

Classification et discrimination : analyse discriminante typologique et applications

Yves Lemoine

▶ **To cite this version:**

Yves Lemoine. Classification et discrimination : analyse discriminante typologique et applications. Mathématiques générales [math.GM]. Université Paul Verlaine - Metz, 1979. Français. NNT : 1979METZ004S . tel-01775610

HAL Id: tel-01775610
https://hal.univ-lorraine.fr/tel-01775610

Submitted on 24 Apr 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

N° d'ordre:

THÈSE

présentée à

L'UNIVERSITÉ DE METZ

pour obtenir

LE DIPLOME DE DOCTEUR DE 3ème CYCLE

BIBLIOTHEQUE UNIVERSITAIRE - METZ
N° inv. 1979008 S
Cote S/M3 79/4
Loc Magasin

par

Yves LEMOINE

Spécialité: **Mathématiques Appliquées**

CLASSIFICATION ET DISCRIMINATION

ANALYSE DISCRIMINANTE TYPOLOGIQUE ET APPLICATIONS

Soutenue le 14 juin 1979 devant la Commission composée de:

Mme	**A.**	**SEC**	Présidente
MM.	**C.**	**CARABATOS**	
	E.	**DIDAY**	Examinateurs
	J. C.	**SIMON**	
	S.	**WATANABE**	

UNIVERSITE DE METZ BIBLIOTHEQUE

Je remercie,

Monsieur le Professeur J.C. SIMON et Monsieur le Professeur S. WATANABE de l'intérêt qu'ils ont bien voulu porter à mon travail,

Madame SEC, Maitre de conférences, pour les précieux conseils qu'elle ma prodigués,

Monsieur CARABATOS, Maître de Conférences, pour les encouragements et le soutien constant qu'il m'a apporté dans le cadre de son laboratoire à l'Université de Metz,

Monsieur DIDAY, Maitre de Conférences, qui est à l'origine de cette étude, pour son accueil à l'IRIA, et ses conseils dans la réalisation de ce travail,

Monsieur MUTEL, pour sa participation à la dernière partie de cette étude,

Les chercheurs et techniciens avec qui j'ai eu le plaisir de travailler à l'Université de Metz et à l'IRIA, pour leur aide et l'atmosphère amicale qu'ils contribuent à faire régner dans les équipes de recherche, ainsi que toutes les personnes qui ont participé à la présentation et à la publication de ce document.

PLAN

INTRODUCTION

NOTATIONS ET RAPPELS

CHAPITRE I : CLASSIFICATION SUR SOUS ESPACE ADAPTATIF

1. Introduction
   1.1. La méthode
   1.2. Les données
   1.3. Notations
   1.4. Proposition

2. Le critère de classification
   2.1. Définition
   2.2. L'optimisation du critère
   2.3. Expression mathématique du critère

3. La représentation
   3.1. Définition
   3.2. Le critère W_R

4. L'algorithme
   4.1. Introduction
   4.2. La fonction de représentation
   4.3. La fonction d'affectation
   4.4. La fonction composée f o g
   4.5. L'algorithme
   4.6. La convergence de l'algorithme

5. Le programme informatique "SESADA"
   5.1. Les entrées
   5.2. Le critère
   5.3. La fonction de représentation
   5.4. Les sorties à chaque itération
   5.5. Les sorties à la convergence

6. Exemples d'applications
   6.1. Exemple 1
   6.2. Exemple 2

7. L'interprétation des résultats
   7.1. Optimum local, optimum global du critère W
   7.2. La validité de la partition de convergence
   7.3. Le choix de la dimension r du sous espace de classification

8. Méthode séquentielle relativement au nombre de classes de la partition
   8.1. Pas 1
   8.2. Pas 2
   8.3. L'algorithme

9. Conclusion

CHAPITRE II : ANALYSE DISCRIMINANTE TYPOLOGIQUE

1. Introduction
   1.1. Le problème de la discrimination
   1.2. L'analyse factorielle discriminante : Une analyse descriptive linéaire
   1.3. L'analyse discriminante décisionnelle linéaire
   1.4. La remise en cause des classes a priori
   1.5. La prise en compte des classes a priori
   1.6. L'analyse discriminante typologique
   1.7. Les données
   1.8. Notations

2. Le critère
   2.1. Définition
   2.2. L'optimisation du critère
   2.3. Le critère et l'inertie inter classe expliquée
   2.4. Expression mathématique du critère
   2.5. Le critère et les facteurs discriminants

3. La représentation
   3.1. Définition
   3.2. Le critère W_R

4. L'algorithme
   4.1. Introduction
   4.2. La fonction de représentation
   4.3. La fonction d'identification
   4.4. La fonction composée f o g
   4.5. L'algorithme
   4.6. La convergence de l'algorithme

5. Le programme informatique "DISCRI"
   5.1. Les entrées
   5.2. Le critère
   5.3. La fonction de représentation
   5.4. Les sorties à chaque itération
   5.5. Les sorties à la convergence

6. Le traitement d'une banque de données
   6.1. Le problème
   6.2. Les types définis a priori
   6.3. La population échantillon : la partition définie a priori
   6.4. L'analyse discriminante typologique relative à la partition a priori
   6.5. La seconde analyse
   6.6. La troisième analyse
   6.7. L'échantillon test
   6.8. Conclusion

7. L'interprétation des résultats
   7.1. La définition de nouvelles familles
   7.2. La règle de décision
   7.3. Le second niveau de discrimination
   7.4. Exemples

8. Conclusion

CHAPITRE III : L'ANALYSE DISCRIMINANTE TYPOLOGIQUE SOUS CONTRAINTE DE MODELISATION LOCALE

1. Introduction
   1.1. Le problème
   1.2. Les données
   1.3. Exemples de modélisation
   1.4. La méthode
   1.5. Notations

2. Le critère
   2.1. Le critère W_1
   2.2. Le critère W_2
   2.3. Le critère W
   2.4. L'optimisation du critère W
   2.5. Expression mathématique du critère W_1
   2.6. Expression mathématique du critère W_2

3. La représentation
   3.1. Définition
   3.2. Le critère W_R

4. L'algorithme
   4.1. Introduction
   4.2. La fonction de représentation
   4.3. La fonction d'affectation
   4.4. La fonction composée f o g
   4.5. L'algorithme
   4.6. La convergence de l'algorithme

5. Le programme informatique "DISCRIML"
   5.1. Les sous programmes relatifs à la modélisation
   5.2. Les entrées
   5.3. Construction d'une partition
   5.4. Le critère
   5.5. La fonction de représentation
   5.6. Les sorties à chaque itération
   5.7. Les sorties à la convergence

6. L'interprétation des résultats
   6.1. Le choix de la meilleure analyse après plusieurs tirages
   6.2. La validation de la partition de convergence
   6.3. Une procédure conversationnelle pour la recherche du compromis "Discrimination-Modélisation"
   6.4. La règle de décision
   6.5. Le renforcement de la représentation

7. Applications
   7.1. Un exemple
   7.2. Application à la reconnaissance des lois d'effort de coupe

8. Conclusion

CONCLUSION GENERALE ET PERSPECTIVES

Annexe 1 : Inertie d'un nuage de points

Annexe 2 : Inertie expliquée par un sous espace

Annexe 3 : Inertie intra classe et inter classe d'une partition

Annexe 4 : Inertie intra classe et inter classe expliquée par un sous espace

Annexe 5 : Proposition

Annexe 6 : Proposition

Sauf mention contraire, les références sont internes au chapitre étudié.

INTRODUCTION

Ce travail se situe dans le domaine de l'Analyse des Données, et plus précisément dans le cadre suivant : décrire, réduire, classer des observations multidimensionnelles dans le but de définir, préciser, expliquer ou reconnaître une structure sur un ensemble d'objets.

On se place successivement dans le cadre de la classification automatique (chapitre I), discrimination (chapitre II), modélisation (chapitre III) :

La classification automatique permet de définir une structure (partition, hiérarchie, classes empiétantes ...) sur une population si on dispose d'une description multidimensionnelle des objets, et si on a fait le choix d'une mesure de ressemblance qui exprime la similitude entre objets. Dans le cas de la recherche d'une partition, on désire optimiser la ressemblance entre objets d'une même classe, et (ou) les distinctions entre objets de classes différentes [8]. On définit dans ce but un critère qui exprime l'homogénéité des classes, et (ou) l'écart entre classes. Le problème est ainsi posé en termes d'optimisation d'un critère [9].

En discrimination et reconnaissance des formes, la structure, une partition en K familles, $\mathcal{C}_1,\ldots,\mathcal{C}_K$, est admise a priori. On étudie une population échantillon pour laquelle, outre une description multidimensionnelle, le groupe d'appartenance des objets est connu. On distingue :

- L'analyse discriminante à but descriptif. Il s'agit de décrire le pouvoir séparateur des paramètres vis à vis des familles définies a priori

- l'analyse discriminante à but décisionnel : il s'agit de définir une règle de classement dans l'une des familles $\mathcal{C}_\ell$ ($1 \leq \ell \leq K$) de tout individu dont on connait la description. La validité de la règle de décision doit ensuite être étudiée.

Si on définit une variable qualitative Y, en associant à chaque famille $\mathcal{C}_\ell$ une modalité Y_ℓ ($\ell=1,\ldots,K$), l'analyse discriminante est l'étude de l'existence et de la forme de la dépendance entre Y (la variable à expliquer), et les paramètres descriptifs des individus (les variables explicatives). Cette approche conduit à considérer l'analyse discriminante linéaire comme un cas particulier de l'analyse canonique ($q \geq 1$ variables à expliquer) [4]

La modélisation de la dépendance entre q_2 variables de sorties et q_1 paramètres d'actions d'un processus se décompose en deux étapes :

- Choix d'une famille H de modèles

- Recherche du modèle h de H le mieux adapté à la population échantillon dont on dispose, au sens d'un indice d'homogénéité.

Si le processus dépend également de paramètres d'entrée $X^1,\ldots,X^p$, on doit restreindre la modélisation à un nombre élevé de sous populations (si on définit t modalités pour chaque variable X^i, on obtient t^p configurations).

L'Analyse Discriminante Typologique a été introduite par E. DIDAY dans [10]. J'ai développé l'étude de cette méthode, qui est présentée au chapitre II. J'ai d'autre part proposé et mis en oeuvre deux extensions de l'Analyse discriminante typologique :

- La classification sur sous espace adaptatif (Chapitre I)
- L'analyse discriminante typologique sous contrainte de modélisation locale (Chapitre III).

La structure à définir (préciser ou expliquer) est une partition de la population échantillon E. On procède par optimisation d'un critère W.

A toute partition $P = (P_1,\ldots,P_K)$ de E est associée une représentation L ; L est unique pour l'ensemble des classes $P_1,\ldots,P_K$. (Une autre approche consiste à définir une représentation sous la forme d'un K-uple $L_1,\ldots,L_K$; L_i est associée à la classe P_i)

Inversement à toute représentation L est associée une partition P de E. Un critère W_R mesure l'adéquation entre une partition et une représentation (W_R généralise W).

Un algorithme itératif converge en un nombre fini d'itérations vers une partition P^* qui réalise un optimum local du critère W.

La représentation L^* associée à P^* est une image de la structure construite. (On obtient en particulier la visualisation de la partition P^* sur quelques axes factoriels).

Les trois méthodes diffèrent par le type de représentation, le mode d'initialisation de l'algorithme, et l'interprétation des résultats qui sont liés au contexte d'application.

Afin de permettre une lecture indépendante des trois parties, chaque méthode est présentée intégralement dans un chapitre. Des définitions et propriétés concernant l'inertie d'un nuage de points sont résumées en avant propos (une présentation détaillée et les démonstrations des résultats sont données aux annexes 1, 2, 3, et 4). Le chapitre I est intitulé

Classification sur sous espace adaptatif

On sait que quelques axes factoriels suffisent à expliquer une très large part de l'inertie d'un nuage de points [4 , 16]. La distinction entre "variables classifiantes" et "variables de bruit", suggère la recherche d'une partition dont les classes soient séparées en projection sur un sous espace

de l'espace des individus. Cette approche est différente de celle de l'analyse factorielle typologique [22] dont le critère à optimiser exprime la somme des inerties des classes expliquées par leur sous espace principal. Ici le sous espace est unique et défini par l'ensemble des classes.
Au chapitre II est présenté

L'analyse discriminante typologique

Située dans le contexte de la discrimination, l'analyse discriminante typologique permet de remédier à une définition subjective, arbitraire ou imprécise des types a priori. La partition définie a priori sur la population échantillon est remise en cause par optimisation d'un critère qui ne tient compte que de la projection des individus sur les premiers axes de l'analyse factorielle discriminante [25 , 29]. Les classes a priori sont ainsi privilégiées dans la classification ce qui apparaît souhaitable; Or, en général, lorsque la partition a priori est contestée, on effectue une analyse inductive (analyse en composantes principales [2], ou des correspondances [1], ou classification) qui tient compte au même titre des paramètres discriminants (ou partiellement discriminants) et des variables de bruit. En particulier, l'approche de l'analyse discriminante typologique est différente de la suivante : Construire une partition, puis réaliser une analyse discriminante relative à celle-ci .

La méthode est non seulement descriptive, mais aussi constructive : En général, la partition induite par l'analyse discriminante typologique fait apparaître des regroupements de classes a priori ou (et) l'éclatement de certaines en groupes plus homogènes, ou (et) la remise en cause de frontières qui les séparent. On en déduit la définition de nouvelles familles interprétables par le spécialiste, constituant une partition de la population, et bien reconnues par quelques facteurs discriminants (une règle de décision numérique complétée par une visualisation est construite).

Les regroupements éventuels de familles a priori peuvent être acceptables du point de vue du spécialiste (cf. le traitement d'une banque de données industrielles); Sinon on réalise localement un second niveau de discrimination. La règle de décision globale se présente alors sous la forme d'un arbre.

En reconnaissance des formes [28] l'analyse discriminante typologique permet l'ajustement éventuel de leur définition, et (ou) leur regroupement, dans le but de rendre simple et fiable leur reconnaissance.

Le traitement d'une banque de données industrielles (contrôle automatique de production en fabrication mécanique) est présenté dans ce chapitre. Cet exemple illustre les aspects descriptif, inductif, et décisionnel de la méthode. Enfin le chapitre III concerne

L'analyse discriminante typologique sous contrainte de modélisation locale

Cette méthode se situe dans le contexte d'une approche locale du problème de la modélisation . De plus dans le cas d'une grande dispersion des données relatives à la modélisation [19], la recherche simultanée de facteurs discriminants dans un ensemble de paramètres explicatifs permet d'assurer la signification physique de la structure construite. (Dans "Sélection typologique de paramètres" [10] il est tenu compte de variables explicatives de chaque classe, et non de la partition). De plus, si la modélisation locale doit être suivie d'une analyse discriminante, il est préférable que la partition soit construite par optimisation d'un critère qui tienne compte simultanément de la modélisation et de la discrimination.

On étudie plus particulièrement :

- le cas du modèle linéaire : les q_1 variables d'actions $z_1,\ldots,z_{q_1}$, et les q_2 variables de sorties $z_{q_1+1},\ldots,z_{q_1+q_2}$ sont liées par une relation du type

$$\alpha_o + \alpha_1 z_1 + \ldots + \alpha_q z_q = 0 \qquad (q=q_1+q_2)$$

On présente une application en modélisation des efforts de coupe en usinage.

- Le cas du modèle de l'inertie [10] : On dispose uniquement de q variables de sorties $z_1,\ldots,z_q$. Le modèle est de la forme :

$$(z_1 = a_1 \; ; \ldots ; \; z_q = a_q)$$

où $a_1,\ldots,a_q$ sont q constantes réelles.

NOTATIONS ET RAPPELS

Dans cette partie :

- On fixe certaines notations (les définitions précises sont rappelées aux annexes 1, 2, 3 et 4.).

- On résume quelques propriétés mathématiques de l'inertie d'un nuage de points, (inertie, inertie expliquée, inertie intra ou inter classe), qui sont utilisées aux chapitres I, II et III. Ces propriétés sont démontrées aux annexes 1, 2, 3 et 4, ou, par exemple, dans [4].

- Enfin, on analyse un problème d'optimisation qui se présente, sous des formes différentes, aux chapitres I, II, III. On propose une solution précise à ce problème.

1) L'espace vectoriel $\mathbb{R}^p$ (Annexe 1)

Il est muni du produit scalaire défini par une matrice Q, symétrique et définie positive, et de la distance d associée. (Dans la suite, toute inertie est définie relativement à cette métrique).

On note

F_r l'ensemble des sous espaces vectoriels de dimension r de $\mathbb{R}^p$.

U_r l'ensemble des familles $\underline{U} = (u_1,\ldots,u_r)$ Q orthonormées de r vecteurs de $\mathbb{R}^p$.

$$u_s' Q u_s = 1 \qquad s = 1,\ldots,r$$

$$u_s' Q u_t = 0 \qquad s = 1,\ldots,r \quad t = 1,\ldots,r \quad s \neq t$$

2) Inertie au point a d'un nuage de points (Annexe 1)

Soit

N un nuage de points dans $\mathbb{R}^p$

V la matrice d'inertie du nuage N

a un point de $\mathbb{R}^p$

On note :

I (a,N) l'inertie au point a du nuage N.

L'inertie I (a,N) est minimale au centre de gravité g du nuage N (Théorème de Huyghens) ; I (g,N) est appelée plus simplement l'inertie du nuage N : On la note I (N)

On a la relation :

$$I(N) = Tr(VQ)$$

3) Inertie expliquée par un sous espace (Annexe 2)

Soit, de plus,

U un sous espace vectoriel de $\mathbb{R}^p$.

On note

I (a,N,U) l'inertie au point a du nuage N, expliquée par le sous espace U.

Pour tout sous espace U, l'inertie I (a,N,U) est minimale au centre de gravité g du nuage N. I (g,N,U) est appelée plus simplement l'inertie du nuage N, expliquée par le sous espace U ; On la note I (N,U).

Pour tout point a, tout sous espace U dans F_r, et toute décomposition de U en somme directe de droites vectorielles Q orthogonales, $D_1, \ldots, D_r$, on a :

$$I(a,N,U) = \sum_{s=1}^{r} I(a,N,D_s) \qquad (0\text{-}1)$$

En particulier :

$$I(N,U) = \sum_{s=1}^{r} I(N,D_s) \qquad (0\text{-}2)$$

L'inertie du nuage N, expliquée par une droite vectorielle D engendrée par le vecteur Q-normé u, s'écrit :

$$I(N,D) = u' Q V Q u \qquad (0\text{-}3)$$

4) Inertie intra classe et inter classe d'une partition (Annexe 3)

On note :

$\mathcal{P}_k$, l'ensemble des partitions du nuage N en K classes

Soit

$a_1,\ldots,a_k$, K points de R^p

$P = (P_1,\ldots,P_k)$ dans $\mathcal{P}_K$.

On note :

$I((a_1,\ldots,a_k), P)$, l'inertie aux points $a_1,\ldots,a_k$ de la partition $P = (P_1,\ldots,P_k)$. (Par définition, $I((a_1,\ldots,a_k), P) = \sum_{\ell=1}^{k} I(a_\ell, P_\ell)$).

L'inertie $I((a_1,\ldots,a_k), P)$ est minimale pour la famille $(g_1,\ldots,g_k)$ des centres de gravité des classes $P_1,\ldots,P_k$ de la partition P ; $I((g_1,\ldots,g_k), P)$ est appelée l'inertie intra classe de la partition P : On la note $I_{intra}(P)$

Soit

P une partition du nuage N

B (P) la matrice d'inertie inter classe de la partition P

On note :

$I_{inter}(P)$, l'inertie inter classe de la partition P

On a les relations :

$$I_{inter}(P) = Tr(B(P)\,Q) \qquad (0\text{-}4)$$

$$I_{intra}(P) + I_{inter}(P) = I(N) \qquad (0\text{-}5)$$

(L'inertie du nuage N se décompose en inerties intra classe et inter classe)

5) Inertie intra classe et inter classe expliquée par un sous espace (Annexe 4)

Soit

$P = (P_1,\ldots,P_k)$ dans $\mathcal{P}_k$

$a_1,\ldots,a_k$, k points de $\mathbb{R}^p$

U un sous espace vectoriel de $\mathbb{R}^p$

On note :

$I((a_1,\ldots,a_k), P, U)$, l'inertie aux points $a_1,\ldots,a_k$ de la partition $P = (P_1,\ldots,P_k)$, expliquée par le sous espace U. (on a : $I((a_1,\ldots,a_k), P, U) = \sum_{\ell=1}^{k} I(a_\ell, P_\ell, U)$).

Pour tout sous espace U, l'inertie $I((a_1,\ldots,a_k), P, U)$ est minimale pour la famille $(g_1,\ldots,g_k)$ des centres de gravité des classes $P_1,\ldots,P_k$ de la partition P ; $I((g_1,\ldots,g_k), P, U)$ est appelée l'inertie intra classe de la partition P expliquée par le sous espace U : on la note $I_{intra}(P,U)$

On note d'autre part :

$I_{inter}(P,U)$ l'inertie inter classe de la partition P expliquée par U.

On a la relation :

$$I_{intra}(P,U) + I_{inter}(P,U) = I(N,U) \qquad (0\text{-}6)$$

(L'inertie du nuage N, expliquée par U, se décompose en inerties intra classe et inter classe, expliquées par U).

6) Expressions mathématiques de l'inertie inter classe et de l'inertie aux points $a_1,\ldots,a_k$, d'une partition, expliquées par un sous espace

Soit

$P = (P_1,\ldots,P_k)$ une partition du nuage N

$g_1,\ldots,g_k$ les centres de gravité des classes $P_1,\ldots,P_k$

$p_1,\ldots,p_k$, les poids des classes $P_1,\ldots,P_k$ définis par :

$$p_\ell = \frac{\text{Nombre d'éléments de } P_\ell}{\text{Nombre d'éléments de N}} \qquad \ell = 1,\ldots,K$$

La matrice d'inertie inter classe de la partition P, B(P), est par définition, (A3. 3.2), la matrice d'inertie du nuage $G = [(g_1,\ldots,g_k),(p_1,\ldots,p_k)]$ des centres de gravités des classes munis des poids $p_1,\ldots,p_k$. On montre (A3.1) que les nuages N et G sont le même centre de gravité g_o. Ainsi :

$$B(P) = \sum_{\ell=1}^{k} (g_\ell - g_o)(g_\ell - g_o)'$$

Proposition 1

Pour tout U dans F_r, et toute base Q orthonormée $\underline{U} = (u_1,\ldots,u_r)$ de U l'inertie inter classe la partition P, expliquée par le sous espace U, s'écrit :

$$I_{inter}(P,U) = \sum_{s=1}^{r} u_s' \; Q \; B(P) \; Q \; u_s \qquad (0\text{-}7)$$

Démonstration

On utilise :

- La décomposition de l'inertie du nuage G expliquée par U en inerties du nuage G expliquées par les droites vectorielles engendrées par $u_1,\ldots,u_r$ (cf. 0-2).

- Puis, pour s = 1,...,r, la relation (0-3) appliquée au nuage G (de matrice d'inertie B (P))

Définition 1

Soit U un sous espace vectoriel de R^p.

Pour tout couple (x,y) de points de R^p, on note x_u et y_u les projections Q orthogonales de x et y sur U.

On pose

$$d_U(x,y) = d(x_u,y_u) = \sqrt{(x_u-y_u)' \; Q \; (x_u-y_u)}$$

Proposition 2

a - d_u est un écart sur $\mathbb{R}^p$, c'est-à-dire

a_1 : $d_u(x,y) = d_u(y,x)$ pour tout couple (x,y) dans $\mathbb{R}^p \times \mathbb{R}^p$

a_2 : $d_u(x,x) = 0$ pour tout x dans $\mathbb{R}^p$

a_3 : $d_u(x,z) \leq d_u(x,y) + d_u(y,z)$ pour tout triplet (x,y,z) dans $\mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^p$ (Inégalité triangulaire)

b - $d_u(x,y) = 0 \iff x-y$ est Q orthogonal au sous espace U.

Démonstration

d est une distance. On en déduit :

a_1 : $d_u(x,y) = d(x_u,y_u) = d(y_u,x_u) = d_u(y,x)$

a_2 : $d_u(x,x) = d(x_u,x_u) = 0$

a_3 : $d_u(x,z) = d(x_u,z_u) \leq d(x_u,y_u) + d(y_u,z_u)$

$d_u(x,z) \leq d_u(x,y) + d_u(y,z)$

b : $d_u(x,y) = 0 \iff d(x_u,y_u) = 0$

$d_u(x,y) = 0 \iff x_u = y_u \iff (x-y)_u = 0$

Proposition 3

Pour toute base Q orthonormée $\underline{U} = (u_1,\ldots,u_r)$ de U, et tout couple (x,y) de points de R^p, on a :

$$d^2_U(x,y) = (x-y)' \; Q \left(\sum_{s=1}^{r} u_s u_s' \right) Q \; (x-y) \qquad (0\text{-}8)$$

Démonstration

Comme $\underline{U}$ est une base de U, et que $u_1,\ldots,u_r$ sont Q normés, on a :

$$x_u = \sum_{s=1}^{r} (x' Q u_s) u_s$$

$$y_u = \sum_{s=1}^{r} (y' Q u_s) u_s$$

On en déduit :

$$d_U^2 (x,y) = \left[\sum_{s=1}^{r} ((x-y)' Q u_s) u_s \right]' Q \left[\sum_{s=1}^{r} ((x-y)' Q u_s) u_s \right]$$

Comme les vecteurs $u_1,\ldots,u_r$ sont Q orthogonaux, on a :

$$d_U^2 (x,y) = \sum_{s=1}^{r} ((x-y)' Q u_s) u_s' Q ((x-y)' Q u_s) u_s$$

$(x-y)' Q u_s$ est un scalaire égal à $u_s' Q (x-y)$. On en déduit :

$$d_U^2 (x,y) = \sum_{s-1}^{r} ((x-y)' Q u_s) u_s' Q u_s (u_s' Q (x-y))$$

Pour $s = 1,\ldots,r$, on a $u_s' Q u_s = 1$.

On en déduit :

$$d_U^2 (x,y) = \sum_{s=1}^{r} (x-y)' Q u_s u_s' Q (x-y)$$

$$d_U^2 (x,y) = (x-y)' Q \left(\sum_{s=1}^{r} u_s u_s' \right) Q (x-y)$$

Proposition 4

a) Pour toute famille $(a_1,\ldots,a_k)$ de points de $\mathbb{R}^p$, et pour tout U dans F_r, l'inertie aux points $a_1,\ldots,a_k$ de la partition P expliquée par le sous espace U s'écrit :

$$I\ ((a_1,\ldots,a_k),\ P,U) = \frac{1}{n} \sum_{\ell=1}^{k} \sum_{x\in P_\ell} d_U^2\ (x,a_\ell)$$

où d_U est l'écart sur R^p associé au sous espace U (définition 1)

b) Soit i(x) l'indice de la classe de la partition P qui contient x. On note a_x, le point de la famille $(a_1,\ldots,a_k)$ d'indice i(x).

On a :

$$I\ ((a_1,\ldots,a_k),\ P,\ U) = \frac{1}{n} \sum_{x\in E} d_U^2\ (x,a_x) \qquad (0\text{-}9)$$

c) Pour toute base Q orthonormée de U, $\underline{U} = (u_1,\ldots,u_r)$

On a :

$$I\ ((a_1,\ldots,a_k),\ P,\ U) = \frac{1}{n} \sum_{\ell=1}^{k} \sum_{x\in P_\ell} (x-a_\ell)'\ Q\ \Big(\sum_{s=1}^{r} u_s\ u_s'\Big)\ Q\ (x-a_\ell) \qquad (0\text{-}10)$$

Démonstration

On déduit a) des définitions de l'inertie $I\ ((a_1,\ldots,a_k,P,U)$ et de l'écart d_U

On déduit b) de a) par transformation de la sommation.

On déduit c) de a) et de la proposition 3.

Au paragraphe 7 on analyse et on donne la solution d'un problème d'optimisation qui sera posé aux chapitres I,II et III.

7.1. Le problème

Soit

N un nuage de points de R^p, K et r deux entiers ($r \leq K-1$).

Il s'agit de définir une fonction g qui associe à toute partition $P = (P_1,\ldots,P_k)$ du nuage N,

- K points $a_1,\ldots,a_k$ de R^p
- une famille Q orthonormée de r vecteurs de R^p, $(u_1,\ldots,u_r)$

tels que

$a_1,\ldots,a_k$ et le sous espace vectoriel U engendré par $u_1,\ldots,u_r$, vérifient :

$$I(N,U) - I((a_1,\ldots,a_k),P,U) = \underset{\substack{(b_1,\ldots,b_k) \in \mathbb{R}^p \times\ldots\times \mathbb{R}^p \\ F \in F_r}}{\text{Max}} \left[I(N,F) - I((b_1,\ldots,b_k)P,F) \right]$$

7.2. Cas particulier

On se place dans l'hypothèse où la matrice Q est égale à l'inverse V^{-1} de la matrice de variance-covariance empirique V (cf. chapitres II et III).

Pour toute droite vectorielle D, l'inertie du nuage N expliquée par D est égale à 1. En effet, soit u un vecteur directeur V^{-1} normé de D, on a, (0-3) :

$$I(N,D) = u' V^{-1} V V^{-1} u = u' V^{-1} u = 1$$

On en déduit, (0-2), que l'inertie expliquée par un sous espace vectoriel F de R^p, est égale à sa dimension.

Dans le problème d'optimisation énoncé ci-dessus, pour tout F dans F_r, l'inertie I (N,F) est constante, égale à r, et, les points $a_1,\ldots,a_k$ et le sous espace U engendré par $u_1,\ldots,u_r$, vérifient simultanément :

$$I(N,U) - I((a_1,\ldots,a_k), P,U) = \underset{\substack{(b_1,\ldots,b_k) \in \mathbb{R}^p \times\ldots\times \mathbb{R}^p \\ F \in F_r}}{\text{Max}} \left[I(N,F) - I((b_1,\ldots,b_k), P,F) \right]$$

égalité qui se ramène à

$$r - I((a_1,\ldots,a_k), P, U) = \max_{\substack{(b_1,\ldots,b_k) \in \mathbb{R}^p \times \ldots \times \mathbb{R}^p \\ F \in F_r}} \left[r - I((b_1,\ldots,b_k), P, F) \right]$$

et

$$I((a_1,\ldots,a_k), P, U) = \min_{\substack{(b_1,\ldots,b_k) \in \mathbb{R}^p \times \ldots \times \mathbb{R}^p \\ F \in F_r}} \left[I((b_1,\ldots,b_k), P, F) \right]$$

7.3. Analyse du problème

Une partition $P = (P_1,\ldots,P_k)$ du nuage N est donnée

Soit B (P) la matrice d'inertie inter classe de P.

Pour tout sous espace F de R^p, l'inertie aux points $b_1,\ldots,b_k$ de la partition P expliquée par F est minimale pour la famille $(g_1,\ldots,g_k)$ des centres de gravité des classes $P_1,\ldots,P_k$ de P.

$$I((g_1,\ldots,g_k), P, F) = \min_{(b_1,\ldots,b_k) \in \mathbb{R}^p \times \ldots \times \mathbb{R}^p} \left[I((b_1,\ldots,b_k), P, F) \right]$$

$I(g_1,\ldots,g_k, P, F)$ est l'inertie intra classe de la partition P, expliquée par F.

On en déduit, d'une part

$$(a_1,\ldots,a_k) = (g_1,\ldots,g_k)$$

et d'autre part,

$$\max_{\substack{(b_1,\ldots,b_k) \in \mathbb{R}^p \times \ldots \times \mathbb{R}^p \\ F \in F_r}} \left[I(N,F) - I((b_1,\ldots,b_r), P, F) \right] = \max_{F \in F_r} \left[I(N,F) - I((g_1,\ldots,g_k), P, F) \right]$$

$$= \max_{F \in F_r} \left[(I(N,F) - I_{intra}(P,F)) \right]$$

Enfin, d'après (0-6)

$$\underset{\substack{(b_1,\ldots,b_k)\in \mathbb{R}^p\times\ldots\times\mathbb{R}^p \\ F\in F_r}}{\text{Max}} \left[I(N,F) - I\,(b_1,\ldots,b_r),P,F)\right] = \underset{F\in F_r}{\text{Max}} \left[I_{\text{inter}}\,(P,F)\right] \qquad (0\text{-}11)$$

L'inertie inter classe de la partition P, expliquée par le sous espace F est l'inertie du nuage G des centres de gravité des classes $P_1,\ldots,P_k$ de P.

La matrice d'inertie inter classe de la partition P, B(P), est la matrice d'inertie du nuage G.

On est donc ramené à la recherche du maximum pour F dans F_r, de l'inertie du nuage G, expliquée par F.

L'analyse de ce problème est présentée en annexe 4. On en déduit une famille Q orthonormée de r vecteurs de R^p, $u_1,\ldots,u_r$, telle que :

le sous espace U engendré par $u_1,\ldots,u_r$ vérifie

$$I_{\text{inter}}\,(P,U) = \underset{F\in F_r}{\text{Max}}\; I_{\text{inter}}\,(P,F)$$

$u_1,\ldots,u_r$ sont vecteurs propres, associés aux r plus grandes valeurs propres de la matrice B(P) Q.

La définition précise des vecteurs $u_1,\ldots,u_r$ est donnée au paragraphe suivant.

7.4. La solution du problème : la fonction g

Soit $P = (P_1,\ldots,P_k)$ une partition de N en k classes

On définit g(P) par :

$$g\,(P) = (g_1,\ldots,g_k,\ u_1,\ldots,u_r)$$

où

$g_1,\ldots,g_k$ sont les centres de gravité des classes $P_1,\ldots,P_k$;

$u_1\ldots,u_r$ sont définis de manière précise à la fin de ce paragraphe.

$g_1,\ldots,g_k$, et le sous espace U engendré par $(u_1,\ldots,u_r)$ vérifient :

$$I\ (N,U) - I\ ((g_1,\ldots,g_k),P,U) = \underset{\substack{(b_1,\ldots,b_k)\in \mathbb{R}^p \times\ldots\times R^p \\ F\ \in\ F_r}}{\text{Max}} \left[I\ (N,F) - I\ ((b_1,\ldots,b_k),\ P,F)\right]$$

La valeur de ce maximum est égal à l'inertie inter classe de la partition P expliquée par le sous espace U :

$$I\ (N,U) - I\ ((g_1,\ldots,g_k),\ P,\ U) = I_{inter}\ (P,U) \qquad (0\text{-}12)$$

L'inertie $I_{inter}\ (P,U)$ se décompose en inerties inter classe expliquées par les droites vectorielles engendrées par $u_1,\ldots,u_r$. Ces inerties sont respectivement égales aux r plus grandes valeurs propres, $\lambda_1 \geq\ldots \geq \lambda_r \geq 0$ de la matrice B(P) Q. (B(P) est la matrice d'inertie inter classe de la partition P).

Ainsi :

$$I\ (N,U) - I\ ((g_1,\ldots,g_k),\ P,\ U) = \sum_{s=1}^{r} \lambda_s \qquad (0\text{-}13)$$

Définition précise des vecteurs $u_1,\ldots,u_r$

Soit B(P) la matrice d'inertie inter classe de la partition P

Soit $\lambda_1 \geq \lambda_2 \geq\ldots\geq \lambda_r \geq 0$ les r plus grandes valeurs propres de la matrice B(P)Q. Une valeur propre multiple λ de multiplicité $m > 1$ apparait comme $\lambda_s = \lambda_{s+1} =\ldots= \lambda_j$, $j = \text{Min}\ (s+m-1,r)$.

. Si λ_s est valeur propre simple

u_s est défini par

$$B(P)\ Q\ u_s = \lambda_s\ u_s$$

$$u_s'\ Q\ u_s = 1$$

$$u_s'\ Q\ e_{i_s} > 0$$

où e_i désigne le $i^{ème}$ vecteur de la base canonique de R^p et où i_s est le plus petit indice i tel que e_i ne soit pas Q orthogonal au sous espace propre E_s associé à λ_s.

. Si λ est valeur propre multiple, de multiplicité m>1,

$\lambda = \lambda_s = \ldots = \lambda_j$ où $j = \text{Min } (s+m-1, r)$, $u_s, \ldots, u_j$ sont définis par :

+ Soit i_s le plus petit indice i tel que e_i ne soit pas Q orthogonal au sous espace propre E_λ associé à λ.

Soit v_s la projection Q orthogonale de e_{i_s} sur le sous espace E_λ.

u_s est défini par :

$$u_s = \frac{v_s}{\sqrt{v_s' \, Q \, v_s}}$$

+ Pour $t = s+1, \ldots, j$

Soit M_t le sous espace vectoriel engendré par $u_s, \ldots, u_t$

Soit $E_\lambda \ominus M_t$ le supplémentaire Q orthogonal de M_t dans E_λ

Soit i_t le plus petit indice i tel que e_i ne soit pas Q orthogonal à $E_\lambda \ominus M_t$.

Soit v_t la projection Q orthogonale de e_{i_t} sur le sous espace $E_\lambda \ominus M_t$

u_t est défini par :

$$u_t = \frac{v_t}{\sqrt{v_t' \, Q \, v_t}}$$

Cas particulier

Dans le cas où Q est choisie égale à V^{-1}, $g_1, \ldots, g_k$ et le sous espace vectoriel U engendré par $u_1, \ldots, u_r$, vérifient également :

$$I((g_1, \ldots, g_k), P, U) = \underset{\substack{(b_1, \ldots, b_k) \in \mathbb{R}^p \times \ldots \times \mathbb{R}^p \\ F \in F_r}}{\text{Min}} I((b_1, \ldots, b_k), P, F) \qquad (0\text{-}14)$$

La valeur de ce minimum est l'inertie intra classe de la partition P expliquée par le sous espace U.

$$I\ ((g_1,\ldots,g_k),\ P,U) = I_{intra}\ (P,U) \qquad (0\text{-}15)$$

Des égalités suivantes

$$I_{intra}\ (P,U) + I_{inter}\ (P,U) = I\ (N,U)$$

$$I\ (N,U) = r$$

$$I_{inter}\ (P,U) = \sum_{s=1}^{r} \lambda_s$$

On déduit la relation

$$I\ ((g_1,\ldots,g_k),\ P,U) = r - \sum_{s=1}^{r} \lambda_s \qquad (0\text{-}16)$$

CHAPITRE I

CLASSIFICATION SUR SOUS ESPACE ADAPTATIF

1 - INTRODUCTION

1.1 - La méthode

Il s'agit d'une méthode de classification non hiérarchique qui fournit une partition en K classes d'une population finie d'objets et un sous espace vectoriel de l'espace des descriptions des objets. Elle procède par optimisation d'un critère qui exprime l'inertie inter classe expliquée par le sous espace.

Les paramètres descriptifs des objets sont pondérés dans l'analyse par le choix d'une métrique et d'un sous espace de classification. La métrique est fixée ; par contre le sous espace est adaptatif.

On définit la notion de représentation associée à une partition. On propose un algorithme itératif du type "nuées dynamiques", [8], et on montre qu'il fait décroitre le critère et converge en un nombre fini d'itérations vers une partition réalisant un optimum local de ce critère.

1.2 - Les données

On dispose d'une population finie E de n individus ou objets, $x_1,\ldots,x_n$, décrits par p paramètres quantitatifs $x^1,\ldots,x^p$. Les individus sont munis de poids égaux, $p_i = \frac{1}{n}$, $i = 1,\ldots,n$.

Les données peuvent être rangées dans une matrice X' comportant n lignes et p colonnes :

La $i^{ème}$ ligne contient les valeurs prises par les p paramètres sur l'individu x_i.

La $j^{ème}$ colonne contient les valeurs prises par le $j^{ème}$ paramètre sur les n individus.

On note x_i^j la valeur prise par le paramètre x^j sur l'individu x_i,

$$X' = \begin{pmatrix} x_1^1 & \cdots & x_1^j & \cdots & x_1^p \\ \vdots & & \vdots & & \vdots \\ x_i^1 & \cdots & x_i^j & \cdots & x_i^p \\ \vdots & & \vdots & & \vdots \\ x_n^1 & \cdots & x_n^j & \cdots & x_n^p \end{pmatrix}$$

On identifie un individu x_i, au vecteur de $\mathbb{R}^p$ des valeurs prises par les p variables sur cet individu. On note :

$$x_i' = (x_i^1, \ldots, x_i^p)$$

On suppose donnée une matrice Q, symétrique, définie positive, de dimension p. L'espace des individus, $\mathbb{R}^p$, est muni du produit scalaire défini par :

$$(x, y) \rightarrow x' \, Q \, y$$

A ce produit scalaire est associé une distance sur R^p, notée d, définie par :

$$d^2(x, y) = (x - y)' \, Q \, (x - y)$$

Le problème du choix de cette métrique est abordé en [4]

On suppose donnés deux entiers, K et r, $(r \leq K-1)$

- K est le nombre maximum de classes de la partition recherchée
- r est la dimension du sous espace de classification.

1.3 - Notations

On note :

$\mathbb{P}_K$, l'ensemble des partitions de E en K classes (disjointes).

Une partition P de $\mathbb{P}_K$ est notée également $(P_1,\ldots,P_K)$

$$E = P_1 \cup \ldots \cup P_K$$

$$P_i \cap P_j = \phi \quad i=1,\ldots,K \; ; \; j=1,\ldots,K \; ; \; i \neq j$$

F_r, l'ensemble des sous espaces vectoriels de $\mathbb{R}^p$ de dimension r.

U_r, l'ensemble des familles Q orthonormées de r vecteurs de $\mathbb{R}^p$.

Un élément de U_r est noté $\underline{U} = (u_1,\ldots,u_r)$.

Les conditions d'orthogonalité et de normalisation s'écrivent :

$$u_s' \; Q \; u_s = 1 \qquad s = 1,\ldots,r$$

$$u_s' \; Q \; u_t = 0 \qquad s = 1,\ldots r \; ; \; t = 1,\ldots r \; ; \; s \neq t$$

On sait [3] que :

Toute famille $\underline{U} = (u_1,\ldots,u_r)$ de U_r engendre un sous espace vectoriel U de $\mathbb{R}^p$ de dimension r.

Ainsi à tout élément $\underline{U}$ de U_r on peut associer, de manière unique, un élément U de F_r.

Réciproquement à tout élément U de F_r on peut associer, de manière non unique, un élément $\underline{U}$ dans U_r [3].

Remarque : Dans la suite de ce travail certaines quantités numériques (inerties, critères...) dont l'expression mathématique fait apparaître un élément $\underline{U}$ de U_r ne dépendent que du sous espace U engendré par $\underline{U}$.

2 - LE CRITERE DE CLASSIFICATION

2.1 - Définition

Soit P une partition de E.

Pour tout sous espace U de R^p, on note I_{inter} (P, U) l'inertie inter classe de la partition P expliquée par U.

On définit un critère W sur l'ensemble des partitions de E par :

$$W(P) = \max_{U \in F_r} \left[I_{inter} (P, U) \right]$$

W(P) mesure l'écartement moyen entre les classes de la partition P, en projection sur un sous espace de dimension r.

