans ce domaine.

## IX.2. LES TARIFS D'AMENAGEMENT

### IX.2.1. CONSTRUCTION DES TARIFS DE CUBAGE A UNE ENTREE

Pour éviter qu'un grand nombre de mesures ne les rendent inutilisables, la plupart des tarifs de cubage donnent, sous la forme de tableaux, de graphiques ou d'équations, le volume moyen des tiges d'un ensemble d'arbres en fonction d'une, deux ou exceptionnellement trois entrées. Dans cette étude, nous ne considérerons que des tarifs à une entrée : le diamètre à hauteur d'homme. Ces tarifs sont généralement désignés sous le nom de "tarifs d'aménagement" car ils sont utilisés pour estimer, d'une manière rapide, le volume des arbres dans le cadre des travaux d'aménagement des exploitations forestières.

Compte tenu des erreurs de mesure sur le terrain et des variations naturelles de la forme des tiges, il n'existe pas de relation fonctionnelle entre le volume  $V$  et le diamètre à hauteur d'homme  $D_{130}$ . La relation entre ces deux grandeurs, de type statistique, est généralement obtenue par une équation de régression, polynomiale ou logarithmique, exprimant le volume  $V$  en fonction du diamètre  $D_{130}$ .

Notre propos n'étant pas la recherche de l'équation la plus appropriée, nous nous sommes arrêté au modèle le plus couramment utilisé en France/1/, en Angleterre/6/ et qui trouve quelques adeptes en Allemagne/5/, à savoir:

$$V = a_0 + a_1 \cdot D_{130}^2 \quad (9.1.)$$

L'équation (9.1.), ajustée aux données de l'échantillon représentatif, sert au calcul du tarif de cubage à une entrée. Celui-ci est le plus souvent présenté sous la forme d'un tableau à deux colonnes: en regard des diamètres indiqués dans la première colonne, on trouve les volumes dans la seconde.

#### IX.2.2. CRITERES DE PRECISION

Notre but étant d'accroître la précision et la fiabilité des tarifs en améliorant la représentativité des échantillons, la détermination du degré de confiance qui peut être accordée à l'équation de régression calculée mérite une attention particulière. Parmi les nombreux paramètres permettant de "mesurer" la précision d'un tarif/7/, nous avons retenu:

##### a) Le coefficient de détermination

Le coefficient de détermination, carré du coefficient de corrélation de la régression, traduit la part des variations du volume  $V$  qui est "expliquée" par l'équation de régression 9.1. linéaire en  $D_{130}^2$ .

##### b) Le coefficient de variation résiduel estimé

L'écart-type résiduel estimé correspond à l'écart-type des différences entre les volumes observés et les volumes estimés des arbres-échantillons. Il permet donc de chiffrer les fluctuations du volume réel par rapport au volume déterminé par la régression. On lui préfère généralement le coefficient de variation rési-

duel, c'est à dire l'écart-type résiduel exprimé en pourcentage de la moyenne des volumes des tiges de l'échantillon.

c) Moyenne et écart-type de la distribution des erreurs de cubage

Un tarif de cubage est généralement destiné à estimer le volume des arbres du peuplement d'où l'échantillon représentatif utilisé pour le construire a été tiré. Cependant, dans cette étude, nous serons amené à appliquer des tarifs à des peuplements différents de ceux pour lesquels ils sont prévus. Dans ces conditions, les deux critères précédents sont d'un intérêt limité. En effet, ils indiquent la manière dont les équations de régression "expliquent" les variations de volume uniquement pour les arbres des échantillons représentatifs. Nous utiliserons donc un troisième critère, basé sur l'analyse des erreurs de cubage dues à l'emploi d'un tarif sur un "échantillon de test", différent de celui utilisé pour le construire / 3 /. (Cet échantillon de test pourra donc provenir du peuplement d'où a été tiré l'échantillon représentatif, ou de tout autre peuplement sur lequel on désire tester les performances du tarif).

La méthode consiste alors à déterminer, pour tous les individus de cet échantillon de test, les différences entre les valeurs observées du volume et les valeurs estimées par l'équation de régression. La moyenne  $\bar{e}$  et l'écart-type  $\sigma_e$  de la distribution de ces erreurs de cubage permet de chiffrer le biais du tarif et le degré de confiance que l'on peut lui attribuer.

Ces trois critères permettront, le moment venu, de comparer la précision et la fiabilité de tarifs obtenus à partir d'arbres-échantillons collectés selon différentes techniques.

### IX.3. TECHNIQUES D'ECHANTILLONNAGE

#### IX.3.1. CHOIX DU TYPE D'ECHANTILLONNAGE

Les mesures effectuées sur les arbres-échantillons servent à calculer les coefficients de l'équation de régression 9.1. Le tarif ainsi obtenu étant destiné au cubage des arbres de toute la population d'où a été tiré l'échantillon, le problème est donc de prélever un échantillon aussi représentatif que possible de cette population.



Lorsqu'on a peu d'information a priori sur la population étudiée, on fait couramment appel aux techniques d'échantillonnage aléatoire simple. Celles-ci consistent à tirer au hasard le nombre d'arbres désiré, généralement de manière exhaustive, c'est à dire de sorte qu'un individu ne puisse apparaître qu'une seule fois. Mais lorsque les individus d'une population plus ou moins hétérogène peuvent être classés en sous-ensembles, ou strates, homogènes, on peut espérer une meilleure représentativité de l'échantillon en utilisant la technique d'échantillonnage aléatoire stratifié/9/. En conservant le principe du tirage aléatoire simple au sein de chaque classe, le principal avantage de cette procédure est de disperser plus uniformément l'échantillon représentatif de la population étudiée.

L'échantillonnage stratifié est utilisé depuis longtemps en foresterie. Les distinctions entre les différentes strates s'y effectuent en fonction du site, du type de forêt (futaie, taillis,...), de la dimension des bois (jeunes, moyens, gros,...) ou encore de la densité du couvert/8/. Nous proposons d'utiliser cette technique de collecte des données en stratifiant l'échantillonnage des cèdres qui nous concernent en fonction des trois types de formes de tige mis en évidence au chapitre précédent. Les arbres-échantillons seront ainsi tirés à l'intérieur des trois classes associées à ces trois types de formes, avec un taux d'échantillonnage éventuellement variable selon les classes.

### IX.3.2. TAILLE DE L'ECHANTILLON

La taille de l'échantillon représentatif nécessaire à la construction d'un tarif de cubage dépend de "l'erreur à craindre" qu'on estime pouvoir accepter sur le volume estimé.