2.2 - L'optimisation du critère

On recherche une partition de E en K classes P_K^* qui réalise le maximum du critère W pour P dans $\mathbb{P}_K$.

$$W(P_K^*) = \max_{P \in \mathbb{P}_K} W(P)$$

Remarque : La recherche du maximum du critère dans l'ensemble $\mathbb{P}$ des partitions de E, conduit à la partition triviale P_E constitué de n classes réduites à un individu.

$$W(P_E) = \max_{P \in \mathbb{P}} W(P)$$

De plus, ce maximum est l'inertie du nuage des objets E, expliquée par son sous espace principal de dimension r. (Analyse en composantes principales [2])

En effet, pour tout sous espace U, et toute partition P, on a (O-6) :

$$I_{inter}\ (P,\ U) + I_{intra}\ (P,\ U) = I(E,U)$$

D'autre part pour tout sous espace U, on a :

$$I_{intra}\ (P_E,\ U) = 0$$

On en déduit :

$$I_{inter}\ (P_E,\ U) = I(E,U)$$

d'où

$$W(P_E) = \underset{U \in F_r}{\text{Max}}\ I(E,U)$$

On est ramené à un problème d'analyse en composantes principales (pour la métrique Q) du nuage des objets E.

2.3 - Expression mathématique du critère

Pour toute partition P de E, de matrice d'inertie inter classe B(P), pour tout U dans F_r, et toute base Q orthonormée $\underline{U} = (u_1,\ldots,u_r)$ du sous espace U, l'inertie inter classe de la partition P expliquée par U s'écrit, (O-7) :

$$I_{inter}\ (P,\ U) = \sum_{s=1}^{r} u_s'\ Q\ B(P)\ Q\ u_s$$

$u_s'\ Q\ B(P)\ Q\ u_s$ est l'inertie inter classe de la partition P expliquée par la droite vectorielle engendrée par le vecteur u_s.

On en déduit :

$$W(P) = \underset{U \in F_r}{\text{Max}} \left[\sum_{s=1}^{r} u_s'\ Q\ B(P)\ Q\ u_s \right]$$

où $(u_1,\ldots,u_r)$ est une base Q orthonormée quelconque de U.

3 - LA REPRESENTATION

Dans ce paragraphe on donne la définition d'une représentation puis d'un critère W_R, qui est une généralisation du critère W nécessaire au déroulement de l'algorithme.

3.1 - Définition

On appelle représentation tout ensemble L constitué de K points (ou centres) de $\mathbb{R}^p$, $a_1,\ldots,a_K$, et d'une famille $(u_1,\ldots,u_r)$ dans U_r.

$L = (a_1,\ldots,a_K, u_1,\ldots,u_r)$

$u_s' Q u_s = 1 \quad s = 1,\ldots,r$

$u_s' Q u_t = 0 \quad s = 1,\ldots,r \;;\; t = 1,\ldots,r \;;\; s \neq t$

On note :

$\underline{A} = (a_1,\ldots,a_K)$

$\underline{U} = (u_1,\ldots,u_r)$

U le sous espace de R^p engendré par $(u_1,\ldots,u_r)$

$L = (\underline{A}, \underline{U})$

L'ensemble des représentations définies ci-dessus est noté $\mathbb{L}_{K,r}$.

3.2 - Le critère W_R

3.2.1 - Définition

On définit un critère W_R sur l'ensemble produit $\mathbb{P}_K \times \mathbb{L}_{K,r}$, de la manière suivante :

Soit :

$P = (P_1,\ldots,P_K)$ dans $\mathbb{P}_K$

$L = (a_1,\ldots,a_K,\ u_1,\ldots,u_r) = (\underline{A},\ \underline{U})$ dans $\mathbb{L}_{K,r}$

U le sous espace engendré par les vecteurs $u_1,\ldots,u_r$

I(E, U), l'inertie du nuage E expliquée par U

$I(\underline{A},\ P,\ U)$, l'inertie aux points $a_1,\ldots,a_K$ de la partition P expliquée par U.

On pose :

$$W_R(P,\ L) = W_R(P,(\underline{A},\ \underline{U})) = I(E,\ U) - I(\underline{A},\ P,\ U) \qquad (3\text{-}1)$$

Remarque : Le critère $W_R(P,\ L)$ et l'inertie $I(\underline{A},\ P,\ U)$ dépendent de la partition P, de la famille de points $\underline{A} = (a_1,\ldots,a_K)$, et du sous espace U. (Ils ne dépendent pas de la base Q orthonormée $\underline{U}$ de U).

L'inertie I(E, U) ne dépend que du sous espace U.

3.2.2 - Cas particulier

Soit $\underline{G} = (g_1,\ldots,g_K)$ la famille des centres de gravité des classes $P_1,\ldots,P_K$.

L'inertie $I(\underline{G},\ P,\ U)$ est l'inertie intra classe de la partition P expliquée par le sous espace U :

$$I(\underline{G},\ P,\ U) = I_{intra}\ (P,\ U)$$

D'autre part, d'après la relation (0-6), on a :

$$I(E,\ U) - I_{intra}\ (P,\ U) = I_{inter}\ (P,\ U)$$

On en déduit :

$$\text{Si } L = (\underline{G},\ \underline{U}),\ W_R(P,\ L) = I_{inter}\ (P,\ U) \qquad (3\text{-}2)$$

3.2.3 - Proposition

Pour toute partition P dans $\mathbb{P}_K$

$$\max_{L \in \mathbb{L}_{K,r}} W_R(P, L) = W(P)$$

Démonstration

D'après la relation (0-11) le problème d'optimisation posé se ramène à la recherche du maximum pour U dans F_r de l'inertie $I_{inter}(P, U)$. Or par définition, (2.1), W(P) est la valeur de ce maximum.

4 - L'ALGORITHME

4.1 - Introduction

On recherche une partition P_K^* de E en K classes qui réalise le maximum pour P dans $\mathbb{P}_K$ du critère W :

$$W(P_K^*) = \max_{P \in \mathbb{P}_K} W(P)$$

Soit, d'après la proposition 3.2.3 :

$$W(P_K^*) = \max_{P \in \mathbb{P}_K} \left(\max_{L \in \mathbb{L}_{K,r}} W_R(P, L) \right)$$

Faute de pouvoir exhiber une partition P_K^* réalisant cet optimum global, on propose un algorithme convergent vers un maximum local du critère W, [8].
Il utilise alternativement deux fonctions, définies en (4.2) et (4.3) :

- g, appelée fonction de représentation, associe à toute partition P de $\mathbb{P}_K$, une représentation g(P) dans $\mathbb{L}_{K,r}$ telle que :

$$W_R(P, g(P)) = \max_{L \in \mathbb{L}_{K,r}} W_R(P, L) \qquad (4\text{-}1)$$

- f, appelée fonction d'affectation, associe à toute représentation L de $\mathbb{L}_{K,r}$, une partition f(L) dans $\mathbb{P}_K$ telle que :

$$W_R(f(L), L) = \max_{P \in \mathbb{P}_K} W_R(P, L) \qquad (4\text{-}5)$$

On montre, (4.4), que pour toute partition P dans $\mathbb{P}_K$

$$W(f \circ g\ (P)) \geq W(P) \qquad (4\text{-}8)$$

L'algorithme est initialisé par une partition P^0 dans $\mathbb{P}_K$. Pour tout $n \geq 0$, on définit :

$$P^{n+1} = f \circ g\ (P^n)$$

D'après (4-8), la suite de terme général $W(P^n)$ est croissante. On montre, (4.6), qu'elle est constante à partir d'un certain rang, et que la suite de partitions (P^n) est alors également constante. La partition P^* ainsi obtenue, comme limite de la suite (P^n), réalise un maximum local du critère W.

4.2 - La fonction de représentation

Il s'agit de définir une fonction g qui associe à toute partition P dans $\mathbb{P}_K$ une représentation L = g(P) dans $\mathbb{L}_{K,r}$, telle que :

$$W_R(P, g(P)) = \max_{L \in \mathbb{L}_{K,r}} W_R(P, L) \qquad (4\text{-}1)$$

Soit :

$P = (P_1, \ldots, P_K)$ dans $\mathbb{P}_K$

$L = (b_1, \ldots, b_K, v_1, \ldots, v_r) = (\underline{B}, \underline{V})$ dans $\mathbb{L}_{K,r}$

F le sous espace vectoriel engendré par les vecteurs $v_1, \ldots, v_r$

$g(P) = (a_1, \ldots, a_K, u_1, \ldots, u_r) = (\underline{A}, \underline{U})$

U le sous espace vectoriel engendré par les vecteurs $u_1, \ldots, u_r$

On a, (3-1) :

$W_R(P, L) = I(E, F) - I(\underline{B}, P, F)$

$W_R(P, g(P)) = I(E, U) - I(\underline{A}, P, U)$

On est ramené au problème d'optimisation étudié au paragraphe 7 de l'avant propos :

Définir une fonction g qui associe à toute partition $P = (P_1,\ldots,P_K)$ de E,

- K points, $a_1,\ldots,a_K$ de $\mathbb{R}^p$

- Une famille Q orthonormée de r vecteurs de $\mathbb{R}^p$, $\underline{U} = (u_1,\ldots,u_r)$

tels que

$a_1,\ldots,a_K$ et le sous espace vectoriel U engendré par $u_1,\ldots,u_r$, vérifient :

$$I(E, U) - I((a_1,\ldots,a_K), P, U) = \max_{\substack{(b_1,\ldots,b_K)\in\mathbb{R}^p\times\ldots\times\mathbb{R}^p \\ F\in F_r}} [I(E, F) - I((b_1,\ldots,b_K), P, F)]$$

On sait, (§ 7.4 de l'avant propos), que la fonction g, définie ci-dessous, convient :

$g(P) = (g_1,\ldots,g_K, u_1,\ldots,u_r)$

où - $g_1,\ldots,g_K$ sont les centres de gravité des classes $P_1,\ldots,P_K$ de la partition P.

- $u_1,\ldots,u_r$ sont les vecteurs propres Q normés, associés aux r plus grandes valeurs propres $\lambda_1 \geq \ldots \geq \lambda_r \geq 0$ de la matrice B(P) Q (où B(P) est la matrice d'inertie inter classe de la partition P).
  La définition précise de $u_1,\ldots,u_r$ est donnée au paragraphe 7.4 de l'avant propos.

Par définition de la fonction g, et d'après (3.2.3) on a :

$$W_R(P, g(P)) = W(P) \qquad (4\text{-}2)$$

D'après la relation (3-2), $W_R(P, g(P))$ est égal à l'inertie inter classe de la partition P expliquée par U, $I_{inter}(P, U)$.

On en déduit d'après (4-2), la relation :

$$W(P) = I_{inter}(P, U) \qquad (4\text{-}3)$$

D'après la relation (0-13), $W_R(P, g(P))$ est égal à la somme des r plus grandes valeurs propres $\lambda_1 \geq \ldots \geq \lambda_r \geq 0$ de la matrice B(P) Q ;

$$W_R(P, g(P)) = \sum_{s=1}^{r} \lambda_s$$

On en déduit d'après (4-2), la relation :

$$W(P) = \sum_{s=1}^{r} \lambda_s \qquad (4\text{-}4)$$

4.3 - La fonction d'affectation

4.3.1 - Définition

Il s'agit de définir une fonction f qui associe à toute représentation L dans $\mathbb{L}_{K,r}$, une partition P = f(L) dans $\mathbb{P}_K$, telle que :

$$W_R(f(L), L) = \max_{P \in \mathbb{P}_K} W_R(P, L) \qquad (4\text{-}5)$$

Soit :

$L = (a_1, \ldots, a_K, u_1, \ldots, u_r) = (\underline{A}, \underline{U})$ dans $\mathbb{L}_{K,r}$

U le sous espace vectoriel engendré par les vecteurs $u_1, \ldots, u_r$

$P = (P_1, \ldots, P_K)$ dans $\mathbb{P}_K$

Par définition du critère W_R, (3.2.1), on a :

$$W_R(P, L) = I(E, U) - I(\underline{A}, P, U) \qquad (4\text{-}6)$$

On en déduit :

$$\underset{P\in\mathbb{P}_K}{\text{Max}} \; W_R(P, L) = I(E, U) - \underset{P\in\mathbb{P}_K}{\text{Min}} \; I(\underline{A}, P, U)$$

D'après (0-9), pour toute partition $P = (P_1,\ldots,P_K)$ dans $\mathbb{P}_K$,

$$I(\underline{A}, P, U) = \frac{1}{n} \sum_{x\in E} d_U^2(x, a_x) \qquad (4\text{-}7)$$

où $\quad d_U^2(x, y) = (x - y)' \; Q \left(\sum_{s=1}^{r} u_s \, u_s' \right) Q(x - y)$

et a_x le centre de la famille $(a_1,\ldots,a_K)$ d'indice $i(x)$, où $i(x)$ est l'indice de la classe de la partition $P = (P_1,\ldots,P_K)$ qui contient x.

On en déduit :

$$\underset{P\in\mathbb{P}_K}{\text{Min}} \; I(\underline{A}, P, U) = \frac{1}{n} \sum_{x\in E} \; \underset{1\le i\le K}{\text{Min}} \; d_U^2(x, a_i)$$

On pose par conséquent :

$$f(L) = (P_1,\ldots,P_K) \quad \text{où}$$

$$P_\ell = \{x \mid d_U^2(x, a_\ell) \le d_U^2(x, a_i) \;\; ; \;\; i = 1,\ldots,K\} \; ; \; \ell = 1,\ldots,K \, .$$

Pour compléter la règle d'affectation d'un individu x, dans l'éventualité où plusieurs a_i réalisent le minimum de $d_U^2(x, a_i)$, on se place dans le contexte de l'algorithme.

Soit :

Une partition Q dans $\mathbb{P}_K$,

La représentation $L = g(Q)$.

On recherche P = f(L) :

Soit x un individu, classé dans Q_j, tel que plusieurs centres a_i réalisent $\underset{1\leq\ell\leq K}{\text{Min}} \; d_U^2(x, a_\ell)$.

On distingue deux éventualités :

Eventualité 1 $\quad d_U^2(x, a_j) = \underset{1\leq\ell\leq K}{\text{Min}} \; d_U^2(x, a_\ell)$

x est affecté à P_j

Eventualité 2 Sinon x est affecté à P_m, où m est le plus petit indice i tel que

$$d_U^2(x, a_i) = \underset{1\leq\ell\leq K}{\text{Min}} \; d_U^2(x, a_\ell)$$

Dans ce cas d'indétermination, la partition P dépend non seulement de la représentation L, mais aussi de la partition Q.

f apparaît ainsi comme une fonction de deux variables :

- une représentation
- une partition utilisée simplement comme contrainte en cas d'indétermination pour l'affectation. (La classe d'appartenance dans la partition Q est prioritaire).

Le rôle de cette seconde variable est secondaire. De plus, dans l'algorithme, (4.5), la partition Q est définie par la représentation L. Aussi, pour simplifier les notations, on écrit f comme une fonction de la seule variable représentation, et on retient la proposition 4.3.2. (On vérifie dans la présentation de l'algorithme, et la démonstration de sa convergence que l'emploi de cette notation n'entraîne aucune inexactitude).

4.3.2 - Proposition

Si les partitions Q et P = f o g (Q) vérifient

$W_R(Q, g(Q)) = W_R(P, g(Q))$ elles sont identiques.

Démonstration

Soit :

$Q = (Q_1, \ldots Q_K)$

$L = g(Q) = (a_1, \ldots, a_K, u_1, \ldots, u_r) = (\underline{A}, \underline{U})$

U le sous espace engendré par $u_1, \ldots, u_r$

$P = f \circ g\ (Q) = (P_1, \ldots, P_K)$

Soit un objet x dans E. On suppose que x appartient à la classe Q_j $(1 \leq j \leq K)$ de la partition Q, et à la classe P_i $(1 \leq i \leq K)$ de la partition P.

On a, par définition de P = f(L)

$$d_U^2(x, a_i) = \min_{\ell=1,\ldots,K} [d_U^2(x, a_\ell)] \leq d_U^2(x, a_j)$$

D'après (4-7) :

$$(d_U^2(x, a_i) < d_U^2(x, a_j)) \Rightarrow (I(\underline{A}, P, U) < I(\underline{A}, Q, U))$$

et d'après (4-6) :

$$(d_U^2(x, a_i) < d_U^2(x, a_j) \Rightarrow W_R(P, g(Q)) > W_R(Q, g(Q))$$

comme la dernière inégalité est contraire à l'hypothèse on en déduit :

$$d^2(x, a_i) = d^2(x, a_j)$$

L'affectation de l'objet x à l'une des classes de la partition P s'est donc effectuée selon la règle définie dans l'éventualité 1 du cas d'indétermination.

On en déduit $i = j$.

Cette démonstration étant valable pour tout x dans E, on en déduit que les partitions P et Q sont identiques.

4.4 - La fonction composée f o g

Proposition

Pour toute partition Q dans $\mathbf{P}_K$.

La partition f o g (Q) = f[g(Q)] vérifie la relation :

$$W(f \circ g\ (Q)) \geq W(Q) \tag{4-8}$$

De plus

$$W(f \circ g\ (Q)) = W(Q) \Rightarrow f \circ g\ (Q) = Q \tag{4-9}$$

Démonstration

D'après (4-2) et la définition de la partition f[g(Q)], on a :

$$W(Q) = W_R(Q,\ g(Q)) \leq W_R(f[g(Q)]\ ,\ g(Q)) \tag{4-10}$$

D'après la définition de la représentation g[f o g(Q)], et (4-2) appliqué à la partition f o g(Q), on a :

$$W_R(f \circ g(Q),\ g(Q)) \leq W_R(f \circ g(Q),\ g[f \circ g(Q)]) = W(f \circ g(Q)) \tag{4-11}$$

Des relations (4-10) et (4-11) on déduit :

$$W(Q) \leq W(f \circ g(Q))$$

De plus, si W(Q) = W(f o g(Q)), on a l'égalité dans la relation (4-10) :

$W_R(Q, g(Q)) = W_R(f[g(Q)], g(Q))$

D'après la proposition 4.3.2, les partitions Q et f o g(Q) sont égales.

4.5 - L'algorithme

L'algorithme est initialisé par une partition P^0 donnée ou tirée au hasard.

On pose, pour tout $n \geq 0$

$L^n = g(P^n)$

$P^{n+1} = f(L^n) = f \circ g(P^n)$

On définit ainsi une suite (P^n) de partitions dans $\mathbb{P}_K$, et une suite (L^n) de représentations dans $\mathbb{L}_{K,r}$.

Soit :

$P^n = (P_1^n, \ldots, P_K^n)$

$g_1^n, \ldots, g_K^n$, les centres de gravité des classes $P_1^n, \ldots, P_K^n$.

On sait, (4.2), que $L^n = g(P^n)$ est de la forme :

$L^n = (g_1^n, \ldots, g_K^n, \underline{U}^n)$

où $\underline{U}^n$, est une famille Q orthonormée de r vecteurs de R^p. On note U^n le sous espace engendré par $\underline{U}^n$.

D'après (4-2) :

$$W_R(P^n, L^n) = W_R(P^n, g(P^n)) = W(P^n) \qquad (4\text{-}12)$$

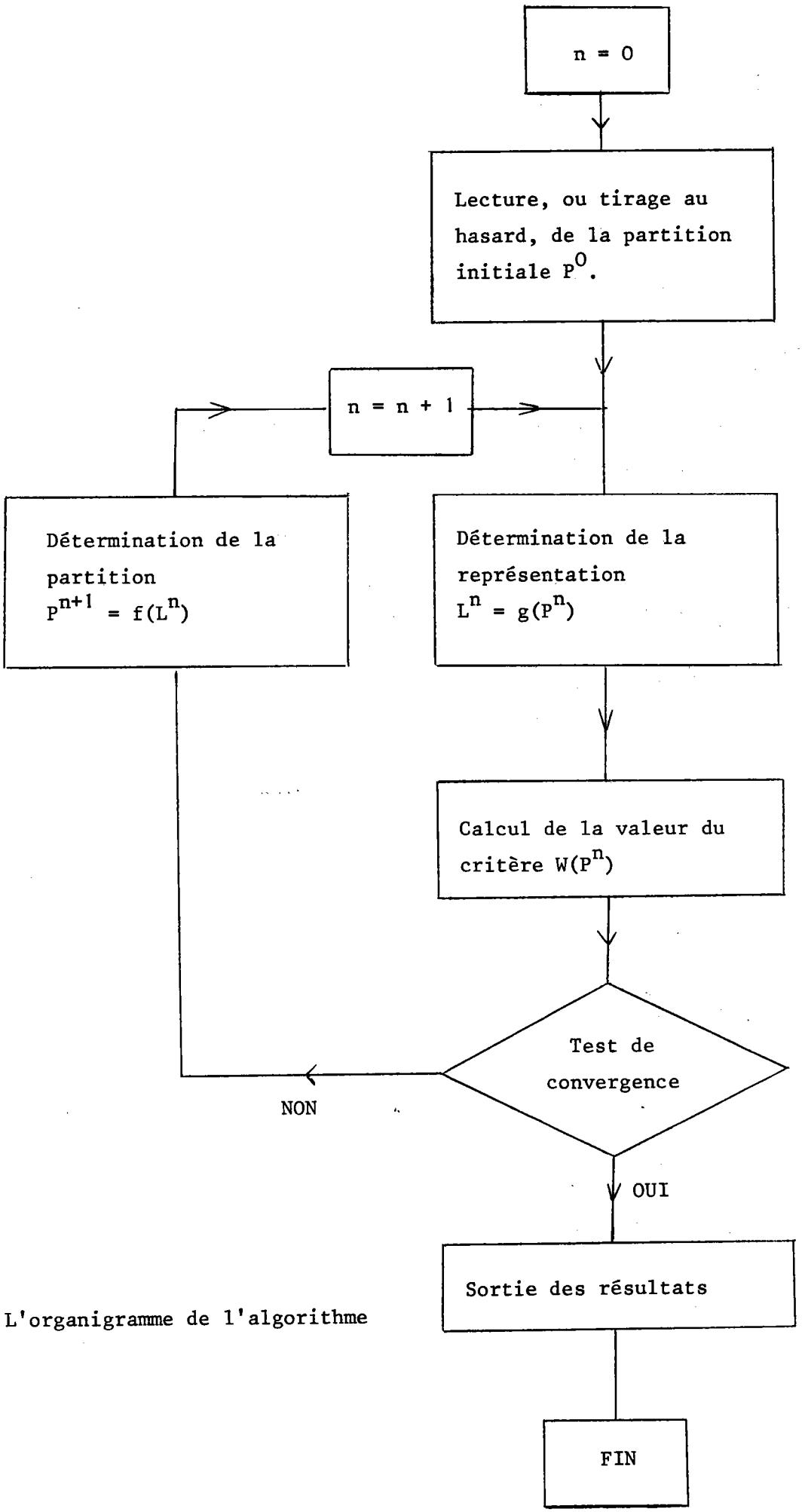


Figure 1 : L'organigramme de l'algorithme

D'après, (4-3) :

$$W(P^n) = I_{inter}(P^n, U^n) \qquad (4\text{-}13)$$

4.6 - La convergence de l'algorithme

Théorème

La suite de terme général $r_n = W(P^n)$ est croissante, et constante à partir d'un certain rang N.

De plus, la suite (P^n, L^n) est constante à partir du rang N.

Démonstration

a) Pour tout $n \geq 0$, on a d'après (4.4) :

$$r_{n+1} = W(P^{n+1}) = W(f \circ g(P^n)) \geq W(P^n) = r_n$$

La suite (r_n) est donc croissante.

b) La suite (r_n) prend ses valeurs dans l'ensemble fini $\{W(P)/P \in \mathcal{P}_K\}$

En effet, comme E est un ensemble fini, l'ensemble $\mathcal{P}_K$, des partitions de E en K classes, est fini.

c) D'après a) et b) il existe un rang N tel que :

$$r_n = r_N \text{ pour tout } n \geq N.$$

d) Pour tout n, $n \geq N$, on a, d'après c) :

$$r_{n+1} = r_n$$

Soit d'après a) :

$W(f \circ g(P^n)) = W(P^n)$

On en déduit d'après (4-9) :

$f \circ g(P^n) = P^n$

e) On déduit de d) :

$P^{n+1} = P^n$ pour tout $n \geq N$

d'où

$P^n = P^N$ pour tout $n \geq N$

f) On déduit de l'égalité $P^{n+1} = P^n$ démontrée en e) :

$L^{n+1} = g(P^{n+1}) = g(P^n) = L^n$ pour tout $n \geq N$

d'où

$L^n = L^N$ pour tout $n \geq N$

Remarque : Dans l'algorithme, la fonction f est utilisée exclusivement dans le contexte suivant :

Détermination de la partition $P^{n+1} = f(L^n)$, où $L^n = g(P^n)$, $n \geq 0$

La partition Q de contrainte qui apparaît dans la définition de la fonction f (4.3.1), est donc parfaitement définie, et f peut être notée comme une fonction d'une variable. La prise en compte de la partition P^n pour définir la partition P^{n+1}, apparaît dans la démonstration de la convergence de l'algorithme par l'utilisation en d) de la relation (4-9), déduite de la proposition 4.3.2.

5 - LE PROGRAMME INFORMATIQUE SESADA

5.1 - Les entrées

Outre le tableau des données (1.2), les entrées du programme SESADA sont les suivantes :

- La matrice Q symétrique définie positive. Le choix de cette matrice, et donc du produit scalaire et de la distance dans $\mathbb{R}^p$ qui lui sont associés, dépend de la nature des données.

- Le nombre maximum de classes désiré : K.

- La dimension du sous espace de classification : r. r doit être choisi inférieur ou égal à K-1.

- En option, la partition initiale P^0. Sinon, elle est tirée au hasard.

Les choix de K, r et éventuellement P^0 sont faits d'après la connaissance des données, par exemple les résultats d'une analyse en composantes principales pour la métrique Q. Dans une première analyse on prendra en général r = K-1.

Remarque : Dans le programme SESADA, il est possible d'initialiser l'algorithme par une représentation L^{00} simple, où par exemple $u_1^0,\ldots,u_r^0$ sont des vecteurs de la base canonique de R^p. On définit la partition initiale par $P^0 = f(L^{00})$, où f est la fonction d'affectation utilisée dans l'algorithme.

5.2 - Le critère

Dans le programme SESADA, on utilise le critère W_R sous la forme suivante :

Soit $P = (P_1,\ldots,P_K)$

$L = (a_1,\ldots,a_K,\ u_1,\ldots,u_r)$

$b_s = Qu_s \qquad s = 1,\ldots,r$

$$W_R(P,\ L) = \frac{1}{n} \sum_{\ell=1}^{K} \sum_{x\in P_\ell} \sum_{s=1}^{r} [(x-a_\ell)'b_s]^2$$

5.3 - La fonction de représentation [16]

a) On sait que $u_1,\ldots,u_r$ sont vecteurs propres associés aux r plus grandes valeurs propres de la matrice B(P) Q, où B(P) est la matrice d'inertie inter classe de la partiton P :

$B(P)\ Q\ u_s = \lambda_s\ u_s \qquad s = 1,\ldots,r$

$\lambda_1 \geq . \ . \ . \geq \lambda_r$

b) La matrice Q est définie positive, donc elle est régulière. On déduit de l'annexe 5 que :

- d'une part, $\lambda_1,\ldots,\lambda_r$ sont les r plus grandes valeurs propres de la matrice Q B(P),

- d'autre part, $b_1,\ldots,b_r$ sont vecteurs propres de la matrice Q B(P) associés aux valeurs propres $\lambda_1,\ldots,\lambda_r$.

Aussi on recherche directement les vecteurs $b_1,\ldots,b_r$.

c) Soit

$g_1,\ldots,g_K$ les centres de gravité des classes $P_1,\ldots,P_K$

$p_1,\ldots,p_K$ les poids des classes $P_1,\ldots,P_K$, $(p_\ell = \frac{\text{Card}P_\ell}{n}\ \ell=1,\ldots,K)$

g_0 le centre de gravité de E

$$C = (\sqrt{p_1}(g_1 - g_0), \ldots, \sqrt{p_K}(g_K - g_0))$$

La matrice d'inertie inter classe de la partition P s'écrit :

$$B(P) = C\ C'$$

On déduit de l'annexe 6 que :

- d'une part, $\lambda_1, \ldots, \lambda_r$ sont les r plus grandes valeurs propres de la matrice C' Q C
- d'autre part, si $a_1, \ldots, a_r$ sont vecteurs propres associés aux valeurs propres $\lambda_1, \ldots, \lambda_r$ de la matrice C' Q C, les vecteurs Q C $a_1, \ldots,$ Q C a_r sont vecteurs propres associés aux valeurs propres $\lambda_1, \ldots, \lambda_r$ de la matrice Q B(P).

d) Conclusion

Finalement, on recherche les r plus grandes valeurs propres $\lambda_1 \ \ldots \ \lambda_r$ de la matrice C' Q C, et r vecteurs propres associés $a_1, \ldots, a_r$. Puis on pose :

$$\tilde{b}_s = Q\ C\ a_s \qquad s = 1, \ldots, r$$

Si λ_s est valeur propre simple, on pose :

$$u_s = Q^{-1}\ \tilde{b}_s \qquad \text{ou} \qquad u_s = -Q^{-1}\ \tilde{b}_s$$

pour respecter la convention donnée dans l'avant propos (§ 7.4).

Si λ_s est valeur propre multiple de multiplicité m, $\lambda_s = \lambda_{s+1} = \ldots = \lambda_{s+m-1}$, u_s est choisi dans E_s, sous espace propre de la matrice B Q associé à λ_s, selon la procédure décrite dans l'avant propos (§ 7.4). E_s est engendré par les vecteurs $Q^{-1}\ \tilde{b}_s, \ldots, Q^{-1}\ \tilde{b}_{s+m-1}$.

Ainsi on effectue à chaque itération la recherche des r plus grandes valeurs propres et des vecteurs propres associés d'une matrice <u>symétrique</u> de <u>dimension K</u>, et non de la matrice Q B(P) qui n'est pas symétrique et de dimension p.(p est en général nettement supérieur à K).

5.4 - Les sorties à chaque itération

Après chaque itération on a les sorties suivantes :

- La valeur du critère W

$W(P^n) = W_R(P^n, g(P^n))$

- Les valeurs propres, $\lambda_1^n, \ldots, \lambda_r^n$

- Les centres de gravité des classes $P_1^n, \ldots, P_K^n$

- La classe d'appartenance dans la partition P^n de chaque individu.

Remarque : Si au cours de l'algorithme la classe de numéro ℓ devient vide, elle n'est pas reconstituée, et dans la partition P^*, la classe P_ℓ^* est vide. (K est donc le nombre maximum de classes désiré).

5.5 - Les sorties à la convergence

Après convergence de l'algorithme on a les sorties suivantes :

- La partition de convergence $P^* = (P_1^*, \ldots, P_K^*)$.

- Les coordonnées des points $g_1^*, \ldots, g_K^*$ et vecteurs $u_1^*, \ldots, u_r^*$ de la représentation $g(P^*)$, ainsi que celles des facteurs $b_1^* = Qu_1^*, \ldots, b_r^* = Qu_r^*$.

- La valeur du critère W

$W(P^*) = W_R(P^*, g(P^*))$

- Les valeurs propres $\lambda_1^*, \ldots, \lambda_r^*$.

- Les représentations graphiques des projections Q orthogonales des centres de gravité $P_1^*,\ldots,P_K^*$, et des individus sur les plans engendrés par u_1^* et u_2^*, u_3^* et u_4^*,... (si r est impair, la dernière représentation est faite avec u_r^* comme vecteur unitaire sur les 2 axes ; les points représentés sont situés sur la diagonale). On utilise le programme de représentation "NUAGE" (ISUP).

6 - EXEMPLES D'APPLICATION

6.1 - Exemple 1

6.1.1 - Les données

On dispose d'une population E de 50 individus décrits par deux paramètres x^1 et x^2 (Tableau 1 et figure 2).

L'espace $\mathbb{R}^2$ est muni du produit scalaire défini par la matrice Q suivante :

$$Q = \begin{pmatrix} \frac{1}{\sigma(x^1)} & 0 \\ 0 & \frac{1}{\sigma(x^2)} \end{pmatrix}$$

où $\sigma(x^1)$ et $\sigma(x^2)$ sont les écarts types des paramètres x^1 et x^2.

On demande une partition de E en 3 classes et un sous espace de classification de dimension 1.

On effectue plusieurs tirages initialisés par des partitions tirées au hasard.

6.1.2 - Les résultats

On retient le tirage dont la partition de convergence correspond à la plus grande valeur du critère W. Cette partition $P^* = (P_1^*, P_2^*, P_3^*)$, ainsi que la droite vectorielle U^* définie par la représentation $g(P^*)$, apparaissant sur la figure 3.

	x^1	x^2
A1	1.0	1.0
A2	.5	1.5
A3	1.5	.5
A4	1.0	2.0
A5	1.5	1.5
A6	2.0	1.0
A7	2.0	1.5
A8	1.5	2.5
A9	2.0	2.5
A10	2.5	2.0
A11	2.5	3.0
A12	3.0	2.0
A13	3.0	2.5
A14	3.0	3.5
A15	3.5	3.0
B10	2.5	7.0
B20	3.0	7.0
B30	3.0	6.0
B40	3.5	6.5
B50	3.5	6.0
B60	3.5	5.0
B70	4.0	6.5
B80	4.0	5.0
B90	4.0	4.0
B10	4.5	6.0
B11	4.5	5.5
B12	4.5	5.0
B13	5.0	6.0
B14	5.0	5.0
B15	5.0	4.0
B16	5.5	5.0
B17	5.5	4.5
B18	6.0	5.0
B19	6.0	3.5
B20	6.5	4.0
B21	7.0	3.0
C1	6.5	6.5
C2	6.5	7.0
C3	6.5	7.5
C4	7.0	6.5
C5	7.0	7.0
C6	7.0	7.5
C7	7.0	8.0
C8	7.5	6.0
C9	7.5	6.5
C10	7.5	7.0
C11	7.5	7.5
C12	7.5	8.0
C13	8.0	6.5
C14	8.0	7.5

Tableau 1 : Les données

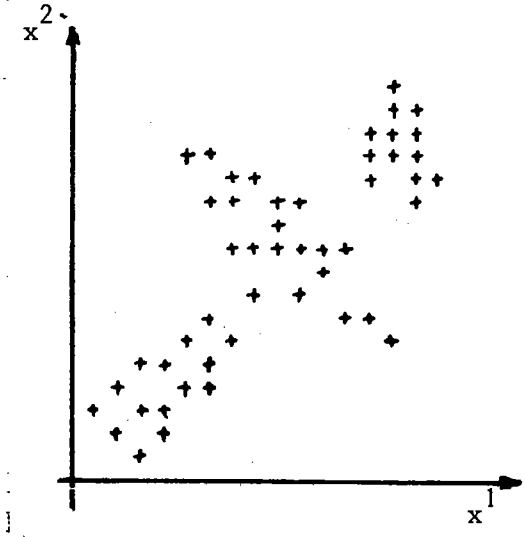


Figure 2 : Les données

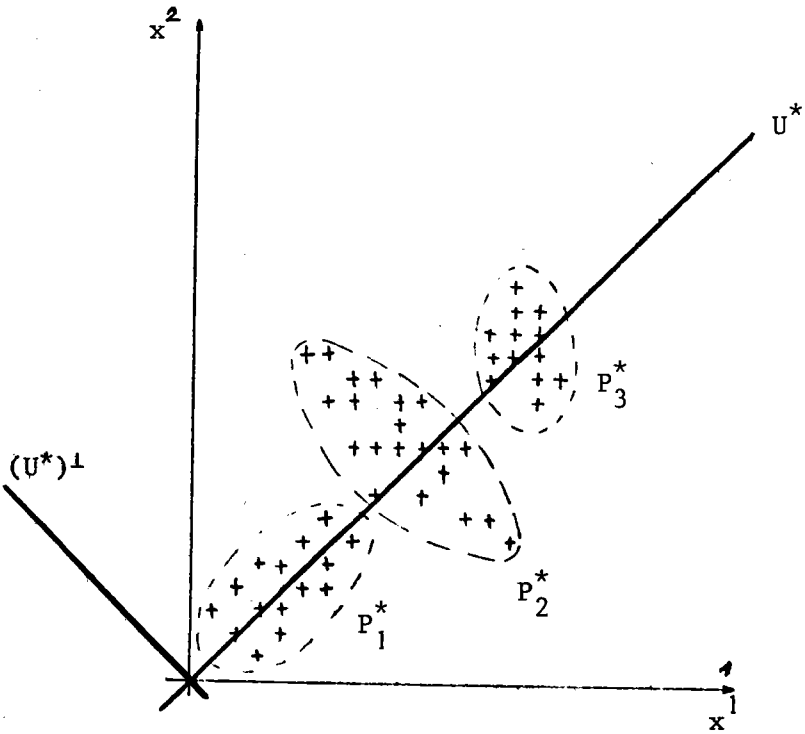


Figure 3 : La partition de convergence P^*
Les sous espaces U^* et $(U^*)^\perp$

On constate que la reconnaissance des classes P_1^*, P_2^*, P_3^* est liée à la reconnaissance du sous espace vectoriel U^*. De plus le sous espace $(U^*)^\perp$ est, relativement à la partition P^*, un sous espace de bruit (figure 3). Cette dernière propriété permet (cf. 7.2) de valider globalement (dans $\mathbb{R}^2$) la partition P^*, que la projection sur U^* a fait apparaitre.

6.2 - Exemple 2

6.2.1 - Les données

On dispose d'une population E de 30 individus décrits par 3 paramètres x^1, x^2, x^3. (Tableau 2 et figure 4).

L'espace R^3 est muni du produit scalaire défini par la matrice inverse de la matrice de covariance empirique des paramètres x^1, x^2, x^3.

On demande une partition de E en 3 classes et un sous espace de classification de dimension 2.

On effectue plusieurs tirages initialisés par des partitions tirées au hasard.

6.2.2 - Les résultats

On retient le tirage dont la partition de convergence correspond à la plus grande valeur du critère W. Cette partition $P^* = (P_1^*, P_2^*, P_3^*)$ apparait dans le tableau 3, et sur la figure 5.

On remarque :

- d'une part (tableau 4), que le sous espace U^* engendré par les vecteurs u_1^* et u_2^* de la représentation $g(P^*)$ privilégie nettement les variables x^1 et x^2.

- d'autre part (figure 4), que la variable x^3 est, relativement à la partition $P^* = (P_1^*, P_2^*, P_3^*)$, une variable de bruit.

AA 1	20.	0.	50.
AA 2	25.	20.	30.
AA 3	40.	15.	45.
AA 4	30.	10.	35.
AA 5	35.	20.	25.
AA 6	25.	5.	0.
AA 7	40.	10.	35.
AA 8	20.	15.	5.
AA 9	25.	10.	30.
AA10	30.	0.	30.
BB 1	0.	40.	55.
BB 2	15.	50.	25.
BB 3	10.	55.	50.
BB 4	5.	45.	40.
BB 5	5.	60.	10.
BB 6	15.	45.	35.
BB 7	0.	45.	60.
BB 8	10.	50.	25.
BB 9	20.	60.	40.
BB10	10.	60.	30.
CC 1	40.	35.	10.
CC 2	45.	50.	20.
CC 3	60.	30.	40.
CC 4	50.	40.	45.
CC 5	45.	40.	15.
CC 6	50.	45.	20.
CC 7	40.	30.	55.
CC 8	45.	45.	15.
CC 9	60.	35.	60.
CC10	55.	40.	45.

Tableau 2 : Les données

CLASS NO 1 CARD. 10									
AA10	AA 9	AA 8	AA 7	AA 6	AA 5	AA 4	AA 3	AA 2	AA 1
CLASS NO 2 CARD. 10									
BB10	BB 9	BB 8	BB 7	BB 6	BB 5	BB 4	BB 3	BB 2	BB 1
CLASS NO 3 CARD. 10									
CC10	CC 9	CC 8	CC 7	CC 6	CC 5	CC 4	CC 3	CC 2	CC 1

Tableau 3 : La partition P^*

$$u_1^* = -\ 12.4\ x^1 + 15.2\ x^2 + 3.5\ x^3$$

$$u_2^* = \ \ 12.5\ x^1 + \ \ 9.7\ x^2 + 1.0\ x^3$$

$$b_1^* = -\ 0.031\ x^1 + 0.038\ x^2 + 0.010\ x^3$$

$$b_2^* = \ \ 0.049\ x^1 + 0.039\ x^2 + 0.003\ x^3$$

Tableau 4 : Les vecteurs u_1^*, u_2^* de la représentation $g(P^*)$, et les facteurs $b_1^* = V^{-1}\ u_1^*$, $b_2^* = V^{-1}\ u_2^*$

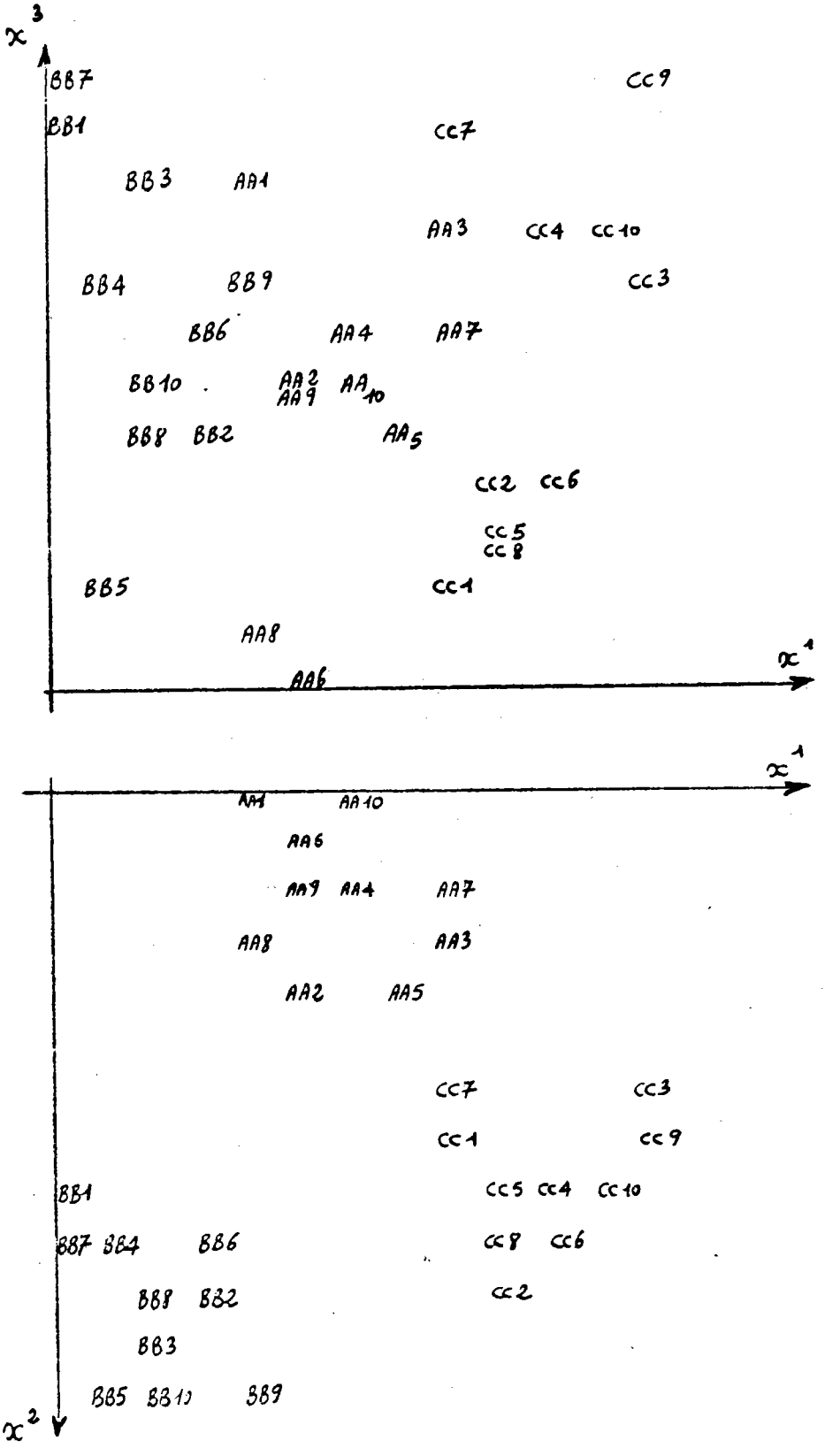


Figure 4 : Les données

7 - L'INTERPRETATION DES RESULTATS

7.1 - Optimum local, optimum global du critère W

La partition P^* de convergence de l'algorithme dépend de la partition initiale P^0. Elle vérifie, (4.6) :

$W(P^0) \leq W(P^*)$

$W(P^*)$ est un optimum local du critère W ;

En général $W(P^*)$ est différent de l'optimum du critère W pour P dans $\mathbb{P}_K$.

$$W(P^*) \neq \underset{P \in \mathbb{P}_K}{\text{Max}} \; [W(P)]$$

Cet optimum global n'est d'ailleurs pas connu.

Si on réalise m analyses (ou tirages) initialisées respectivement par les partitions P^{01}, P^{02},...,P^{0m}, on obtient m partitions de convergences P^{*1},...,P^{*m} qui réalisent m optima locaux du critère W.

La méthode des formes fortes [8] permet d'isoler des groupements d'individus classés ensemble dans les m partitions P^{*1},...,P^{*m}. De plus, l'arbre des connexités descendantes précise les proximités entre ces groupements stables [8].

On peut aussi choisir parmi les partitions P^{*1},...,P^{*m} celle qui réalise la meilleure valeur du critère W ; Cette procédure est justifiée pour toute méthode de classification qui procède par optimisation d'un critère. Dans la suite la partition chosie de cette manière après m tirages est encore notée P^*.

7.2 - La validité de la partition de convergence

Soit

P^* la partition de convergence

U^* le sous espace vectoriel engendré par les vecteurs $u_1^*,\ldots,u_r^*$ de la représentation $L^* = g(P^*)$

$(U^*)^\perp$ le supplémentaire Q orthogonal de U^* dans R^p

Il faut s'assurer que la projection du nuage E sur le sous espace $(U^*)^\perp$ ne dédie pas les regroupements d'individus $P_1^*,\ldots,P_K^*$, c'est-à-dire que :

- soit $(U^*)^\perp$ est, relativement à la partition P^*, un sous espace de bruit (cf. exemples : 6.1, 6.2),

- soit les proximités entre projections sur $(U^*)^\perp$ des individus confirment les regroupements $P_1^*,\ldots,P_K^*$.

Dans ces conditions, la partition P^* (qui réalise avec le sous espace U^* un maximum local de l'inertie inter classe expliquée par un sous espace de dimension r) est validée relativement à l'ensemble des p paramètres initiaux $x^1,\ldots,x^p$. De plus, le sous espace U^* est validé comme sous espace classifiant relativement à P^*, dans le sens où il induit (par projection) une partition non remise en cause par son supplémentaire $(U^*)^\perp$.

La stabilité de la partition P^* peut, par exemple, être testée par visualisation de la projection du nuage E sur des plans vectoriels inclus dans $(U^*)^\perp$.

Si la dimension r de U^* vérifie :

$r < K-1$

On peut aussi réaliser une seconde analyse initialisée par la partition P^*, avec pour dimension du sous espace de classification r+1.

7.3 - Le choix de la dimension r du sous espace de classification

Soit P^* une partition validée selon 7.2.

La suite des valeurs propres $\lambda_1^*,\ldots,\lambda_r^*$ et surtout (5.5), la visualisation des projections des individus sur les plans vectoriels engendrés par (u_1^*, u_2^*), etc... permettent d'apprécier "le pouvoir classifiant" des droites vectorielles engendrées par les vecteurs u_r^*, u_{r-1}^*,... et donc associés aux plus faibles valeurs propres λ_r^*, λ_{r-1}^*,...

Eventuellement, on réalise une analyse initialisée par P^*, avec pour dimension du sous espace de classification r-1 ou r-2,...

On peut de cette manière tester la stabilité de P^*, ou, plus généralement, faire apparaitre une suite de partitions de E, P_r^*, P_{r-1}^*,..., en K_r, K_{r-1},... classes, obtenues respectivement avec des sous espaces de classification de dimensions r, r-1,... et liées par la relation de finesse entre partitions, [4].

Par exemple : P_r^* partition de E en K_r classes est plus fine que P_{r-1}^* partition de E en K_{r-1} classes ($K_{r-1} < K_r$) ; P_r^* est obtenue comme partition de convergence d'une analyse initialisée par P_{r-1}^*, la dimension supplémentaire du sous espace de classification ayant permis l'éclatement de s classes ($1 \leq s \leq K_{r-1}$) en $s + (K_r - K_{r-1})$ classes.

8 - METHODE SEQUENTIELLE RELATIVEMENT AU NOMBRE DE CLASSES DE LA PARTITON

Dans ce paragraphe la dimension r du sous espace de classification, est choisie égale à K-1.

8.1 - Pas 1

Pour tout entier K, $K \geq 2$, la méthode de classification sur sous espace adaptatif associe à toute partition P^0 dans $\mathbb{P}_K$, une partition P^* dans $\mathbb{P}_K$ qui

réalise un local du critère W défini en 2.1 (avec r = K-1).

En résumé :

$C\ell_K \quad \mathbb{P}_K \longrightarrow \mathbb{P}_K$

$C\ell_K \quad P^0 \longrightarrow P^* = C\ell_K(P^0)$

8.2 - Pas 2

Soit un critère C d'homogénéité ou de séparation des classes d'une partition (C peut prendre en compte le nuage des individus dans $\mathbb{R}^p$, ou sa projection sur un sous espace vectoriel donné).

Soit un ensemble T_K de transformations T sur les partitions de E en K classes, ($K \geq 2$), qui associent à toute partition $Q = (Q_1,\ldots,Q_K)$ dans $\mathbb{P}_K$ une partition $T(Q)$ dans $\mathbb{P}_{K+1}$. Par exemple :

T_K peut être réduit à l'ensemble des transformations élementaires du type "Scinder une classe Q_i en deux sous classes".

T_K peut être l'ensemble des transformations définies par :

1) Scinder chaque classe en deux sous classes (le découpage trivial $Q_i \rightarrow (Q_i, \emptyset)$ étant permis K-1 fois au plus)

2) Réunir (par une procédure hiérarchique par exemple) les ensembles d'individus ainsi définis, de manière à constituer K+1 classes.

Soit

Une partition Q dans $\mathbb{P}_K$.

L'ensemble $T_K(Q) = \{T(Q) \mid T \in T_K\}$ inclu dans $\mathbb{P}_{K+1}$.

Une partition Q_+ dans $T_K(Q)$ qui réalise l'optimum du critère C sur $T_K(Q)$.

On note Cr_K la procédure qui associe ainsi à toute partition Q dans $\mathbb{P}_K$ une partition Q_+ dans $\mathbb{P}_{K+1}$.

Remarque : Pour toute partition $Q^* = C\ell_K(Q^0)$ on dispose (5.5) des graphiques des projections des individus sur le sous espace U^* défini par la représentation $L^* = g(Q^*)$. L'examen de cette visualisation des classes de la partition Q^* est une autre procédure de définition d'une partition en K+1 classes déduite de Q^*.

8.3 - L'algorithme

On suppose donnée ou tirée au hasard une partition P_2^0 de E en 2 classes.

Pour tout $K \geq 2$, on pose

$P_K^* = C\ell_K(P_K^0)$ (Pas 1)

$P_{K+1}^0 = Cr_K(P_K^*)$ (Pas 2)

Le test d'arrêt de l'algorithme est constitué par la donnée d'un nombre maximum de classes, ou (et) est inclu dans la procédure de découpage des classes.

9 - CONCLUSION

La classification sur sous espace adaptatif permet la réduction de la dimension de l'espace des descriptions des objets dans l'optique de la classification.

La visualisation de la partition de convergence sur quelques axes factoriels donne une image de la structure multidimensionnelle du nuage. Il apparaît des groupements d'individus homogènes en projection sur le sous espace ; une procédure d'interaction "homme-machine" (avec augmentation simultanée du nombre K de classes de la partition, et de la dimension r du sous espace de représentation) permet alors la construction séquentielle d'une partition en K^* classes expliquée par un sous espace de dimension r^*.

Le sous espace de représentation pourrait être défini selon un critère différent de l'inertie inter classe expliquée [30].

Si on impose au sous espace de classification d'être engendré par r variables initiales, on fait de la sélection de paramètres (éventuellement pas à pas).

CHAPITRE II

ANALYSE DISCRIMINANTE TYPOLOGIQUE

1 - INTRODUCTION

1.1 - Le problème de la discrimination

Un ensemble $\mathcal{E}$ d'individus ou objets est réparti en K parties (appelées familles) $\mathcal{C}_1,\ldots,\mathcal{C}_K$ définies a priori. On dispose d'autre part d'une description multidimensionnelle des objets. (On suppose dans ce travail que les paramètres sont quantitatifs. Dans le cas de données qualitatives on peut avoir recours à un codage préalable [26]).

Il s'agit de décrire le pouvoir séparateur des paramètres descriptifs des objets vis à vis des familles a priori (Analyse discriminante descriptive); Puis, éventuellement, de construire une règle d'identification de la famille d'appartenance d'un objet (Analyse discriminante décisionnelle).

On dispose d'une population échantillon finie E de $\mathcal{E}$, pour laquelle le groupe d'appartenance des objets est connu. L'ensemble des objets de E qui sont issus de la famille $\mathcal{C}_\ell$ constitue un sous ensemble de E noté T_ℓ : $T_1,\ldots,T_K$ sont les classes d'une partition de E, dite partition a priori.

1.2 - L'analyse factorielle discriminante : Une analyse descriptive linéaire [25,29]

On dispose du nuage d'objets E dans $\mathbb{R}^p$ (p est le nombre de paramètres) répartis en K classes. $\mathbb{R}^p$ est muni de la métrique définie par V^{-1}, inverse de la matrice de variance covariance empirique des paramètres.

On recherche l'axe (puis le plan,...) dans R^p qui sépare le mieux les K classes en projection, au sens du maximum de l'inertie inter classe expliquée.

La visualisation des projections des classes $T_1,\ldots,T_K$ sur le sous espace discriminant ainsi défini permet d'apprécier le pouvoir séparateur de l'ensemble des p paramètres.

Avec l'analyse factorielle discriminante, on tient compte des liens linéaires entre variables initiales pour séparer les K classes a priori.

1.3 - L'analyse discriminante décisionnelle linéaire [25, 29]

Il s'agit de la règle d'affectation au centre le plus proche pour une distance quadratique. Elle consiste à définir par des séparations hyperplanes K zones dans R^p, chaque région étant associée à une classe (Dans le cas de deux classes on a par exemple l'hyperplan de FISHER [13]). Le choix de la métrique de Mahalanobis (définie par l'inverse de la matrice de variance covariance intra classe de la partition a priori) se justifie par des considérations géométriques et probabilistes [29]; Par contre la métrique V^{-1} est liée à l'analyse factorielle discriminante et présente l'avantage d'être indépendante de la partition a priori. (Pour ces deux raisons, dans l'analyse discriminante typologique on a choisi la métrique V^{-1}).

1.4 - La remise en cause des classes a priori

La définition d'une partition a priori de la population échantillon E constitue une hypothèse de structure. Il est souhaitable de pouvoir s'en affranchir localement, afin de prendre en compte les groupements homogènes d'objets distincts des classes a priori. Cette remise en cause de la partition a priori se justifie d'ailleurs dans de nombreux cas, et ce pour différentes raisons :

a) Le problème posé peut être étudié selon différentes approches. Par exemple :

En contrôle de production, les types a priori peuvent être définis par des défauts de fonctionnement ou par des défauts de fabrication.

En sciences de la nature une étude peut être abordée de différents points de vue : Physique, chimie, Biologie ...

En médecine, il faut distinguer le symptôme et la cause.

Les types a priori doivent induire une partition de la population (en classes disjointes), et leur définition implique en général le choix d'une seule approche. Sinon on multiplie le nombre de classes.

Ce choix, parfois subjectif, doit pouvoir être remis en cause. En effet, il est possible qu'un groupement d'individus C, homogène et (ou) isolé dans la population échantillon E, et distinct de toute classe a priori, soit interprétable selon une approche différente. Par exemple, dans le cas de la banque de données industrielles dont l'étude est présentée dans ce chapitre, le spécialiste définit la partition a priori selon les types de défauts de fonctionnement; L'analyse discriminante typologique met en évidence un groupement d'individus qui s'interprête facilement selon l'approche "défaut de fabrication".

b) Les frontières entre les classes a priori peuvent être arbitraires (déterminées par des seuils) ou imprécises. Par exemple dans le cas de 2 classes du type "présence" "absence" il n'existe pas, en général, de frontière objective et nette. La définition des classes a priori est alors soumise au choix d'un seuil numérique, si le phénomène étudié est mesurable; sinon, le classement des individus, (ceux de la population échantillon en particulier) résulte d'une interprétation humaine. Cette dernière n'est pas toujours indépendante de l'expérimentateur, ni de l'ordre d'apparition des individus. Par exemple, pour la détection de défauts de fonctionnement de pièces mécaniques par écoute du bruit émis (cf. § 6) il n'existe pas de références précises pour l'affectation à une classe. Elle dépend en partie de comparaisons entre individus, (Un bruit très léger sera mal perçu après l'écoute d'un bruit marqué); En particulier les premiers objets étudiés prennent l'importance de références.

Pour ces raisons, les frontières entre classes doivent parfois être remises en cause. Plus généralement il peut apparaître que la partition a priori est trop fine, ou au contraire trop grossière, dans le sens où aux zones frontières entre classes a priori il est souhaitable de faire correspondre de nouvelles classes (portant éventuellement l'étiquette "indécision").

1.5 - La prise en compte des classes a priori

La donnée de la partition a priori apporte cependant une information dont il est souhaitable de tenir compte pour apprécier la séparabilité des K familles dans $\mathbb{R}^p$.

On sait par exemple, que l'analyse factorielle discriminante (1,2) induit une meilleure séparation des classes a priori que des méthodes inductives (l'analyse en composantes principales, [2], ou l'analyse des correspondances [1]). Comme le montrent les figures 1 et 2 dans le cas de la séparation de deux classes, (les individus dont décrits par deux variables x^1 et x^2), les sous espaces discriminants et principaux peuvent être totalement différents. On sait, [29], que l'analyse factorielle discriminante est équivalente à l'analyse en composantes principales du nuage G des centres de gravité des classes munis des poids des classes, pour la métrique définie par l'inverse de la matrice de covariance empirique des paramètres. Les axes factoriels d'inertie minimum du nuage G, qui apportent peu d'information pour la discrimination, peuvent être des axes de forte dispersion du nuage E : Dans le cas ils nuisent à la reconnaissances des classes a priori

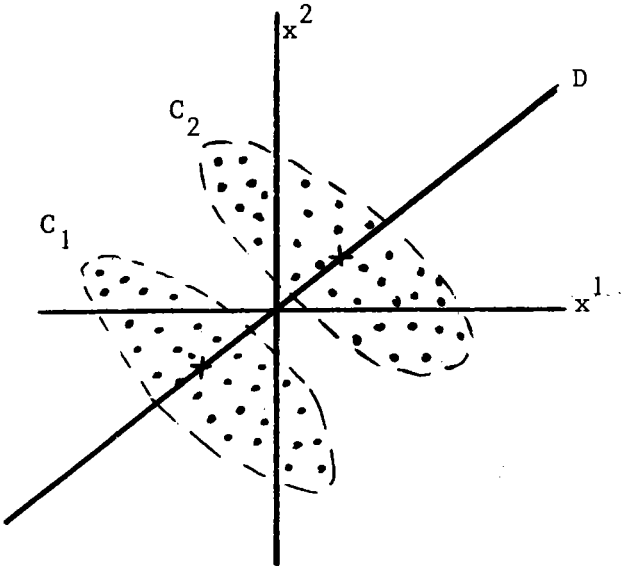


Figure 1

D : le premier axe de l'analyse factorielle discriminante relative à la partition (C_1, C_2) du nuage E.

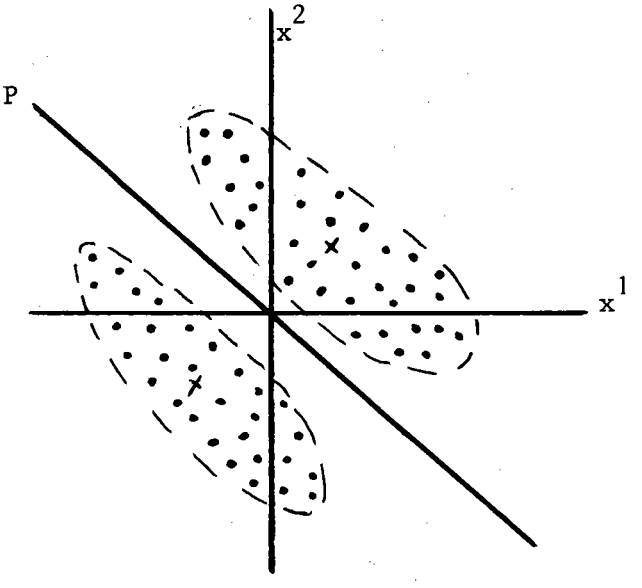


Figure 2

P : le premier axe de l'analyse en composantes principales du nuage E réunion des classes C_1 et C_2.

1.6 - L'analyse discriminante typologique (A.D.T.)

L'analyse discriminante typologique propose une démarche médiane entre discrimination et classification, alors que ces deux domaines sont généralement opposés. La partition a priori est remise en cause localement, afin de tenir compte de la structure réelle du nuage des individus dans $\mathbb{R}^p$. Le but recherché est la construction d'une partition déduite de celle des types a priori dont les classes soient plus homogènes ou mieux séparées, dans l'optique de la discrimination ; En effet, en général certaines classes a priori sont homogènes et (ou) isolées, et leur reconnaissance est aisée, d'autres sont très dispersées ou même non séparables par la méthode utilisée [21] : par exemple, en reconnaissance des caractères A et B sont bien séparés, ce qui n'est pas le cas pour U et V. Ces dernières classes nuisent à une reconnaissance fiable des premières.

L'espace $\mathbb{R}^p$ des objets est muni pour toute l'analyse de la métrique V^{-1}. Par conséquent l'A.D.T. se distingue des méthodes qui utilisent

- une métrique fixée mais qui dépend d'une partition de E (la métrique de Mahalanobis par exemple).

- une métrique adaptative [12]

- des métriques locales fixées [27]

- des métriques locales adaptatives [14]

L'A.D.T procède à une transformation linéaire adaptative des variables, afin de séparer information utile et nuisible à la discrimination [7]. L'A.D.T se distingue ainsi :

- d'une part des méthodes ayant recours à la sélection de paramètres [7, 10, 20]
- d'autre part des méthodes autorisant la transformation non linéaire des variables [15].

Du point de vue de l'analyse discriminante décisionnelle, l'A.D.T. procède par affectation au centre le plus proche (pour la métrique V^{-1}) en projection sur le sous espace induit par la transformation des variables (seule l'information utile est prise en compte). Il ne s'agit pas de la règle de décision

classique de l'analyse discriminante linéaire (1.3), dont l'interprétation comme un cas particulier de la règle Bayesienne [21 ,29] fait apparaître les insuffisances du point de vue probabiliste.

L'A.D.T. procède par optimisation d'un critère. Elle met en oeuvre alternativement une technique de Discrimination et une technique de Classification.

Elle induit une partition P*. Une classe de cette nouvelle partition distincte de toute classe a priori, peut cependant être interprétable par le spécialiste comme on l'a vu ci-dessus. En particulier il peut s'agir d'une sous classe homogène d'une classe a priori dispersée, ou constituée de plusieurs ilots. Dans le cas où la méthode induit des réunions de classes a priori, on procède éventuellement à un deuxième niveau de discrimination (Si toutefois le pouvoir discriminant des paramètres n'est pas en cause). En effet, dans certains cas une perte de finesse relativement à la partition a priori est acceptable pour le spécialiste (cf. § 6).

De plus, on obtient

- une règle de décision simple, relative à la partition construite, issue directement de la méthode.

- une représentation sur quelques axes factoriels discriminants de la population échantillon, qui illustre l'éclatement ou la réunion de classes a priori (§ 6.6).Cette représentation permet aussi le classement de nouveaux individus par visualisation sur le même sous espace Dans ces conditions, pour le classement d'un individu dont la projection se situe dans une zone frontière entre classes on peut tenir compte a posteriori d'un coût de mauvaise affectation. Il peut aussi être créé une zone de non décision, afin d'assurer la fiabilité du classement.

1.7 - Les données

Les individus de E sont munis de poids égaux :

$$p_i = \frac{1}{n} \qquad i=1,\ldots,n$$

Ils sont décrits par un ensemble d de p paramètres quantitatifs :

$$x^1, \ldots, x^p.$$

On note x_i^j la valeur prise par le paramètre x^j sur l'individu x_i. On note y_i le numéro de la classe de la partition a priori qui contient l'individu x_i.

Les données peuvent être rangées dans une matrice X' comportant n lignes et (p+1) colonnes

$$X' = \begin{pmatrix} x_1^1 & \cdots & x_1^j & \cdots & x_1^p & y_1 \\ \vdots & & \vdots & & \vdots & \vdots \\ x_i^1 & \cdots & x_i^j & \cdots & x_i^p & y_i \\ \vdots & & \vdots & & \vdots & \vdots \\ x_n^1 & \cdots & x_n^j & \cdots & x_n^p & y_n \end{pmatrix}$$

On identifie un individu x_i, au vecteur de R^p des valeurs prises par les p paramètres sur cet individu. On note :

$$x_i' = (x_i^1, \ldots, x_i^p)$$

L'espace R^p est muni du produit scalaire défini par la matrice symétrique définie positive V^{-1}, inverse de la matrice d'inertie du nuage $(x_1,\ldots,x_n)$ de R^p

$$V = \frac{1}{n} \bar{X}\, \bar{X}'$$

où

$$\bar{X} = \begin{pmatrix} x_1^1 - \bar{x}^1 & \cdots\cdots & x_1^p - \bar{x}^p \\ \vdots & & \vdots \\ x_n^1 - \bar{x}^1 & \cdots\cdots & x_n^p - \bar{x}^p \end{pmatrix}$$

en notant $\bar{x}^1,\ldots,\bar{x}^p$ les moyennes sur E des paramètres $x^1,\ldots,x^p$.

V est aussi la matrice de covariance empirique des paramètres $x^1,\ldots,x^p$.

(Si V n'est pas régulière, X n'est pas de rang p, et on se restreint à un sous ensemble de l'ensemble des paramètres J).

On suppose choisi la dimension r ($r \leq K-1$) du sous espace discriminant (cf. § 5.1)

1.8 - Notations

On note :

$\mathbb{P}_K$ l'ensemble des partitions de E en K classes (disjointes)

Une partition P de $\mathbb{P}_K$ est notée également $(P_1,\ldots,P_K)$

$E = P_1 \cup \ldots \cup P_K$

$P_i \cap P_j = \emptyset \quad i = 1,\ldots,K \; ; \quad j = 1,\ldots,K \; ; \quad i \neq j$

F_r l'ensemble des sous espaces vectoriels de $\mathbb{R}^p$ de dimension r

U_r l'ensemble des familles V^{-1} orthonormées de r vecteurs de $\mathbb{R}^p$.

2 - LE CRITERE

2.1.- Définition

Soit P une partition de E,

Pour tout sous espace U de $\mathbb{R}^p$, soit $I_{intra}(P,U)$ l'inertie intra classe de la partition P, expliquée par U.

On définit un critère W sur l'ensemble des partitions de E par :

$$W(P) = \min_{U \in F_r} I_{intra}(P,U) \qquad (2\text{-}1)$$

W mesure l'homogénéité moyenne des classes de la partition, en projection sur un sous espace de dimension r.

2.2 - L'optimisation du critère

On recherche une partition de E en K classes, P_K^* qui réalise le minimum du critère W pour P dans $\mathbb{P}_K$:

$$W(P_K^*) = \min_{P \in \mathbb{P}_K} W(P) \qquad (2\text{-}2)$$

Remarque : La recherche du minimum du critère dans l'ensemble $\mathbb{P}$ des partitions de E, conduit à la partition triviale en n classes réduites à un individu, P_E. En effet, pour tout sous espace U de $\mathbb{R}^p$, l'inertie intra classe de la partition E, expliquée par U est nulle.

$$W(P_E) = \min_{P \in \mathbb{P}} W(P) = 0$$

2.3 - Le critère et l'inertie inter classe expliquée

Pour toute partition P de E, et pour tout sous espace U de $\mathbb{R}^p$, la somme des inerties intra classe et inter classe de P, expliquées par U, est égale à l'inertie du nuage E, expliquée par U, (0-6).

$$I_{intra}(P,U) + I_{inter}(P,U) = I(E,U)$$

Pour toute base V^{-1} orthonormée, $(u_1,\ldots,u_r)$, du sous espace U, on a, d'après (O-2) et (O-3) :

$$I(E,U) = \sum_{s=1}^{r} u_s' V^{-1} V V^{-1} u_s$$

$$I(E,U) = \sum_{s=1}^{r} u_s' V^{-1} u_s = r \qquad (2\text{-}3)$$

On en déduit :

$$I_{intra}(P,U) = r - I_{inter}(P,U) \qquad (2\text{-}4)$$

et d'après (2-1),

$$W(P) = r - \operatorname*{Max}_{U \in F_r} (I_{inter}(P,U)) \qquad (2\text{-}5)$$

L'optimisation relativement à P des critères W et $\operatorname*{Max}_{U\in F_r} (I_{inter}(P,U))$, le critère retenu au chapitre I, sont donc équivalentes.

Cette propriété est liée au choix du produit scalaire défini par la matrice V^{-1}. Dans les chapitres II et III on retient le critère du minimum de l'inertie intra classe expliquée. On utilise cependant ce critère sous les deux formes équivalentes :

$$W(P) = \operatorname*{Min}_{U \in F_r} (I_{intra}(P,U))$$

$$W(P) = r - \operatorname*{Max}_{U \in F_r} (I_{inter}(P,U))$$

2-4 - Expression mathématique du critère

Soit $P = (P_1,\ldots,P_K)$ une partition dans $\mathbb{P}_K$, $g_1,\ldots,g_K$ les centres de gravité des classes $P_1,\ldots,P_K$, U un sous espace vectoriel de $\mathbb{R}^p$.

On a, d'après le paragraphe 6 (Proposition 4) de l'avant propos :

$$I_{intra}(P,U) = \frac{1}{n} \sum_{\ell=1}^{k} \sum_{x \in P_\ell} d_U^2(x,g_\ell) \qquad (2\text{-}6)$$

où d_U est l'écart sur $\mathbb{R}^p$ défini par :

$$d_U(x,y) = \sqrt{(x_U - y_U)' \, V^{-1} \, (x_U - y_U)}$$

où x_U et y_U sont les projections V^{-1} orthogonales de x et y sur le sous espace U.

On montre au paragraphe 6 (Proposition 3) de l'avant propos que pour toute base V^{-1} orthonormée, $(u_1,\ldots,u_r)$, du sous espace U, on a :

$$d_U^2(x,y) = (x-y)' \, V^{-1} \, \left(\sum_{s=1}^{r} u_s \, u_s' \right) V^{-1} \, (x-y) \qquad (2\text{-}7)$$

On en déduit :

$$I_{intra}(P,U) = \frac{1}{n} \sum_{\ell=1}^{k} \sum_{x \in P_\ell} (x-g_\ell)' \, V^{-1} \left(\sum_{s=1}^{r} u_s \, u_s' \right) V^{-1} \, (x-g_\ell) \qquad (2\text{-}8)$$

pour toute base V^{-1} orthonormée $(u_1,\ldots,u_r)$ du sous espace U, soit

$$I_{intra}(P,U) = \frac{1}{n} \sum_{\ell=1}^{k} \sum_{x \in P_\ell} \sum_{s=1}^{r} (x-g_\ell)' \, V^{-1} \, u_s \, u_s' \, V^{-1} \, (x-g_\ell) \qquad (2\text{-}9)$$

Les sommations étant finies, elles peuvent être permutées. De plus, comme $(x-g_\ell)' \, V^{-1} \, u_s$ et $u_s' \, V^{-1} \, (x-g_\ell)$ sont des scalaires,

$$(x-g_\ell)' \, V^{-1} \, u_s \, u_s' \, V^{-1} \, (x-g_\ell) = u_s' \, V^{-1} \, (x-g_\ell)(x-g_\ell)' \, V^{-1} \, u_s$$

On en déduit :

$$I_{intra}(P,U) = \frac{1}{n} \sum_{s=1}^{r} \sum_{\ell=1}^{K} \sum_{x \in P_\ell} u_s' \, V^{-1} \, (x-g_\ell)(x-g_\ell)' \, V^{-1} \, u_s$$

ou encore

$$I_{intra}(P,U) = \frac{1}{n} \sum_{s=1}^{r} u_s' V^{-1} \left(\sum_{\ell=1}^{K} \sum_{x \in P_\ell} (x-g_\ell)(x-g_\ell)' \right) V^{-1} u_s \qquad (2\text{-}10)$$

Des relations (2-8), (2-9) et (2-10), on déduit les expressions suivantes du critère W :

Pour toute partition $P = (P_1,\ldots,P_K)$ dans $\mathbb{P}_K$, (On note $g_1,\ldots,g_K$ les centres de gravité des classes $P_1,\ldots,P_K$, et, pour tout sous espace U de $\mathbb{R}^p$, $(u_1,\ldots,u_r)$ une base V^{-1} orthonormée quelconque de U). On a :

$$W(P) = \underset{U \in F_r}{\text{Min}} \left(\frac{1}{n} \sum_{\ell=1}^{K} \sum_{x \in P_\ell} (x-g_\ell)' V^{-1} \left(\sum_{s=1}^{r} u_s u_s' \right) V^{-1} (x-g_\ell) \right)$$

$$W(P) = \underset{U \in F_r}{\text{Min}} \left(\frac{1}{n} \sum_{s=1}^{r} \sum_{\ell=1}^{K} \sum_{x \in P_\ell} u_s' V^{-1} (x-g_\ell)(x-g_\ell)' V^{-1} u_s \right)$$

$$W(P) = \underset{U \in F_r}{\text{Min}} \left(\frac{1}{n} \sum_{s=1}^{r} u_s' V^{-1} \left(\sum_{\ell=1}^{K} \sum_{x \in P_\ell} (x-g_\ell)(x-g_\ell)' \right) V^{-1} u_s \right)$$

2.5 - Le critère et les facteurs discriminants

2.5.1 - Facteur associé à un vecteur de $\mathbb{R}^p$

On sait (Ann.2.5) que pour tout vecteur u, V^{-1} normé, de $\mathbb{R}^p$, on peut identifier le vecteur $b=V^{-1}u$ à la forme linéaire qui associe à tout vecteur y de $\mathbb{R}^p$ sa projection V^{-1} orthogonale sur la droite engendrée par u, égale à $y' V^{-1} u$.

Si on note

$b' = (b_1,\ldots,b_p)$

$y' = (y_1,\ldots,y_p)$

On a :

$$b : \mathbb{R}^p \to \mathbb{R}$$

$$b : y \to y' b = \sum_{j=1}^{p} y_j b_j$$

Pour tout individu x_i de E, i=1,...,n,

$$x_i = (x_i^1,...,x_i^p) = (x^1(x_i),...,x^p(x_i))$$

on a :

$$x_i' b = b' x_i = \sum_{j=1}^{p} b_j x^j(x_i)$$

On peut identifier b à la combinaison linéaire des paramètres x^j, $\sum_{j=1}^{p} b_j x^j$. b est appelé un facteur.

2.5.2 - Propriétés d'une famille de facteurs associée à une famille V^{-1} orthonormée de vecteurs.

Pour toute famille V^{-1} orthonormée de vecteurs de $\mathbb{R}^p$, $(u_1,...,u_r)$, on pose :

$$b_s = V^{-1} u_s \qquad s=1,...,r \tag{2-11}$$

et

$$b_s' = (b_{s,1},...,b_{s,p}) \qquad s=1,...,r$$

b_s est le facteur défini par :

$$b_s = \sum_{j=1}^{p} b_{s,j} x^j$$

Des relations

$$u_s' V^{-1} u_s = 1 \qquad s=1,...,r$$

$$u_s' V^{-1} u_t = 0 \qquad s=1,...,r; \quad t=1,...,r; \quad s \neq t$$

On déduit :

$$b_s' V b_s = u_s' V^{-1} V V^{-1} u_s = 1 \qquad s=1,\ldots,r$$

$$b_s' V b_t = u_s' V^{-1} V V^{-1} b_t = 0 \qquad s=1,\ldots,r \; ; \; t=1,\ldots,r \; ; \; s\neq t$$

La famille $(b_1,\ldots,b_r)$ est V orthonormée.

On montre, [25], que $b_s' V b_s$ est la variance empirique du facteur b_s, et $b_s' V b_t$ la covariance empirique des facteurs b_s et b_t.

Les facteurs $b_1,\ldots,b_r$ sont de variance unité et de covariance nulle.

2.5.3 - Bijection entre l'ensemble des familles V^{-1} orthonormées dans $\mathbb{R}^p$, et l'ensemble des familles V orthonormée dans $\mathbb{R}^p$

D'après le paragraphe 2.5.2, la transformation $\left(u \to b = V^{-1} u\right)$ fait correspondre à toute famille V^{-1} orthonormée $(u_1,\ldots,u_r)$, une famille V orthonormée, $(b_1,\ldots,b_r)$.

Réciproquement, la transformation $\left(b \to u = V b\right)$ fait correspondre à toute famille V orthonormée $(b_1,\ldots,b_r)$, une famille V^{-1} orthonormée, $(u_1,\ldots,u_r)$.

En effet, des relations

$$b_s' V b_s = 1 \qquad s=1,\ldots,r$$

$$b_s' V b_t = 0 \qquad s=1,\ldots,r \; ; \; t=1,\ldots,r \; ; \; s\neq t$$

On déduit :

$$u_s' V^{-1} u_s = b_s' V V^{-1} V b_s = 1 \qquad s=1,\ldots,r$$

$$u_s' V^{-1} u_t = b_s' V V^{-1} V b_t = 0 \qquad s=1,\ldots,r \; ; \; t=1,\ldots,r \; ; \; s\neq t$$

2.5.4 - Le critère W et la variance intra classe des facteurs

D'après (2-6) et (2-7), pour toute partition P, on a :

$$I_{intra}(P,U) = \frac{1}{n} \sum_{\ell=1}^{k} \sum_{x \in P_\ell} d_U^2(x,g_\ell)$$

où $d_U^2(x,y)$ s'écrit, pour toute base V^{-1} orthonormée, $(u_1,\ldots,u_r)$, du sous espace U :

$$d_U^2(x,y) = (x-y)' V^{-1} \left(\sum_{s=1}^{r} u_s u_s' \right) V^{-1} (x-y)$$

$$d_U^2(x,y) = \sum_{s=1}^{r} (x-y)' V^{-1} u_s u_s' V^{-1} (x-y)$$

Soit $(b_1,\ldots,b_r)$, la famille de facteurs associée à la base $(u_1,\ldots,u_r)$ du sous espace U par :

$$b_s = V^{-1} u_s \qquad s=1,\ldots,r$$

On a :

$$d_U^2(x,y) = \sum_{s=1}^{r} (x-y)' b_s b_s' (x-y)$$

$$d_U^2(x,y) = \sum_{s=1}^{r} [(x-y)' b_s]^2$$

On en déduit une expression de l'inertie $I_{intra}(P,U)$:

$$I_{intra}(P,U) = \frac{1}{n} \sum_{\ell=1}^{k} \sum_{x \in P_\ell} \sum_{s=1}^{r} [(x-g_\ell)' b_s]^2$$

$$I_{intra}(P,U) = \sum_{s=1}^{r} \frac{1}{n} \sum_{\ell=1}^{k} \sum_{x \in P_\ell} [(x-g_\ell)' b_s]^2$$

$\frac{1}{n} \sum_{x \in P_\ell} [(x-g_\ell)' b_s]^2$ est la variance du facteur b_s dans la classe P_ℓ

La somme de ces variances pour $\ell=1,\ldots,K$,

$$\frac{1}{n} \sum_{\ell=1}^{K} \sum_{x \in P_\ell} [(x-g_\ell)' b_s]^2$$

est la variance intra classe du facteur b_s [25] : On la note $Var_{intra}(P,b_s)$.

On a ainsi, pour toute partition P, tout sous espace U, et toute base V^{-1} orthonormée $(u_1,\ldots,u_r)$ de U :

$$I_{intra}(P,U) = \sum_{s=1}^{r} Var_{intra}(P,b_s) \qquad (2\text{-}12)$$

où b_s est le facteur défini par $b_s = V^{-1} u_s$, s=1,...,r.

En particulier pour tout vecteur u de R^p, l'inertie intra classe expliquée par la droite vectorielle U engendrée par u est égale à la variance intra classe du facteur $b = V^{-1}u$

$$I_{intra}(P,U) = Var_{intra}(P,b)$$

Soit F_r^*, l'ensemble des familles V orthonormées de r facteurs sur $\mathbb{R}^p$ D'après 2.5.3, l'application $u \to b = V^{-1}u$ est une bijection de F_r sur F_r^*. On en déduit :

$$W(P) = \min_{(b_1,\ldots,b_r)\in F_r^*} \sum_{s=1}^{r} Var_{intra}(P,b_s)) \qquad (2\text{-}13)$$

2.5.5 - Le critère W et la variance inter classe de facteurs

D'après (2-5), pour toute partition P, on a :

$$W(P) = r - \max_{U\in F_r} (I_{inter}(P,U))$$

Soit B(P), la matrice d'inertie inter classe de la partition P; Pour toute base V^{-1} orthonormée $(u_1,\ldots,u_r)$ du sous espace U on a d'après (0-7) :

$$I_{inter}(P,U) = \sum_{s=1}^{r} u_s' V^{-1} B(P) V^{-1} u_s$$

Soit $(b_1,\ldots,b_r)$, la famille de facteurs associée à la base $(u_1,\ldots,u_r)$ du sous espace U, par

$$b_s = V^{-1} u_s \qquad s=1,\ldots,r$$

On a :

$$I_{inter}(P,U) = \sum_{s=1}^{r} b_s' \, B(P) \, b_s$$

$b_s' \, B(P) \, b_s$ est la variance inter classe du facteur b_s, [25]; On la note $Var_{inter}(P,b_s)$.

On a ainsi :

$$I_{inter}(P,U) = \sum_{s=1}^{r} Var_{inter}(P,b_s) \qquad (2\text{-}14)$$

En particulier pour tout vecteur u de R^p l'inertie inter classe expliquée par la droite vectorielle U engendrée par u est égale à la variance inter classe du facteur $b = V^{-1} u$

$$I_{inter}(P,U) = Var_{inter}(P,b)$$

Du paragraphe 2.5.3 et de la relation (2-14) on déduit :

$$W(P) = r - \max_{(b_1,..,b_r)\in F_r^*} \sum_{s=1}^{r} Var_{inter}(P,b_s) \qquad (2\text{-}15)$$

2.5.6 - Variances intra classe, inter classe et totale d'un facteur : Propriété - Optimisation

D'après (10-3), pour tout vecteur u de R^p, l'inertie du nuage E expliquée par la droite vectorielle engendrée par u s'écrit :

$$I(E,U) = u' \, V^{-1} \, V \, V^{-1} \, u \qquad (= u' \, V^{-1} \, u)$$

Soit le facteur $b = V^{-1} u$. On a

$$I(E,U) = b' \, V \, b$$

$b' \, V \, b$ est la variance du facteur b sur E, ou variance totale du facteur b. On la note $Var(b)$.

De la décomposition de l'inertie expliquée par une droite vectorielle en inerties intra classe et inter classe expliquées

$$I(E,U) = I_{intra}(P,U) + I_{inter}(P,U)$$

et des relations

$$I(E,U) = \text{Var } b$$

$$I_{intra}(P,U) = \text{Var}_{intra}(P,b)$$

$$I_{inter}(P,U) = \text{Var}_{inter}(P,b)$$

On déduit la décomposition de la variance d'un facteur en variance intra classe et inter classe :

$$\text{Var}_{intra}(P,b) + \text{Var}_{inter}(P,b) = \text{Var } (b) \qquad (2\text{-}16)$$

On a, [25] :

$$\text{Var } (b) = b' \, V \, b.$$

Si le facteur b est V normé, ($b' \, V \, b = 1$), on déduit de (2-16) la relation suivante :

$$\text{Var}_{intra}(P,b) + \text{Var}_{inter}(P,b) = 1 \qquad (2\text{-}17)$$

Plus généralement, si Var (b) $\neq 0$, on déduit de (2.16) la relation suivante :

$$\frac{\text{Var}_{intra}(P,b)}{\text{Var } (b)} + \frac{\text{Var}_{inter}(P,b)}{\text{Var } (b)} = 1 \qquad (2\text{-}18)$$

Ces deux rapports sont indépendants de la norme du facteur b, puisque les fonctions

$b \rightarrow Var_{intra}(P,b)$

$b \rightarrow Var_{inter}(P,b)$

$b \rightarrow Var\ (b)$

sont homogènes de degré 2 [25]

2.5.7 - Les facteurs discriminants d'une partition

Des relations (2-17) et (2-18), on déduit que pour toute partition P de E, les problèmes a), b), c), et d) sont équivalents :

a) Rechercher un facteur V normé qui minimise $Var_{intra}(P,b)$

b) Rechercher un facteur V normé qui maximise $Var_{inter}(P,b)$

c) Rechercher un facteur non constant qui minimise le rapport $\frac{Var_{intra}(P,b)}{Var(b)}$

d) Rechercher un facteur non constant qui maximise le rapport $\frac{Var_{inter}(P,b)}{Var\ (b)}$

Plus précisément :

Les problèmes a) et b) ont les mêmes solutions

Les problèmes c) et d) ont les mêmes solutions

Toute solution de a) et b) est solution de c) et d). De toute solution de c) et d), on peut déduire par V normalisation une solution de a) et b).

Définition : Tout facteur V normé, b_1, solution de a), b), c) et d) est appelé premier facteur discriminant de la partition P.

On définit les facteurs discriminants d'ordre supérieur par récurrence, en imposant une contrainte de V orthogonalité :

Si les facteurs discriminants $b_1,\ldots,b_{s-1}$, d'ordre $1,\ldots,s-1$, sont définis, tout facteur V normé b_s, qui réalise les conditions équivalentes $C_{1,s}$ et $C_{2,s}$ est appelé facteur discriminant d'ordre s de la partition P.