Les forestiers s'en tiennent généralement à une erreur à craindre de  $\pm 10\%$  au seuil de probabilité de 95%/8/. Les limites de confiance  $l_1$  et  $l_2$  à 95%, calculées à partir d'un volume observé dont la moyenne  $\bar{V}$  est supposée distribuée normalement sont, en négligeant la correction de population finie:

$$l_1 = \bar{V} + \frac{2 \sigma_V}{\sqrt{m}} \quad \text{et} \quad l_2 = \bar{V} - \frac{2 \sigma_V}{\sqrt{m}}$$

$\sigma_V$  étant l'écart-type des volumes observés de la population et  $m$  désignant la taille de l'échantillon représentatif.

Pour utiliser ces relations, il faut disposer d'estimations de la moyenne  $\bar{V}$  et de l'écart-type  $\sigma_V$  des volumes de la population. Celles-ci peuvent être déterminées à partir d'un échantillon obtenu par tirage aléatoire simple.

Lorsque l'échantillonnage est stratifié, il a été montré que le meilleur tirage consiste à prendre dans chaque classe  $C_i$ ,  $i=1,2,3$ , un nombre d'arbres-échantillons  $m_i$  proportionnel à  $P(C_i) \cdot \sigma_V^i$ ,  $\sigma_V^i$  étant l'écart-type des volumes dans la  $i$ ème classe et  $P(C_i)$  la probabilité a priori de la classe  $C_i$  / 4 /.

Nous allons maintenant appliquer ces principes généraux d'échantillonnage à la construction de tarifs d'aménagement destinés à l'estimation des volumes de peuplements sur pied dans les massifs forestiers du Rif.

#### IX.4. INFLUENCE DE LA STABILITE DE LA FORME DES CEDRES SUR LA PRECISION DES TARIFS DE CUBAGE

Au chapitre précédent, nous avons mis en évidence dans le Rif, sur la base de l'analyse de la forme des tiges, deux ensembles de massifs forestiers, relativement homogènes et bien différenciés. On peut se demander quel est l'intérêt pratique de séparer ainsi les forêts sur des critères de stabilité de la forme des arbres. La comparaison de la précision de tarifs spécifiques, établis pour cuber les peuplements de chacun de ces ensembles forestiers, à la précision d'un tarif unique applicable à toutes les cédraies du Rif doit permettre de justifier la séparation de ces deux ensembles.

Nous commencerons donc par construire trois tarifs, notés  $T_A$ ,  $T_B$  et  $T_C$ , sur la base de données obtenues par échantillonnage simple au sein des trois ensembles de massifs forestiers suivants:

- a) Ensemble A, comprenant les massifs forestiers  
 N° 1 : Tizirène  
 N° 2 : Bab Chiker

- b) Ensemble B , comprenant les massifs forestiers  
 N° 3 : Ghommara  
 N° 4 : Tidiqhine  
 N° 5 : Beni Khennous  
 N° 6 : Jbel Lerz
- c) Ensemble C , constitué par la réunion des ensembles A et B et comprenant donc les six cédraies étudiées dans le Rif.

Comme il est d'usage/ 2 /, le tirage simple dans chaque ensemble est effectué de manière à ce que la distribution des diamètres à hauteur d'homme pour les arbres de l'échantillon représentatif soit semblable à celle de ces diamètres pour tout l'ensemble forestier considéré. Selon une technique éprouvée/ 8 /, la distribution des diamètres des arbres au sein de chacun des ensembles A , B et C a été déterminée d'après l'inventaire d'un grand nombre de "placettes" réparties dans toutes les cédraies. (20 placettes pour l'ensemble A , 80 pour l'ensemble B et 100 pour l'ensemble C ). Dans un but de simplification des mesures sur le terrain, la plage de variation des diamètres est divisée en catégories d'égale étendue, de 5 en 5 cm. La figure 9.1. représente les trois histogrammes des fréquences de ces catégories ainsi obtenus.

Le nombre de cèdres à tirer pour constituer l'échantillon représentatif de chaque ensemble forestier est prédéterminé en fonction de la précision souhaitée (  $\pm 10\%$  au seuil de signification de 95% ) et de l'estimation de la dispersion des volumes. Pour atteindre ce nombre, on tire donc aléatoirement dans chaque catégorie de diamètre un nombre d'arbres-échantillons proportionnel à l'ordonnée de l'histogramme correspondant à cette catégorie.

Les échantillons représentatifs des ensembles forestiers ont tous été constitués à partir des arbres répertoriés dans le fichier morphométrique présenté au chapitre précédent (paragraphe VIII,3.). L'ensemble des arbres figurant dans ce fichier, après élimination des individus aberrants, est appelé "échantillon morphométrique". Tous les tirages aléatoires d'arbres répertoriés dans ce fichier ont été réalisés en utilisant un programme de génération de séquences de nombres aléatoires.

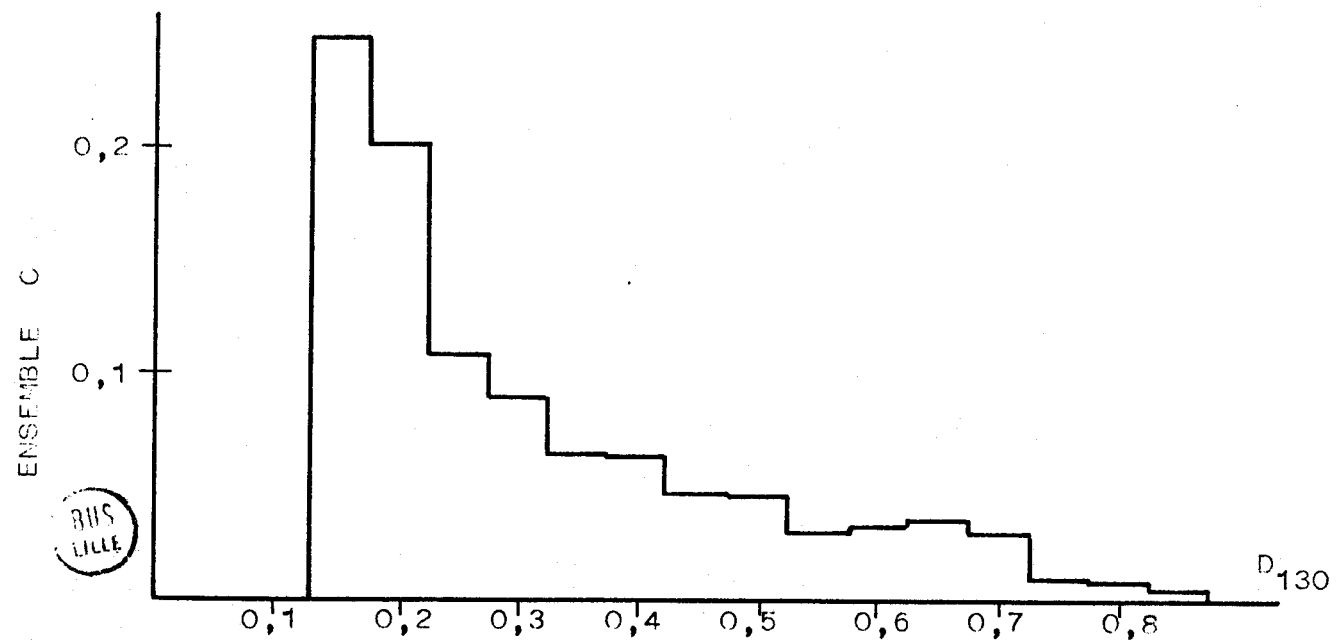
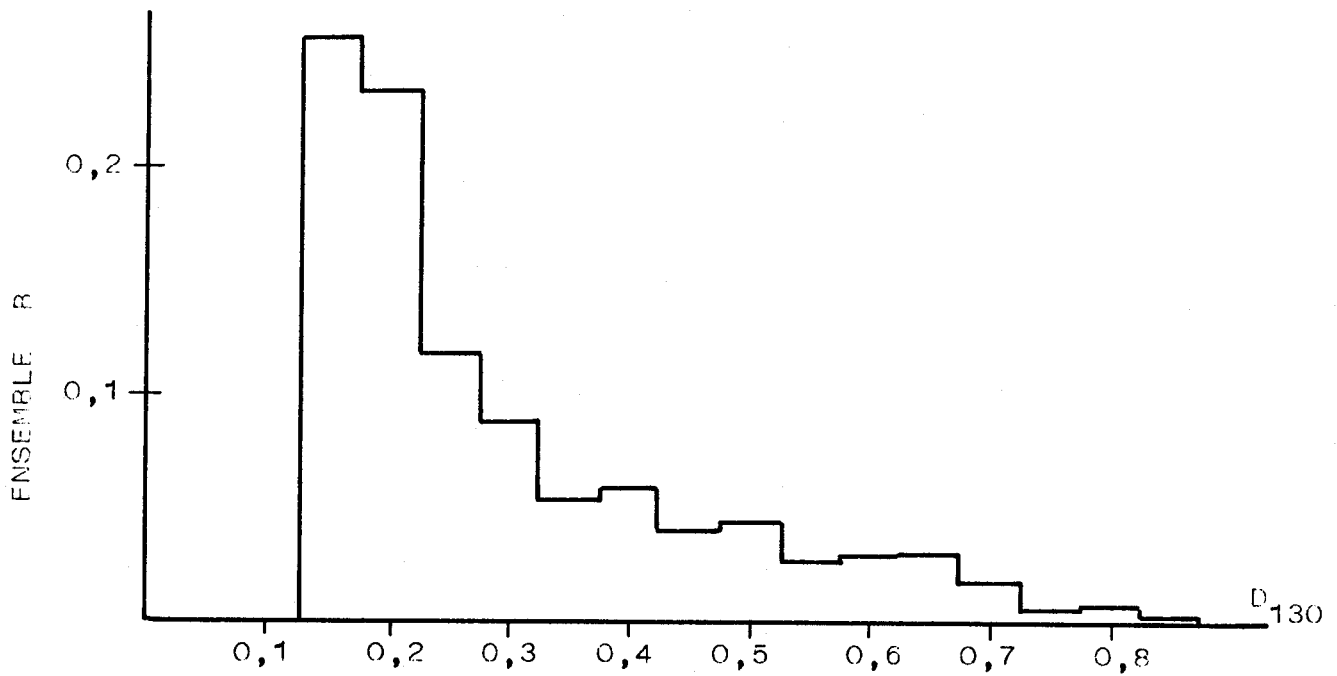
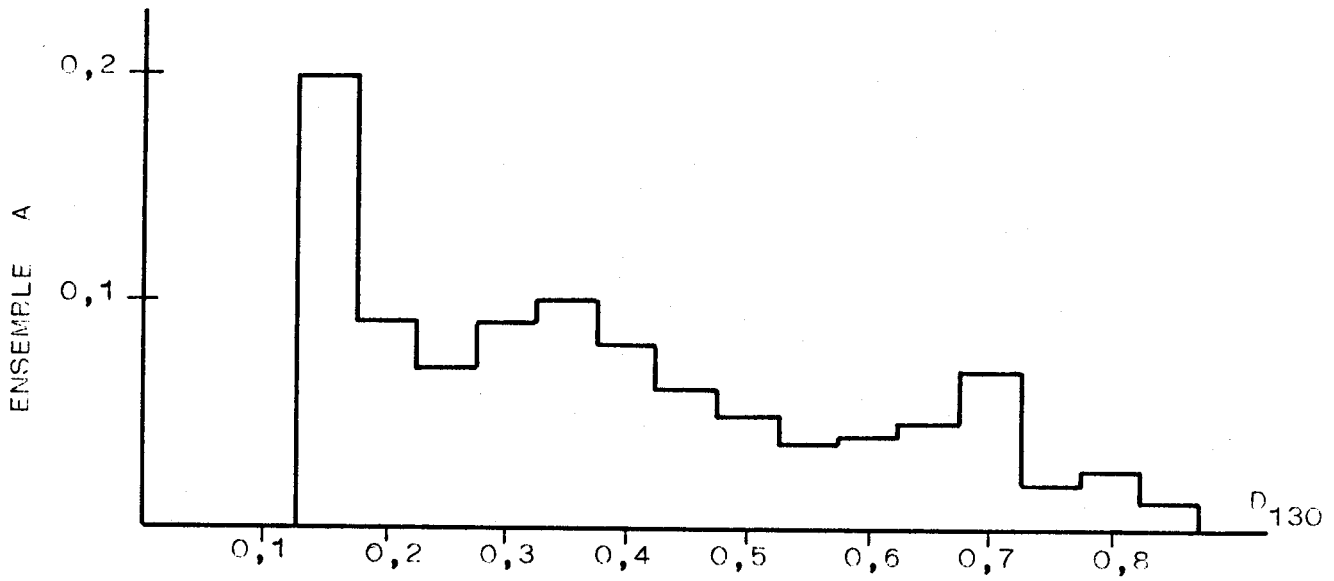


Figure 9.1. Histogrammes des fréquences des catégories de diamètre à hauteur d'homme pour les ensembles forestiers A, B et C.