$C_{1,s}$: b_s réalise le minimum de la variance intra classe, $Var_{intra}(P,b)$, sous contrainte de V orthogonalité avec $b_1,\ldots,b_{s-1}$

$C_{2,s}$: b_s réalise le maximum de la variance inter classe, $Var_{inter}(P,b)$, sous contrainte de V orthogonalité avec $b_1,\ldots,b_{s-1}$.

Le facteur b_s vérifie aussi les conditions équivalentes $C_{3,s}$ et $C_{4,s}$.

$C_{3,s}$: b_s réalise le minimum du rapport $\dfrac{Var_{intra}(P,b)}{Var(b)}$, sous contrainte de V orthogonalité avec $b_1,\ldots,b_{s-1}$.

$C_{4,s}$: b_s réalise le maximum du rapport $\dfrac{Var_{inter}(P,b)}{Var(b)}$, sous contrainte de V orthogonalité avec $b_1,\ldots,b_{s-1}$.

2.5.8 - Recherche des facteurs discriminants

De la relation (2-12) et du fait que l'application $u \to b = V^{-1}u$ est une bijection de F_r sur F_r^*, on déduit que pour toute partition P, et tout r, $1 \leq r \leq K-1$

$$\underset{U\in F_r}{\text{Min}} \; I_{intra}(P,U) = \underset{(b_1,\ldots,b_r)\in F_r^*}{\text{Min}} \; \left(\sum_{s=1}^{r} \text{Var}_{intra}(P,b_s) \right) \qquad (2\text{-}19)$$

et que, aux vecteurs $u_1,\ldots,u_r$ définis au paragraphe 7.4 de l'avant propos, l'application V^{-1} fait correspondre les r facteurs discriminants la partition P, $b_1 = V^{-1}u_1,\ldots, b_r = V^{-1}u_r$.

De (2 - 1) et (2-19) on déduit que W (P) est égal à la somme des variances intra classe des r premiers facteurs discriminants de la partition P (ou r diminué de la somme des variances inter classe de ces facteurs).

3 - LA REPRESENTATION

Dans ce paragraphe on donne la définition d'une représentation puis du critère W_R qui est une généralisation du critère W nécessaire au déroulement de l'algorithme.

3.1 - Définition

Elle est analogue à celle du I.3.1. On appelle représentation, tout ensemble L constitué de K points (ou centres) de R^p, $a_1,\ldots,a_K$, et d'une famille $(u_1,\ldots,u_r)$ de U_r

$$L = (a_1,..,a_K, u_1,\ldots,u_r)$$

$$u_s' V^{-1} u_s = 1 \qquad s=1,\ldots,r$$

$$u_s' V^{-1} u_t = 0 \qquad s=1,\ldots,r;\ t=1,\ldots,r\ ;\ s\neq t$$

On note :

$\underline{A} = (a_1,\ldots,a_K)$

$\underline{U} = (u_1,\ldots,u_r)$

U le sous espace de R^p engendré par $(u_1,\ldots,u_r)$

$L = (\underline{A},\underline{U})$

L'ensemble des représentations définies ci-dessus est notée $L_{K,r}$.

3.2 - Le critère W_R .

Soit $P = (P_1,\ldots,P_K)$ dans $\mathbb{P}_K$

$L = (a_1,\ldots,a_K,u_1,\ldots,u_r) = (\underline{A},\underline{U})$ dans $L_{K,r}$

U le sous espace engendré par les vecteurs $u_1,\ldots,u_r$

$I(\underline{A},P,U)$ l'inertie aux points $a_1,\ldots,a_K$ de la partition P, expliquée par U.

3.2.1 - Définition

On définit un critère W_R sur l'ensemble produit $\mathbb{P}_K \times L_{K,r}$ par :

$$W_R(P,L) = I(\underline{A},P,U) \qquad (3\text{-}1)$$

Remarque : $W_R(P,(\underline{A},\underline{U}))$ ne dépend pas de la famille V^{-1}-orthonormée $\underline{U}$, mais seulement de l'espace vectoriel U engendré par cette famille.

3.2.2 - Cas particulier

Soit $\underline{G} = (g_1,\ldots,g_K)$ la famille des centres de gravité des classes $P_1,\ldots,P_K$.

L'inertie $I(\underline{G},P,U)$ est l'inertie intra classe de la partition P expliquée par le sous espace U :

$$I(\underline{G},P,U) = I_{intra}(P,U)$$

On en déduit :

$$\text{Si } L = (\underline{G},\underline{U}), \quad W_R(P,L) = I_{intra}(P,U) \tag{3-2}$$

3.2.3 - Proposition

Pour toute partition P dans $\mathbb{P}_K$

$$\min_{L \in L_{K,r}} W_R(P,L) = W(P)$$

Démonstration : On sait, (A3.2.2), que pour tout sous espace U, l'inertie $I(\underline{A}, P,U)$ est minimale pour la famille $\underline{G}$ des centres de gravité des classes :

$$\min_{\underline{A}} I(\underline{A},P,U) = I(\underline{G},P,U)$$
$$= I_{intra}(P,U)$$

On en déduit :

$$\min_{L \in L_{K,r}} W_R(P,L) = \min_{(\underline{A},U) \in L_{K,r}} [I(\underline{A},P,U)]$$
$$= \min_{U \in F_r} I_{intra}(P,U)$$
$$= W(P)$$

4 - L'ALGORITHME

4.1 - Introduction

L'algorithme utilise alternativement deux fonctions définies en (4.2)

et (4.3) :

- g appelée fonction de représentation, associe à toute partition P de $\mathbb{P}_K$, une représentation g(P) dans $L_{K,r}$ telle que :

$$W_R(P,g(P)) = \min_{L \in L_{K,r}} W_R(P,L)$$

- f appelée fonction d'affectation, associe à toute représentation L de $L_{K,r}$ une partition f(L) dans $\mathbb{P}_K$ telle que :

$$W_R(f(L),L) = \min_{P \in \mathbb{P}_K} W_R(P,L)$$

On montre, (4.4), que pour toute partition P dans $\mathbb{P}_K$, on a la relation :

$$W(f \circ g\ (P)) \leq W(P) \qquad (4\text{-}7)$$

L'algorithme est initialisé par une partition P^o connue, la partition a priori par exemple.

Pour tout $n \geq 0$, on définit

$$P^{n+1} = f \circ g\ (P^n)$$

D'après (4-7) la suite de terme général $W(P^n)$ est décroissante. On montre, (4.6), qu'elle est constante à partir d'un certain rang, et que la suite de partitions (P^n) est alors également constante. La partition P^* limite de la suite (P^n), est meilleure au sens du critère W que la partition initiale P^o :

$$W(P^*) \leq W(P^o)$$

4.2 - La fonction de représentation

Le problème est de définir une fonction g qui associe à toute partition P dans $\mathbb{P}_K$ une représentation $L = g(P)$ dans $L_{K,r}$, telle que :

$$W_R(P,g(P)) = \min_{L \in L_{K,r}} W_R(P,L)$$

Soit

$P = (P_1,\ldots,P_K)$ dans $\mathbb{P}_K$

$L = (b_1,\ldots,b_K,v_1,\ldots,v_r) = (\underline{B},\underline{V})$ dans $L_{K,r}$

F le sous espace vectoriel engendré par les vecteurs $v_1,\ldots,v_r$.

$g(P) = (a_1,\ldots,a_K,u_1,\ldots,u_r) = (\underline{A},\underline{U})$

U le sous espace vectoriel engendré par les vecteurs $u_1,\ldots,u_r$.

On a par définition, (3.1) :

$W_R(P,L) = I(\underline{B},P,F)$

$W_R(P,g(P)) = I(\underline{A},P,U)$

On est ramené au problème d'optimisation étudié au paragraphe 7 de l'avant propos, dans le cas particulier où $Q = V^{-1}$, présenté au paragraphe 7.2 :

Définir une fonction g qui associe à toute partition $P = (P_1,\ldots,P_k)$ de E,

- K points $a_1,\ldots,a_K$ de $\mathbb{R}^p$

- une famille V^{-1}orthonormée de r vecteurs de $\mathbb{R}^p$, $\underline{U} = (u_1,\ldots,u_r)$

tels que :

$a_1,\ldots,a_K$, et le sous espace vectoriel U engendré par $u_1,\ldots,u_r$ vérifient :

$$I((a_1,\ldots,a_K),\ P,U) = \min_{\substack{(b_1,\ldots,b_K)\in\mathbb{R}^p\times\ldots\times\mathbb{R}^p \\ F\in F_r}} [I((b_1,\ldots,b_K),P,F)]$$

D'après le paragraphe 7.4 de l'avant propos, on pose, pour toute partition $P = (P_1,\ldots,P_K)$ de E,

$$g(P) = (g_1,\ldots,g_K,u_1,\ldots,u_r)$$

où

- $g_1,\ldots,g_K$ sont les centres de gravité des classes $P_1,\ldots,P_K$.
- $u_1,\ldots,u_r$ sont vecteurs propres Q normés associés aux r plus grandes valeurs propres $\lambda_1 \geq \ldots \geq \lambda_r \geq 0$ de la matrice $B(P)\ V^{-1}$ (où B(P) est la matrice d'inertie inter classe de la partition P). La définition précise de $u_1,\ldots,u_r$ est donnée au paragraphe 7.4 de l'avant propos.

Par définition de la fonction g, et d'après 3.2.3

$$W_R(P,g(P)) = W(P) \qquad (4\text{-}1)$$

D'après (0-15)

$$W_R(P,g(P) = I_{intra}(P,U) \qquad (4\text{-}2)$$

On déduit de (4-1) et (4.2)

$$W(P) = I_{intra}(P,U) \qquad (4\text{-}3)$$

D'après (0-16)

$$I_{intra}(P,U) = r - \sum_{s=1}^{r} \lambda_s$$

On en déduit d'après (4-3)

$$W(P) = r - \sum_{s=1}^{r} \lambda_s \qquad (4\text{-}4)$$

4.3 - La fonction d'affectation

4.3.1 - Définition

Il s'agit de définir une fonction f qui associe à toute représentation L dans $L_{K,r}$ une partition f(L) dans $\mathbb{P}_K$, telle que :

$$W_R(f(L),L) = \min_{P\in\mathbb{P}_K} W_R(P,L) \qquad (4\text{-}5)$$

Soit

$$L = (a_1,\ldots,a_K,u_1,\ldots,u_r) = (\underline{A},\underline{U}) \text{ dans } L_{K,r}$$

U le sous espace vectoriel engendré par les vecteurs $u_1,\ldots,u_r$

$$P = (P_1,\ldots,P_K) \text{ dans } \mathbb{P}_K$$

On a par définition, (3.2.1.) :

$$W_R(P,L) = I(\underline{A},P,U)$$

D'après (0-9) pour toute partition $P = (P_1,\ldots,P_K)$ dans $\mathbb{P}_K$,

$$I(\underline{A},P,U) = \frac{1}{n} \sum_{x\in E} d_U^2(x,a_x) \qquad (4\text{-}6)$$

où $d_U^2(x,y) = (x-y)' \; V^{-1} \left(\sum_{s=1}^{r} u_s \, u_s' \right) V^{-1} \; (x-y)$

et a_x le centre de la famille $(a_1,\ldots,a_K)$ d'indice i(x), où i(x) est l'indice de la classe de la partition $P = (P_1,\ldots,P_K)$ qui contient x.

On en déduit :

$$\min_{P\in\mathbb{P}_K} I(\underline{A},P,U) = \frac{1}{n} \sum_{x\in E} \min_{1\le i\le K} d_U^2(x,a_i)$$

On pose par conséquent :

$$f(L) = (P_1,\ldots,P_K) ,$$

où $P_\ell = \{x \mid d_U^2(x,a_\ell) \leq d_U^2(x,a_i) ; \quad i=1,\ldots,K\} \quad \ell=1,\ldots,K$

Pour compléter la règle d'affectation d'un individu x dans l'éventualité où plusieurs centres a_i réalisent le minimum de $d_U^2(x,a_i)$ on se place dans le contexte de l'algorithme :

Une partition Q est donnée dans $\mathbb{P}_K$. On recherche la partition $P = f(L)$ où $L = g(Q)$.

Soit un individu x qui appartient à Q_j et tel que plusieurs centres a_i réalisent $\underset{1\leq\ell\leq K}{\text{Min}} \; d_U^2(x,a_\ell)$.

On distingue deux éventualités :

Eventualité 1 : Si $d_U^2(x,a_j) = \underset{1\leq\ell\leq K}{\text{Min}} \; d_U^2(x,a_\ell)$

x est affecté à P_j

Eventualité 2 : Sinon x est affecté à P_m où m est le plus petit indice i tel que

$$d_U^2(x,a_i) = \underset{1\leq\ell\leq K}{\text{Min}} \; d_U^2(x,a_\ell)$$

Dans ce cas d'indétermination, la partition P dépend non seulement de la représentation L, mais aussi de la partition Q.

f apparaît ainsi comme une fonction de deux variables :

- une représentation

- une partition utilisée simplement comme contrainte en cas d'indétermination pour l'affectation. (La classe d'appartenance dans la partition Q est prioritaire).

Le rôle de cette seconde variable est secondaire. De plus, dans l'algorithme, (4-5), la partition Q est définie par la représentation L. Aussi, pour simplifier les notations, on écrit f comme une fonction de la seule variable représentation, et on retient la proposition 4.3.2. (On vérifie dans la présentation de l'algorithme, et la démonstration de sa convergence que l'emploi de cette notation n'entraîne aucune inexactitude).

4.3.2 - Proposition

Si les partitions Q et $P = f \circ g\ (Q)$ vérifient

$W_R(Q,g(Q)) = W_R(P,g(Q))$ elles sont identiques.

Démonstration

Soit :

$Q = (Q_1,\ldots,Q_K)$

$L = g(Q) = (a_1,\ldots,a_K,\ u_1,\ldots,u_r) = (\underline{A},\ \underline{U})$

U le sous espace engendré par $u_1,\ldots,u_r$

$P = f \circ g\ (Q) = (P_1,\ldots,P_K)$

Soit un objet x dans E. On suppose que x appartient à la classe Q_j $(1 \leq j \leq K)$ de la partition Q, et à la classe P_i $(1 \leq i \leq K)$ de la partition P.

On a par définition de P = f(L)

$$d_U^2(x,a_i) = \min_{\ell=1,..,K} d_U^2(x,a_\ell) \leq d_U^2(x,a_j)$$

D'après (4-6) :

$$(d_U^2(x,a_i) < d_U^2(x,a_j)) \Rightarrow (I(\underline{A},P,U) < I(\underline{A},Q,U))$$

Comme $W_R(P,g(Q)) = I(\underline{A},P,U)$, on en déduit

$$(d_U^2(x,a_i) < d_U^2(x,a_j)) \Rightarrow W_R(P,g(Q)) < W_R(Q,g(Q))$$

Comme la dernière inégalité est contraire à l'hypothèse, on en déduit :

$$d^2(x,a_i) = d^2(x,a_j)$$

L'affectation de l'objet x à l'une des classes de la partition P s'est donc effectuée selon la règle définie dans l'éventualité 1 du cas d'indétermination.

On en déduit i=j.

Cette démonstration étant valable pour tout x dans E, on en déduit que les partitions P et Q sont identiques.

4.4 - La fonction composée f ∘ g

Proposition.

Pour toute partition Q dans $\mathbb{P}_K$, la partition f ∘ g (Q) = f[g(Q)] vérifie :

$$W(f \circ g\ (Q)) \leq W(Q) \qquad (4\text{-}7)$$

de plus,

$$W(f \circ g\ (Q)) = W(Q) \iff f \circ g\ (Q) = Q \qquad (4\text{-}8)$$

Démonstration

D'après (4-1) et la définition de la partition f[g(Q)], on a :

$$W(Q) = W_R(Q,g(Q)) \geq W_R(f[g(Q)], g(Q)) \qquad (4\text{-}9)$$

D'après la définition de la représentation g[f ∘ g(Q)] et (4-1) appliqué à la partition f ∘ g(Q), on a :

$$W_R(f \circ g(Q), g(Q)) \geq W_R(f \circ g(Q), g[f \circ g(Q)]) = W(f \circ g(Q)) \qquad (4\text{-}10)$$

Des relations (4-9) et (4-10), on déduit :

$$W(Q) \geq W(f \circ g(Q))$$

De plus, si $W(Q) = W(f \circ g(Q))$, on a l'égalité dans la relation (4 9) :

$$W_R(Q,g(Q)) = W_R(f[g(Q)],g(Q))$$

D'après la proposition 4.3.2, les partitions Q et $f \circ g(Q)$ sont identiques.

4.5 - L'algorithme

L'algorithme est initialisé par une partition P^o connue, la partition a priori par exemple.

Pour tout $n \geq 0$, on pose :

$$L^n = g(P^n)$$

$$P^{n+1} = f(L^n) = f \circ g(P^n)$$

On définit ainsi une suite (P^n) de partitions dans $\mathbb{P}_K$, et une suite (L^n) de représentations dans $L_{K,r}$ (cf. fig. 1 : L'organigramme de l'algorithme). Soit

$$P^n = (P_1^n,\ldots,P_K^n)$$

et

$g_1^n,\ldots,g_K^n$ les centres de gravité des classes $P_1,\ldots,P_K$.

On sait, (4.2), que $L^n = g(P^n)$ est de la forme :

$$L^n = (g_1^n,\ldots,g_K^n,\underline{U}^n)$$

où $\underline{U}^n$ est une famille V^{-1} orthonormée de r vecteurs de $\mathbb{R}^p$:

$$\underline{U}^n = (u_1^n,\ldots,u_r^n).$$

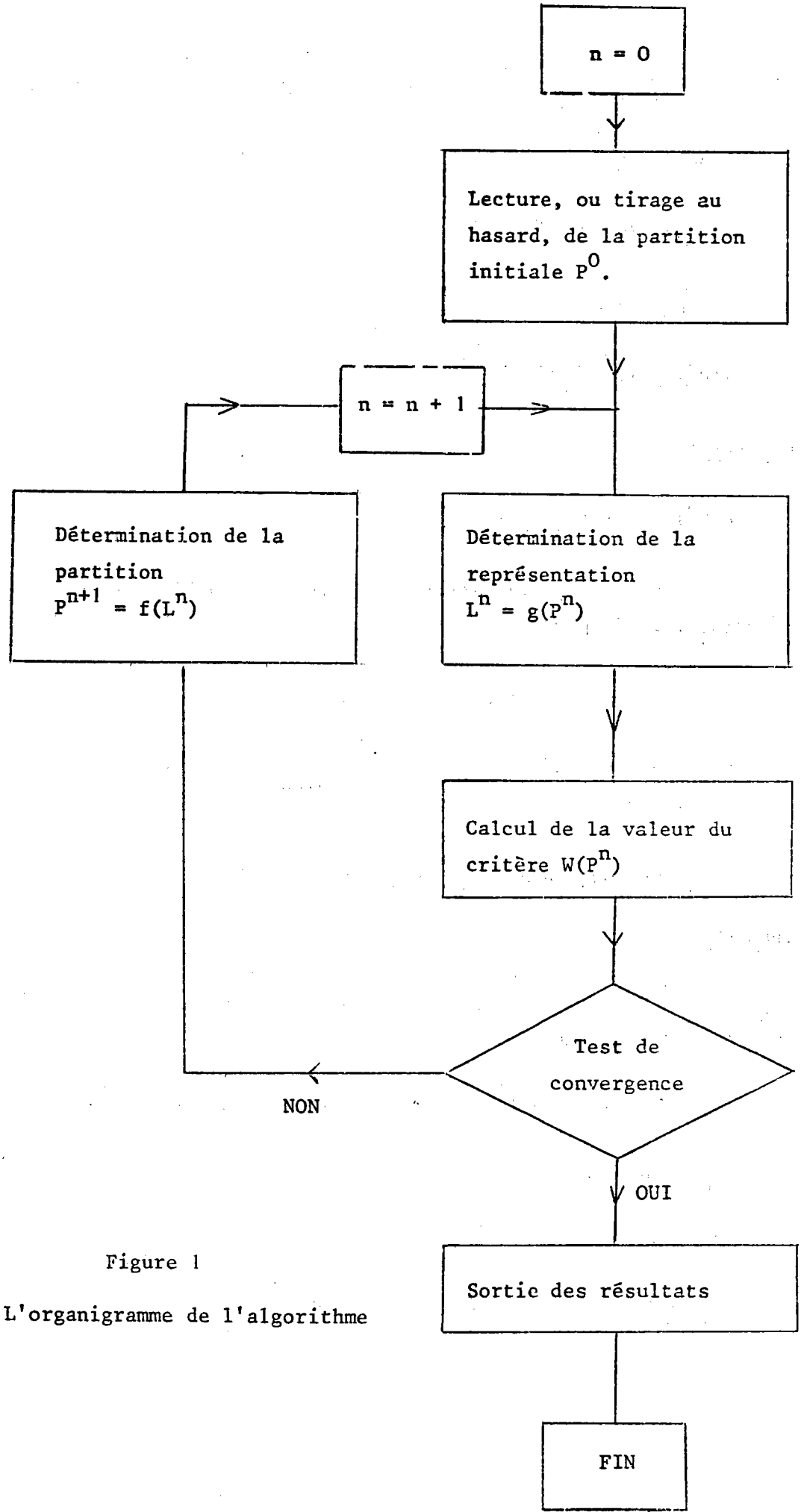


Figure 1

L'organigramme de l'algorithme

On note U^n le sous espace vectoriel de $\mathbb{R}^p$ engendré par $\underline{U}^n$.

D'après (4-1), on a :

$$W_R(P^n,L^n) = W_R(P^n,g(P^n)) = W(P^n) \qquad (4-11)$$

D'après (4-3), on a :

$$W(P^n) = I_{intra}(P^n,U^n) \qquad (4-12)$$

D'après (2-4)

$$W(P^n) = r - I_{inter}(P^n,U^n) \qquad (4-13)$$

Soit $(b_1^n,\ldots,b_r^n)$ la famille de facteurs associée à la famille $\underline{U}^n = (u_1^n,\ldots,u_r^n)$ par :

$$b_s^n = V^{-1}\, u_s^n \qquad s=1,\ldots,r \ ; \ n \geq 0$$

D'après 2.5.8 $b_1^n,\ldots,b_r^n$ sont les r premiers facteurs discriminants de la partition P^n.

4.6 - La convergence de l'algorithme

Théorème

La suite de terme général $r_n = W(P^n)$ est décroissante, et constante à partir d'un certain rang N.

De plus, la suite (P^n,L^n) est constante à partir du rang N.

Démonstration

a) Pour tout $n \geq 0$, on a d'après (4-4)

$$r_{n+1} = W(p^{n+1}) = W(f \circ g(P^n)) \leq W(P^n) = r_n$$

La suite (r_n) est donc décroissante.

b) La suite (r_n) prend ses valeurs dans l'ensemble fini $\{W(P) \mid P \in \mathbb{P}_K\}$

En effet comme E est un ensemble fini, l'ensemble $\mathbb{P}_K$ des partitions de E en K classes est fini.

c) D'après a) et b) il existe un rang N, tel que :

$r_n = r_N$ pour tout $n \geq N$.

d) Pour tout n, $n \geq N$, on a d'après c)

$$r_{n+1} = r_n$$

Soit d'après a)

$$W(f \circ g(P^n)) = W(P^n)$$

On en déduit d'après (4.4)

$$f \circ g(P^n) = P^n$$

e) On déduit de d) l'égalité

$$P^{n+1} = P^n \qquad \text{pour tout } n, \quad n \geq N$$

d'où

$$P^n = P^N \qquad \text{pour tout } n, \quad n \geq N$$

f) De l'égalité $P^{n+1} = P^n$ $(n \geq N)$ démontrée en e), on déduit que :

$$L^{n+1} = g(P^{n+1}) = g(P^n) = L^n \qquad \text{pour tout } n, \ n \geq N$$

d'où

$$L^n = L^N \qquad \text{pour tout } n \geq N$$

Remarque :

Dans l'algorithme la fonction f est utilisée exclusivement dans le contexte suivant :

Détermination dans la partition $P^{n+1} = f(L^n)$
où $L^n = g(P^n)$, $n \geq 0$.

La partition Q de contrainte qui apparaît dans la définition de la fonction f, (4.3.1) est donc parfaitement définie, et f peut être notée comme une fonction d'une variable. La prise en compte de la partition P^n pour définir la partition P^{n+1}, apparaît dans la démonstration de la convergence de l'algorithme par l'utilisation en d) de la relation (4-9), déduite de la proposition 4.3.2.

5 - LE PROGRAMME INFORMATIQUE DISCRI

5.1 - Les entrées

Outre le tableau des données (1.7) , les entrées du programme DISCRI sont les suivantes :

- la partition initiale : P^o

Dans une première analyse, on choisit pour P^o la partition définie a priori. Pour des analyses complémentaires P^o est choisi en fonction des résultats obtenus précédemment (cf. 6.5, 6.6).

(Le nombre maximum de classes demandé K est égal au nombre de classes de la partition P^o).

- la dimension du sous espace discriminant : r.

r doit être choisi inférieur ou égal à K-1. Dans une première analyse r est en général (si $K \leq 5$) choisi égal à K-1. Pour des analyses complémentaires il est choisi en fonction des résultats obtenus précédemment. (cf. § 6 et 7).

5.2 - Le critère

Dans le programme DISCRI on utilise le critère W_R sous la forme suivante : Soit

$$P = (P_1,\ldots,P_K)$$

$$L = (a_1,\ldots,a_K,u_1,\ldots,u_r)$$

$$b_s = V^{-1}u_s \quad , \qquad s=1,\ldots,r$$

$$W_R(P,L) = \frac{1}{n} \sum_{\ell=1}^{K} \sum_{x\in P_\ell} \sum_{s=1}^{r} [(x-a_\ell)'b_s]^2$$

5.3 - La fonction de représentation [16]

a) On sait que les vecteurs $u_1,\ldots,u_r$ sont vecteurs propres associés aux r plus grandes valeurs propres de la matrice $B(P)\ V^{-1}$, où $B(P)$ est la matrice d'inertie inter classe de la partition P

$$B(P)\ V^{-1}\ u_s = \lambda_s\ u_s \qquad s=1,\ldots,r$$

$$\lambda_1 \geq \ldots \geq \lambda_r$$

b) On déduit de l'annexe 5 que :

- d'une part $\lambda_1,\ldots,\lambda_r$ sont les r plus grandes valeurs propres de la matrice $V^{-1}\ B(P)$

- d'autre part, $b_1,\ldots,b_r$ sont vecteurs propres de la matrice $V^{-1}\ B(P)$ associés aux valeurs propres $\lambda_1,\ldots,\lambda_r$.

c) Soit

$g_1,\ldots,g_K$ les centres de gravité des classes $P_1,\ldots,P_K$

$p_1,\ldots,p_K$ les poids des classes $P_1,\ldots,P_K$, $(p_\ell = \frac{\text{Card } P_\ell}{n} = \ell = 1,\ldots,K)$

g_o le centre de gravité de E

$$C = (\sqrt{p_1}(g_1-g_o),\ldots,\sqrt{p_K}(g_K-g_o))$$

La matrice d'inertie inter classe de la partition P s'écrit :

$$B(P) = C\ C'$$

On déduit de l'annexe 6 que :

- d'une part, $\lambda_1,\ldots,\lambda_r$ sont les r plus grandes valeurs propres de la matrice $C'\ V^{-1}\ C$

- d'autre part, si $a_1,\ldots,a_r$ sont vecteurs propres associés aux valeurs propres $\lambda_1,\ldots,\lambda_r$ de la matrice $C'\ V^{-1}\ C$, les vecteurs $V^{-1}\ C\ a_1,\ldots,\ V^{-1}C\ a_r$ sont vecteurs propres associés aux valeurs propres $\lambda_1,\ldots,\lambda_r$ de la matrice $V^{-1}\ B(P)$.

d) Conclusion

Finalement, on recherche les r plus grandes valeurs propres $\lambda_1 \geq \ldots \geq \lambda_r$ de la matrice $C'\ V^{-1}\ C$, et r vecteurs propres associés $a_1,\ldots,a_r$. Puis on pose :

$$\tilde{b}_s = V^{-1}\ C\ a_s \qquad s=1,\ldots,r$$

Si λ_s est valeur propre simple, on pose

$$u_s = V\ \tilde{b}_s \quad \text{où} \quad u_s = -V\ \tilde{b}_s$$

pour respecter la convention donnée dans l'avant propos au § 7.4

Si λ_s est valeur propre multiple de multiplicité m, $\lambda_s = \lambda_{s+1} = \ldots = \lambda_{s+m-1}$, b_s est choisi dans E_s, sous espace propre de la matrice $B\ V^{-1}$ associé à λ_s, selon la procédure décrite dans l'avant propos au § 7.4. E_s est engendré par les vecteurs $V\ b_s,\ldots,V\ b_{s+m-1}$.

Ainsi on effectue à chaque itération la recherche des r plus grandes valeurs propres et des vecteurs propres associés d'une matrice symétrique de dimension K, et non de la matrice Q B(P) qui n'est pas symétrique et de dimen-

sion p. (p est en général nettement supérieur à K).

5.4 - Les sorties à chaque itération

Après chaque itération on a les sorties suivantes :

- la valeur du critère W

$W(P^n) = W_R(P^n, g(P^n))$

- les valeurs propres, $\lambda_1^n,\ldots,\lambda_r^n$

- les centres de gravité des classes $P_1^n,\ldots,P_K^n$

- la classe d'appartenance dans la partition P^n de chaque individu.

Remarque :

Si au cours de l'algorithme la classe de numéro ℓ est vide, elle n'est pas reconstituée et dans la partition P^*, la classe P_ℓ^* est vide (K est donc le nombre maximum de classes désiré).

5.5 - Les sorties à la convergence

Après convergence de l'algorithme on a les sorties suivantes :

- la partition de convergence $P^* = (P_1^*,\ldots,P_K^*)$

- les coordonnées des points $g_1^*,\ldots,g_K^*$ et vecteurs $u_1^*,\ldots,u_r^*$ de la représentation $g(P^*)$, ainsi que celles des facteurs $b_1^* = V^{-1}u_1^*,\ldots,b_r^* = V^{-1}u_r^*$.

- la valeur du critère W

$W(P^*) = W_R(P^*,g(P^*))$

- les valeurs propres $\lambda_1^*,\ldots,\lambda_r^*$

- les représentations graphiques des projections V^{-1} orthogonales des centres de gravité $P_1^*,\ldots,P_K^*$, et des individus sur les plans engendrés par u_1^* et u_2^*, u_3^* et u_4^*, ... (si r est impair, la dernière représentation est faite avec u_r^* comme vecteur unitaire sur les 2 axes; les points représentés sont situés sur la diagonale). On utilise le programme de représentation "NUAGE" (ISUP).

Au paragraphe 6 on présente le traitement d'une banque de données industrielles qui illustre les points suivants :

- Interprétation des résultats d'une analyse discriminante typologique

- Mode d'initialisation d'analyses discriminantes typologiques complémentaires.

- Fonction de décision issue de la méthode

Au paragraphe 7, ces points sont abordés dans un contexte général.

6 - LE TRAITEMENT D'UNE BANQUE DE DONNEES*

6.1 - Le problème

L'étude a pour but la mise en place d'un contrôle automatique de la qualité de production de machines tournantes en bout de chaine de fabrication. Il s'agit de construire une règle de décision pour une reconnaissance fiable des machines défectueuses, et alors du type de défaut présent en vue de localiser la (ou les) pièce (s) incriminées.

---

* Cette application a été réalisée en collaboration avec G. CELEUX

6.2 - Les types définis a priori

Etant donné la complexité des machines tournantes étudiées, le contrôle de production est fait par détection de défauts de fonctionnement.

Une machine peut présenter l'un des deux types de défauts D_1 et D_2 (D_1 et D_2 s'excluent). Ce défaut peut être sensible lors du fonctionnement dans un seul sens de rotations (R_+ ou bien R_-) ou dans les deux (R_+ et R_-).

Le spécialiste définit ainsi 7 catégories disjointes et exhaustives de machines :

Catégorie	Défaut en fonct. dans le sens R_+	Défaut en fonct. dans le sens R_-
$\mathfrak{D}_o$	Néant	Néant
$\mathfrak{D}_{1+}$	D_1	Néant
$\mathfrak{D}_{1-}$	Néant	D_1
$\mathfrak{D}_1$	D_1	D_1
$\mathfrak{D}_{2+}$	D_2	Néant
$\mathfrak{D}_{2-}$	Néant	D_2
$\mathfrak{D}_2$	D_2	D_2

$\mathfrak{D}_o, \mathfrak{D}_{1+}, \mathfrak{D}_{1-}, \mathfrak{D}_1, \mathfrak{D}_{2+}, \mathfrak{D}_{2-}$, et $\mathfrak{D}_2$ constituent les types définis a priori sur la population des machines.

Le contrôle de qualité d'une machine se déroule de la manière suivante : Un spécialiste décelle par simple écoute l'existence éventuelle d'un défaut lors de deux simulations de fonctionnement (une dans chaque sens de rotation). Ce contrôle est subjectif car les premières machines contrôlées dans une série jouent un rôle d'étalons, et une lente détérioration de la fabrication est difficilement perceptible.

D'autre part, un défaut doit être jugé présent ou absent, alors qu'il peut apparaître avec différents degrés de gravité. En particulier, en moyenne, les défauts D_1 et D_2 sont faiblement perceptibles dans la configuration R_+, et il n'est pas possible au spécialiste de déceler de manière fiable un défaut léger de ce type. Aussi les frontières entre les catégories $\mathcal{D}_{1+}$ et $\mathcal{D}_0$, $\mathcal{D}_{1-}$ et $\mathcal{D}_1$, $\mathcal{D}_{2+}$ et $\mathcal{D}_0$, $\mathcal{D}_{2-}$ et $\mathcal{D}_2$ sont subjectives.

6.3 - La population échantillon - La partition définie a priori

La population échantillon E est constituée de 113 machines tournantes. Aux 7 catégories définies a priori par le spécialiste il correspond 7 classes d'objets de E, $T_1,\ldots,T_7$, qui constituent une partition T de E.

Catégories	Classes de E	Effectifs
$\mathcal{D}_0$	T_1	32
$\mathcal{D}_{1+}$	T_2	12
$\mathcal{D}_{1-}$	T_3	12
$\mathcal{D}_1$	T_4	21
$\mathcal{D}_{2+}$	T_5	12
$\mathcal{D}_{2-}$	T_6	12
$\mathcal{D}_2$	T_7	12

Les effectifs des classes de la population échantillon E respectent les proportions de machines de chaque catégorie dans $\mathcal{E}$ (sauf évidemment en ce qui concerne la classe T_1 échantillon de la famille $\mathcal{D}_0$ des machines sans défauts).

Le fonctionnement des machines est décrit par 24 paramètres quantitatifs extraits d'un spectre de fréquence [5]

6.4 - L'analyse discriminante typologique relative à la partition a priori

L'analyse est initialisée par la partition en 7 classes, $T = (T_1,\ldots,T_7)$ définie a priori. On choisit r=5.

L'algorithme converge en 5 itérations vers la partition $T^* = (T_1^*,\ldots,T_7^*)$. On précise dans le tableau 1 la composition des classes de la partition T^* relativement aux classes de la partition initiale T. On remarque en particulier :

T \ T^*	T_1^*	T_2^*	T_3^*	T_4^*	T_5^*	T_6^*	T_7^*
T_1	25	2				3	2
T_2		11				1	
T_3		1	10			1	
T_4	2	1	3	14	1		
T_5		1			9	1	1
T_6						8	4
T_7						2	10

Tableau 1 : La partition initiale T et la partition de convergence T^*

La classe T_4^*, incluse dans la classe T_4, et la dispersion des objets du complémentaire dans T_4 de T_4^*.

De même, la classe T_1^*, constituée en presque totalité d'objets de la classe T_1, et la dispersion des objets de T_1 non classés dans T_1^*.

Les classes T_6 et T_7, incluses en totalité dans la réunion des classes T_6^* et T_7^*; D'autre part, chacune de ces dernières classes n'est pas interprétable vis à vis des types a priori.

L'étude des projections des centres de gravité des classes de la partition T^* (figures 1 et 2) permet de préciser ces résultats :

On observe que les classes T_6^* et T_7^* sont proches en projection sur les 3 premiers axes discriminants de la partition T^*.

On peut faire la même remarque a propos des classes T_1^*, T_2^* et T_5^*.

Interprétation des résultats :

D'après le tableau 1, la réunion des classes T_6^* et T_7^* s'interprète comme la réunion des classes a priori T_6 et T_7. Or ces deux dernières classes respectivement définies par la présence du défaut D_2 dans la configuration R_- seulement, et dans les deux configurations. La réunion des classes T_6 et T_7 s'interprète simplement comme l'ensemble des objets de E présentant le défaut D_2 dans la configuration R_- au moins.

Le spécialiste reconnaît en la classe T_4^* (incluse dans T_4) l'ensemble des machines de E présentant un défaut caractéristique de fabrication (les types a priori sont définis selon l'approche de fonctionnement).

La stabilité médiocre de la classe T_1 (machines sans défaut) est le reflet du problème évoqué plus haut de la notion floue de machine sans défaut.

Ces résultat et interprétations suggèrent la définition de la partition suivante de E en 5 classes :

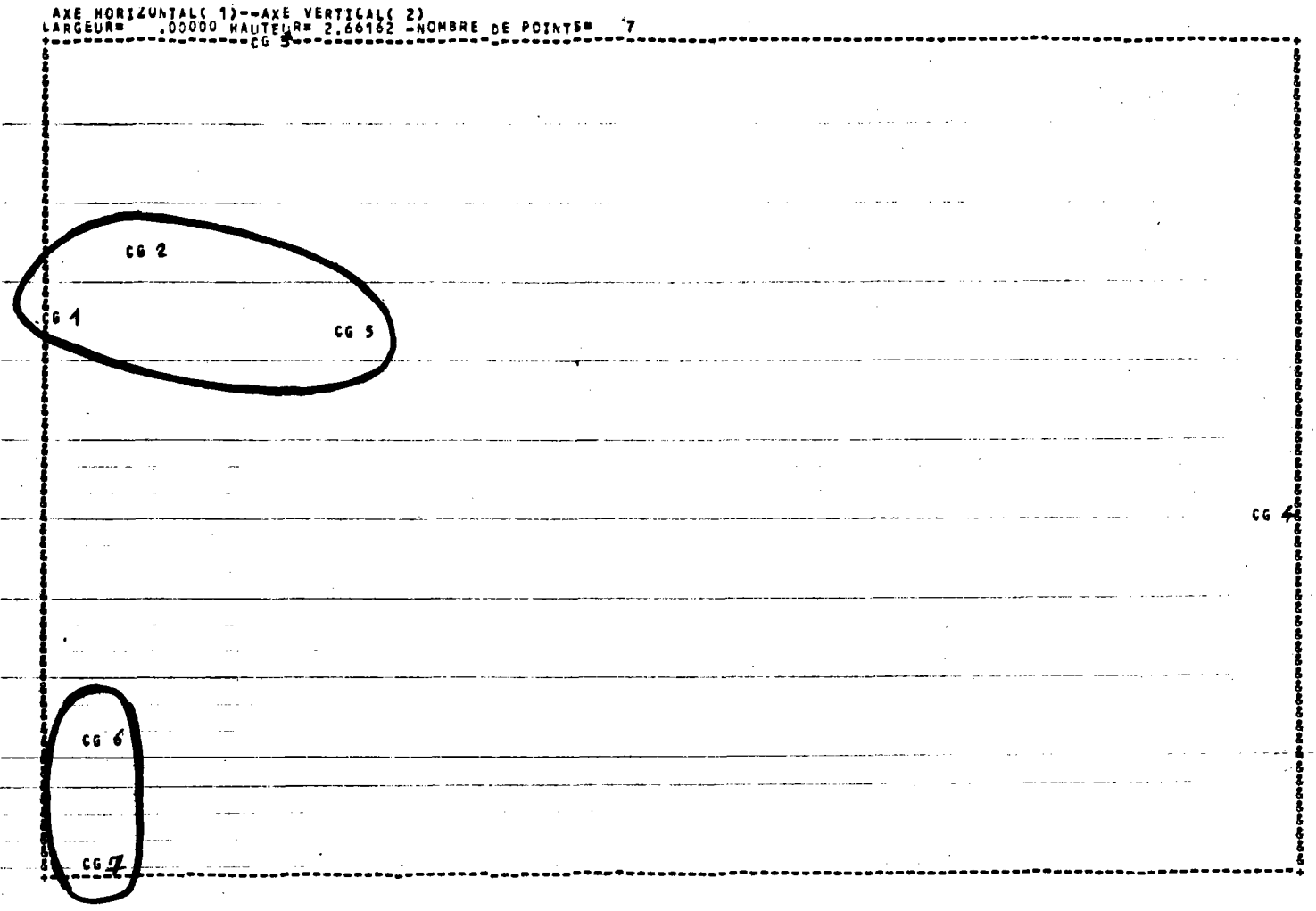


Figure 1 : Les projections des centres de gravité CG1,...,CG7 des classes $T_1^*,\ldots,T_7^*$ sur les axes u_1^* et u_2^* issus de la représentation $L^* = g\ (T^*)$

B. U. METZ

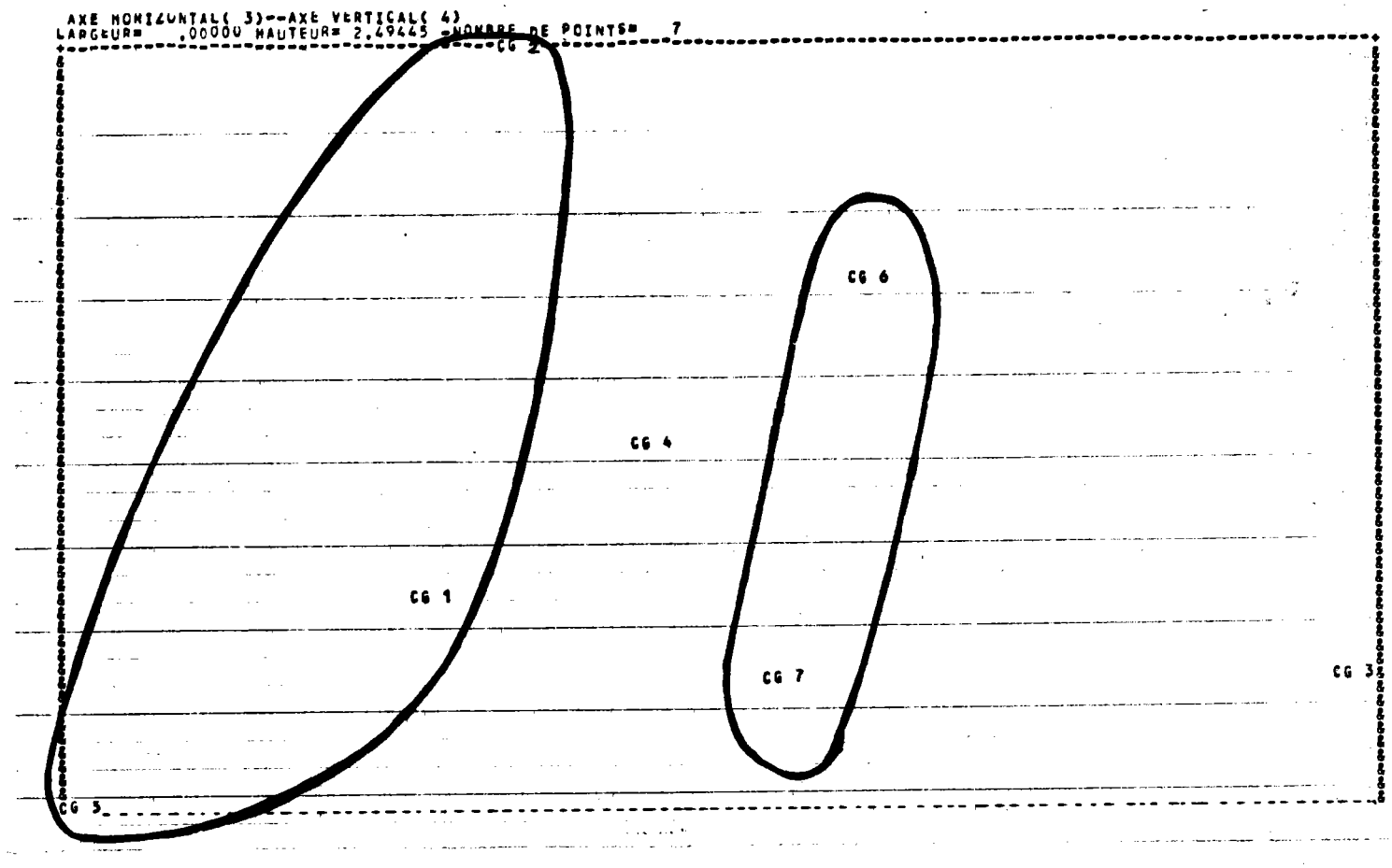


Figure 2 : Les projections des centres de gravité CG1,...,CG7 des classes $T_1^*,\ldots,T_7^*$ sur les axes u_3^* et u_4^* issus de la représentation $L^* = g\ (T^*)$.