Compte tenu de la faible représentation des petits bois et des gros bois dans cet échantillon morphométrique, construit à l'occasion de coupes d'exploitation, l'échantillonnage pour le tarif  $T_A$  a dû être limité aux arbres dont le diamètre à 1,30m est compris entre 22,5cm et 87,5cm. Ces deux diamètres constituent donc les limites de validité de ce tarif. Pour la même raison, dans les échantillons tirés pour la construction des tarifs  $T_B$  et  $T_C$  ne figurent que des arbres dont les diamètres sont compris entre 17,5cm et 87,5cm.

Les coefficients des trois équations de régression du type 9.1., ajustés sur les données des trois échantillons ainsi tirés des ensembles forestiers A, B et C, apparaissent dans le tableau 9.1.

On ne peut, à ce niveau, utiliser les valeurs des coefficients de détermination et des coefficients de variation résiduelle pour comparer le tarif général  $T_C$  aux tarifs  $T_A$  et  $T_B$ . En effet, ce type de critère de précision est bien adapté à la comparaison de tarifs différents, construits sur le même échantillon représentatif. Mais ici, les tarifs  $T_A$ ,  $T_B$  et  $T_C$  sont calculés à partir d'échantillons différents qui n'ont donc pas les mêmes propriétés statistiques. Il est préférable, dans ce cas, d'utiliser le critère présenté au paragraphe IX.2.2.c). Nous allons tirer de chacun des échantillons morphométriques provenant des ensembles A, B et C, plusieurs échantillons de test ne contenant que des arbres qui n'ont pas été utilisés pour construire les tarifs. Compte-tenu du nombre d'arbres disponibles dans les échantillons morphométriques, nous avons tiré quatre échantillons de test de 50 individus pour l'ensemble forestier A et quatre échantillons de test de 100 individus pour chacun des deux autres ensembles.

L'analyse de la distribution des erreurs de cubage résultant de l'application des différents tarifs à ces échantillons de test fournit un critère de précision objectif. Afin de pouvoir comparer les résultats de ces tests, la moyenne  $\bar{e}$  des erreurs de cubage et l'écart-type  $\sigma_e$  de leur distribution sont exprimés en valeurs relatives, c'est à dire en pourcentage du volume réel moyen  $V$  des arbres de l'échantillon de test considéré.

	TARIF	EQUATION DU TARIF V en m <sup>3</sup> ; D <sub>130</sub> en m	TAILLE DE L'ECHANTILLON REPRESENTATIF	COEFFICIENT DE DETERMINATION	COEFFICIENT DE VARIATION RESIDUEL
ECHANTILLONS SIMPLES	T <sub>A</sub>	$V = -0,172 + 6,582 D_{130}^2$	225	0,785	40%
	T <sub>B</sub>	$V = -0,086 + 6,416 D_{130}^2$	331	0,873	39%
	T <sub>C</sub>	$V = -0,101 + 6,320 D_{130}^2$	356	0,838	43%
ECHANTILLONS STRATIFIES	T' <sub>A</sub>	$V = -0,324 + 7,446 D_{130}^2$	Classe1: 82 Classe2: 100 Classe3: 43	0,875	33%
	T' <sub>B</sub>	$V = -0,159 + 7,507 D_{130}^2$	Classe1: 105 Classe2: 80 Classe3: 146	0,920	31%

Tableau 9.1. Comparaison des différents tarifs de cubane applicables aux cédraines du Rif



La figure 9.2. représente ces résultats de manière schématique. Pour chaque test, on a indiqué sur un axe, en valeurs relatives, le volume réel moyen des arbres de l'échantillon :  $\bar{V}$  (100%) et la valeur du volume moyen estimé :  $\frac{\bar{V} + \bar{e}}{\bar{V}} = 100\% + \frac{\bar{e}}{\bar{V}}$ . Centré sur ce dernier point, on a reporté un segment de longueur double de l'écart-type des erreurs de cubage, également exprimé en valeur relative. Les extrémités de ce segment ont donc pour abscisses :

$$100\% + \frac{\bar{e}}{\bar{V}} \pm \frac{\sigma_{\bar{e}}}{\bar{V}}$$

Le milieu de ce segment indique ainsi le biais de l'estimation des volumes et sa longueur fournit une indication précieuse quant à la fiabilité de cette estimation.

On constate, en examinant la figure 9.2., que l'application du tarif général  $T_C$  aux quatre échantillons de test issus de chacun des ensembles A, B et C donne un biais et une dispersion des erreurs de cubage toujours plus importants que lorsqu'on applique, aux mêmes échantillons, les tarifs  $T_A$  ou  $T_B$  qui leur sont spécifiquement destinés. Notons que l'application de ces deux tarifs aux échantillons de test tirés de l'ensemble C ne présente aucun intérêt.

Le tableau 9.2. indique, pour chaque ensemble forestier, les moyennes des erreurs de cubage relatives  $\bar{e}/\bar{V}$  et des écarts-types relatifs  $\sigma_{\bar{e}}/\bar{V}$  obtenus pour les quatre échantillons de test, en appliquant, d'une part le tarif général, d'autre part le tarif spécifique propre à l'ensemble considéré.

TYPE DE TARIF	ENSEMBLE A		ENSEMBLE B		ENSEMBLE C	
	erreur de cubage relative moyenne	écart type relatif moyen	erreur de cubage relative moyenne	écart type relatif moyen	erreur de cubage relative moyenne	écart type relatif moyen
TARIF GENERAL	10,8%	19,5%	11,8%	21,3%	11,8%	21,0%
TARIF SPECIFIQUE	10,0%	16,6%	8,8%	21,1%	-	-

Tableau 9.2. Comparaison du tarif général et des tarifs spécifiques

Ces résultats montrent l'intérêt pratique de construire un tarif de cubage spécifique pour chaque ensemble de massifs forestiers au sein duquel la forme des cèdres présente une relative stabilité. Nous allons maintenant montrer comment il est possible d'améliorer davantage la précision des tarifs spécifiques en tenant compte des différents types de forme présents dans chacun des ensembles forestiers A et B.

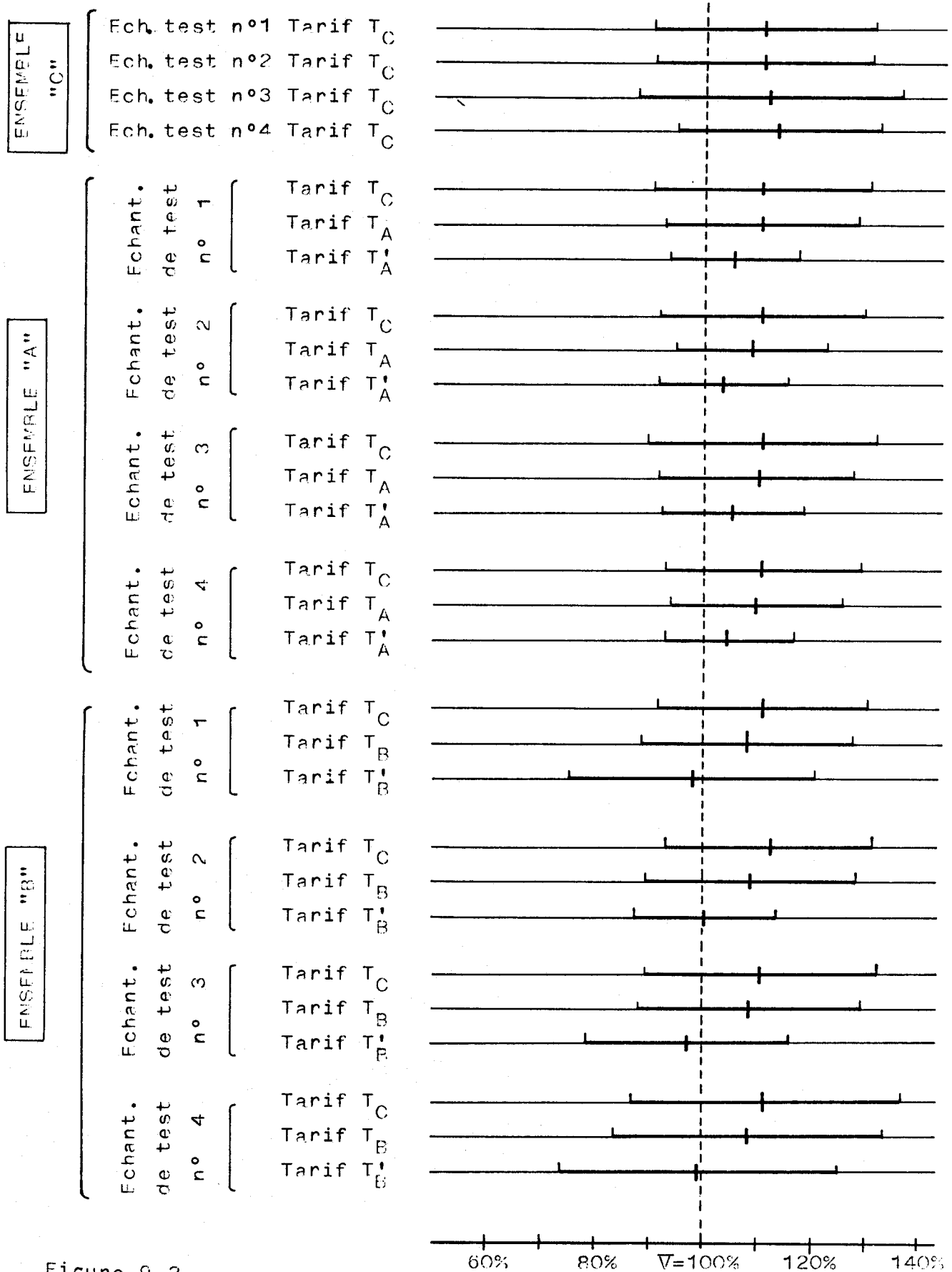


Figure 9.2.

Comparaison des performances des différents tarifs de cubage appliqués aux échantillons de test tirés des ensembles A, B et C.

Segment centré sur la valeur estimée du volume moyen des arbres de l'échantillon de test et de longueur double de l'écart-type de la distribution des erreurs de cubage (grandeurs exprimées en valeurs relatives).

BUS  
LILLE

$\bar{v}/V$   $\sigma_e/V$



IX.5. INFLUENCE DU POLYMORPHISME DES CEDRES SUR LA PRECISION  
DES TARIFS DE CUBAGE

Les deux tarifs  $T_A$  et  $T_B$  ont été obtenus sans tenir compte du polymorphisme des cèdres. Les différences de forme constatées nous suggèrent de stratifier l'échantillonnage au sein de chacun des ensembles forestiers A et B. L'échantillon représentatif est alors constitué en tirant, dans l'échantillon morphométrique de l'ensemble forestier considéré, les arbres par type de forme. Pour chaque type de forme, on prend un nombre d'individus proportionnel au produit du nombre de cèdres de ce type présents dans l'échantillon morphométrique par l'écart-type de la distribution de leurs volumes.

Un contrôle rapide des histogrammes des fréquences des catégories de diamètre à 1,30m des cèdres de chaque type de forme, pour chaque ensemble forestier, ne laisse apparaître qu'une faible corrélation entre la forme et la grosseur des arbres (cf. fig. 9.3.). L'échantillonnage est donc conduit de manière à ce que les arbres-échantillons de chaque type de forme aient des diamètres à hauteur d'homme distribués de la même manière que ceux de tout l'ensemble forestier dont ils proviennent. Pour chaque type de forme, les arbres-échantillons sont donc tirés aléatoirement par catégorie de diamètres. Le nombre d'individus tirés dans chaque catégorie est proportionnel à la valeur correspondante de l'histogramme des diamètres de l'ensemble forestier considéré (cf. figure 9.1.).

Le tableau 9.1. montre que, pour les deux ensembles forestiers A et B, les tarifs  $T'_A$  et  $T'_B$  obtenus par échantillonnage stratifié sont nettement meilleurs que les tarifs  $T_A$  et  $T_B$  construits avec les mêmes nombres d'arbres-échantillons, mais par échantillonnage simple.

Les résultats de l'application des tarifs  $T'_A$  et  $T'_B$  aux échantillons de test tirés des ensembles forestiers A et B font également ressortir cette supériorité (cf. figure 9.2.). L'intérêt économique de la stratification est évident quand on considère, pour chaque ensemble forestier, les moyennes des erreurs de cubage relatives  $\bar{e}/\bar{V}$  et des écarts-types  $\sigma_e/\bar{V}$  obtenus pour les quatre échantillons de test, en appliquant d'une part le tarif construit par échantillonnage simple, d'autre part celui résultant de l'échantillonnage stratifié (cf. tableau 9.3.).