$C = (C_1, C_2, C_3, C_4, C_5)$

$C_1 = T_1 \cap T_1^*$

$C_2 = (T_1 - C_1) \cup T_2 \cup T_5$

$C_3 = T_3 \cup (T_4 - T_4^*)$

$C_4 = T_4^*$

$C_5 = T_6 \cup T_7$

La définition de la classe C_3 est justifiée par celles des classes a priori T_3 et T_4 :

T_3 : Défaut D_1 dans la configuration R_- seulement

T_4 : Défaut D_1 dans les configurations R_+ et R_-

La définition de la classe C_2 tient compte de la moindre gravité (du point de vue du spécialiste) de la présence des défauts D_1 ou D_2 dans la configuration R_+ seulement, auxquelles sont associées respectivement les classes T_2 et T_5. Les machines considérées a priori comme sans défaut, mais non classées dans T_1^*, sont naturellement adjointes à ces machines présentant un défaut léger.

6.5 - La seconde analyse

L'analyse discriminante typologique est initialisée par la partition $C = (C_1, \ldots, C_5)$ définie ci-dessus. On choisit r=4.

L'algorithme converge en 2 itérations vers une partition $C^* = (C_1^*, \ldots, C_5^*)$ analogue à la partition C au point de vue de l'interprétation des classes :

C_1^* est incluse dans la classe a priori T_1

C_4^* est incluse dans la classe a priori T_4

C_2^* s'interprète comme $(T_1 - C_1^*) \cup T_2 \cup T_5$

C_3^* s'interprète comme $T_3 \cup (T_4 - C_4^*)$

C_5^* s'interprète comme $T_6 \cup T_7$

Cette analyse confirme les résultats de la première : la définition des sous classes des classes a priori T_1 et T_4 est affinée.

6.6 - La troisième analyse

L'analyse discriminante typologique est initialisée par la partition $F = (F_1,\ldots,F_5)$ suivante déduite de la partition C^* obtenue en 6.5.

$F_1 = C_1^*$

$F_4 = C_4^*$

$F_2 = (T_1 - F_1) \cup T_2 \cup T_5$

$F_3 = T_3 \cup (T_4 - F_4)$

$F_5 = T_6 \cup T_7$

On choisit r=4.

L'algorithme converge dès la 1ère itération (on note F^* la partition de convergence).

Le tableau 2 permet d'apprécier la stabilité de la partition F. Dans le tableau 3 on précise pour les objets des classes initiales F_2, F_3, F_5 l'affectation aux classes de la partition F selon l'appartenance aux types a priori.

	F_1^*	F_2^*	F_3^*	F_4^*	F_5^*	
F_1	20					20
F_2		30	1		5	36
F_3		4	16			20
F_4				13		13
F_5	1				23	24
	21	34	17	13	28	

Tableau 2

		F_1^*	F_2^*	F_3^*	F_4^*	F_5^*	
F_1	$\mathcal{R}_o$	20					20
F_2	$\mathcal{R}_o$		9			3	12
	$\mathcal{R}_{1+}$		10	1		1	12
	$\mathcal{R}_{2+}$		11			1	12
F_3	$\mathcal{R}_{1-}$		2	10			12
	$\mathcal{R}_1$		2	6			8
F_4	$\mathcal{R}_{1+-}$				13		13
F_5	$\mathcal{R}_{2-}$					12	12
	$\mathcal{R}_2$	1				11	12
		21	34	17	13	28	

Tableau 3

Les représentations des projections sur les 4 premiers axes discriminants de la partition F^*, des centres de gravité des classes (figures 3 et 4), et des objets de E (figures 5 et 6) illustrent et complètent les résultats précédents :

- le premier axe discriminant sépare la classe F_4^* de l'ensemble des 4 autres (figure 3)

- les classes F_3^* et F_5^* sont opposées sur le second (figure 3)

- la classe F_2^* se distingue sur le troisième (figure 4)

- la classe F_1^* apparaît sur le quatrième (figure 4)

- sur les figures 5 et 6 il apparaît des groupements homogènes d'individus et des zones frontières entre classes.

6.7 - L'échantillon test

6.7.1 - Introduction

Il n'a pas été possible d'obtenir l'échantillon test dont la constitution était initialement prévue. Aussi on a réparti par tirage au hasard (en respectant les proportions d'objets de chaque famille a priori) les 113 machines de E en :

- une population de base E_b de 85 objets

- une population test E_t de 28 objets

6.7.2 - L'étude de la population E_b

On a repris la démarche effectuée aux paragraphes 6.4, 6.5 et 6.6 sur la population de base E_b. On obtient des résultats analogues à ceux obtenus sur la population E, à savoir :

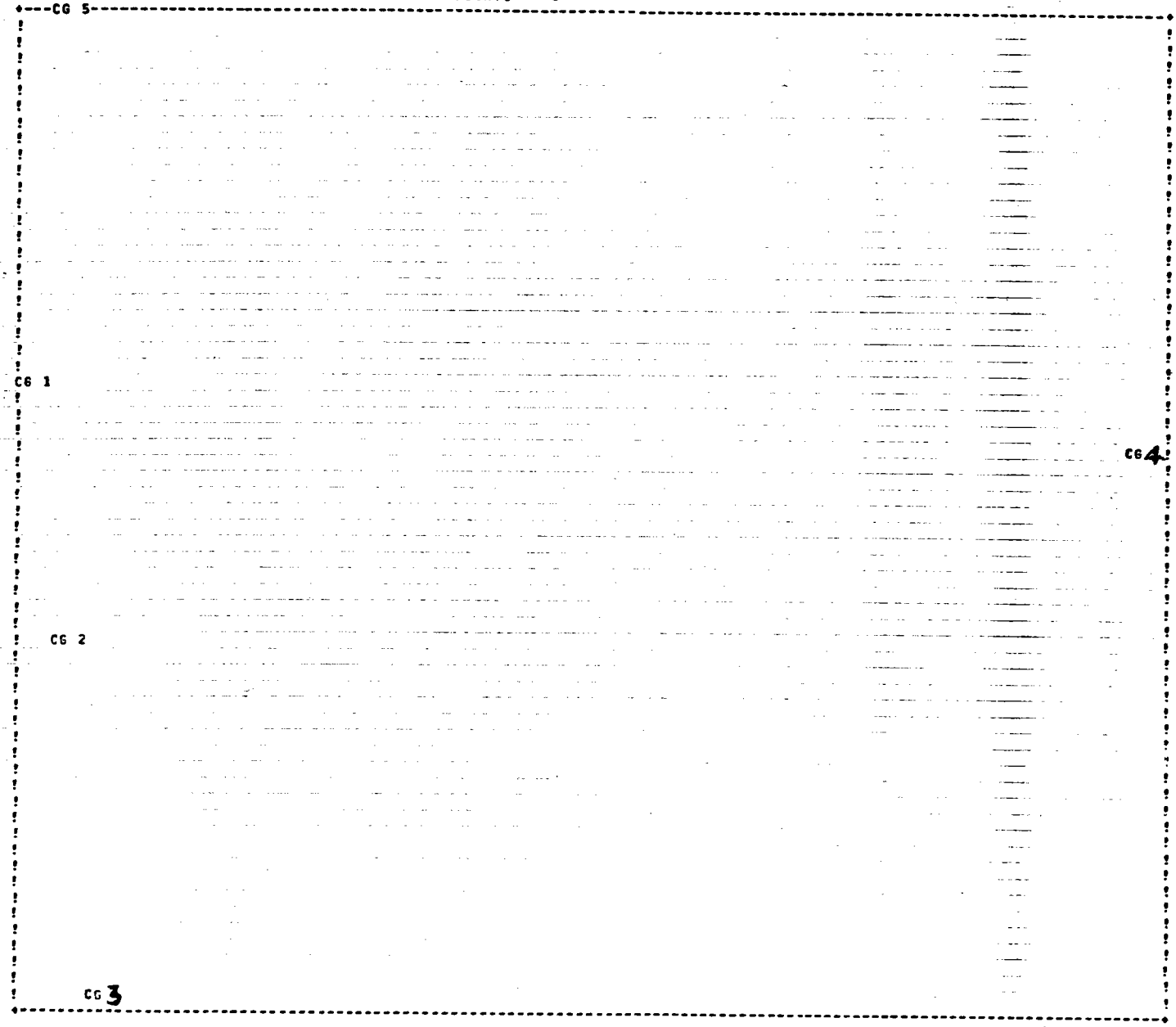


Figure 3 : Les projections des centres de gravité CG1,...,CG5 des classes $F_1^*,\ldots,F_5^*$ sur les axes u_1^* et u_2^* de la représentation $L^* = g\ (F^*)$.

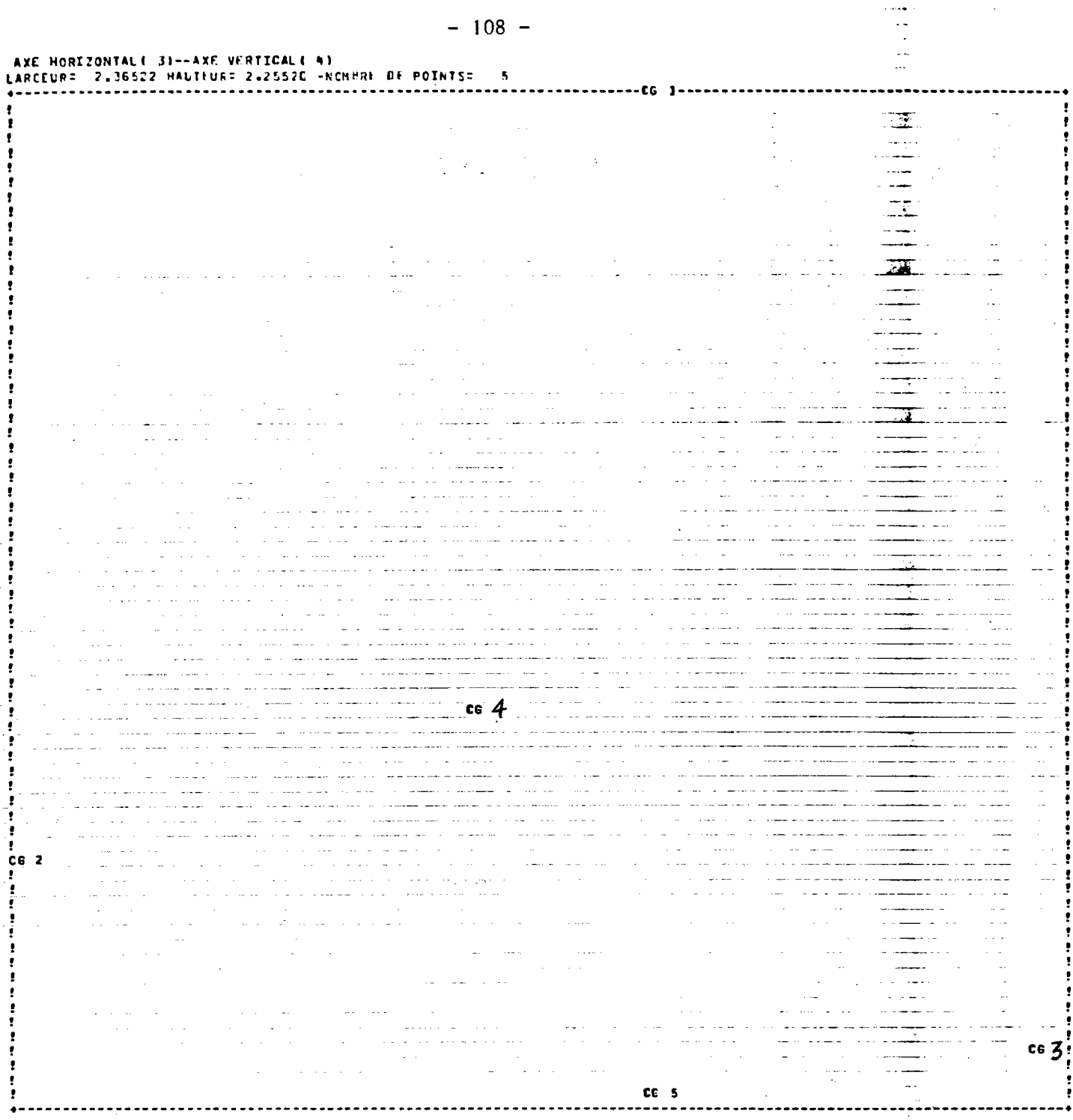


Figure 4 : Les projections des centres de gravité CG1,...,CG5 des classes $F_1^*,\ldots,F_5^*$ sur les axes u_3^* et u_4^* de la représentation $L^* = g(F^*)$

Figure 5 : Les projections des machines des classes F_1, F_2, F_3 et F_5 sur les axes u_1^* et u_2^* de la représentation $L^* = g(F^*)$. La classe F_4 est nettement séparée sur u_1^* (cf. figure 3) ; ses éléments ne sont pas projetés ici.
Les identificateurs des individus sont B, ST, SP, STP, CT, CP, CTP. Ils correspondent dans l'ordre aux classes a priori $T_1,\ldots, T_7$).
Séparation des classes F_3 et F_5 (cf. figure 6) ; Zone A <—> classe F_5
Zone B <—> classe F_3

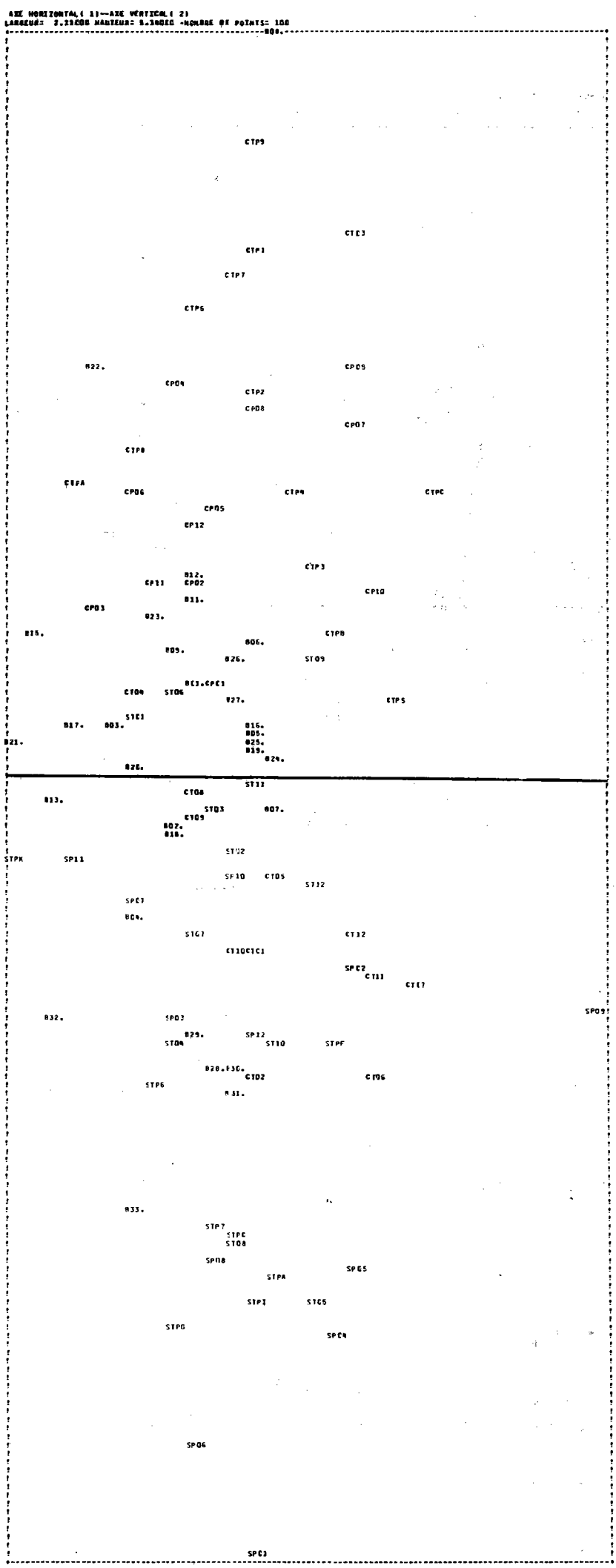

AXE HORIZONTAL(1)--AXE VERTICAL(2)
LARGEUR= 2.21E08 HAUTEUR= 8.30E10 -NOMBRE DE POINTS= 100
Zone A
Zone B

Figure 5

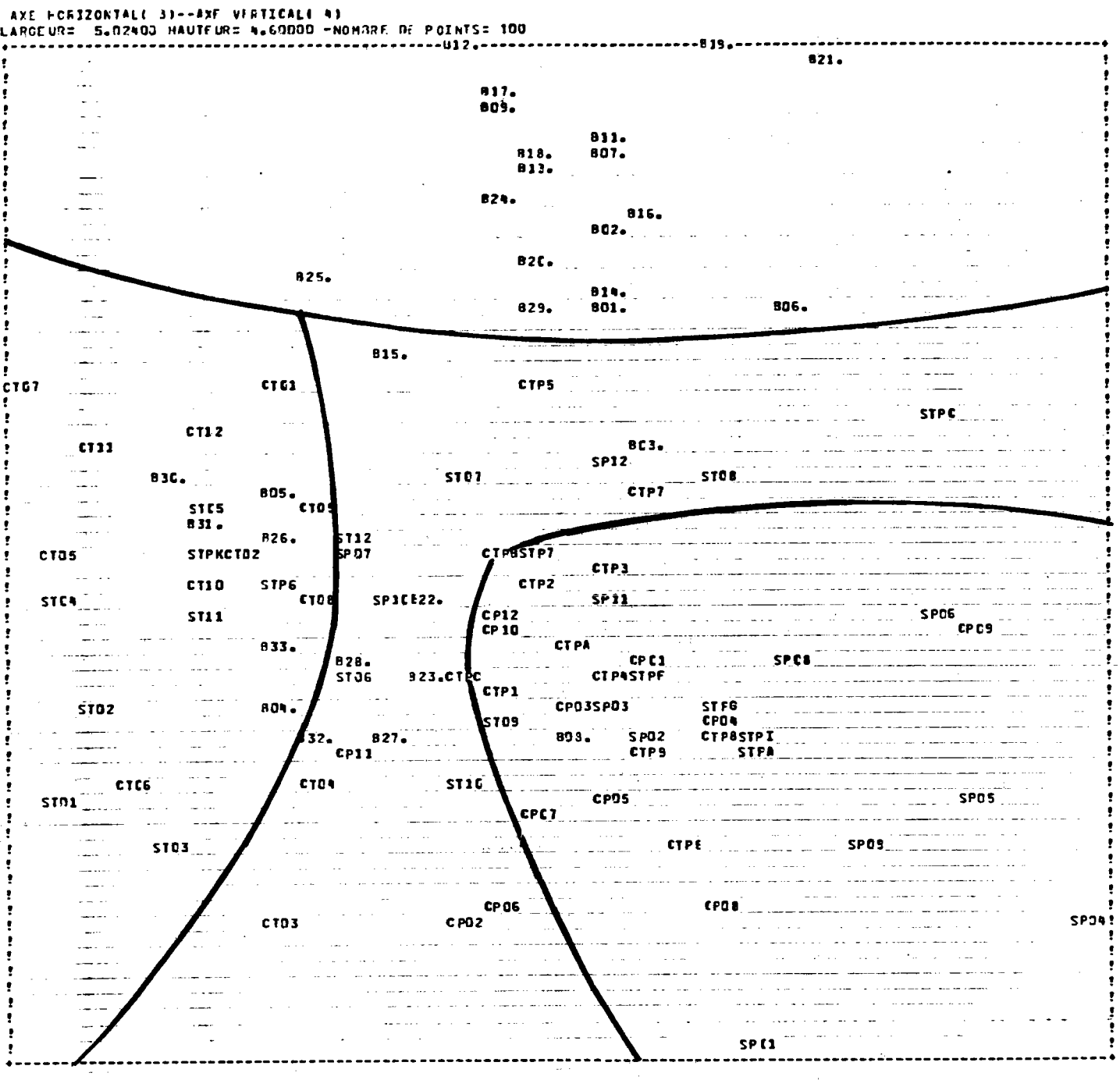
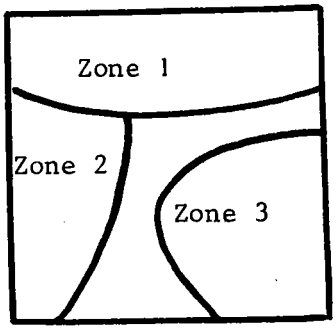


Figure 6 : Les projections des machines des classes F_1,F_2,F_3 et F_5 sur les axes u_3^* et u_4^* de la représentation $L^* = g(F^*)$



Zone 1 <=> classe F_1

Zone 2 <=> classe F_2

Zone 3 <=> classe F_3 ou classe F_5 (cf. figure 5)

Soit $T_b = (T_{b1},\ldots,T_{b7})$ la partition a priori de la population E_b.

Catégories	Classes de E_b	Effectifs
$\mathfrak{D}_o$	T_{b1}	24
$\mathfrak{D}_{1+}$	T_{b2}	9
$\mathfrak{D}_{1-}$	T_{b3}	9
$\mathfrak{D}_1$	T_{b4}	16
$\mathfrak{D}_{2+}$	T_{b5}	9
$\mathfrak{D}_{2-}$	T_{b6}	9
$\mathfrak{D}_2$	T_{b7}	9

Il apparait que les 7 classes de la partition T_b sont mal séparées par les paramètres descriptifs des objets: En particulier les classes T_{b2} et T_{b5} d'une part, T_{b6} et T_{b7} d'autre part sont mal séparées en projection sur les premiers axes discriminants de la partition T_b.

On définit par conséquent la partition $F_b = (F_{b1},\ldots,F_{b5})$ de E_b en 5 classes :

$F_{b1} = T_{b1}$

$F_{b2} = T_{b2} \cup T_{b5}$

$F_{b3} = T_{b3}$

$F_{b4} = T_{b4}$

$F_{b5} = T_{b6} \cup T_{b7}$

Remarque :

Les effectifs des classes de la partition T_b ne permettent plus la définition de sous classes.

L'analyse discriminante typologique initialisée par la partition F_b de E_b, (r=4), et qui converge dès la 1ère itération vers la partition F_b^* (Tableau 4), confirme les résultats de l'analyse précédente.

	F^*_{b1}	F^*_{b2}	F^*_{b3}	F^*_{b4}	F^*_{b5}	
F_{b1}	18	2			4	24
F_{b2}		16			2	18
F_{b3}		1	8			9
F_{b4}	1	1	1	13		16
F_{b5}				1	17	18
	19	20	9	14	23	

Tableau 4

6.7.3 - Les familles induites par l'analyse discriminante typologique

D'après les résultats de 6.7.2 on est amené à considérer non plus les 7 familles a priori ($\mathfrak{D}_o$, $\mathfrak{D}_{1+}$, $\mathfrak{D}_{1-}$, $\mathfrak{D}_1$, $\mathfrak{D}_{2+}$, $\mathfrak{D}_{2-}$ et $\mathfrak{D}_2$), mais les 5 familles suivantes :

$$\mathfrak{F}_1 = \mathfrak{D}_o$$

$$\mathfrak{F}_2 = \mathfrak{D}_{1+} \cup \mathfrak{D}_{2+}$$

$$\mathfrak{F}_3 = \mathfrak{D}_{1-}$$

$$\mathfrak{F}_4 = \mathfrak{D}_1$$

$$\mathfrak{F}_5 = \mathfrak{D}_{2-} \cup \mathfrak{D}_2$$

6.7.4 - La règle de décision

Soit

$L = g(F_b) = (g_1,\ldots,g_5,u_1,\ldots,u_4)$ la représentation associée à la partition F_b de E_b.(g_ℓ est le centre de gravité de la classe $F_{b\ell}$, $\ell=1,\ldots,5$)

$b_1,\ldots,b_4$ les facteurs associés aux vecteurs $u_1,\ldots,u_4$ ($b_i = V^{-1}u_i$, $i=1,\ldots,4$)

D_e la matrice de décision définie par :

$$D_e = \sum_{\ell=1}^{4} b_\ell \, b'_\ell$$

La règle de décision consiste à affecter tout nouvel objet x à la famille $\mathfrak{F}_\ell$ ($1 \leq \ell \leq 5$) si :

$$(x-g_\ell)' \, D_e \, (x-g_\ell) \leq (x-g_i)' \, D_e \, (x-g_i) \quad \text{pour } i=1,\ldots,5.$$

	$\mathcal{F}_1$	$\mathcal{F}_2$	$\mathcal{F}_3$	$\mathcal{F}_4$	$\mathcal{F}_5$	
F_1	4	3			1	8
F_2	2	4				6
F_3		1	2			3
F_4		2	1	2		5
F_5			1		5	6
	6	10	4	2	6	

Tableau 5

En ligne, la famille d'origine
En colonne, la famille de classement

On déduit de 6.7.2 (l'algorithme converge à la 1ère itération) que le nombre d'objets bien classés dans la population de base E_b par cette règle, est égal au nombre d'objets figurant sur la diagonale du tableau 4 : En effet, il s'agit du tableau croisé des cardinaux des classes des partitions F_b (associé aux types $\mathcal{F}_1, \ldots, \mathcal{F}_5$) et $f \circ g(F_b) = f(L)$. On a par conséquent 72 objets bien classés.

6.7.5 - L'affectation des objets de l'échantillon test

La règle de décision énoncée en 6.7.4 est appliquée aux 28 objets de la population test E_t. Les résultats sont portés dans le tableau 5. Cette règle

est satisfaisante, compte tenu des résultats obtenus en 6.6 concernant l'homogénéité des familles $\mathfrak{D}_o$ et $\mathfrak{D}_1$ (qui apparaissent également dans l'étude de la population de base E_t) dont il n'a pas été tenu compte dans la définition des types $\mathfrak{F}_1, \ldots, \mathfrak{F}_5$.

6.8 - Conclusion

D'après (6.4), (6.5) et (6.6) il apparaît du point de l'analyse descriptive des données que :

a) les 7 familles définies a priori, $\mathfrak{D}_o$, $\mathfrak{D}_{1+}$, $\mathfrak{D}_{1-}$, $\mathfrak{D}_1$, $\mathfrak{D}_{2+}$, $\mathfrak{D}_{2-}$, $\mathfrak{D}_2$, ne sont pas séparables par l'analyse discriminante linéaire. (G. CELEUX [5] a réalisé sur cette population E des analyses factorielles des correspondances et des analyses discriminantes pas à pas dont les résultats confirment ceux du §. 6.4).

b) la famille $\mathfrak{D}_o$ n'est pas globablement homogène. Certains objets classés dans $\mathfrak{D}_o$ sont difficilement séparables des familles $\mathfrak{D}_{1+}$ ou $\mathfrak{D}_{2+}$ qui par ailleurs sont proches entre elles.

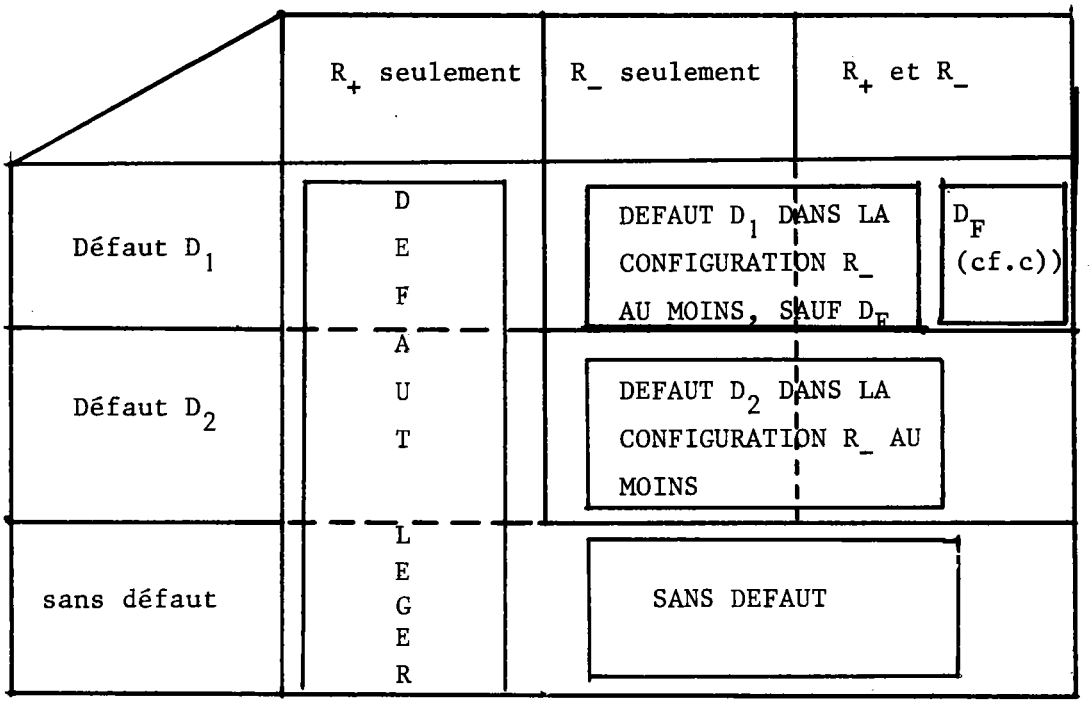
c) La famille $\mathfrak{D}_1$ n'est pas homogène. Elle est constituée :

- d'une part, d'une sous famille D_F interprétable par le spécialiste selon l'approche défaut de fabrication du contrôle de production

- d'autre part, d'objets proches de la famille D_{1-}

d) Les familles $\mathfrak{D}_{2-}$ et $\mathfrak{D}_2$ sont proches. Leur réunion est séparée des 5 autres familles par l'analyse discriminante linéaire.

Du point de vue de l'analyse discriminante décisionnelle, les résultats des paragraphes 6.6 et 6.7 montrent que 4 facteurs discriminants permettent de reconnaître les familles encadrées suivantes (présentées dans un tableau afin de souligner leur interprétation par le spécialiste) :



7 - L'INTERPRETATION DES RESULTATS

7.1.- La définition de nouvelles familles

L'analyse discriminante typologique initialisée par la partition a priori P^o induit une partition P^*, dont les K classes sont reconnues par r facteurs discriminants. (En effet, on a : $f(L^*) = P^*$). On recherche l'interprétation (vis à vis des types a priori) de chaque classe P^*_ℓ . En particulier, on peut avoir

- P^*_ℓ est égale à une classe a priori

- P^*_ℓ est incluse dans une classe a priori

- P^*_ℓ est égale à la réunion de deux classes a priori.

On en déduit (cf. 6.5.) une partition de E en K_1 classes (K_1 peut être différent de K), $C = (C_1,\ldots,C_{K_1})$.

Pour $\ell=1,\ldots,K_1$, la classe C_ℓ est :

- égale à une classe P_i^o de la partition P^o
- ou incluse dans une classe P_i^o de la partition P^o
- ou enfin, égale à la réunion de plusieurs classes ou sous classes de la partition P^o.

L'analyse discriminante typologique initialisée par la partition $C = (C_1,\ldots,C_{K_1})$ induit une partition C^* en K_1 classes.

Les classes de la partition C qui sont restées stables dans l'algorithme, sont conservées. Les autres sont refondues en classes plus homogènes. Eventuellement le nombre d'axes discriminants r est ajusté.

Finalement, on déduit de la partition a priori T en K classes, une partition F de E en K^* classes, (cf. 6.6), $F = (F_1,\ldots,F_{K^*})$ et la dimension r^* telles que :

- les classes F_ℓ $(1 \leq \ell \leq K^*)$ soient interprétables par le specialiste (Eventuellement certaines classes pourront faire l'objet d'un second niveau de discrimination, cf. 7.3).

- la stabilité de la partition F par l'analyse discriminante typologique soit satisfaisante. En particulier, la représentation g(F) induit une partition $f \circ g(F)$ peu différente de F; c'est à dire, la représentation g(F) permet la reconnaissance de la partition F par la fonction d'affectation f.

A la partition $F = (F_1,\ldots,F_{K^*})$ de E il correspond une partition $\mathcal{F}_1,\ldots\mathcal{F}_{K^*}$ de $\mathcal{E}$, en K^* catégories ou formes déduites des familles a priori et dont les échantillons $F_1,\ldots,F_{K^*}$ sont séparés par la règle de décision $\mathcal{R}$ définie au paragraphe suivant .

7.2 - La règle de décision $\mathfrak{R}$

7.2.1 - Définition

Soit :

$L = g(F) = (g_1,\ldots,g_{K^*},\ u_1,\ldots,u_{r^*})$

$b_i = V^{-1}\ u_i \qquad i=1,\ldots,r^*$

$D_e = \sum_{i=1}^{r^*} b_i\ b_i'$

L'objet x est affecté à la famille $\mathfrak{F}_\ell$ si :

$(x-g_\ell)'\ D_e\ (x-g_\ell) \leq (x-g_i)'\ D_e\ (x-g_i) \qquad i=1,\ldots,K^*$

On a :

$(x-g_i)'\ D_e\ (x-g_i) = (x-g_i)\ V^{-1}\ (\sum_{s=1}^{r^*} u_s u_s')\ V^{-1}\ (x-g_i)$

On en déduit d'après (0-8) que la règle $\mathfrak{R}$ consiste en l'affectation à la famille dont le centre est le plus proche en projection (V^{-1}-orthogonale) sur le sous espace vectoriel U engendré par, $u_1,\ldots,u_{r^*}$, pour la métrique V^{-1}.

La règle $\mathfrak{R}$ peut être complétée par la visualisation de la projection de l'objet x à classer sur le sous espace U engendré par $u_1,\ldots,u_r$. En effet, la projection des objets de E sur U permet de définir des zones frontières entre les classes $F_1,\ldots,F_K$ et donc entre les familles $\mathfrak{F}_1,\ldots,\mathfrak{F}_K$ (figures 5 et 6). Pour un objet x dont la projection sur U appartient à une zone frontière, la règle de décision $\mathfrak{R}$ est peu fiable. Le spécialiste peut dans ce cas décider de l'affectation de x en tenant compte a postériori d'un coût de mauvais classement.

Il peut aussi être défini une zone de non affectation si le spécialiste considère qu'il est préférable dans la phase opérationnelle de créer une famille de "non classés" plutôt que d'augmenter fortement le risque de mau-

vais classements. Cette démarche est adaptée en particulier aux applications médicales, pour lesquelles une erreur de classement, peut être infiniment grave de conséquences, sans que son coût puisse être chiffré et infléchir en conséquence une règle de décision.

La règle de décision $\mathfrak{R}$ doit être validée sur un échantillon test, c'est à dire un ensemble d'objets qui n'ont pas participé à l'ellaboration de $\mathfrak{R}$, et dont on connaît la famille d'appartenance.

7.2.2 - Comparaison avec la règle de décision bayesienne [29]

Soit $G_1,\ldots,G_K$ les K groupes définis a priori. La mise en oeuvre de cette règle suppose connues ou estimées les quantités suivantes :

- $f_1(x),\ldots,f_K(x)$, densités de probabilités dans $\mathbb{R}^p$ pour chaque groupe

- $q_1,\ldots,q_K$, probabilités a priori d'appartenance à un groupe

- $C(j/i)$, coût de classement d'un objet appartenant au groupe G_i dans le groupe G_j (on a : $C(j)/j) = 0$ pour $j=1,\ldots,K$; En général $C(j/i) \neq C(i/j)$).

L'objet x est affecté par la règle de décision bayesienne au groupe G_ℓ

$$\sum_{j=1}^{K} C(\ell,j)q_j f_j(x) \leq \sum_{j=1}^{K} C(i,j)q_j f_j(x) \qquad i=1,\ldots,K$$

On démontre, [29], que cette règle minimise le coût moyen a posteriori de l'affectation de l'objet x à l'un des groupes G_ℓ $(1 \leq \ell \leq K)$ qui s'écrit :

$$\sum_{\ell=1}^{K} \sum_{x\in P_\ell} \sum_{j=1}^{K} C(\ell,j)\, q_j\, f_j(x)$$

On démontre, [29], que sous les 4 hypothèses suivantes

H1 - Multinormalité des distributions dans chaque groupe

H2 - Egalité des matrices de variance covariance dans les groupes

H3 - Equiprobabilité a priori des K groupes

H4 - Egalité des coûts : $c(i/j) = 1, \quad i\neq j$

la règle de Bayes est équivalente à l'affectation au groupe G_r dont le centre de gravité est le proche dans R^p au sens de la métrique définie par V^{-1} inverse de la matrice de variance - covariance totale. La règle de Bayes dans ce cas particulier est différente de la règle $\mathcal{R}$ (7.2.1) qui procède à l'affectation au groupe dont le centre de gravité est le plus proche en projection sur le sous espace U pour la métrique V^{-1}.

La règle de Bayes est rarement applicable dans sa généralité :

- Les distributions dans les groupes ne sont en général pas connues, et une bonne estimation est souvent difficile, surtout si le nombre de variables est élevé.

- il est difficile de chiffrer les coûts de mauvais classement.

Dans le cas où on fait l'hypothèse de normalité des classes, la prise en compte de matrices de variance covariance différentes selon les groupes est un atout si leur estimation est fiable, ce qui suppose l'accès à de larges échantillons de chaque famille a priori.

Dans l'analyse discriminante typologique seule la matrice de variance-covariance totale V est estimée sur l'échantillon E (dont le cardinal est K fois supérieur au cardinal moyen des échantillons des K familles a priori). D'autre part, on a remarqué, (7.2.1), que la règle $\mathcal{R}$ pouvait être modulée par la prise en compte a posteriori de coûts de mauvais classements. (De même, on peut considérer les probabilités d'appartenance aux différentes familles). Par contre la règle $\mathcal{R}$ ne permet pas de tenir compte de la non égalité éventuelle des matrices de variance-covariance dans les familles a priori.

L'utilisation de la règle $\mathcal{R}$ est justifiée sur la partition des catégories induites par l'analyse discriminante typologique, parce qu'elle est issue de cette méthode. De plus, elle est simple, non paramétrique et présente l'avantage de ne pas tenir compte des paramètres de bruit. Cependant on peut considérer l'analyse discriminante typologique comme une méthode d'analyse discriminante descriptive et constructive d'une nouvelle partition, et utiliser dans la phase décisionnelle, la règle de Bayes.

7.3 - Le second niveau de discrimination

Si la partition $\mathcal{F} = (\mathcal{F}_1,\ldots,\mathcal{F}_{K^*})$ de $\mathcal{E}$ est jugée trop grossière par le spécialiste, on réalise un second niveau de discrimination. Pour toute catégorie $\mathcal{F}_\ell$ $(1 \leq \ell \leq K^*)$ qui apparaît comme une réunion de familles $\mathcal{F}_{\ell,1},\ldots,\mathcal{F}_{\ell,K}$ qui doivent être séparées, on réalise une analyse discriminante pour laquelle :

- la population étudiée est $\mathcal{F}_\ell$
- les familles définies a priori sont $\mathcal{F}_{\ell,1},\ldots,\mathcal{F}_{\ell,K}$

Globalement la règle de décision se présente sous la forme d'un arbre.

7.4.- Exemples

7.4.1 - Exemple 1

Soit P la partition en 3 classes du nuage E représenté sur la figure 7. (Les objets sont décrits par deux paramètres).

L'Analyse Discriminante Typologique initialisée par la partition a priori P (on choisit r = 1) converge vers une partition $P^* = (P_1^*,P_2^*,P_3^*)$. Le sous espace de représentation associé est la droite D^* (cf. figure 8). On remarque que :

- Les classes P_3 et P_3^* sont égales
- Chacune des classes P_1^* et P_2^* est constituée en proportions comparables d'éléments des classes a priori P_1 et P_2
- En projection sur D^*, la classe P_3^* est séparée de l'ensemble des deux classes P_1^* et P_2^*.

On en déduit la partition en deux classes $T = (T_1,T_2)$ suivante :

$$T_1 = P_1 \cup P_2$$

$$T_2 = P_3$$

T traduit la structure du nuage E, relativement aux classes définies a priori (cf. figure 9).