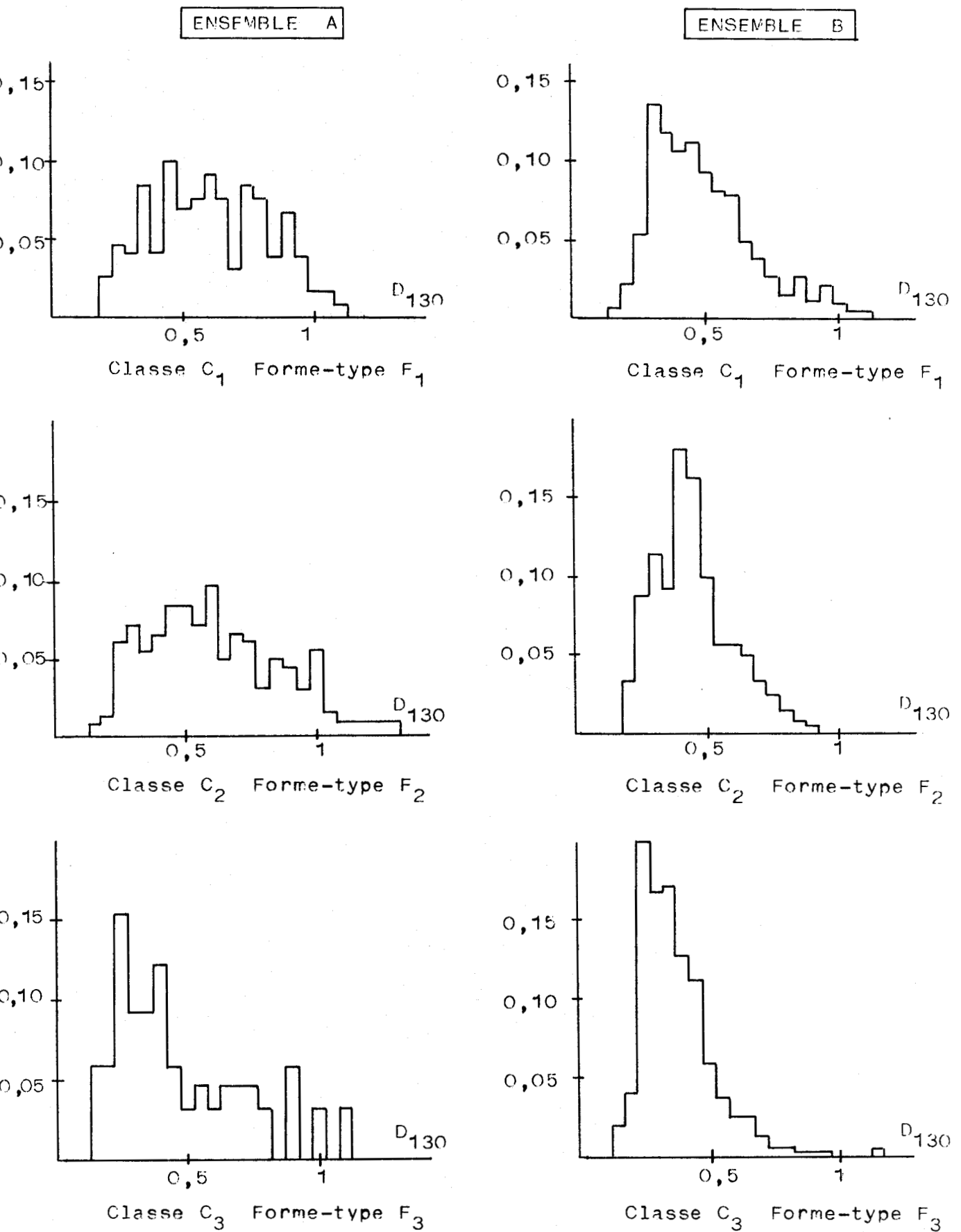


Figure 9.3. Histogrammes des fréquences des catégories de diamètres à hauteur d'homme pour les trois classes associées aux trois types de formes mis en évidence dans les deux ensembles A et B.

ECHANTILLONNAGE	ENSEMBLE A		ENSEMBLE B	
	erreur de cubage relative moyenne	écart type relatif moyen	erreur de cubage relative moyenne	écart type relatif moyen
SIMPLE	10,0%	16,6%	8,8%	21,1%
STRATIFIE	4,8%	12,1%	1,0%	20,1%

Tableau 9.3. Comparaison des tarifs obtenus par échantillonnage simple et échantillonnage stratifié.

Ces différents résultats font apparaître de manière pertinente l'intérêt pratique immédiat de la stratification de l'échantillonnage en fonction de la forme des arbres, qui apporte un gain de précision encore plus important que celui résultant de l'analyse de la stabilité des formes.

#### IX.6. CONCLUSION

Les résultats présentés dans ce chapitre mettent clairement en évidence le rôle déterminant que peuvent jouer les techniques de reconnaissance des formes et de classification automatique dans la construction des tables de cubage forestières.

Le regroupement des massifs forestiers en ensembles homogènes sur la base de l'analyse de la stabilité de la forme des arbres a d'abord permis d'obtenir, pour chacun des ensembles, des tarifs spécifiques plus précis qu'un tarif unique applicable à tous les massifs.

D'autre part, le polymorphisme des cèdres du Maroc est suffisamment marqué pour que, lors de la construction des tarifs de cubage, il soit nécessaire de stratifier l'échantillonnage en fonction du type de forme des arbres sélectionnés.

La méthodologie suivie peut paraître lourde du fait du nombre important de traitements numériques nécessaires à sa mise en œuvre (cf. figure 9.3.). Compte-tenu de l'enjeu économique de cette étude, elle se trouve toutefois amplement justifiée par les gains en précision et en fiabilité des tarifs obtenus. Le haut degré de signification des améliorations constatées permet d'être affirmatif quant à la nécessité d'utiliser, pour les cédraies du Rif, deux tarifs d'aménagement spécifiques établis sur la base d'un échantillonnage stratifié en fonction de la forme des arbres. La croissance des arbres des cédraies humides et fraîches du Rif occidental se distingue ainsi de celle des cèdres du Rif central, humide et froid.