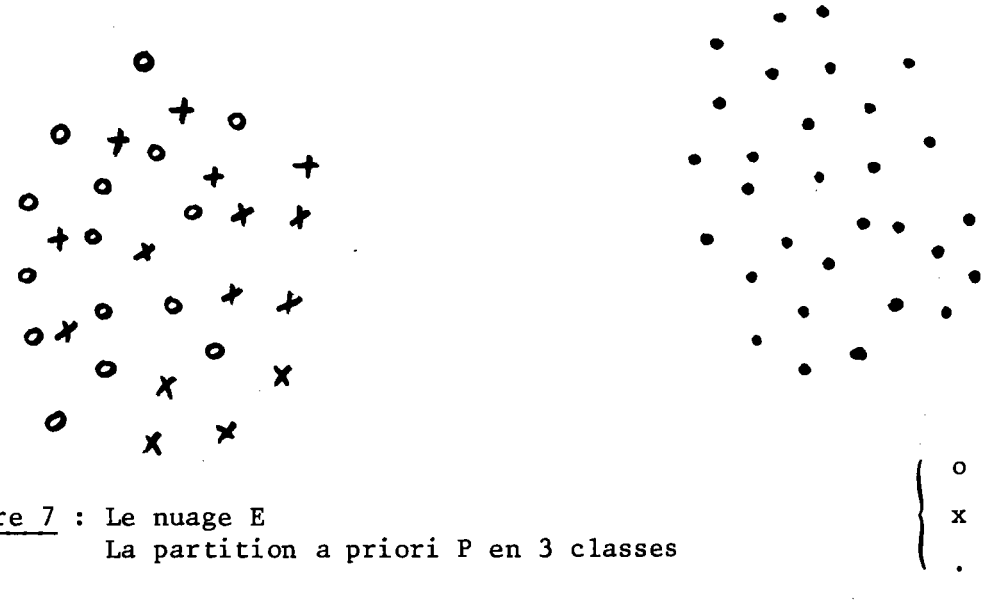


Figure 7 : Le nuage E
La partition a priori P en 3 classes

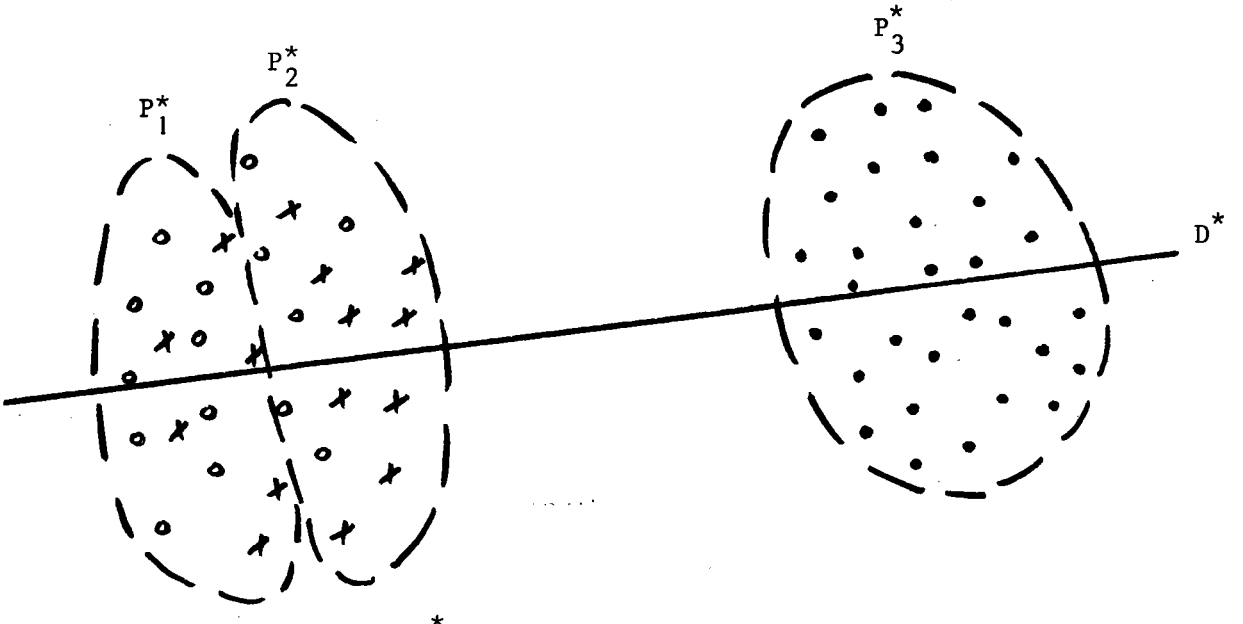


Figure 8 : La partition P^*
La droite discrimiante D^*

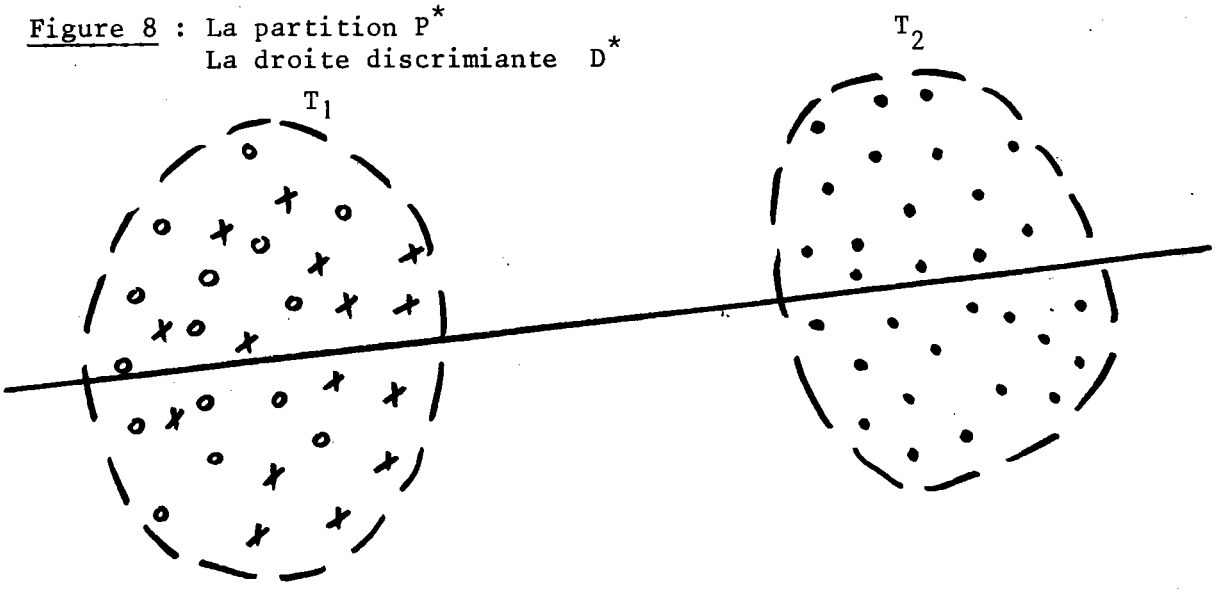


Figure 9 : La partition T = (T_1, T_2).
Les classes T_1 et T_2 sont séparées par une droite discriminante.

7.4.2 - Exemple 2

Soit Q la partition en 2 classes du nuage N représenté sur la figure 10 (les objets sont décrits par deux paramètres).

L'Analyse Discriminante Typologique initialisée par la partition a priori Q (on choisit r = 1) converge vers une partition $Q^* = (Q_1^*, Q_2^*)$ dont les classes ne sont pas interprétables relativement aux classes définies a priori Q_1 et Q_2. (cf. figure 11).

La projection des objets sour la droite de représentation U^* suggère l'éclatement de la classe Q_2 en deux sous classes, Q_{21} et Q_{22}.

On en déduit la partition en 3 classes $C = (C_1, C_2, C_3)$ suivante :

$$C_1 = Q_1$$
$$C_2 = Q_{21} \subset Q_2$$
$$C_3 = Q_{22} \subset Q_2$$

qui traduit la structure du nuage E relativement aux classes définies a priori (cf. figure 12)

Une droite discriminante sépare les classes C_1, C_2 et C_3, et par conséquent les familles a priori Q_1 et Q_2.

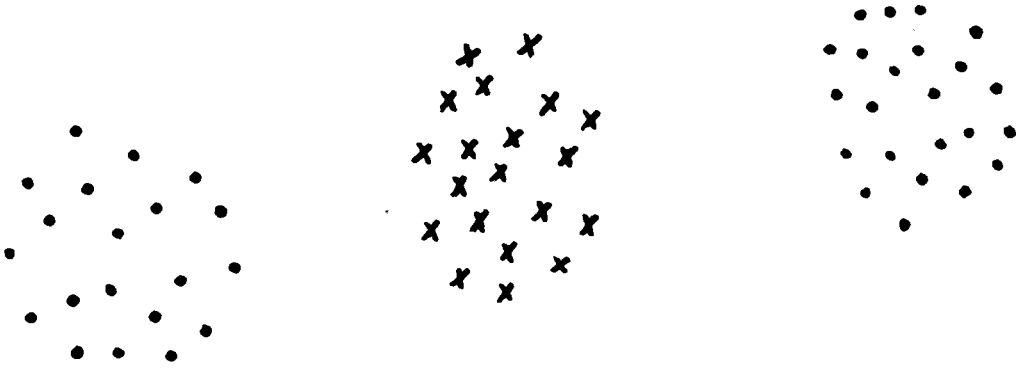


Figure 10 : Le nuage N
La partition a priori Q en 2 classes $\begin{cases} x\ Q_1 \\ .\ Q_2 \end{cases}$

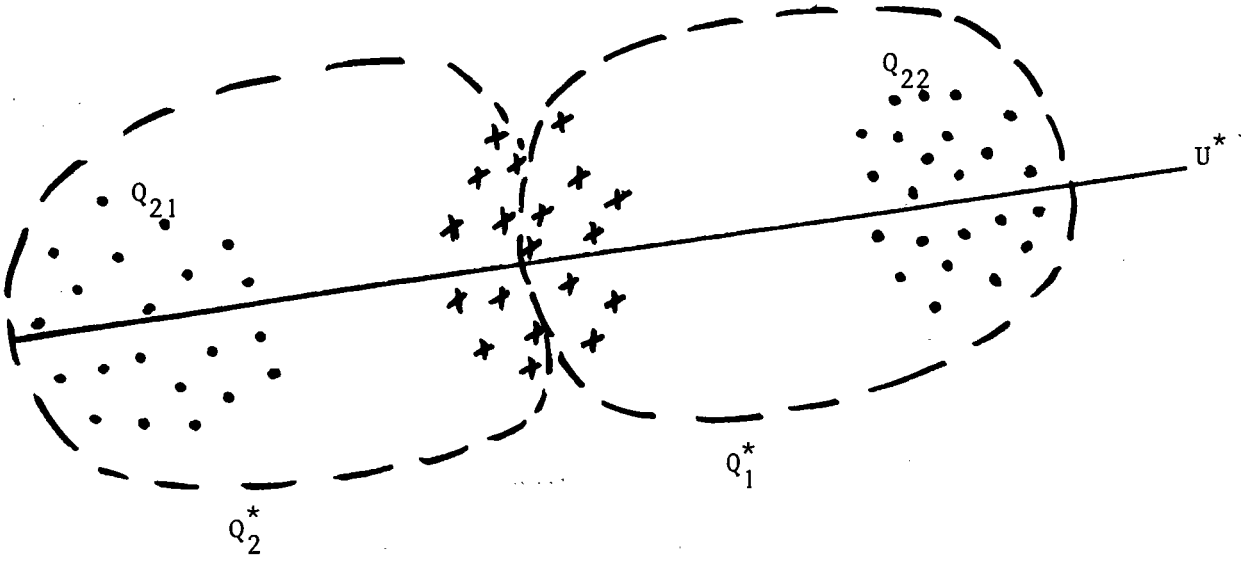


Figure 11 : La partition $Q^* = (Q_1^*, Q_2^*)$
La droite discriminante U^*

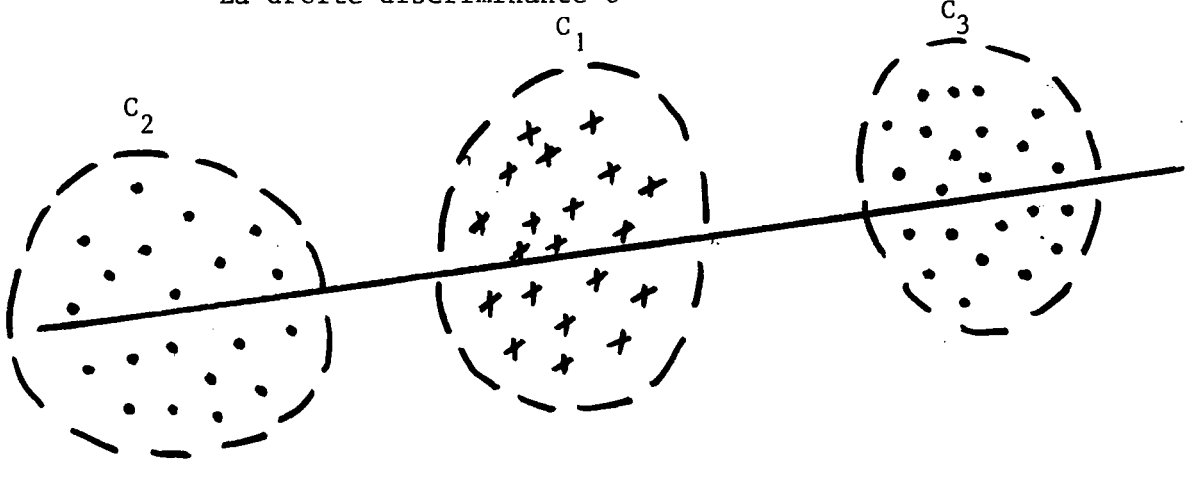


Figure 12 = La partition $C = (C_1, C_2, C_3)$.

Les classes C_1, C_2 et C_3 sont séparées par une droite discriminante

8 - CONCLUSION

Avec l'analyse discriminante typologique on aborde le problème de la reconnaissance de formes imprécises par une étude descriptive et inductive dans le cadre de la transformation linéaire des variables. La méthode permet de valider la structure définie a priori, ou d'en déduire des formes plus pertinentes du point de vue de la classification et de la discrimination.

L'analyse discriminante typologique fournit une règle de décision pour la reconnaissance des formes qu'elle induit; On en déduit en particulier une reconnaissance fiable de certaines familles a priori. Pour les autres on réalise éventuellement un 2ème niveau local de discrimination. On peut aussi déduire des nouvelles formes une initialisation privilégiée de l'analyse discriminante locale [11].

Par ailleurs la visualisation sur quelques axes factoriels du problème de classement d'un nouvel objet permet de prendre en compte des coûts de mauvais classement et des probabilités a priori d'appartenance aux classes, dans le cas où l'affectation géométrique au centre le plus proche semble peu fiable.

Selon des exigences liées au problème réel posé et selon les données dont on dispose, on peut réaliser sur la partition induite par l'analyse discriminante typologique une analyse discriminante pas à pas [20 , 25] ou bayesienne [21, 29].

Dans l'Analyse Discriminante Typologique on pourrait envisager de substituer à l'inertie intra classe expliquée, d'autres critères liés à l'analyse discriminante ; le sous espace de représentation d'une partition pourrait, par exemple, privilégier une zone frontière entre classes [23].

CHAPITRE III

L'ANALYSE DISCRIMINANTE TYPOLOGIQUE SOUS CONTRAINTE DE MODELISATION LOCALE

1 - INTRODUCTION

1.1 - Le problème

On étudie un processus décrit par :

- p paramètres d'entrée $x^1,\ldots,x^p$
- q_1 variables d'actions $z^1,\ldots,z^{q_1}$
- q_2 variables de sortie $z^{q_1+1},\ldots,z^q$; $q=q_1+q_2$

On note : $J_1 = \{x^1,\ldots,x^p\}$ $J_2 = \{z^1,\ldots,z^q\}$

On suppose que pour toute réalisation du processus dans une configuration fixée dans J_1 (dont la définition est équivalente à la donnée des valeurs des paramètres $x^1,\ldots, x^p$) les variables de J_2 sont liées par un modèle h issu d'une famille H.

Il s'agit de déterminer K familles de configurations, $\mathcal{C}_1,\ldots,\mathcal{C}_K$, et K modèles de H, $h_1,\ldots,h_K$, tels que, pour $\ell=1,\ldots,K$, on ait :

Pour toute configuration c dans $\mathcal{C}_\ell$ et toute réalisation du processus dans la configuration c, les variables de J_2 soient liées par le modèle h_ℓ.

Dans la phase opérationnelle (après construction de la modélisation locale) avant chaque nouvelle réalisation du processus, on connaît la configuration décrite par J_1; la connaissance du modèle associé permet d'optimiser le choix des paramètres d'actions au sens d'un critère de coût [19].

Le découpage de chaque paramètre x^j (j=1,...,p) en n_j modalités, et la définition des familles de configurations par tri croisé sur les modalités présente l'inconvénient de multiplier le nombre des familles qui s'élève à $n_1 \times n_2 \times \ldots \times n_p$.

On suppose ici que les paramètres x^j (j=1,...,p) sont quantitatifs; les K familles de configurations peuvent être définies par K centres dans $\mathbb{R}^p$, et l'affectation de toute configuration au centre le plus proche pour une métrique donnée. L'analyse factorielle discriminante permet de définir un sous espace U dans $\mathbb{R}^p$ optimal pour la séparation des K familles (au sens du maximum du rapport des inerties inter classe et totale expliquées). Les K familles peuvent alors être définies par K centres et affectation de toute configuration au centre le plus proche <u>après projection sur U</u>. (La métrique est défini par l'inverse de la matrice de variance covariance des paramètres $x^1,...,x^p$), (cf. Ch. II). Cette approche permet d'isoler dans J_1 l'information utile à la reconnaissance des classes.

1.2 - <u>Les données</u>

On dispose d'un ensemble E de n réalisations du processus $x_1,...,x_n$, pour lesquelles on connait les valeurs prises par :

- les p paramètres d'entrée $x^1,...,x^p$
- les q_1 variables d'action $x^1,...,x^{q_1}$
- les q_2 variables de sortie $x^{q_1+1},...,x^q$ $\quad q=q_1+q_2$

Les données peuvent être rangées dans une matrice à n lignes et p+q colonnes

$$\begin{pmatrix} x_1^1 \ldots\ldots\ldots x_1^p & z_1^1 \ldots\ldots\ldots z_1^q \\ \vdots & \\ x_i^1 \ldots\ldots\ldots x_i^p & z_i^1 \ldots\ldots\ldots z_i^q \\ \vdots & \\ x_n^1 \ldots\ldots\ldots x_n^p & z_n^1 \ldots\ldots\ldots z_n^q \end{pmatrix}$$

J_1 $\qquad\qquad$ J_2

On suppose que les n réalisations du processus ont des poids égaux : $p_i = \frac{1}{n}$, i=1,...,n.

Pour i=1,...,n, on note :

$x_i' = (x_i^1,\ldots,x_i^p)$

$z_i' = (z_i^1,\ldots,z_i^q)$

On suppose la donnée par le spécialiste :

- d'une famille H de modèles admissibles

- d'une application δ qui à toute partie A de E et tout modèle h de H associe une mesure de l'homogénéité de A relativement à h :

$\delta \quad \mathcal{P}(E) \times H \to R_+$

$\delta \quad (A,h) \to \delta(A,h)$

On suppose que pour tout couple (A_1,A_2) de parties de E, et tout modèle h dans H, on a :

$$\delta(A_1 \cup A_2,h) = \delta(A_1,h) + \delta(A_2,h)$$

En particulier, pour toute partie A de E, et tout modèle h de H, on a :

$$\delta(A,h) = \sum_{x\in A} \delta(x,h) \qquad (1\text{-}1)$$

- d'une fonction notée g_M qui a toute partie A non vide de E fait correspondre un modèle noté $g_M(A)$ dans H vérifiant :

$$\delta(A,g_M(A)) = \min_{h\in H} \delta(A,h) \qquad (1\text{-}2)$$

1.3 - Exemples de modélisation

1.3.1 - Le modèle linéaire

H est la famille des modèles

$$\alpha_o + \alpha_1 z^1 + \ldots + \alpha_q z^q = 0$$

où $\alpha_o, \alpha_1, \ldots, \alpha_q$ sont réels et $\sum_{i=1}^{q} \alpha_i^2 = 1$.

L'indice δ est défini par :

$$\delta(A) = \sum_{x \in A} (\alpha_o + \alpha_1 z^1(x) + \ldots + \alpha_q z^q(x))^2$$

Pour toute partie A de E, le modèle $g_M(A)$ est défini par une régression orthogonale sur A [16]. (En cas d'indétermination dans la détermination des coefficients de cette régression des conventions assurent à g_M la qualité de fonction).

1.3.2 - Le modèle de l'inertie

Une métrique M est donnée dans $\mathbb{R}^q$. H est la famille des modèle I(a,) inertie au point a de R^q (Annexe 1).

L'indice δ est défini par :

$$\delta(A) = I(a,A)$$

Pour toute partie A de E, le modèle $g_M(A)$ est le modèle I(g,) où g est le centre de gravité de A. (Annexe 1).

1.4 - La méthode

L'espace des configurations est identifié à l'espace vectoriel $\mathbb{R}^p$. Il est muni du produit scalaire défini par la matrice V^{-1}, inverse de la matrice de covariance empirique des paramètres $x^1, \ldots, x^p$

$$V = \frac{1}{n} X' X$$

où

$$\bar{X} = \begin{pmatrix} x_1^1 - \bar{x}^1 & \cdots\cdots & x_1^p - \bar{x}^p \\ \vdots & & \vdots \\ x_n^1 - \bar{x}^1 & \cdots\cdots & x_n^p - \bar{x}^p \end{pmatrix}$$

en notant $\bar{x}^1,\ldots,\bar{x}^p$ les moyennes sur E des paramètres $x^1,\ldots,x^p$.

V est aussi la matrice d'inertie du nuage dans R^p, des individus $x_1,\ldots,x_n$ caractérisés par les paramètres $x^1,\ldots,x^p$.

Si V n'est pas régulière, X n'est pas de rang p, et on se restreint à ensemble de paramètres sous ensemble de J_1.

On recherche :

- une partition P de E en K classes : $P = (P_1,\ldots,P_K)$

- une famille $\underline{H}$ de K modèles de H : $\underline{H} = (h_1,\ldots,h_K)$

- un sous espace U de dimension r de l'espace des configurations identifié à $\mathbb{R}^p$.

tels que le triplet $(P,\underline{H},U)$ minimise un critère W qui rende compte <u>simultanément</u>

- de l'homogénéité des classes P_ℓ de la partition P vis à vis du modèle h_ℓ qui leur est associé ($\ell=1,\ldots,K$)

- et de la séparation des classes de la partition P en projection sur le sous espace U des paramètres explicatifs.

L'approche est différente de celle des méthodes multi critère [18], et bien sûr de celle qui consiste à effectuer une modélisation locale [6] puis une analyse discriminante.

Finalement, de la population échantillon E, on déduit une famille $\underline{H}^*$ de K modèles, K centres et un sous espace U^* dans $\mathbb{R}^p$ qui constituent une structure construite dans le cadre

- de la modélisation locale sur J_2

- et de la reconnaissance des classes par les paramètres explicatifs de J_1 (une règle de décision géométrique simple permet l'affectation d'un modèle à une nouvelle réalisation du processus).

1.5. Notations

On note :

$\mathbb{P}_K$ l'ensemble des partitions de E en K classes (disjointes)
Une partition P de $\mathbb{P}_K$ est notée également $(P_1,\ldots,P_K)$

$E = P_1 \cup \ldots \cup P_K$

$P_i \cap P_j = \emptyset \quad i = 1,\ldots,K \; ; \quad j = 1,\ldots,K \; ; \quad i \neq j$

F_r l'ensemble des sous espaces vectoriels de R^p de dimension r

U_r l'ensemble des familles V^{-1} orthonormées de r vecteurs de R^p

F_r^* l'ensemble des familles V orthonormées de r formes linéaires ou facteurs sur $\mathbb{R}^p$

2 - LE CRITERE

2.1 - Le critère W_1

Soit P une partition de E,

Pour tout sous espace U de R^p, soit $I_{intra}(P, U)$ l'inertie intra classe de la partition P, expliquée par U.

On définit un critère W_1 sur l'ensemble des partitions de E par :

$$W_1(P) = \underset{U \in F_r}{\text{Min}} \; I_{intra}(P, U) \qquad (2\text{-}1)$$

Le critère W_1 mesure l'homogénéité moyenne des classes de la partition, en projection sur un sous espace de dimension r de l'espace des paramètres explicatifs J_1.

Le critère W_1 est identique au critère W du chapitre II.

On déduit de la relation (2-5) du paragraphe II.2.3, l'expression suivante du critère W_1

$$W_1(P) = r - \underset{U \in F_r}{\text{Max}} \; [I_{inter}(P, U)] \qquad (2\text{-}2)$$

où $I_{inter}(P, U)$ est l'inertie inter classe de la partition P expliquée par le sous espace U.

On déduit des relations (2-13) et (2-15) des paragraphes II.2.5.4 et II.2.5.5 les expressions suivantes du critère W_1, en termes de variances intra classe ou inter classe de facteurs discriminants ;

$$W_1(P) = \underset{(b_1, \ldots, b_r) \in F_r^*}{\text{Min}} \; [\sum_{s=1}^{r} \text{Var}_{intra}(P, b_s)] \qquad (2\text{-}3)$$

$$W_1(P) = r - \max_{(b_1,\ldots,b_r)\in F_r^*} \left[\sum_{s=1}^{r} \mathrm{Var}_{inter}(P, b_s) \right] \qquad (2\text{-}4)$$

(F_r^* est l'ensemble des familles V orthonormées de r facteurs sur $\mathbb{R}^p$).

2.2 - Le critère W_2

Soit $P = (P_1,\ldots,P_K)$ une partition dans $\mathbb{P}_K$.

Pour toute classe P_ℓ, $\ell=1,\ldots,K$, et tout modèle h dans H, on connaît l'indice d'homogénéité de la classe P_ℓ relativement au modèle h, $\delta(P_\ell, h)$.

On définit un critère W_2 sur l'ensemble $\mathbb{P}_K$ des partitions de E en K classes par :

$$W_2(P) = \min_{(h_1,\ldots,h_K)\in H\times\ldots\times H} \left(\sum_{\ell=1}^{K} \delta(P_\ell, h_\ell) \right) \qquad (2\text{-}5)$$

Le critère W_2 s'écrit plus simplement :

$$W_2(P) = \sum_{\ell=1}^{K} \min_{h\in H} \delta(P_\ell, h) \qquad (2\text{-}6)$$

Le critère W_2 exprime l'homogénéité moyenne des classes de la partition P relativement au modèle optimal associé à chacune d'elle.

2.3 - Le critère W

On définit un critère W sur l'ensemble $\mathbb{P}_K$ des partitions de E en K classes en posant :

$$W(P) = W_1(P) + \beta\, W_2(P) \qquad (2\text{-}7)$$

où β est une constante positive donnée (β est choisie de manière à équilibrer les poids des critères W_1 et W_2 (cf. § 6.2, 6.3).

Le critère W donne une mesure pondérée de l'homogénéité des classes de la partition vis à vis des paramètres explicatifs et de la modélisation.

2.4 - L'optimisation du critère W

On recherche une partition de E en K classes, P_K^* qui réalise le minimum du critère W pour P dans $\mathbb{P}_K$:

$$W(P_K^*) = \min_{P \in \mathbb{P}_K} W(P)$$

2.5 - Expression mathématique du critère W_1

On montre comme au paragraphe 2.4 du chapitre II, les propriétés et relations suivantes :

Soit

$P = (P_1, \ldots, P_K)$ une partition dans $\mathbb{P}_K$

$g_1, \ldots, g_K$ les centres de gravité des classes $P_1, \ldots, P_K$

U un sous espace vectoriel de R^p.

Pour tout x dans R^p, x_U la projection V^{-1} orthogonale de x sur U.

d_U l'écart sur R^p défini par

$$d_U(x, y) = \sqrt{(x_U - y_U)' \, V^{-1} (x_U - y_U)}$$

On a :

$$I_{intra}(P, U) = \frac{1}{n} \sum_{\ell=1}^{K} \sum_{x \in P_\ell} d_U^2(x, g_\ell)$$

Pour toute base V^{-1} orthonormée, $(u_1, \ldots, u_r)$, du sous espace U, on a :

$$I_{intra}(P, U) = \frac{1}{n} \sum_{\ell=1}^{K} \sum_{x \in P_\ell} (x - g_\ell)' V^{-1} \left(\sum_{s=1}^{r} u_s u_s' \right) V^{-1} (x - g_\ell)$$

$$= \frac{1}{n} \sum_{s=1}^{r} u_s' V^{-1} \left(\sum_{\ell=1}^{K} \sum_{x \in P_\ell} (x - g_\ell)(x - g_\ell)' \right) V^{-1} u_s$$

On en déduit les expressions suivantes du critère W_1 (qui ne dépendent pas de la base V^{-1} orthonormée $(u_1,\ldots,u_r)$ du sous espace U) :

$$W_1(P) = \operatorname*{Min}_{U \in F_r} \left(\frac{1}{n} \sum_{\ell=1}^{K} \sum_{x \in P_\ell} (x - g_\ell)' V^{-1} \left(\sum_{s=1}^{r} u_s u_s' \right) V^{-1} (x - g_\ell) \right) \qquad (2\text{-}8)$$

$$W_1(P) = \operatorname*{Min}_{U \in F_r} \left(\frac{1}{n} \sum_{s=1}^{r} u_s' V^{-1} \left(\sum_{\ell=1}^{K} \sum_{x \in P_\ell} (x - g_\ell)(x - g_\ell)' \right) V^{-1} u_s \right) \qquad (2\text{-}9)$$

Par la transformation ;

$$u \to b = V^{-1} u$$

qui associe à toute une famille V^{-1} orthonormée $(u_1,\ldots,u_r)$ de r vecteurs de $\mathbb{R}^p$, une famille V orthonormée $(b_1 = V^{-1}u_1,\ldots,b_r = V^{-1}u_r)$ de r facteurs sur $\mathbb{R}^p$, (II.2.5), on déduit de (2-8) et (2-9) les relations suivantes :

$$W_1(P) = \operatorname*{Min}_{(b_1,\ldots,b_r) \in F_r^*} \left[\frac{1}{n} \sum_{\ell=1}^{K} \sum_{x \in P_\ell} (x - g_\ell)' \left(\sum_{s=1}^{r} b_s b_s' \right) (x - g_\ell) \right] \qquad (2\text{-}10)$$

$$= \operatorname*{Min}_{(b_1,\ldots,b_r) \in F_r^*} \left[\frac{1}{n} \sum_{s=1}^{r} b_s' \left(\sum_{\ell=1}^{K} \sum_{x \in P_\ell} (x - g_\ell)(x - g_\ell)' \right) b_s \right] \qquad (2\text{-}11)$$

(F_r^* est l'ensemble des familles V orthonormées de r facteurs sur R^p)

$b_s' \left(\sum_{\ell=1}^{K} \sum_{x \in P_\ell} (x-g_\ell)(x-g_\ell)' \right) b_s$ est égal à la variance intra classe du facteur b_s.

Dans la définition du critère W_1, l'optimisation sur le sous espace U dans F_r est équivalente à l'optimisation sur la famille de facteurs V orthogonaux $(b_1,\ldots,b_r)$ dans F_r^*.

2.6 - Expression mathématique du critère W_2

Par définition du critère W_2, et d'après la relation (1-1), pour toute partition P dans P_K, on a :

$$W_2(P) = \sum_{\ell=1}^{K} \operatorname{Min}_{h \in H} \left(\sum_{x \in P_\ell} \delta(x, h) \right) \qquad (2\text{-}12)$$

3 - LA REPRESENTATION

3.1 - Définition

On appelle représentation tout ensemble L constitué de :

- K points (ou centres) de R^p, $a_1, \ldots, a_K$
- une famille $(u_1, \ldots, u_r)$ de U_r
- K modèles de la famille H, $h_1, \ldots, h_K$

$$L = (a_1, \ldots, a_K, u_1, \ldots, u_r, h_1, \ldots, h_K)$$

$$u_s' V^{-1} u_s = 1 \quad s = 1, \ldots, r$$

$$u_s' V^{-1} u_t = 0 \quad s = 1, \ldots, r \; ; \; t = 1, \ldots, r \; ; \; s \neq t.$$

On note :

$\underline{A} = (a_1, \ldots, a_K)$

$\underline{U} = (u_1, \ldots, u_r)$

U le sous espace de R^p engendré par $u_1, \ldots, u_r$

$\underline{H} = (h_1, \ldots, h_K)$

$L = (\underline{A}, \underline{U}, \underline{H})$

L'ensemble des représentations définies ci-dessus est notée $L_{K,r}$.

3.2 - Le critère W_R

W_R est une généralisation du critère W qui est utilisée dans l'algorithme.

3.2.1. Définition

Soit :

$P = (P_1,\ldots,P_K)$ dans $\mathbb{P}_K$

$L = (a_1,\ldots,a_K, u_1,\ldots,u_r, h_1,\ldots,h_K) = (\underline{A}, \underline{U}, \underline{H})$ dans $L_{K,r}$

U le sous espace engendré par les vecteurs $u_1,\ldots,u_r$

$I(\underline{A}, P, U)$ l'inertie aux points $a_1,\ldots,a_K$ de la partition P, expliquée par le sous espace U.

On pose :

$$W_D(P, \underline{A}, \underline{U}) = I(\underline{A}, P, U) \tag{3-1}$$

($W_D(P, \underline{A}, \underline{U})$ ne dépend pas de la famille V^{-1} orthonormée $\underline{U}$, mais seulement de l'espace U engendré par cette famille).

$$W_M(P, \underline{H}) = \sum_{\ell=1}^{K} \delta(P_\ell, h_\ell) \tag{3-2}$$

$$W_R(P, L) = W_R(P, \underline{A}, \underline{U}, \underline{H}) = W_D(P, \underline{A}, \underline{U}) + \beta\ W_M(P, \underline{H}) \tag{3-3}$$

où β est la constante positive introduite pour la définition du critère W (2-7).

Le critère W_D ne dépend ni de la famille $\underline{H}$ de modèles, ni des paramètres $z^1,\ldots,z^q$ de J_2. W_D concerne la discrimination des classes de la partition par les paramètres explicatifs $x^1,\ldots,x^P$ de J_1.

Le critère W_M ne dépend ni des familles $\underline{A}$ de centres et $\underline{U}$ de vecteurs dans R^P, ni des paramètres explicatifs $x^1,\ldots,x^P$ de J_1. W_M concerne la modélisation locale relative aux paramètres de J_2.

3.2.2 - Cas particulier

Soit $\underline{G} = (g_1,\ldots,g_K)$ la famille des centres de gravité des classes $P_1,\ldots,P_K$.

L'inertie $I(\underline{G}, P, U)$ est l'inertie intra classe de la partition P expliquée par le sous espace U :

$$I(\underline{G}, P, U) = I_{intra}(P, U)$$

On en déduit :

Si $L = (\underline{G}, \underline{U}, \underline{H})$

$$W_R(P, L) = I_{intra}(P, U) + \beta \sum_{\ell=1}^{K} \delta(h_\ell, P_\ell) \qquad (3\text{-}4)$$

3.2.3 - Proposition

Pour toute partition P dans $\mathbb{P}_K$

$$\min_{L\in L_{K,r}} W_R(P, L) = W(P)$$

Démonstration

La minimisation selon le triplet $(\underline{A}, \underline{U}, \underline{H})$ du critère W_R se ramène à la minimisation selon le couple $(\underline{A}, \underline{U})$ du critère W_D et à la minimisation en $\underline{H}$ du critère W_M. En effet :

$$\min_{L\in L_{K,r}} W_R(P, L) = \min_{L=(\underline{A}, \underline{U}, \underline{H})\in L_{K,r}} (W_D(P, \underline{A}, \underline{U}) + \beta W_M(P, \underline{H}))$$

$$\min_{L\in L_{K,r}} W_R(P, L) = \min_{(\underline{A}, \underline{U})} W_D(P, \underline{A}, \underline{U}) + \beta \min_{\underline{H}} W_M(P, \underline{H}) \qquad (3\text{-}5)$$

On sait, (A3.2.2), que pour tout sous espace U, l'inertie $I(\underline{A}, P, U)$ est minimale pour la famille $\underline{G}$ des centres de gravité des classes

$$\min_{\underline{A}} I(\underline{A}, P, U) = I(\underline{G}, P, U) = I_{intra}(P, U)$$

On en déduit que :

$$\min_{(\underline{A}, \underline{U})} W_D(P, \underline{A}, \underline{U}) = \min_{\underline{A}, U} I(\underline{A}, P, U) = \min_{U \in F_r} I_{intra}(P, U) \qquad (3\text{-}6)$$

De (2-1) et (3 6), on déduit :

$$\min_{(\underline{A}, \underline{U})} W_D(P, \underline{A}, \underline{U}) = W_1(P) \qquad (3\text{-}7)$$

D'autre part, par définition du critère W_2, (2-5), on a :

$$\min_{\underline{H}} W_M(P, \underline{H}) = W_2(P) \qquad (3\text{-}8)$$

De (3-5), (3-7) et (3-8), on déduit le résultat :

$$\min_{L \in L_{K,r}} W_R(P, L) = W_1(P) + \beta W_2(P) = W(P)$$

4 - L'ALGORITHME

4.1 - Introduction

L'algorithme utilise alternativement deux fonctions définies en (4.2) et (4.3) :

- g appelée fonction de représentation, associe à toute partition P dans $\mathbb{P}_K$, une représentation g(P) dans $L_{K,r}$ telle que :

$$W_R(P, g(P)) = \min_{L \in L_{K,r}} W_R(P, L)$$

- f appelée fonction d'affectation, associe à toute représentation L de $L_{K,r}$ une partition f(L) dans $\mathbb{P}_K$ telle que :

$$W_R(f(L), L) = \min_{P \in \mathbb{P}_K} W_R(P, L)$$

On montre, (4.4), que pour toute partition P dans $\mathbb{P}_K$, on a la relation :

$$W(f \circ g(P)) \leq W(P) \qquad (4\text{-}15)$$

L'algorithme est initialisé par une partition P^0.

Pour tout $n \geq 0$, on définit :

$$P^{n+1} = f \circ g(P^n)$$

D'après (4-15) la suite de terme général $W(P^n)$ est décroissante. On montre, (4.6), qu'elle est constante à partir d'un certain rang, et que la suite de partitions (P^n) est alors également constante. La partition P^* limite de la suite P^n est meilleure au sens du critère W que la partition initiale P^0 :

$$W(P^*) \leq W(P^0)$$

4.2 - La fonction de représentation

Il s'agit de définir une fonction g qui associe à toute partition P dans $\mathbb{P}_K$ une représentation $L = g(P)$ dans $L_{K,r}$ telle que :

$$W_R(P, g(P)) = \min_{L \in L_{K,r}} W_R(P, L) \qquad (4\text{-}1)$$

Soit

$P = (P_1, \ldots, P_K)$ dans $\mathbb{P}_K$

$L = (b_1, \ldots, b_K, v_1, \ldots, v_r, m_1, \ldots, m_K) = (\underline{B}, \underline{V}, \underline{M})$

F le sous espace engendré par les vecteurs $v_1, \ldots, v_r$

$g(P) = (a_1, \ldots, a_K, u_1, \ldots, u_r, h_1, \ldots, h_K) = (\underline{A}, \underline{U}, \underline{H})$

U le sous espace engendré par les vecteurs $u_1, \ldots, u_r$

La relation (4-1) s'écrit :

$$W_D(P, \underline{A}, \underline{U}) + \beta\ W_M(P, \underline{H}) = \min_{(\underline{B}, \underline{V}, \underline{M}) \in L_{K,r}} [W_D(P, \underline{B}, \underline{V}) + \beta W_M(P, \underline{M})]$$

On en déduit :

$$W_D(P, \underline{A}, \underline{U}) = \min_{(\underline{B}, \underline{V})} W_D(P, \underline{B}, \underline{V}) \qquad (4\text{-}2)$$

$$W_M(P, \underline{H}) = \min_{\underline{M}} W_M(P, \underline{M}) \qquad (4\text{-}3)$$

a) Optimisation du critère W_D

Par définition du critère W_D, (3-1), la relation (4-2) s'écrit :

$$I(\underline{A}, P, U) = \min_{(\underline{B}, F)} I(\underline{B}, P, F)$$

Il s'agit du problème d'optimisation étudié au paragraphe 7 de l'avant propos, dans le cas particulier où $Q = V^{-1}$ (§ 7.2) :

Définir une fonction g_D qui associe à toute partition $P = (P_1, \ldots, P_K)$ de E,

- K points $a_1, \ldots, a_K$ de $\mathbb{R}^p$
- une famille Q orthonormée de r vecteurs de $\mathbb{R}^p$.

$\underline{U} = (u_1, \ldots, u_r)$

tels que :

$a_1, \ldots, a_K$, et le sous espace vectoriel U engendré par $u_1, \ldots, u_r$ vérifient :

$$I((a_1, \ldots, a_K), P, U) = \min_{\substack{(b_1, \ldots, b_K) \in \mathbb{R}^p \times \ldots \times \mathbb{R}^p \\ F \in F_r}} [I((b_1, \ldots, b_K), P, F)]$$

D'après le paragraphe 7.4 de l'avant propos, pour toute partition $P = (P_1,\ldots,P_K)$ de E, on pose :

$$g_D(P) = (g_1,\ldots,g_K,\ u_1,\ldots,u_r)$$

où

- $g_1,\ldots,g_K$ sont les centres de gravité des classes $P_1,\ldots,P_K$
- $u_1,\ldots,u_r$ sont vecteurs propres Q normés associés aux r plus grandes valeurs propres $\lambda_1 \geq \ldots \geq \lambda_r \geq 0$ de la matrice $B(P)\ V^{-1}$ (où B(P) est la matrice d'inertie inter classe de la partition P. La définition précise des vecteurs $u_1,\ldots,u_r$ est donnée au paragraphe 7.4 de l'avant propos.

La fonction g_D est identique à la fonction g du Chapitre II.

b) Optimisation du critère W_M

Par définition du critère W_M, (3-2), la relation (4-3) s'écrit :

$$\sum_{\ell=1}^{K} \delta(P_\ell,\ h_\ell) = \min_{(m_1,\ldots,m_K)\in H\times\ldots\times H} \left[\sum_{\ell=1}^{K} \delta(P_\ell,\ m_\ell)\right]$$

On en déduit que pour tout ℓ, $\ell=1,\ldots,K$, on a :

$$\delta(P_\ell,\ h_\ell) = \min_{h\in H} [\delta(P_\ell,\ h)]$$

Or, on a supposé au paragraphe 1.2 la donnée d'une fonction g_M qui à toute partie A de A associe un modèle $g_M(A)$ dans H tel que :

$$g_M(A) = \min_{h\in H}\ \delta(A,\ h)$$

Ainsi, pour $\ell=1,\ldots,K$, on pose

$$h_\ell = g_M(P_\ell)$$

c) Optimisation du critère W_R

D'après a) et b) pour toute partition $P = (P_1,\ldots,P_K)$, on pose :

$$g(P) = (g_1,\ldots,g_K,\ u_1,\ldots,u_r,\ h_1,\ldots,h_K)$$

où $(g_1,\ldots,g_K,\ u_1,\ldots,u_r) = g_D(P)$

et $(h_1,\ldots,h_K) = (g_M(P_1),\ldots,g_M(P_K))$

D'après (O-15) et (O-16) :

$$W_D(P,\ g_D(P)) = I_{intra}\ (P,\ U) = r - \sum_{s=1}^{r} \lambda_s \qquad (4\text{-}4)$$

(où $\lambda_1,\ldots,\lambda_r$ sont les r plus grandes valeurs propres de la matrice $B(P)V^{-1}$).

Par définition de la fonction g, et d'après 3.2.3 :

$$W_R(P,\ g(P)) = W(P) \qquad (4\text{-}5)$$

Des relations (4-4) et (4-5) on déduit :

$$W(P) = I_{intra}\ (P,\ U) + \beta \sum_{\ell=1}^{K} \delta\,(P_\ell,\ g_M(P_\ell)) \qquad (4\text{-}6)$$

$$W(P) = r - \sum_{s=1}^{r} \lambda_s + \beta \sum_{\ell=1}^{K} \delta\,(P_\ell,\ g_M(P_\ell)) \qquad (4\text{-}7)$$

4.3 - La fonction d'affectation

4.3.1 - Définition

Il s'agit de définir une fonction f qui associe à toute représentation L dans $L_{K,r}$ une partition f(L) dans $\mathbb{P}_K$ telle que :

$$W_R(f(L),\ L) = \operatorname*{Min}_{P\in\mathbb{P}_K} W_R(P,\ L) \qquad (4\text{-}8)$$

Soit

$L = (a_1,\ldots,a_K,\ u_1,\ldots,u_r,\ h_1,\ldots,h_K) = (\underline{A},\ \underline{U},\ \underline{H})$ dans $L_{K,r}$

U le sous espace vectoriel engendré par les vecteurs $u_1,\ldots,u_r$

$P = (P_1,\ldots,P_K)$ dans $\mathbb{P}_K$

Par définition des critères W_R (3-3), W_D (3-1) et W_M (3-2), on a :

$$W_R(P,\ L) = I(\underline{A},\ P,\ U) + \beta \sum_{\ell=1}^{K} \delta\,(P_\ell,\ h_\ell) \tag{4-9}$$

D'après (0-9), on a :

$$I(\underline{A},\ P,\ U) = \frac{1}{n} \sum_{x\in E} d_U^2\,(x,\ a_x) \tag{4-10}$$

où $$d_U^2(x,\ y) = (x-y)'V^{-1}\ \Big(\sum_{s=1}^{r} u_s\ u_s'\Big)\ V^{-1}(x-y) \tag{4-11}$$

et a_x le centre de la famille $(a_1,\ldots,a_K)$ d'indice $i(x)$, où $i(x)$ est l'indice de la classe de la partition $P = (P_1,\ldots,P_K)$ qui contient x.

D'autre part, d'après (1-1), on a :

$$\delta\,(P_\ell,\ h_\ell) = \sum_{x\in P_\ell} \delta\,(x,\ h_\ell) \tag{4-12}$$

Si on note h_x le modèle de la famille $(h_1,\ldots,h_K)$ d'indice $i(x)$, on a d'après (4-12) :

$$\sum_{\ell=1}^{K} \delta\,(P_\ell,\ h_\ell) = \sum_{x\in E} \delta\,(x,\ h_x) \tag{4-13}$$

D'après les relations (4-9), (4-10) et (4-13) on a :

$$W_R(P,\ L) = \sum_{x\in E} \Big(\frac{1}{n}\, d_U^2\,(x,\ a_x) + \beta\ \delta\,(x,\ h_x)\Big) \tag{4-14}$$

On en déduit :

$$\min_{P \in \mathcal{P}_K} W_R(P, L) = \sum_{x \in E} \left[\min_{1 \le i \le K} \left(\frac{1}{n} d_U^2 (x, a_i) + \beta \, \delta (x, h_i) \right) \right]$$

On pose par conséquent :

$f(L) = (P_1, \ldots, P_K)$ où, pour $\ell = 1, \ldots, K$

$$P_\ell = \left\{ x \middle| \frac{1}{n} d_U^2 (x, a_\ell) + \beta \, \delta (x, h_\ell) \le \frac{1}{n} d_U^2 (x, a_i) + \beta \, \delta (x, h_i) \; ; \; i = 1, \ldots, K \right\}$$

Cette définition est analogue à celle du Chapitre II (II.4.3) ; $\frac{1}{n} d_U^2 (x, a)$ est remplacé par $\frac{1}{n} d_U^2 (x, a) + \beta \, \delta (x, h)$.

Pour compléter la règle d'affectation d'un individu x dans l'éventualité où plusieurs couples (a_i, h_i) réalisent le minimum de $\frac{1}{n} d_U^2 (x, a_i) + \beta \, \delta (x, h_i)$, on se place dans le cadre de l'algorithme.

Une partition Q est donnée dans $\mathcal{P}_K$. On recherche la partition $P = f(L)$ où $L = g(Q)$.

Soit un individu x qui appartient à Q_j et tel que plusieurs couples (a_i, h_i) réalisent

$$\min_{1 \le \ell \le K} \left[\frac{1}{n} d_U^2 (x, a_\ell) + \beta \, \delta (x, h_\ell) \right]$$

On distingue deux éventualités :

Eventualité 1 :

$$\text{Si} \quad \frac{1}{n} d_U^2 (x, a_j) + \beta \, \delta (x, h_j) = \min_{1 \le \ell \le K} \left[\frac{1}{n} d_U^2 (x, a_\ell) + \beta \, \delta (x, h_\ell) \right]$$

x est affecté à P_j.

Eventualité 2 :

Sinon x est affecté à P_m, où m est le plus petit indice i tel que :

$$\frac{1}{n} d_U^2 (x, a_i) + \beta \delta (x, h_i) = \underset{1 \leq \ell \leq K}{\text{Min}} \left[\frac{1}{n} d_U^2 (x, a_\ell) + \beta \delta (x, h_\ell)\right]$$

Dans ce cas d'indétermination, la partition P dépend non seulement de la représentation L, mais aussi de la partition Q.

f apparaît ainsi comme une fonction de deux variables :

- une représentation
- une partition utilisée simplement comme contrainte en cas d'indétermination pour l'affectation. (La classe d'appartenance dans la partition Q est prioritaire).

Le rôle de cette seconde variable est secondaire. De plus, dans l'algorithme, (4.5), la partition Q est définie par la représentation L. Aussi, pour simplifier les notations, on écrit f comme une fonction de la seule variable représentation, et on retient la proposition 4.3.2. (On vérifie dans la présentation de l'algorithme, et la démonstration de sa convergence que l'emploi de cette notation n'entraîne aucune inexactitude).

4.3.2 - Proposition

Si les partitions Q et P = f o g (Q) vérifient

$W_R(Q, g(Q)) = W_R(P, g(Q))$ elles sont identiques

Démonstration

Soit :

$Q = (Q_1, \ldots, Q_K)$

$L = g(Q) = (a_1,\ldots,a_K,\ u_1,\ldots,u_r,\ h_1,\ldots,h_K) = (\underline{A},\ \underline{U},\ \underline{H})$

U le sous espace engendré par $u_1,\ldots,u_r$

$P = f \circ g\ (Q) = (P_1,\ldots,P_K)$

Soit un objet x dans E. On suppose que x appartient à la classe Q_j $(1 \leq j \leq K)$ de la partition Q, et à la classe P_i $(1 \leq i \leq K)$ de la partition P.

On a par définition de P = f(L)

$$\frac{1}{n}\ d_U^2\ (x,a_i) + \beta\ \delta\ (x,h_i) = \min_{\ell=1,\ldots,K} \left(\frac{1}{n}\ d_U^2\ (x,a_\ell) + \beta\ \delta\ (x,h_\ell)\right)$$

$$\leq \frac{1}{n}\ d_U^2\ (x,a_j) + \beta\ \delta\ (x,h_j)$$

et d'après (4-14)

$$\frac{1}{n}\ d_U^2\ (x,a_i) + \beta\ \delta\ (x,h_i) < \frac{1}{n}\ d_U^2\ (x,a_j) + \beta\ \delta\ (x,h_j) \Rightarrow W_R\ (P,g(Q)) < W_R(Q,g(Q))$$

Comme la dernière inégalité est contraire à l'hypothèse on en déduit :

$$\frac{1}{n}\ d_U^2\ (x,a_i) + \beta\ \delta\ (x,h_i) = \frac{1}{n}\ d_U^2\ (x,a_j) + \beta\ \delta\ (x,h_j)$$

L'affectation de l'objet x à l'une des classes de la partition P s'est donc effectuée selon la règle définie dans l'éventualité 1 du cas d'indétermination.

On en déduit i = j.

Cette démonstration étant valable pour tout x dans E, on en déduit que les partitions P et Q sont identiques.

4.4 - La fonction composée f o g

Proposition

Pour toute partition Q dans $\mathbb{P}_K$.

La partition f o g (Q) = f[g(Q)] vérifie la relation :

$$W(f \circ g\ (Q)) \leq W(Q) \tag{4-15}$$

De plus

$$(W(f \circ g\ (Q)) = W(Q)) <=> (f \circ g\ (Q) = Q) \tag{4-16}$$

Démonstration

D'après (4-5) et la définition de la partition f[g(Q)], on a :

$$W(Q) = W_R(Q,\ g(Q)) \geq W_R(f[g(Q)],\ g(Q)) \tag{4-17}$$

D'après la définition de la représentation g[f o g (Q)], et (4-5) appliqué à la partition f o g(Q), on a :

$$W_R(f \circ g(Q),\ g(Q)) \geq W_R(f \circ g(Q),\ g[f \circ g\ (Q)]) = W(f \circ g(Q)) \tag{4-18}$$

Des relations (4-17) et (4-18) on déduit :

$$W(Q) \geq W(f \circ g\ (Q))$$

De plus, si W(Q) = W(f o g(Q)), on a l'égalité dans la relation (4-17) :

$W_R(Q, g(Q)) = W_R(f \circ g(Q), g(Q))$

D'après la proposition 4.3.2 les partitions Q et f o g (Q) sont égales.

4.5 - L'algorithme

L'algorithme est initialisé par une partition P^0 donnée ou tirée au hasard.

Pour tout $n \geq 0$, on pose :

$L^n = g(P^n)$

$P^{n+1} = f(L^n) = f \circ g\ (P^n)$

On définit ainsi une suite (P^n) de partitions dans $\mathbb{P}_K$, et une suite (L^n) de représentations dans $L_{K,r}$.

Soit $P^n = (P_1^n, \ldots, P_K^n)$

et $g_1^n, \ldots, g_K^n$ les centres de gravité des classes $P_1^n, \ldots, P_K^n$.

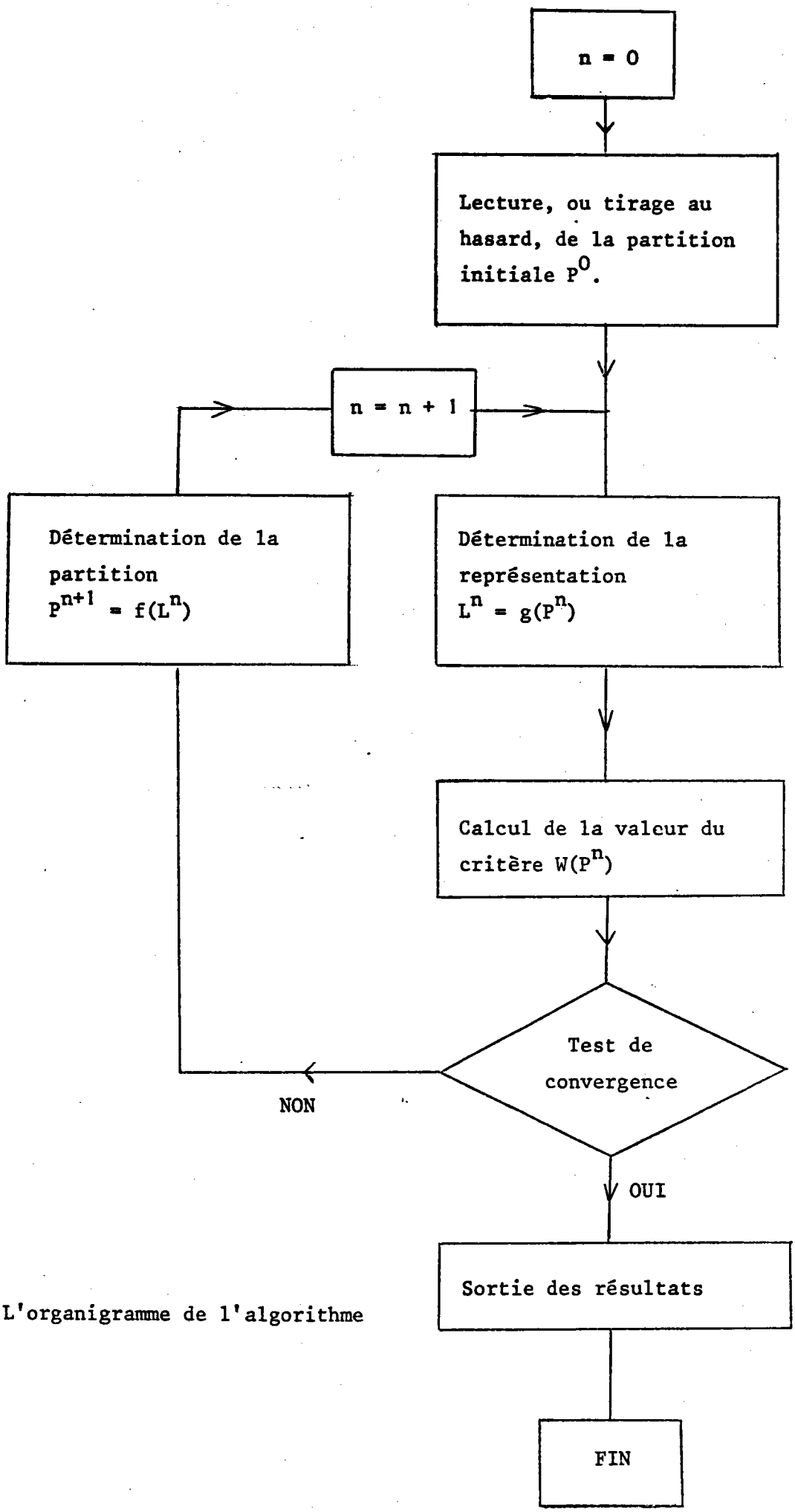
On sait, (4.2), que $L^n = g(P^n)$ est de la forme :

$L^n = (g_1^n, \ldots, g_K^n, \underline{U}^n, \underline{H}^n)$

où $\underline{U}^n$ est une famille V^{-1} orthonormée de r vecteurs de R^p, $\underline{U}^n = (u_1^n, \ldots, u_r^n)$, (On note U^n le sous espace vectoriel de R^p engendré par U^n) ; $\underline{H}^n$ est la famille de modèles $(h_1^n, \ldots, h_K^n) = (g_M(P_1^n), \ldots, g_M(P_K^n))$.

D'après (4-5), on a :

$W_R(P^n, L^n) = W_R(P^n, g(P^n)) = W(P^n)$



L'organigramme de l'algorithme

D'après (4-6), on a :

$$W(P^n) = I_{intra}(P^n, U^n) + \beta \sum_{\ell=1}^{K} \delta(P_\ell^n, g_M(P_\ell^n))$$

4.6 - La convergence de l'algorithme

Théorème

La suite de terme général $r_n = W(P^n)$ est décroissante, et constante à partir d'un certain rang N.

De plus, la suite (P^n, L^n) est constante à partir du rang N.

Démonstration

a) Pour tout $n \geq 0$, on a d'après (4.4) :

$$r_{n+1} = W(P^{n+1}) = W(f \circ g\ (P^n)) \leq W(P^n) = r_n$$

La suite (r_n) est donc décroissante.

b) La suite (r_n) prend ses valeurs dans l'ensemble fini $\{W(P) \mid P \in \mathbb{P}_K\}$

En effet comme E est un ensemble fini, l'ensemble $\mathbb{P}_K$ des partitions de E en K classes est fini.

c) D'après a) et b) il existe un rang N, tel que ;

$$r_n = r_N \text{ pour tout } n \geq N$$

d) Pour tout n, $n \geq N$, on a d'après c)

$$r_{n+1} = r_n$$

Soit d'après a)

$$W(f \circ g\ (P^n)) = W(P^n)$$

On en déduit d'après (4.4)

$f \circ g\ (P^n) = P^n$

e) On déduit de d) l'égalité

$P^{n+1} = P^n$ pour tout n, $n \geq N$

d'où

$P^n = P^N$ pour tout n, $n \geq N$

f) De l'égalité $P^{n+1} = P^n$ $(n \geq N)$ démontrée en e), on déduit que :

$L^{n+1} = g(P^{n+1}) = g(P^n) = L^n$ pour tout n, $n \geq N$

d'où

$L^n = L^N$ pour tout $n \geq N$

Remarque : Dans l'algorithme la fonction f est utilisée exclusivement dans le contexte suivant :

Détermination dans la partition $P^{n+1} = f(L^n)$ où $L^n = g(P^n)$, $n \geq 0$

La partition Q de contrainte qui apparaît dans la définition de la fonction f, (4.3.1) est donc parfaitement définie, et f peut être notée comme une fonction d'une variable. La prise en compte de la partition P^n pour définir la partition P^{n+1}, apparaît dans la démonstration de la convergence de l'algorithme par l'utilisation en d) de la relation (4-16) déduite de la proposition 4.3.2.

5 - LE PROGRAMME INFORMATIQUE DISCRIML

5.1 - Les sous programmes relatifs à la modélisation

Le programme DISCRIML appelle deux sous programmes propres au type de modélisation retenu :

- Le sous programme DELTA : Il calcule pour tout modèle h dans la famille H de modèles et toute partie A de E, l'indice d'homogénéité de A relativement au modèle h, $\delta(A, h)$.

- Le sous programme MODELE : il détermine pour toute partie A de E, un modèle $h^*(A)$ de H qui vérifie

$$\delta(A, h^*(A)) = \underset{h \in H}{\text{Min}}\ \delta(A, h)$$

5.2 - Les entrées

Outre le tableau des données (1.2), les entrées du programme DISCRIML sont les suivantes :

- la pondération β
- le nombre maximum de classes désirées : K
- la dimension du sous espace discriminant : r $\quad (r \leq K-1)$
- la partition initiale P^0 lue, tirée au hasard ou déterminée selon la procédure proposée ci-dessous.

5.3 - Construction d'une partition

On dispose :

- du sous programme MODELE qui à toute partie A de E associe un modèle $h^*(A)$ qui minimise l'indice $\delta(A, h)$ pour h dans H,

- de la fonction $\tilde{f}$ qui associe à toute famille $\underline{H} = (h_1, \ldots, h_K)$ de modèles une partition de E en K classes $P = (P_1, \ldots, p_K) = \tilde{f}(\underline{H})$ définie par :

$$P_\ell = \{x \in E / \delta(x, h_\ell) \leq \delta(x, h_i) \quad i=1,\ldots,K\} \quad \ell=1,\ldots,K$$

(En cas d'égalité x est affecté à la classe de plus petit indice).

La fonction $\tilde{f}$ est déduite de la fonction d'affectation f de l'algorithme. (on pose $\beta=0$)

Soit :

- un entier K', $\quad 0 \leq K' \leq K$

- K' modèles dans H, $\quad h_1, \ldots, h_{K'}$

- Et, si $K' \neq K$, $(K - K')$ sous populations disjointes $E_{K'+1},\ldots,E_K$ de E (mais ne constitutant pas une partition de E).

Si $K' \neq K$, on pose pour $\ell = K'+1,\ldots,K$

$$h_\ell = h^*(E_\ell)$$

On connait ainsi, quelque soit K', une famille $\underline{H} = (h_1,\ldots,h_K)$ de K modèles de H.

On en déduit une partition P^0 dans $\mathbb{P}_K$ en posant :

$$P^0 = \tilde{f}(\underline{H})$$

Cette procédure permet d'initialiser l'algorithme par une famille de K' modèles et de K - K' sous populations homogènes, choisis par le spécialiste pour leur qualité de référence ou d'après les résultats d'analyses antérieures (cf. 6.2).

5.4 - Le critère

Dans le programme DISCRIML on utilise le critère W_R sous la forme suivante :

Soit $P = (P_1,\ldots,P_K)$

$$L = (a_1,\ldots,a_K,\ u_1,\ldots,u_r,\ h_1,\ldots,h_K)$$

$$b_s = V^{-1} u_s \qquad s = 1,\ldots,r$$

$$W_R(P, L) = \frac{1}{n} \sum_{\ell=1}^{K} \sum_{x\in P_\ell} \sum_{s=1}^{r} [(x-a_\ell)'b_s]^2 + \beta \sum_{\ell=1}^{K} \sum_{x\in P_\ell} \delta(x, h_\ell)$$

5.5 - La fonction de représentation [16]

Pour la détermination des vecteurs $u_1,\ldots,u_r$ de la représentation $L = g(P)$ d'une partition P, on se ramène (cf. II.5.3) à la diagonalisation d'une matrice symétrique de dimension K.

5.6 - Les sorties à chaque itération

Après chaque itération on dispose des sorties suivantes :

- les valeurs des critères $W(P^n)$, $W_1(P^n)$ et $W_2(P^n)$
- les valeurs propres $\lambda_1^n,\ldots,\lambda_r^n$
- la famille des centres de gravité des classes $g_1^n,\ldots,g_K^n$
- la famille des modèles $h_1^n,\ldots,h_K^n$
- la classe d'appartenance dans la partition P^n de chaque individu de E.

5.7 - Les sorties à la convergence de l'algorithme

- Les valeurs des critères $W(P^*)$, $W_1(P^*)$, $W_2(P^*)$
- les valeurs propres $\lambda_1^*,\ldots,\lambda_r^*$
- la famille des centres de gravité des classes $g_1^*,\ldots,g_K^*$
- la famille des modèles $h_1^*,\ldots,h_K^*$
- la classe d'appartenance de chaque individu dans la partition P^*, et pour chaque classe P_ℓ^* $(1 \leq \ell \leq K)$ la liste des individus qu'elle contient
- les représentations graphiques des projections sur les r premiers axes discriminants de la partition P^* (qui sont engendrés par les vecteurs $u_1^*,\ldots,u_r^*$ de la représentation $L^* = g(P^*)$) des centres de gravité des classes et des individus. On utilise le programme de représentation "NUAGE" (ISUP).

6 - L'INTERPRETATION DES RESULTATS

6.1 - Le choix de la meilleure analyse après plusieurs tirages

Dans le cas où l'analyse est initialisée par une partition tirée au hasard, ou si la connaissance à priori des données suggère plusieurs configurations initiales, on effectue successivement t analyses ou tirages. Chaque tirage induit à la convergence de l'algorithme une partition P^{*i} qui réalise un optimum local du critère W. Les partitions P^{*i}, i=1,...,t, peuvent être ordonnées par valeur de critère $W(P^{*i})$ croissante.

Si $W(P^{*i_1}) \leq W(P^{*i_2}) \leq \ldots \leq W(P^{*i_t})$

P^{*i_1} est la meilleure partition au sens du critère W, P^{*i_2} la seconde...

La validation des partitions P^{*i}, i=1,...,t, (cf. 6.2), est entreprise dans l'ordre $i_1, i_2, \ldots, i_t$. On pourra retenir finalement la (ou les) première(s) partition(s) ainsi validée(s).

6.2 - La validation de la partition de convergence

La partition $P^* = (P_1^*, \ldots, P_K^*)$ et sa représentation $L^* = (g_1^*, \ldots, g_K^*, u_1^*, \ldots, u_r^*, h_1^*, \ldots, h_K^*)$ doivent pour être validées vérifier les 3 conditions suivantes énoncées en 6.1.1, 6.1.2 et 6.1.3.

6.2.1 - La condition 1

Condition 1 : Les centres $g_1^*, \ldots, g_K^*$ et les facteurs discriminants $b_1^* = V^{-1} u_1^*, \ldots, b_r^* = V^{-1} u_r^*$ permettent la reconnaissance des classes $P_1^*, \ldots, P_K^*$, par la règle de décision $\mathcal{R}$ suivante :

Soit $D_e = \sum_{s=1}^{r} b_s^* b_s^{*'}$

L'individu x est affecté à la classe P^* si :

$$(x-g_\ell^*)' D_e (x-g_\ell) \leq (x-g_i^*) D_e (x-g_i) \qquad i=1,\ldots,K$$

Cette règle se déduit de la fonction d'affectation f de l'algorithme en remplaçant β par 0. La fonction f tient compte de la discrimination par les paramètres explicatifs de J_1 et de la modélisation sur J_2, la règle de décision $\mathcal{R}$ de la discrimination seulement.

On détermine le pourcentage d'individus de E affectés par la règle $\mathcal{R}$ à leur classe d'appartenance dans P^*. (Ces individus sont dits bien classés par $\mathcal{R}$).

Si la règle de décision $\mathcal{R}$ ne permet pas une bonne reconnaissance de la partition P^*, il est nécessaire de réaliser une analyse complémentaire (cf. 6.3).

6.2.2 - La condition 2

Condition 2 : La règle $\mathcal{R}$ est validée sur un échantillon test E_t, dans le sens suivant :

Soit E_t un ensemble d'individus n'appartenant pas à E pour lesquels on connaît les valeurs prises par les paramètres de J_1 et J_2.

Soit $\ell_1(x)$ et $\ell_2(x)$ les indices qui minimisent respectivement les quantités

$$(x-g_\ell^*)' \, D_e \, (x-g_\ell^*)$$

et

$$\frac{1}{n} (x-g_\ell^*)' \, D_e \, (x-g_\ell^*) + \beta \, \delta(x, h_\ell^*)$$

et qui correspondent respectivement à l'affectation par la règle $\mathcal{R}$ et par la fonction f.

$$\text{Soit } \rho = \frac{\text{Card } \{x \in E_t \mid \ell_1(x) = \ell_2(x)\}}{\text{Card } E_t}$$

ρ s'interprète comme un pourcentage de bien classés sur la population test E_t. (En effet, comme $f(L^*) = P^*$ l'affectation par la fonction f d'un nouvel individu est référence).

Si le pourcentage de bien classés ρ n'est pas satisfaisant, il convient comme en 6.1.1 de réaliser une analyse complémentaite (cf. 6.3).

6.2.3 - La condition 3

Condition 3 : Pour $\ell=1,\ldots,K$, la classe P_ℓ^* est homogène relativement au modèle h_ℓ^*.

Le spécialiste juge, selon la qualité requise de la modélisation, acceptable ou non, l'attribution du modèle h_ℓ^* à chaque élément de la classe P_ℓ^*. ($\delta(P_\ell^*, h_\ell^*)$ est un indice d'homogénéïté moyenne).

Si au moins une classe ne vérifie pas la condition 3, la partition P^* doit être remise en cause dans une analyse complémentaire (cf. 6.3).

6.3 - Une procédure conversationnelle pour la recherche du compromis "Discrimination - Modélisation"

La partition P^* est construite par optimisation du critère $W = W_1 + \beta W_2$ où W_1 (resp. W_2) est une mesure de la qualité globale ou moyenne de la discrimination de ses K classes par les r premiers facteurs discriminants dans J_1 (resp. de la modélisation locale).

Les conditions de validation de la partition P^* (6.2) sont locales ; elles sont relatives à chaque classe de la partition P^*. Elles peuvent être vérifiées pour certaines classes sans l'être pour toutes. Dans ce cas la partition P^* n'est pas validée, mais pour l'initialisation d'une analyse ultérieure il convient de privilégier toute classe P^*_ℓ vérifiant les conditions d'homogénéité vis-à-vis du modèle h^*_ℓ et de séparation par la règle de décision $\mathfrak{R}$. Il a été introduit dans ce sens en 5.3 une procédure générale de définition d'une partition initiale.

En général, il sera nécessaire de réaliser une suite de quelques analyses pour construire une partition validée au sens de 6.2. L'initialisation de chaque nouvelle analyse peut s'accompagner de l'une ou de plusieurs des modifications suivantes :

- correction de la pondération β : son accroissement peut remédier à une médiocre modélisation, et sa diminution à une médiocre discrimination,

- changement du nombre de classes maximum désiré : en effet, à la définition de la partition initiale, il peut apparaître judicieux d'effectuer la réunion de classes de P^* ou au contraire de scinder certaines en sous classes, afin de remédier par exemple à une médiocre discrimination dans le premier cas, à une médiocre modélisation dans le second.

- Ajustement de la dimension r du sous espace discriminant : son accroissement peut améliorer la puissance de la règle de décision. Par contre, si les facteurs discriminants d'ordre r, r-1,... ne sont pas significatifs, leur suppression est souhaitable.

Cette démarche conversationnelle permet l'intervention du spécialiste. Selon ses objectifs il peut privilégier la discrimination ou la modélisation locale. Sa connaissance du phénomène étudié est un complément précieux au calcul de critères mathématiques pour la validation d'une partition, ou plus encore l'initialisation d'une analyse nouvelle.

6.4 - La règle de décision

Si la partition de convergence de l'analyse, P^*, est validée (dans le sens où elle vérifie les conditions 1, 2 et 3 de 6.2), on déduit de la représentation $L^* = g(P^*)$ la règle de décision suivante, notée $\mathcal{R}_M(L^*)$

Soit $L^* = (g_1^*,\ldots,g_K^*, u_1^*,\ldots,u_r^*, h_1^*,\ldots,h_K^*)$

$$b_s^* = V^{-1} u_s^*$$

$$D_e = \sum_{s=1}^{r} b_s^* b_s^{*'}$$

Si l'individu x vérifie la relation

$$(x-g_\ell^*)' D_e (x-g_\ell^*) \leq (x-g_i^*)' D_e (x-g_i^*) \qquad i=1,\ldots,K$$

on lui affecte le modèle h_ℓ^*.

6.5 - Le renforcement de la représentation

La représentation L^* et la règle de décision $\mathcal{R}_M(L^*)$ dont elle est issue sont construits par étude de la population échantillon E.

Dans la phase opérationnelle, on peut envisager une procédure séquentielle de validation et de renforcement de la représentation ; En effet toute nouvelle réalisation du processus peut être adjointe à la population échantillon.

Soit

L^* la représentation issue de la population échantillon E

E_+ une population complémentaire

$\underline{E} = E \cup E_+$

$\underline{P^\circ}$ la partition de $\underline{E}$ image par la fonction d'affectation f de la représentation L^*

$\underline{L^*}$ la représentation obtenue à la convergence de la méthode appliquée à $\underline{E}$ et initialisée par $\underline{P^\circ}$

On substitue la représentation $\underline{L^*}$ à L^* comme image de la structure de l'ensemble $\mathcal{E}$ des réalisations potentielles du processus.

7 - APPLICATIONS

7.1 - Un exemple

7.1.1 - Les données

Soit un ensemble E de 24 réalisations (Tableau 1) d'un processus décrits par :

- un ensemble J_1 de 5 paramètres d'entrée, $X_1,\ldots,X_5$
- un ensemble J_2 de 2 paramètres d'actions, Z_1, Z_2
- une variable de sortie notée Z_3.

	X_1	X_2	X_3	X_4	X_5	Z_1	Z_2	Z_3
A 1	1.0	2.0	1.0	16.0	26.0	1.0	1.0	2.0
A 2	2.0	.0	2.0	20.0	31.0	1.0	9.0	2.0
A 3	.0	1.0	.0	25.0	12.0	9.0	9.0	2.0
A 4	2.0	1.0	1.0	32.0	23.0	5.0	5.0	2.0
A 5	1.0	2.0	1.0	17.0	7.0	9.0	1.0	2.0
A 6	1.0	.0	.0	12.0	13.0	10.0	6.0	2.1
A 7	.0	2.0	2.0	33.0	10.0	12.0	11.0	1.8
A 8	2.0	.0	1.0	25.0	25.0	13.0	13.0	2.1
B 1	12.0	21.0	10.0	19.0	12.0	1.0	1.0	4.0
B 2	13.0	20.0	12.0	4.0	16.0	1.0	9.0	4.0
B 3	11.0	22.0	11.0	13.0	29.0	9.0	9.0	4.0
B 4	12.0	21.0	12.0	32.0	31.0	5.0	5.0	4.0
B 5	10.0	21.0	10.0	24.0	15.0	9.0	1.0	4.0
B 6	12.0	20.0	10.0	29.0	7.0	2.0	12.0	4.2
B 7	13.0	22.0	11.0	12.0	36.0	4.0	12.0	3.8
B 8	11.0	21.0	12.0	21.0	24.0	6.0	5.0	4.3
C 1	22.0	12.0	22.0	26.0	29.0	1.0	1.0	6.0
C 2	21.0	10.0	21.0	37.0	18.0	1.0	9.0	6.0
C 3	23.0	11.0	22.0	5.0	22.0	9.0	9.0	6.0
C 4	20.0	11.0	20.0	3.0	32.0	5.0	5.0	6.0
C 5	22.0	12.0	21.0	19.0	12.0	9.0	1.0	6.0
C 6	21.0	10.0	21.0	36.0	3.0	12.0	3.0	6.3
C 7	21.0	11.0	20.0	32.0	15.0	11.0	4.0	6.1
C 8	22.0	12.0	20.0	27.0	16.0	11.0	4.0	5.9

Tableau 1 : Les données

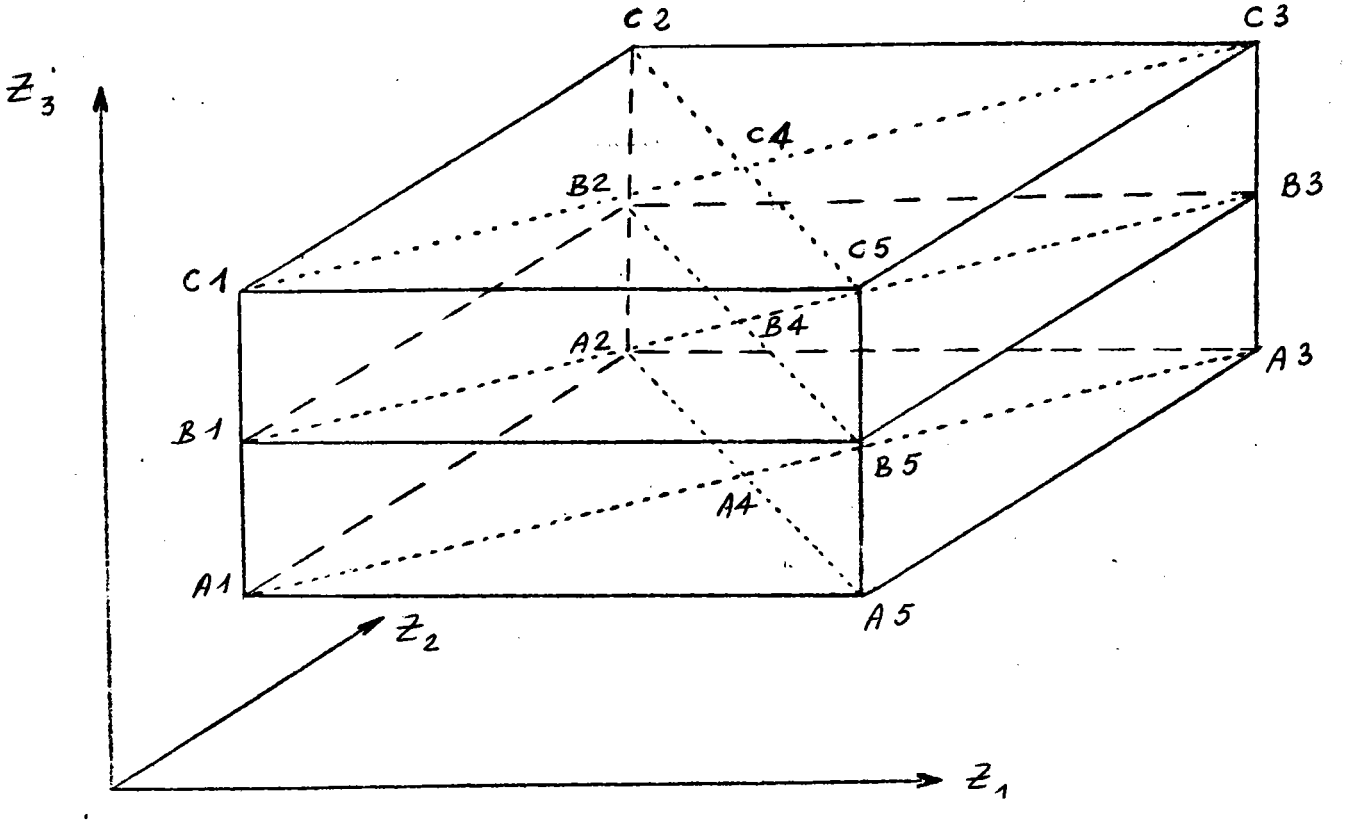


Figure 1 : Les données

Pour une configuration d'entrée $X_1,\ldots,X_5$ donnée, les paramètres Z_1, Z_2 et Z_3 sont liés par un modèle linéaire.

$(\alpha_1 Z_1 + \alpha_2 Z_2 + \alpha_3 Z_3 + \alpha_0 = 0)$

On recherche :

- une partition de E en 3 classes, $P = (P_1, P_2, P_3)$,
- deux facteurs dans R^5, b_1, b_2,
- trois modèles h_1, h_2, h_3 dans la famille H des modèles linéaires,

tel que :

1) les deux facteurs b_1 et b_2 soient discriminants de la partition P.

2) la classe P_ℓ soit homogène vis-à-vis du modèle h_ℓ, ℓ=1, 2, 3.

7.1.2 - Les analyses

L'analyse discriminante typologique sous contrainte de modélisation locale (K=3, r=2, β=1) initialisée par une partition tirée au hasard. On effectue plusieurs tirages et on retient (6.1) celui dont la partition de convergence réalise le minimum du critère.

7.1.3 - Les résultats

La partition P^* obtenue est :

Classe 1 : B1, B2, B3, B4, B5, B6, B7, B8

Classe 2 : A1, A2, A3, A4, A5, A6, A7, A8

Classe 3 : C1, C2, C3, C4, C5, C6, C7, C8

On a :

$W_1(P^*) = 2.06 \ 10^{-2}$

$W_2(P^*) = 1.25 \ 10^{-2}$

Les facteurs discriminants b_1^* et b_2^* sont à une constante multiplicative près :

$b_1^*(X_1,\ldots,X_5) = X_1 + 1.31\ X_2 + 0.96\ X_3 + 0.02\ X_4 + 0.07\ X_5$

$b_2^*(X_1,\ldots,X_5) = X_1 - 5.11\ X_2 + 3.49\ X_3 - 0.05\ X_4 - 0.06\ X_5$

(Le poids des paramètres X_4 et X_5 est négligeable).

Les modèles h_1^*, h_2^*, h_3^* s'écrivent :

h_1^* $\qquad Z_3 = 4.0$

h_2^* $\qquad Z_3 = 2.0$

h_3^* $\qquad Z_3 = 6.0$

On remarque (figure 1) que 15 des 24 points de R^3 (Z_1, Z_2, Z_3) sont situés sur les faces d'un parallélipipède. Plusieurs partitions et triplets de modèles apparaissant ainsi sur J_2. La prise en compte simultanée de la discrimination par J_1 lève l'indétermination.

7.2 - Application à la reconnaissance des lois d'effort de coupe*

7.2.1 - Le problème

La réalisation d'une pièce mécanique par tournage longitudinal est une opération industrielle pour laquelle on distingue :

- les paramètres d'entrée (explicatifs) qui permettent la description

---

* Cette application a été réalisée en collaboration avec B. MUTEL.

- de la pièce à usiner,
- de la machine outil utilisée,
- du montage de la pièce sur la machine outil.

- les paramètres de commande

  - vitesse de coupe,
  - vitesse d'avance,
  - profondeur de passe.

- les paramètres de sorties

  - force(s) de coupe.

Le spécialiste fait habituellement l'hypothèse suivante : Pour une configuration d'entrée donnée (description de l'usinage à réaliser, des caractéristiques de la machine outil disponible, et du montage effectué), les paramètres de commande Z_1, Z_2, Z_3 et la variable de sortie Z_4 sont liés par un modèle linéaire :

$$\alpha_1 Z_1 + \alpha_2 Z_2 + \alpha_3 Z_3 + \alpha_4 Z_4 + \alpha_0 = 0$$

(La connaissance de ce modèle permet d'optimiser le choix des paramètres de commande).

Le spécialiste dispose d'un catalogue de modèles définis selon un tri déterministe des usinages à réaliser, auquel il souhaiterait substituer une classification des usinages $P = (P_1,\ldots,P_K)$ telle que :

- quelques facteurs discriminants (issus de l'ensemble des variables d'entrée) permettent la reconnaissance des classes $P_1,\ldots,P_K$,

- chaque classe P_ℓ $(1 \leq \ell \leq K)$ soit homogène relativement à un modèle linéaire h_ℓ.

7.2.2 - Les données

Il a été effectué dans notre laboratoire, LPMC (Metz) (par M. BURNER R.) une série d'environ 500 mesures d'effort de coupe en tournage

longitudinal pour différents types de matériau de la pièce et géométrie de l'outil.

7.2.3 - Modélisation locale en fonction de la géométrie de l'outil et de la pièce

Pour cette application toutes les données sont relatives à l'usinage de pièces de même matériau (35 CD 4).

On retient les paramètres d'entrée suivant (cf. Tableau 2 bis) :

γ, Z, $\emptyset$, V_b

La modélisation concerne les variables suivantes (cf. Tableau 2 bis)

$Z_1 = \text{Log } V$, $Z_2 = \text{Log } A$, $Z_3 = \text{Log } S$, $Z_4 = \text{Log } F_Z$

On écrit un modèle sous la forme :

$$Z_4 = \alpha_0 + \alpha_1 Z_1 + \alpha_2 Z_2 + \alpha_3 Z_3$$

Une procédure conversationnelle (6.3) conduit à choisir : K=6, r=2, β=100.

On réalise plusieurs analyses initialisées par des partitions différentes. Les résultats sont sensiblement identiques, ce qui prouve la stabilité de la solution.

On obtient ainsi la partition $P^* = (P_1^*,\ldots,P_6^*)$ et la représentation $L^* = (g_1^*,\ldots,g_6^*, u_1^*, u_2^*, h_1^*,\ldots,h_6^*)$ présentée dans le tableau 3.

ESSAI

Matériau Pièce NF :
Obtention brut
Traitement

Outil Fabricant
Nuance
Forme

	COMPOSITION - 10^{-4}								caractéristiques						Porte-outil mm			Fixation Plaquette		OUTIL		
	C	Mn	Si	Ni	Cr	Mo	W	Cu	R	H_B	Re	σ	R_S		ℓ	B	H		r^N	K°	α°	γ°
1	6	9	12	15	18	21	24	27	30	36	39	42	45	48	51	54	57	60	62	65	68	71	74

IDENTIFICATION

R : résistance à la rupture daN/mm²
H_o : Dureté Brinell
R_e : Limite d'élasticité daN/mm²
σ : Limite de fatigue daN/mm²
R_S : Allongement
ℓ : longueur du porte à faux mm
B : base porte-outil mm
H : Hauteur porte-outil mm

r : rayon pointe de l'outil
K : angle de direction principale
α : angle de dépouille
γ : angle d'attaque
T_y : type d'usinage
Pri : Prise de pièce
LT : longueur totale de la pièce mm
Z : position outil/extrêmité pièce mm

Ø : diamètre pièce mm
L : lubrification
s : stabilité machine
K_W : puissance à la broche kw
c : for du copeau
V_b : usure en dépouille

1	6	8	10	13	16	19	24	30	32	35	41	44	47	50	58	62	66	70					
	Ty	Pri	LT	Z	Ø		L	s	kw		V m/mn	$\frac{1}{10}^{A}$ mm	$\frac{1}{100}^{S}$ mm		F_x	F_y daN	F_z	V_b $\frac{1}{100}$ mm					

lubrification opération | type machine | Conditions de coupe | Résultats

	γ	Z	ϕ	V_b
g_1^*	- 4.9	134	72.9	2.0
g_2^*	1.8	150	74.9	2.3
g_3^*	3.1	105	48.9	9.1
g_4^*	11.1	122	74.7	2.0
g_5^*	10.4	127	60.8	6.6
g_6^*	- 5.6	129	60.6	6.6

$$u_1^* = - 14.4\ \gamma + 0.1\ Z + 2.2\ \phi + 6.5\ V_b$$

$$u_2^* = - 0.3\ \gamma + 0.1\ Z - 5.1\ \phi + 18\ V_b$$

	α_0	α_1	α_2	α_3
h_1^*	5.7	- 0.08	1.41	0.69
h_2^*	10.7	- 1.23	0.92	0.79
h_3^*	22.3	- 0.03	0.59	0.68
h_4^*	5.9	0.05	0.83	0.77
h_5^*	7.7	- 0.05	0.95	0.74
h_6^*	18.6	- 0.18	1.40	0.34

Tableau 3 : La représentation $L^* = g(P^*)$

On a :

$W_1(P^*) = 0.40$

$\beta\ W_2(P^*) = 0.28$

$\lambda_1^* = 0.83$

$\lambda_2^* = 0.78$

Le plan engendré par u_1^* et u_2^* explique 96 % de l'inertie du nuage dans l'espace des paramètres d'entrée J_1.