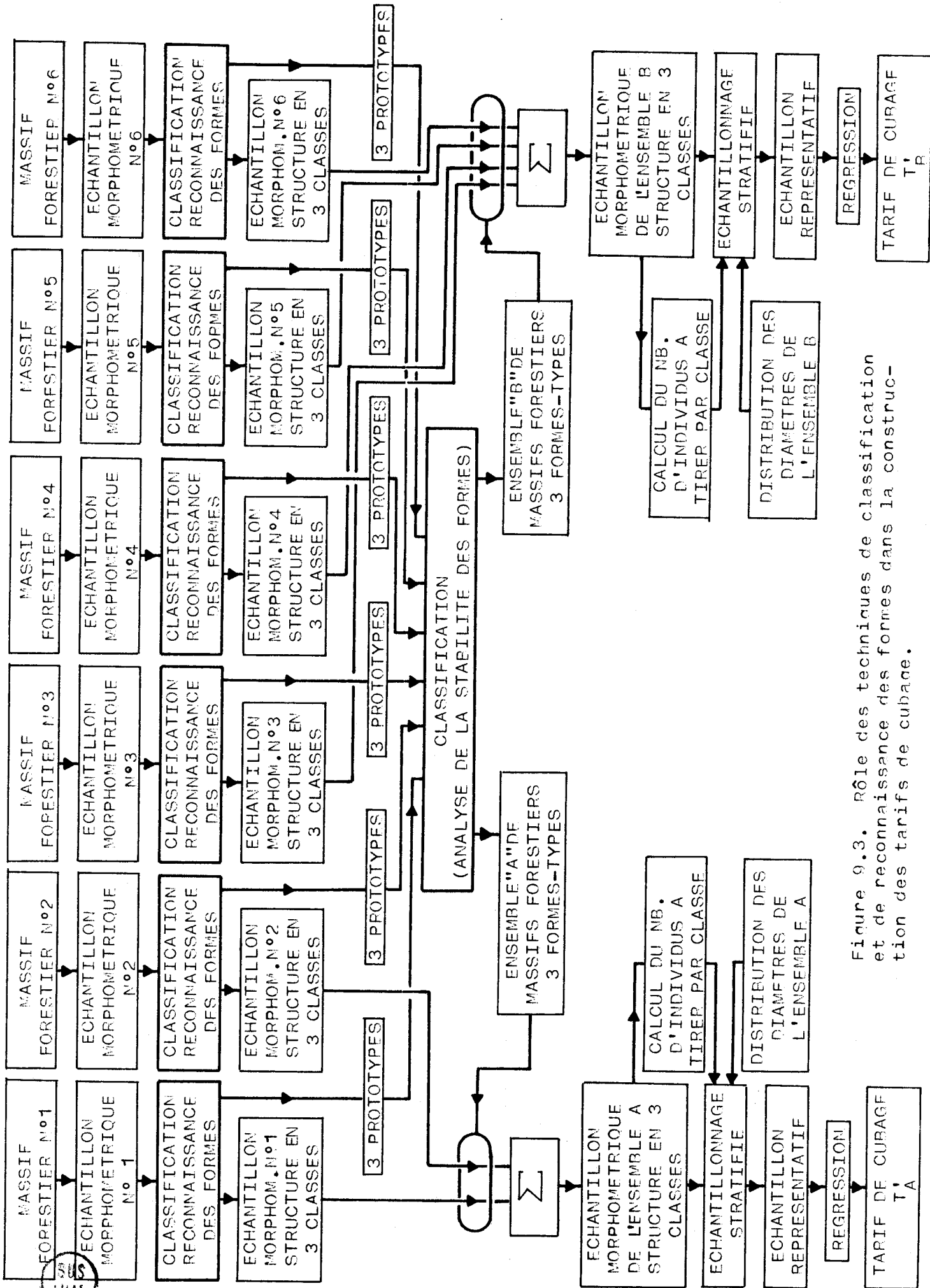


Figure 9.3. Rôle des techniques de classification et de reconnaissance des formes dans la construction des tarifs de cubage.

REFERENCES (chapitre IX)

- /1/ ABADIE, J. et AYRAL, P.  
 "Méthode de calcul du volume des peuplements sur pied dans les places d'essais de sylviculture. I- Définitions et présentation par P. Ayrat. II- Etudes mathématiques et statistiques par J. Abadie"  
 Annales ENEF p. 1-135 (1956)
- /2/ BOUCHON, J.  
 "Les tarifs de cubage"  
 ENGREF Nancy (1974)
- /3/ CAZES, P.  
 "Méthodes de régression. I- La régression sous contraintes"  
 Les cahiers de l'analyse des données Vol.3 n°2 p. 147-165 (1978)
- /4/ COCHRAN, W.G.  
 "Sampling techniques"  
 Wiley, 2nd ed. New York (1963)
- /5/ DITTMAR, O.  
 "Eine studie über die "graphischen" massener mittlungsverfahren speidelsche massenkurve und kopetzky. Gehrhardt'she massenlinie an hand einer kiefernkaahlschlag. Vollsektionierung."  
 Archiv für Forstwesen Vol. 2 p. 130-145 (1958)
- /6/ HUMMEL, F.C.  
 "The volume - basal area line - a study in forest mensuration"  
 Forestry Commission Bul. N° 24 84 p. (1955)
- /7/ PALM, R. et RONDEUX, J.  
 "A propos de deux problèmes rencontrés dans la construction de tarifs de cubage: le choix des équations de régression et l'échantillonnage des arbres"  
 Bull. Rech. Agro. Gembloux . Vol. 11, n° 1-2, p. 243-257 (1976)
- /8/ PARDE, J.  
 "Dendrométrie"  
 E.N.E.F. Nancy (1961)
- /9/ SNEDECOR, G.W. et COCHRAN, W.G.  
 "Méthodes statistiques"  
 Ass. de Coord. tech. Ann. Ch. 6 (1971)

## CONCLUSION GENERALE

Dans différents domaines scientifiques, la démarche fondamentale du chercheur consiste à recueillir des données expérimentales pour chercher ensuite à les organiser. Après l'observation et la description des données, les premiers essais d'interprétation et de compréhension consistent en leur classification en catégories se prêtant à la découverte de lois. Bien souvent, les informations collectées proviennent de plusieurs sources aléatoires et le problème de la classification des observations se trouve posé en termes statistiques.

Dans ce travail, nous avons montré comment l'introduction du concept de convexité fournit un cadre général pour aborder l'aspect statistique de la classification automatique. La même approche mathématique permet de résoudre de manière originale et efficace des problèmes d'analyse des mélanges aussi bien que des problèmes de recherche de groupements.

Servant de base à toutes les procédures de classification présentées dans ce mémoire, nous avons proposé une technique non paramétrique qui permet, à partir des observations, de déterminer localement le sens de la convexité de la fonction de densité sous-jacente. Nous plaçant dans le cas très général où l'analyste ne dispose d'aucune autre information que celle qui peut être extraite de l'échantillon qui lui est soumis, cette technique fournit une description succincte et concise de la distribution des observations en termes de convexité.

L'application de cette technique à la détection des modes de la distribution d'un ensemble d'observations pour y rechercher des groupements est immédiate. Le caractère local de l'analyse de la convexité des fonctions de densité permet de mettre en évidence des domaines modaux de forme et de taille très variées.

Mais c'est dans le cadre de l'analyse des mélanges gaussiens que l'utilisation du concept de convexité apparaît la plus féconde. Sans aucune hypothèse restrictive, l'analyse de la convexité de la fonction de densité d'un mélange à partir des observations permet d'abord de détecter les classes en présence. De l'étude des domaines où la fonction de densité est concave, on peut ensuite déterminer, pour chaque composante mise en évidence, des valeurs approchées du vecteur-moyenne, de la matrice de covariance et de la probabilité a priori. La construction de fonctions de décision sur la base de ces valeurs approchées permet ensuite d'optimiser le processus de classification. En effet, l'expérience montre que la combinaison de cette technique paramétrique aux fondements non paramétriques de la procédure conduit à des taux d'erreur très proches des taux d'erreur théoriques minimum atteignables.

L'utilisation de techniques non paramétriques pour l'analyse de la convexité des fonctions de densité a posé deux problèmes majeurs, d'ailleurs bien connus de leurs utilisateurs.

Le premier est apparu au niveau des difficultés de mise en oeuvre de la technique d'estimation par la méthode du noyau, qui, bien que présentant des avantages certains sur le plan méthodologique, conduisait à des temps de calcul prohibitifs dès que la dimension des données dépassait quelques unités. Nous avons été conduits à proposer un algorithme rapide d'estimation qui amène un gain de temps substantiel par rapport à la procédure classique. Par delà son application immédiate dans le cadre de cette étude, cet algorithme peut contribuer à redonner un nouvel intérêt à cette technique d'estimation très générale.

Le second problème que nous avons rencontré est lié à la dégradation des performances de l'estimateur lorsque la taille de l'échantillon disponible vient à diminuer. Pour traiter les problèmes de forte dimension pour lesquels on ne dispose que de

petits échantillons, nous avons été amenés à proposer des variantes des méthodes générales d'analyse des mélanges et de recherche de groupements. Les algorithmes mis au point en substituant à l'analyse de la convexité de la fonction de densité elle-même, l'analyse de la convexité de ses densités marginales, permettent ainsi à l'analyste d'aborder, par nos méthodes, la classification d'échantillons de taille et de dimension très variées.

Mais se limiter à ne présenter qu'un éventail de méthodes de classification serait éviter d'approfondir plus avant les véritables motivations des recherches en classification automatique. La simple description d'une distribution d'observations ne fournit, en aucune façon, toute l'information nécessaire pour résoudre un problème particulier.

Pour illustrer l'intérêt pratique que peuvent présenter les méthodes proposées, nous les avons utilisées pour résoudre un problème concret dans le domaine forestier, à savoir l'établissement de tables de cubage. Une méthodologie basée sur des techniques de reconnaissance de la forme des arbres a été mise au point pour construire les tarifs de cubage destinés à l'estimation du volume des peuplements forestiers. Cette nouvelle approche, qui fait largement appel aux procédures de classification automatique présentées dans ce mémoire, a été appliquée aux cèdres du Maroc. Les premiers résultats font apparaître une amélioration sensible de la précision et de la fiabilité des tarifs d'aménagement ainsi obtenus par rapport à ceux construits selon les méthodes conventionnelles.

Ces résultats, très encourageants, montrent la voie à suivre pour construire, pour le cèdre, des tarifs d'exploitation à entrées multiples en tenant compte du polymorphisme de cet arbre. L'enjeu économique d'un tel travail justifie un effort dans ce sens. Il serait également souhaitable que la méthodologie proposée pour le cèdre puisse être testée pour le cubage d'autres essences forestières.

En dehors de ce vaste champ de recherches appliquées, dans lequel de nombreuses voies restent ouvertes, ce travail laisse en suspens un certain nombre de questions qui méritent d'être approfondies.



Les méthodes d'analyse des mélanges que nous avons proposées ne sont applicables théoriquement que pour des populations normales. D'une manière générale, dans les problèmes où on ne dispose que de très peu d'information a priori sur les phénomènes mis en jeu, on est amené à juger la validité du choix d'un modèle a posteriori, par le succès ou l'échec de l'utilisation qui en est faite. C'est l'attitude que nous avons adoptée pour traiter le problème de la construction des tables de cubage forestières et les résultats obtenus justifient, à eux seuls, le choix du modèle gaussien.

Mais la statistique classique, qui a été bien concurrencée depuis l'apparition des techniques d'analyse des données et de classification automatique, propose de nombreux tests de contrôle de validité des modèles théoriques. A l'instar des statisticiens, il serait donc souhaitable de s'interroger sur la validité des résultats et des modèles en adaptant les tests de la statistique aux problèmes de la classification automatique. Il s'agit là d'un champ d'investigation étendu où on ne trouve actuellement que quelques rares travaux portant sur la validité des résultats des procédures de recherche de groupements. C'est dans cette direction, où de très nombreuses questions restent ouvertes, que nous concentrerons nos prochains efforts de recherche en classification automatique.

Les limitations des méthodes de classification faisant appel au concept de convexité viennent du fait qu'elles ne se prêtent pas à des schémas récurrents. Cependant, grâce à l'algorithme d'estimation que nous avons proposé, il est possible d'envisager une remise à jour immédiate de la description d'une fonction de densité en termes de convexité, dès qu'une nouvelle observation est disponible. La mise au point d'une telle procédure permettrait de prendre en compte l'information apportée par chaque nouvelle observation pour préciser la structure d'un mélange ou la position et la forme des domaines modaux d'une distribution. Pratiquement l'avantage d'un schéma itératif se traduirait par une réduction du temps d'exécution de la procédure et une diminution de l'encômbrement en mémoire, puisqu'il ne serait plus nécessaire de mémoriser toutes les observations.

D'une façon générale, on peut conclure que le concept de convexité apporte un nouvel outil qui s'adapte bien aux problèmes de classification automatique, particulièrement dans les problèmes où on dispose de très peu d'information a priori sur la structure des données à analyser et où la seule information disponible est celle qui peut être extraite de l'ensemble des observations.