L'affectation par la règle de décision $\mathfrak{D}$(6.2.1) utilisant les centres $g_1^*,\ldots,g_6^*$ et les facteurs discriminants $b_1^* = V^{-1}\ u_1^*$ et $b_2^* = V^{-1}\ u_2^*$ induit sur E 86 % de bien classés. (Elle est de plus validée sur un échantillon test). Plus précisement le pourcentage de bien classés est selon la classe d'appartenance P^*, ℓ=1,...,6.

ℓ	1	2	3	4	5	6
ρ	92	90	71	92	95	88

La partition P^* et la représentation $L^* = g(P^*)$ répondent aux exigences du spécialiste.

La fonction de décision (6.4) :

Pour réaliser l'usinage d'une pièce x en acier 35 CD 4, de diamètre 100 mm, avec Z = 250 mm, en utilisant un outil d'angle d'attaque $\gamma = 11°$ et d'usure en dépouille $V_b = 0.1$ mm, on retient h_5 ; en effet, le centre g_5 vérifie :

$$(x-g_5^*)'\ D_e\ (x-g_5^*) \leq (x-g_i^*)'\ D_e\ (x-g_i^*) \qquad i=1,\ldots,6$$

où D_e est la matrice de dimension 4 définie par :

$$D_e = (V^{-1}\ u_1^*)(V^{-1}\ u_1^*)' + (V^{-1}\ u_2^*)(V^{-1}\ u_2^*)'$$

7.2.4 - Modélisation locale en fonction du matériau de la pièce

Pour cette application les paramètres explicatifs sont (cf. Tableau 2 bis) R, R_e, σ, Z, $\emptyset$, V_b.

La modélisation concerne les variables suivantes (cf. Tableau 2 bis)

$Z_1 = \text{Log } V$, $Z_2 = \text{Log } A$, $Z_3 = \text{Log } S$, $Z_4 = \text{Log } F_Z$

Un modèle est écrit sous la forme suivante :

$$Z_4 = \alpha_0 + \alpha_1 Z_1 + \alpha_2 Z_2 + \alpha_3 Z_3$$

Une procédure conversationnelle (6.3) conduit à choisir K=3, r=2, β=100.

On obtient la partition $P^* = (P_1^*, P_2^*, P_3^*)$ et la représentation $L^* = (g_1^*, g_2^*, g_3^*, u_1^*, u_2^*, h_1^*, h_2^*, h_3^*)$ (cf. Tableaux 4, 5 et 6). Cette solution est obtenue pour des partitions initiales différentes.

	α_0	α_1	α_2	α_3
h_1^*	23.8	- 0.07	0.64	0.70
h_2^*	8.3	- 0.09	0.84	0.84
h_3^*	21	- 0.17	0.98	0.61

Tableau 4 : Les coefficients des modèles h_1^*, h_2^*, h_3^*

$$u_1^* = 0.17\ R + 2.67\ R_e + 1.9\ \sigma$$

$$u_2^* = 19.2\ R - 5.9\ R_e + 3.2\ \sigma$$

Les coordonnées de u_1 et u_2 selon Z, $\emptyset$ et V_b sont négligeables.

Tableau 5 : Les vecteurs u_1^*, u_2^*

	g_1^*	g_2^*	g_3^*
$b_1^* = V^{-1}\ u_1^*$	3.88	1.64	1.88
$b_2^* = V^{-1}\ u_2^*$	12.3	11.4	13.8

Tableau 6 : La séparation des centres de gravité des classes par les facteurs discriminants.

8 - CONCLUSION

Cette méthode se situe dans le cadre d'une approche locale de la modélisation, et de la recherche, dans une optique prévisionnelle, de facteurs explicatifs de la structure qui est définie.

De nombreux types de modélisation peuvent être envisagés (il suffit au spécialiste de fournir les programmes MODELE et DELTA relatifs à la famille de modèles retenue). En particulier, J_2 peut être un espace de courbes ou de formes, ou traduire l'évolution dans le temps.

Le développement d'une procédure complète d'apprentissage devrait permettre en outre :

- un choix optimal du nombre de classes K

- la sélection d'un sous ensemble de paramètres de J_1, qui suffise à la reconnaissance des classes dans le cadre de l'analyse discriminante linéaire. (Cette sélection permettrait de limiter la prise de données dans la phase opérationnelle).

La pondération β serait rendue adaptative, dans le but d'assurer que la structure construite par optimisation d'un critère unique, soit stable du point de vue de la modélisation sur J_2 et de la reconnaissance par des facteurs de J_1.

CONCLUSION GENERALE ET PERSPECTIVES

Cette étude illustre l'intérêt de la notion de représentation globale d'une partition du point de vue de l'analyse descriptive des données et du point de vue de la reconnaissance des classes. La transformation linéaire des variables présente les avantages d'une mise en oeuvre simple, sans estimations de paramètres, et de la stabilité des résultats [17].

La visualisation des résultats est un apport précieux dans le cadre d'un dialogue Homme-Machine.

Les trois méthodes présentées sont adaptées au traitement de grands tableaux de données aux niveaux de la taille mémoire nécessaire et du temps de calcul.

Outre l'étude de nouveaux types de modélisation, il apparait les directions de recherche suivantes :

- le développement d'une procédure d'apprentissage pour un choix adaptatif de paramètres de structure, en particulier :
  - le nombre de classes de la partition
  - la dimension du sous espace de représentation
- le traitement de variables qualitatives ou mixtes (qualitatives et quantitatives) par codage des variables qualitatives dans l'optique de l'analyse discriminante.

ANNEXE 1

INERTIE D'UN NUAGE DE POINTS

On se place dans l'espace vectorial R^p. On note de la même manière un point, ou vecteur de R^p, x, et le vecteur colonne de ses p coordonnées dans la base canonique de R^p.

Soit Q une matrice carrée de dimension p symétrique, définie positive;

a) L'application définie par

$$R^p \times R^p \to R$$

$$(x,y) \to x' Q y$$

est un produit scalaire.

b) L'application définie par

$$R^p \times R^p \to R$$

$$(x,y) \to d(x,y) = \sqrt{(x-y)' Q(x-y)}$$

est une distance.

1. Définition.

Soit N un ensemble, ou nuage, de n points de R^p, $x_1,\ldots,x_n$, munis respectivement des masses $p_1,\ldots,p_n$, $(p_i > 0,\ i=1,\ldots,n\ ;\ \sum_{i=1}^{n} p_i = 1)$.

Soit a un point de R^p,

On appelle inertie au point a du nuage N, et on note I(a,N), le réel positif

$$\sum_{i=1}^{n} p_i\, d^2(x_i,a) = \sum_{i=1}^{n} p_i (x_i-a)' Q (x_i-a)$$

2. Théorème de Huyghens

L'inertie d'un nuage de points $N = [(x_1,\ldots,x_n),(p_1,\ldots,p_n)]$ est minimale au point g, appelé centre de gravité de N, et défini par :

$$g = \sum_{i=1}^{n} p_i x_i$$

$I(g,N) = \underset{a}{\text{Min}}\ I(a,N)$ s'appelle l'inertie du nuage N, et se note I(N).

Démonstration :

Soit a un point de R^p, et g le centre de gravité de N,

$$I(a,N) = \sum_{i=1}^{n} p_i (x_i-g+g-a)' Q (x_i-g+g-a)$$

$$= \sum_{i=1}^{n} p_i (x_i-g)' Q (x_i-g) + 2 \sum_{i=1}^{n} p_i (x_i-g)' Q (g-a)$$

$$+ (g-a)' Q (g-a)$$

On a :

$$\sum_{i=1}^{n} p_i (x_i-g)' Q (g-a) = \left(\sum_{i=1}^{n} p_i x_i - g\right)' Q (g-a) = 0$$

On en déduit que pour tout point a dans R^p,

$$I(a,N) = I(g,N) + d^2(g,a)$$

ce qui prouve le résultat.

3. Matrice d'inertie.

On appelle matrice d'inertie au point a du nuage $N[(x_1,\ldots,x_n),(p_1,\ldots,p_n)]$, et on note V(a,N), la matrice carrée de dimension p,

$$V(a,N) = \sum_{i=1}^{n} p_i (x_i-a)(x_i-a)'$$

Si g est le centre de gravité du nuage N, V(g,N) est appelé matrice d'inertie du nuage N.

4. Proposition.

Pour toutes matrices A et B de dimensions respectives (n_1,n_2) et (n_2,n_1) on démontre que les matrices produit AB et BA ont même trace.

En particulier, pour tout couple (x,y) de points de R^p, le réel $(x-y)'\ Q\ (x-y)$ s'écrit :

$$(x-y)'\ Q\ (x-y) = Tr[(x-y)'\ Q\ (x-y)] = Tr[(x-y)(x-y)'\ Q]$$

On en déduit une nouvelle expression de l'inertie I(a,N),

$$I(a,N) = \sum_{i=1}^{n} p_i\ Tr[(x_i-a)(x_i-a)'Q]$$

$$I(a,N) = Tr[(\sum_{i=1}^{n} p_i(x_i-a)(x_i-a)')Q]$$

$$I(a,N) = Tr[V(a,N)\ Q]$$

en particulier, l'inertie du nuage N s'écrit

$$I(N) = Tr[V(g,N)\ Q]\ . \qquad (1)$$

ANNEXE 2

INERTIE EXPLIQUEE PAR UN SOUS ESPACE

1. Sous espaces vectoriels Q orthogonaux [3]

On dit que deux vecteurs x et y de R^p sont Q orthogonaux si $x' Q y = 0$.

Soit U un sous espace vectoriel de R^p. L'ensemble des vecteurs de R^p Q-orthogonaux à tous les vecteurs de U est un sous espace vectoriel de R^p, noté $U^\perp$.

Soit x un vecteur de R^p, et

$$\delta = \underset{y \in U}{\text{Min}} \; (x-y)' \, Q \, (x-y)$$

On démontre qu'il existe un vecteur z de U, et un seul, tel que

$$\delta = (x-z)' \, Q \, (x-z)$$

z est appelé projection Q-orthogonale de x sur U.

De plus on montre que z est le seul vecteur y de U tel que (x-y) soit Q-orthogonal à U.

Tout vecteur x de R^p s'écrit donc, de manière unique :

$$x = x_1 + x_2, \qquad x_1 \in U, \qquad x_2 \in U^\perp.$$

x_1 (resp. x_2) est la projection Q-orthogonale de x sur U (resp. $U^\perp$).

On dit que R^p est somme directe des sous espaces Q-orthogonaux U et $U^\perp$. On note :

$$R^p = U \oplus U^\perp$$

2. Définition.

Soit $N = [(x_1,\ldots,x_n),(p_1,\ldots,p_n)]$, un nuage de points de R^p
a, un point de R^p
U, un sous espace vectoriel de R^p .

On note $x_{1,1},\ldots,x_{n,1}$, a_1 les projections Q-orthogonales sur U des points $x_1,\ldots,x_n,a$.

On appelle inertie au point a du nuage N, expliquée par le sous espace U, et on note I(a,N,U), l'inertie au point a_1 du nuage $N_1 = [(x_{1,1},\ldots,x_{n,1}),(p_1,\ldots,p_n)]$.

3. Théorème de Huyghens (pour les inerties expliquées)

Soit N un nuage de points de R^p
U un sous espace vectoriel de R^p.

L'inertie au point a du nuage N, expliquée par le sous espace U, atteint son minimum pour le centre de gravité g du nuage N.

$$I(g,N,U) = \operatorname{Min}_a I(a,N,U)$$

s'appelle l'inertie du nuage N, expliquée par le sous espace U.

On note I(N,U) cette inertie.

Démonstration :

$$I(a,N,U) = \sum_{i=1}^{n} p_i (x_{i,1}-a_1)' Q (x_{i,1}-a_1)$$

où $x_{i,1},\ldots,x_{i,n},a_1$ sont les prjections Q-orthogonales sur U de $x_1,\ldots,x_n,a$.

I(a,N,U) est l'inertie au point a_1 du nuage $N_1 = (x_{1,1},\ldots,x_{n,1})$. D'après (A1.2), elle sera minimale pour g_1, centre de gravité du nuage

$$g_1 = \sum_{i=1}^{n} p_i \, x_{i,1}$$

g_1 est la projection Q-orhtogonale sur U du centre de gravité g du nuage N.

L'inertie I(a,N,U) sera donc minimale pour tout a de la forme $a = g_1 + a_2$ avec a_2 dans $U^{\perp}$; en particulier $g = g_1 + g_2$ convient.

4. Théorème de Pythagore - Applications

4.1. Théorème de Pythagore

Soit N un nuage de points de R^p
a un point de R^p
U un sous espace vectoriel de R^p

L'inertie au point a du nuage N est égale à la somme des inerties au point a du nuage N expliquées par les sous espaces U et $U^{\perp}$.

$$I(a,N) = I(a,N,U) + I(a,N,U^{\perp})$$

Démonstration :

$$I(a,N) = \sum_{i=1}^{n} p_i (x_i - a)' \, Q \, (x_i - a)$$

Soit $x_i = x_{i,1} + x_{i,2}$ et $a = a_1 + a_2$, $x_{i,1} \in U$, $a_1 \in U$, $x_{i,2} \in U^{\perp}$, $a_2 \in U^{\perp}$ $i=1,\ldots,n$.

$$I(a,N) = \sum_{i=1}^{n} p_i ((x_{i,1} - a_1) + (x_{i,2} - a_2))' \, Q \, ((x_{i,1} - a_1) + (x_{i,2} - a_2))$$

$$I(a,N) = I(a,N,U) + I(a,N,U^{\perp}) + 2 \sum_{i=1}^{n} p_i (x_{i,1} - a_1)' \, Q \, (x_{i,2} - a_2)$$

Pour tout i, $i=1,\ldots,n$, $x_{i,1} - a_1$ appartient au sous espace U, $x_{i,2} - a_2$ appartient au sous espace U : Ainsi,

$$(x_{i,1} - a_1)' \, Q \, (x_{i,2} - a_2) = 0$$

d'où le résultat.

4.2. Proposition

Pour toute décomposition d'un sous espace vectoriel U de R^p en somme directe des sous espaces Q-orthogonaux U_1 et U_2, $U = U_1 \oplus U_2$, on a la décomposition de l'inertie expliquée

$$I(a,N,U) = I(a,N,U_1) + I(a,N,U_2)$$

Démonstration :

Il suffit d'appliquer le théorème de Pythagore à la projection du nuage N sur U.

4.3. Proposition

Pour toute décomposition d'un sous espace vectoriel U de R^p en somme directe de droites vectorielles Q-orthogonales $D_1,\ldots,D_r$, $U = D_1 \oplus \ldots \oplus D_r$, on a la décomposition de l'inertie expliquée,

$$I(a,N,U) = \sum_{s=1}^{r} I(a,N,D_s)$$

En particulier :

$$I(N,U) = \sum_{s=1}^{r} I(N,D_s)$$

Démonstration :

Elle s'effectue par récurrence en utilisant 4.2.

4.4. Proposition

Soit F_s l'ensemble des sous espaces vectoriels de dimension s de R^p, $(s=1,\ldots,p)$;

Tout sous espace U_o de dimension s, $(s \geq 2)$, tel que

$$I(a,N,U_o) = \max_{U \in F_s} I(a,N,U)$$

contient un sous espace V_o de dimension (s-1) tel que

$$I(a,N,V_o) = \max_{V \in F_{s-1}} I(a,N,V)$$

Démonstration

Soit U_o dans F_s vérifiant $I(N,U_o) = \max_{U \in F_s} I(N,U)$ et V_1 dans F_{s-1} vérifiant $I(N,V_1) = \max_{V \in F_{s-1}} I(N,V)$

$$\dim U_o + \dim V_1^{\perp} = s + (p-(s-1)) = p+1$$

L'intersection des sous espaces U_o et $V_1^{\perp}$ est par conséquent de dimension supérieure à 1.

Soit D une droite vectorielle incluse dans $U_o \cap V_1^{\perp}$

Soit V_o le supplémentaire Q-orthogonal de D dans U_o

$$U_o = D \oplus V_o \qquad V_o \in F_{s-1}$$

Soit U_1 la somme directe

$$U_1 = D \oplus V_1 \qquad U_1 \in F_s$$

On a, d'après le théorème de Pythagore pour les inerties expliquées :

$$I(N,U_o) = I(N,D) + I(N,V_o) \qquad (1)$$

$$I(N,U_1) = I(N,D) + I(N,V_1) \qquad (2)$$

On a d'autre part :

$$I(N,U_1) \leq I(N,U_o) = \max_{U \in F_s} I(N,U)$$

On déduit alors de (1) et (2)

$$\max_{V \in F_{s-1}} I(N,V) = I(N,V_1) \leq I(N,V_o)$$

ce qui prouve que V_o sous espace vectoriel de U_o réalise le maximum pour V dans F_{s-1} de I(N,V).

5. Inertie expliquée par une droite vectorielle.

Soit D une droite vectorielle, et u un vecteur unitaire de D (pour la métrique Q), c'est à dire :

$$u' Q u = 1.$$

Tout vecteur x de R^p se décompose, de manière unique, sous la forme, (A2.1),

$$x = \lambda u + x_2 , \qquad \lambda \in R, \; x_2 \in D^{\perp}$$

On en déduit :

$$(Qu)'x = u' Q (\lambda u + x_2) = \lambda$$

Qu s'interprète comme la forme linéaire qui associe à tout vecteur de R^p sa coordonnée suivant u.

L'inertie au point a d'un nuage $N = [(x_1,\ldots,x_n),(p_1,\ldots,p_n)]$ expliquée par D s'écrit :

$$I(a,N,D) = \sum_{i=1}^{n} p_i [(x_i-a)' Q u] u' Q [(x_i-a)' Q u] u$$

$(x_i-a)' Q u$ est un scalaire, égal à $u' Q (x_i-a)$.

On en déduit :

$$I(a,N,D) = \sum_{i=1}^{n} p_i \, u' Q (x_i-a) \, u' Q u (x_i-a)' Q u$$

Comme u est Q normé ($u' Q u = 1$), on a :

$$I(a,N,D) = \sum_{i=1}^{n} p_i \, u' Q (x_i-a)(x_i-a)' Q u$$

$$I(a,N,D) = u' Q V(a,N) Q u$$

où $V(a,N) = \sum_{i=1}^{n} p_i(x_i-a)(x_i-a)'$ est la matrice d'inertie au point a du nuage N, (A1.3).

Remarque :

Si le vecteur u n'est pas supposé unitaire, on a :

$$I(a,N,D) = \frac{u' Q V(a,N) Q u}{u' Q u}$$

6. Recherche du maximum de l'inertie d'un nuage expliquée par un sous espace de dimension donnée.

6.1. Introduction

Soit U_r dans F_r, $(2 \leq r \leq p-1)$, tel que

$$I(N,U_r) = \underset{U \in F_r}{\text{Max}} \; I(N,U)$$

D'après (4.4), appliqué avec s=r,...,2, il existe des sous espaces de R^p, $U_{r-1},\ldots,U_1$, de dimensions respectives r-1,...,1, tels que :

a) $U_r \supset U_{r-1} \supset \ldots \supset U_2 \supset U_1$

b) $I(N,U_s) = \underset{U \in F_s}{\text{Max}} \; I(N,U) \qquad s=r,\ldots,1$

U_1 est une droite vectorielle. On la note D_1.

Soit D_s le supplémentatire Q-orthogonal de U_{s-1} dans U_s (s=2,...,r). D_s est une droite vectorielle

$$U_{s-1} \oplus D_s = U_s$$

Avec ces notations,

$$U_s = D_s \oplus \ldots \oplus D_1 \qquad s=r,\ldots,1$$

La droite vectorielle D_1 vérifie, d'après b) ci dessus,

$$I(N,D_1) = \underset{D \in F_1}{\text{Max}} \; I(N,D)$$

6.2. Proposition

Les droites vectorielles $D_2,\ldots,D_r$ vérifient :

$$I(N,D_{s+1}) = \underset{D\in F(D_1\oplus\ldots\oplus D_s)^{\perp}}{\text{Max}} I(N,D) \qquad s=1,\ldots,r-1$$

où $F(D_1\oplus\ldots\oplus D_s)^{\perp}$ désigne l'ensemble des droites vectorielles de R^p, Q-orthogonales à chacune des droites $D_1,\ldots,D_s$, et donc à leur somme directe U_s.

Démonstration :

On suppose, pour un entier s, $1 \leq s \leq r-1$

$$I(N,D_{s+1}) < \underset{D\in F(D_1\oplus\ldots\oplus D_s)^{\perp}}{\text{Max}} I(N,D) \qquad 1 \leq s \leq r-1$$

Soit D_o dans $F(D_1\oplus\ldots\oplus D_s)^{\perp}$ tel que

$$I(N,D_{s+1}) < I(N,D_o)$$

Soit $U_o=U_s \oplus D_o$. U_o appartient à F_{s+1}, et d'après (4.2)

$$I(N,U_o) = I(N,U_s) + I(N,D_o)$$

On en déduit :

$$I(N,U_o) > I(N,U_s) + I(N,D_{s+1}) = I(N,U_{s+1})$$

Soit

$$I(N,U_o) > \underset{U\in F_{s+1}}{\text{Max}} I(N,U)$$

ce qui est absurde.

6.3. Résumé

Si le sous espace U_r de dimension r vérifie

$$I(N,U_r) = \underset{U\in F_r}{\text{Max}} I(N,U)$$

Il existe r droites vectorielles Q-orthogonales $D_1,\ldots,D_r$ telles que :

- $U_r = D_1 \oplus \ldots \oplus D_r$
- $D_1,\ldots,D_r$ réalisent les maxima successifs, sous contrainte de Q orthogonalité, de l'inertie du nuage N, expliquée par une droite vectorielle.
- De plus, (4.3),

$$I(N,U_r) = \sum_{s=1}^{r} I(N,D_s)$$

6.4. Proposition

Soit N un nuage de points dans R^P, de matrice d'inertie V, Soit D une droite vectorielle de vecteur directeur u et D_o la droite vectorielle engendrée par le vecteur v = VQu.

On a la relation :

$$I(N,D) \leq I(N,D_o)$$

De plus :

$$(I(N,D) = I(N,D_o)) \iff (u \text{ est vecteur propre de la matrice VQ. } VQu = \lambda u)$$

et dans ces conditions,

$$I(N,D) = \lambda$$

(la démonstration est donnée en [4]).

6.5. Proposition

a) La matrice VQ est diagonalisable

b) Ses valeurs propres sont positives ou nulles

c) Les sous espaces propres associés à deux valeurs propres distinctes sont Q-orthogonaux.

(la démonstration est donnée en [4])

6.6. Notations

Soit (6.5)

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r \geq 0,$$

les r plus grandes valeurs propres de la matrice VQ, (une valeur propre λ de multiplicité m > 1 apparaît comme $\lambda_s = \lambda_{s+1} = \ldots = \lambda_j$, $j = \text{Min}(s+m-1,r)$.).

$$E_1, E_2, \ldots, E_r$$

les sous espaces propres associés aux valeurs propres $\lambda_1, \lambda_2, \ldots, \lambda_r$ (si λ est valeur propre multiple, de multiplicité m > 1, le sous espace propre associé E_λ, de dimension m, apparaît comme $E_s = E_{s+1} = \ldots = E_j$, $j = \text{Min}(s+m-1,r)$).

6.7. Proposition

Soient les droites vectorielles $D_1,\ldots,D_r$ définis en 6.3. Avec les notations introduites en 6.6, on a :

Pour tout s, s=1,...,r, la droite vectorielle D_s est incluse dans le sous espace propre E_s, (si λ_s est valeur propre simple, D_s et E_s sont égaux), et

$$I(N,D_s) = \lambda_s$$

Démonstration :

On note v_s un vecteur directeur de D_s, s=1,...,r.

Pour s=1,...,r, D_s réalise le maximum de l'inertie I(N,D) expliquée par une droite vectorielle, sous contrainte (pour s=2,...,r) de Q orthogonalité avec les droites $D_1,\ldots,D_{s-1}$. Cette dernière condition implique pour les vecteurs directeurs $v_1,\ldots,v_s$, la relation :

$$v_s' \, Q \, v_t = 0 \qquad s=1,\ldots,r \; ; \; t=1,\ldots,s-1 \qquad (3)$$

Soit D_s^* la droite vectorielle engendrée par le vecteur $v_s^* = VQ\ v_s$, (s=1,...,r).

D'après (6.4), on a, pour s=1,...,r :

$$I(N,D_s) \leq I(N,D_s^*) \qquad (4)$$

Montrons par récurrence la propriété P_s :

(P_s) : la droite vectorielle D_s est incluse dans E_s
(conséquences, $VQ\ v_s = \lambda_s\ v_s$
$I(N,D_s) = \lambda_s$)

a) Cas s=1 On déduit de (4)

$$I(N,D_1) = \underset{D\in F_1}{\text{Max}}\ I(N,D) = I(N,D_1^*)$$

On en déduit, (6.4), que v_1 est vecteur propre de la matrice VQ,

$$VQ\ v_1 = \lambda v_1 \qquad (5)$$

et que de plus

$$I(N,D_1) = \lambda\ . \qquad (6)$$

Comme toute droite vectorielle D incluse dans E_1 vérifie, (6.4)

$$I(N,D) = \lambda_1 \geq \lambda,$$

on a

$$\lambda = \lambda_1 \qquad (7)$$

Conclusions. $VQ\ v_1 = \lambda_1\ v_1$ (5) et (7)

Le sous espace propre E_1 contient le vecteur v_1, et donc la droite vectorielle D_1. De plus

$$I(N,D_1) = \lambda_1 \qquad (6) \text{ et } (7)$$

b) Cas $1 < s \leq r$

D'après l'hypothèse de récurrence, on a :

$$VQ\ v_t = \lambda_t\ v_t \qquad t = 1,\ldots,s-1 \tag{8}$$

D'après (4)

$$I(N,D_s) = \underset{D\in F(D_1\oplus\ldots\oplus D_{s-1})}{\text{Max}} I(N,D) \leq I(N,D_s^*) \tag{9}$$

On déduit de (8) et de (3) , pour t=1,...,s-1 :

$$v_s^{*\prime}\ Q\ v_t = v_s'\ QV\ Q\ v_t = \lambda_t v_s' Q v_t = 0$$

Ainsi, la droite D_s est orthogonale aux droites $D_1,\ldots,D_{s-1}$. On en déduit alors de (9),

$$I(N,D_s) = I(N,D_s^*).$$

D'après (6.4), v_s est valeur propre de la matrice VQ

$$V\ Q\ v_s = \lambda\ v_s \tag{10}$$

et

$$I(N,D_s) = \lambda\ . \tag{11}$$

Comme toute droite vectorielle D incluse dans E_s vérifie

$$I(N,D) = \lambda_s$$

et comme toute droite D' incluse dans un sous espace propre associé à une valeur propre $\lambda_i > \lambda_s$ ne vérifie pas les conditions de Q orthogonalité avec $D_1,\ldots,D_{s-1}$, on a :

$$\lambda = \lambda_s \tag{12}$$

Conclusions $VQ\ v_s = \lambda_s\ v_s$ (10) et (12)

Le sous espace propre E_s contient le vecteur v_s et donc la droite vectorielle D_s. De plus

$$I(N,D_s) = \lambda_s \qquad (11) \text{ et } (12)$$

6.8. Proposition

1. La famille $(u_1,\ldots,u_r)$ définie ci-dessous est Q - orthonormée.

2. Le sous espace vectoriel F_r engendré par cette famille réalise le maximum pour U dans F_r de l'inertie expliquée I (N,U).

Définition précise de la famille $(u_1,\ldots,u_r)$

Les vecteurs $u_1,\ldots,u_r$ sont vecteurs propres, associés aux r plus grandes valeurs propres de la matrice V Q.

Plus précisément (on utilise les notations introduites en 6.6) :

a) Si λ est valeur propre simple, u_s est défini par :

$$V\,Q\,u_s = \lambda_s\,u_s$$

$$u_s'\,Q\,u_s = 1$$

$$u_s'\,Q\,e_{is} > 0$$

où e_i désigne le $i^{ème}$ vecteur de la base canonique de R^p

et où i_s est le plus petit indice i tel que e_i ne soit pas Q orthogonal à E_s.

b) Si λ est valeur propre multiple, de multiplicité $m>1$, $\lambda = \lambda_s = \ldots \lambda_j$ où $j = \text{Min}\,(s+m-1,r)$, $u_s,\ldots,u_j$ sont définis de la manière suivante :

a) Soit i_s le plus petit indice i tel que e_i ne soit pas Q orthogonal au sous espace propre E_λ

Soit v_s la projection Q orthogonale de e_{is} sur E_λ u_s est défini par

$$u_s = \frac{v_s}{\sqrt{v_s'\,Q\,v_s}}$$

b) Pour t = s+1,...,j :

Soit M_t le sous espace de R^p engendré par les vecteurs $u_s', \ldots, u_{t-1}$

Soit $E_\lambda \ \theta \ M_t$ le supplémentaire Q orthogonal de M_t dans E_λ

Soit i_t le plus petit indice i tel que e_i ne soit pas Q orthogonal à $E_\lambda \ \theta \ M_t$

Soit v_t la projection Q orthogonale de e_{i_t} sur le sous espace $E_\lambda \ \theta \ M_t$

u_t est défini par

$$u_t = \frac{v_t}{v_t' \, Q \, v_t}$$

Démonstration :

1. Les vecteurs u_s $(1 \leq s \leq r)$ associés à des vecteurs propres différentes sont Q orthogonaux (6.5). Dans le cas d'une valeur multiple $\lambda_s = \ldots = \lambda_j$, les vecteurs $u_s, \ldots, u_j$ sont choisis Q orthogonaux. Chaque vecteur u_s, s = 1,...,r, est normé.

La famille $(u_1, \ldots, u_r)$ est donc Q-orthonormée

2. D'après (6.3) et (6.7) on a

$$\underset{U \in F_r}{\text{Max}} \quad I(N,U) = \lambda_1 + \ldots + \lambda_r$$

U_r est somme directe des droites vectorielles Q orthogonales engendrées par $u_1, \ldots, u_r$, vecteurs propres de la matrice VQ.

D'après (4.3.) et (6.4) on a :

$$I(N,U_r) = \lambda_1 + \ldots + \lambda_r$$

On en déduit

$$I(N,U_r) = \underset{U \in F_r}{\text{Max}} \quad I(N,U)$$

Remarque

Le sous espace U_r est inclu dans la somme directe des sous espaces propres $E_1 \oplus \ldots \oplus E_r$.

U_r et $E_1 \oplus \ldots \oplus E_r$ sont égaux (et donc U_r ne dépend pas de la famille $(u_1,\ldots,u_r)$) sauf dans le cas suivant :

λ_s est la valeur multiple ($s \leq r$), de multiplicite m, $m > r-s+1$. Dans ce cas U_r ne dépand pas de $u_1,\ldots,u_{s-1}$, mais dépand du choix de $u_s,\ldots,u_r$ dans $E_s = \ldots = E_r$ de dimension m.

ANNEXE 3

INERTIES INTRA CLASSE ET INTER CLASSE D'UNE PARTITION

1. Introduction

On suppose que les points $x_1,\ldots,x_n$ du nuage $N=[(x_1,..,x_n),(p_1,..,p_n)]$ sont répartis en K groupes, $P_1,\ldots,P_K$. A chaque classe P_ℓ de la partition $(P_1,\ldots,P_K)$ du nuage N est associé un nuage de points, noté également P_ℓ.

Pour $\ell=1,\ldots,K$, on note $\underline{p}_\ell$ le poids de la classe P_ℓ,

$$\underline{p}_\ell = \sum_{x_i \in P_\ell} p_i \qquad \left(\text{on a } \sum_{\ell=1}^{K} \underline{p}_\ell = 1\right)$$

et g_ℓ le centre de gravité de la classe P_ℓ,

$$g_\ell = \frac{1}{\underline{p}_\ell} \sum_{x_i \in P} p_i\, x_i$$

Proposition :

Les nuages N et $G = [(g_1,\ldots,g_K),(p_1,\ldots,p_K)]$ ont même centre de gravité.

Démonstration :

Soit g_o le centre de gravité du nuage G

$$g_o = \sum_{\ell=1}^{K} \underline{p}_\ell\, g_\ell = \sum_{\ell=1}^{K} \underline{p}_\ell \cdot \frac{1}{\underline{p}_\ell} \sum_{x_i \in P_\ell} p_i\, x_i$$

$$g_o = \sum_{i=1}^{n} p_i\, x_i$$

On en déduit que g_o est centre de gravité du nuage N

2. Inertie aux points $(a_1,\ldots,a_K)$ d'une partition $P = (P_1,\ldots,P_K)$

2.1. Définition :

On appelle inertie aux points $a_1,\ldots,a_K$, de la partition $P = (P_1,\ldots,P_K)$, et on note $I((a_1,\ldots,a_K),P)$ la somme, pour $\ell = 1,\ldots,K$, des inerties $I(a_\ell,P_\ell)$, où $I(a_\ell,P_\ell)$ est l'inertie au point a_ℓ de la classe P_ℓ.

$$I((a_1,\ldots,a_K),P) = \sum_{\ell=1}^{K} I(a_\ell,P_\ell)$$

Si $(g_1,\ldots,g_K)$ sont les centres de gravité des classes $P_1,\ldots,P_K$, l'inertie $I((g_1,\ldots,g_K),P)$ s'appelle l'inertie intra classe de la partition P, et se note $I_{intra}(P)$.

2.2. Proposition :

Pour toute partition $P = (P_1,\ldots,P_K)$ du nuage N,

$$I_{intra}(P) = I((g_1,\ldots,g_K),P) = \min_{a_1,\ldots,a_K} I((a_1,\ldots,a_K),P).$$

Démonstration

On déduit de la définition (2.1) de l'inertie $I((a_1,\ldots,a_K),P)$ que :

$$\min_{a_1\ldots a_K} I((a_1,\ldots,a_K),P) = \sum_{\ell=1}^{K} \min_{a} I(a,P_\ell)$$

D'autre part, d'après le théorème de Huyghens le minimum de l'inertie $I(a,P_\ell)$ est atteint pour le centre de gravité de la classe P_ℓ.

On en déduit :

$$\min_{a_1,\ldots,a_K} I((a_1,\ldots,a_K),P) = \sum_{\ell=1}^{K} I(g_\ell,P_\ell) = I((g_1,\ldots,g_K),P)$$

3. Inertie inter classe d'une partition

3.1. Définition :

On appelle inertie inter classe de la partition $P = (P_1,\ldots,P_K)$, et on note $I_{inter}(P_1,\ldots,P_K)$ ou $I_{inter}(P)$, l'inertie du nuage $G = [(g_1,\ldots,g_K),(p_1,\ldots,p_K)]$ des centres de gravité des classes $P_1,\ldots,P_K$, munis des poids de ces classes.

3.2. Définition :

On appelle matrice d'inertie inter classe de la partition $P = (P_1,\ldots,P_K)$, et on note $B(P_1,\ldots,P_K)$ ou $B(P)$, la matrice d'inertie (A1.3) du nuage $G = (g_1,\ldots,g_K, \underline{p}_1 ,\ldots,\underline{p}_K)$ des centres de gravité des classes $P_1,\ldots,P_K$.

3.3. Proposition :

Pour toute partition P, on a :

$$I_{inter}(P) = Tr\ (B(P)\ Q)$$

Démonstration :

On applique la relation (1) du §.4 de l'annexe 1 au nuage G.

4. Inerties intra classe, inter classe et totale

Proposition :

Pour toute partition P du nuage N, La somme des inerties intra classe et inter classe de la partition P est égale à l'inertie du nuage N.

Démonstration :

Soit $G = [(g_1,\ldots,g_K),(\underline{p}_1 ,\ldots,\underline{p}_K)]$ le nuage des centres de gravité des classes $P_1,\ldots,P_K$, d'une partition P.

$$I_{intra}(P) = \sum_{\ell=1}^{K} \sum_{x_i \in P_\ell} p_i (x_i-g_\ell)'\ Q\ (x_i-g_\ell)$$

$$I_{inter}(P) = \sum_{\ell=1}^{K} \underline{p}_\ell\ \ (g_\ell-g_o)'\ Q\ (g_\ell-g_o)$$

où g_o est le centre de gravité des nuages G et N (3.1)

$$g_o = \sum_{\ell=1}^{K} \underline{p}_\ell \quad g_\ell = \sum_{i=1}^{n} p_i \, x_i$$

L'inertie $I(g_o,N)$ du nuage N se décompose de la manière suivante :

$$I(g_o,N) = \sum_{i=1}^{n} p_i (x_i-g_o)' \, Q \, (x_i-g_o)$$

$$I(g_o,N) = \sum_{\ell=1}^{K} \sum_{x_i \in P_\ell} p_i (x_i-g_\ell+g_\ell-g_o)' \, Q \, (x_i-g_\ell+g_\ell-g_o)$$

$$I(g_o,N) = I_{intra}(P) + I_{inter}(P) + 2 \sum_{\ell=1}^{K} \sum_{x_i \in P_\ell} p_i (x_i-g_\ell)' \, Q \, (g_\ell-g_o)$$

Enfin, le dernier terme, qui s'écrit,

$$\sum_{\ell=1}^{K} \left(\sum_{x_i \in P_\ell} p_i \, x_i - \underline{p}_\ell \, g_\ell \right)' \, Q \, (g_\ell-g_o)$$

est nul par définition de g_ℓ, $\ell=1,\ldots,K$.

ANNEXE 4

INERTIE INTRA CLASSE ET INTER CLASSE

EXPLIQUEES PAR UN SOUS ESPACE

Soit $P = (P_1,\ldots,P_K)$ une partition d'un nuage $N = [(x_1,\ldots,x_n),(p_1,\ldots,p_K)]$ de $\mathbb{R}^P$.

Soit U un sous espace vectoriel de $\mathbb{R}^P$.

Soit $N_1 = [(y_1,\ldots,y_n),(p_1,\ldots,p_n)]$ le nuage des projections Q-orthogonales sur U, $y_1,\ldots,y_n$, des points $x_1,\ldots,x_n$.

1. Définition :

On appelle inertie aux points $a_1,..,a_K$ de la partition P du nuage N, expliquée par le sous espace U, et on note $I((a_1,..,a_K),P,U)$, l'inertie de la partition P du nuage N_1 aux points $b_1,\ldots,b_K$, projections Q-orhtogonales sur U des points $a_1,\ldots,a_K$.

Si $g_1,\ldots,g_K$ sont les centres de gravité des classes $P_1,\ldots,P_K$, l'inertie $I((g_1,\ldots,g_K),P,U)$ s'appelle l'inertie intra classe de la partition P expliquée par le sous espace U, et se note $I_{intra}(P,U)$.

2. Définition :

On appelle inertie inter classe de la partition P du nuage N, expliquée par le sous espace U, et on note $I_{inter}(P,U)$, l'inertie inter classe du nuage N_1.

3. Proposition :

Pour toute partition $P = (P_1,\ldots,P_K)$ du nuage N, et tout sous espace U,

$$I_{intra}(P,U) = I((g_1,..,g_K),P,U) = \min_{a_1,..,a_K} I((a_1,..,a_K),P,U)$$

Démonstration

On a par définition

$$I_{intra}(P,U) = I((g_1,\ldots,g_K),P,U)$$

En appliquant la proposition 2.2. de l'annexe 3 au nuage N_1 on déduit la relation

$$I((g_1,\ldots,g_K),P,U) = \min_{a_1,..,a_K} I((a_1,\ldots,a_K),P,U)$$

4. Proposition :

Pour toute partition P du nuage N, et tout sous espace U, la somme des inerties intra classe et inter classe de la partition P, expliquées par U est égale à l'inertie du nuage N, expliquée par U.

$$I_{intra}(P,U) + I_{inter}(P,U) = I(N,U)$$

Démonstration :

On applique la proposition du § 4 de l'annexe 3 au nuage N_1.

ANNEXE 5

Soit B et Q deux matrices carrées. On suppose que Q est régulière (on note Q^{-1} son inverse).

Les matrices BQ et QB ont mêmes valeurs propres. Pour toute valeur propre des matrices BQ et BQ, on a les relations suivantes entre vecteurs propres :

si u est vecteur propre de la matrice BQ, b = Qu est vecteur propre de la matrice QB.

si b est vecteur propre de la matrice QB, $u = Q^{-1}b$ est vecteur propre de la matrice BQ.

(La vérification est immédiate)

ANNEXE 6

Soit B et Q deux matrices carrées de dimension p. On suppose que Q est régulière, et que la matrice B se décompose sous la forme

$$B = C\ C'$$

où C est une matrice à p lignes et K colonnes.

Les matrices QB et C'QC ont les mêmes valeurs propres non nulles. (QB est de dimension p, C'QC est de dimension K. De plus C'QC est une matrice symétrique). Pour toute valeur propre λ non nulle des matrices QB et CQC, on a les relations suivantes entre vecteurs propres :

si b est vecteur propre de la matrice QB, a = C'b est vecteur propre de la matrice C'QC.

si a est vecteur propre de la matrice C'QC, b = QCa est vecteur propre de la matrice QB.

(La vérification est immédiate).

BIBLIOGRAPHIE

[1] BENZECRI J.P. - "L'analyse des données"
Tome 2, L'analyse des correspondances DUNOD - Paris 1973.

[2] BERTIER P., et BOURROCHE J.M. - "Analyse des données multidimensionnelles"
Presses Universitaires de France - Paris 1975.

[3] CAGNAC G., RAMIS E. et COMMEAU J. - "Traité de Mathématiques Spéciales"
MASSON - Paris.

[4] CAILLIEZ F. et PAGES J.P. - "Introduction à l'Analyse des Données"
SMASH - Paris 1976.

[5] CELEUX G. - "Contribution statistique à la reconnaissance automatique de spectres de machines toutnantes"
Thèse de 3° cycle - Université PARIS VI 1978.

[6] CHARLES C. -"Regression typologique et reconnaissance des formes"
Thèse de 3° cycle - Université Paris IX 1977.

[7] DER MEGREDITCHIAN G. - "Quelques aspects pratiques de la sélection des prédicteurs en analyse discriminante"
Journées de Statistiques Paris 1979.

[8] DIDAY E. - "Une nouvelle méthode en classification automatique et reconnaissance des formes"
Revue de Statistique Appliquée, Vol. XIX, N° 2, 1971.

[9] DIDAY E. - "Optimisation en classification automatique et reconnaissance des formes"
Note Scientifique n° 6 - Supplément au bulletin de l'IRIA n° 12 - 1972.

[10] DIDAY E. - "Sélection typologique de paramètres"
Rapport de rechecher n° 188, IRIA-LABORIA 1976.

[11] DIDAY E. - "Analyse canonique du point de vue de la classification automatique"
Rapport de recherche n° 293, IRIA-LABORIA 1978.

[12] DIDAY E., GOVAERT G., et LEMOINE Y. - "A new kind of representation in clustering"
Proc of the fourth international joint conference on pattern recognition Kyoto 1978.

[13] FISHER R.A. - "The use of multiple measurements in taxinomie problems"
Annals of Eugenics, Vol. 7, 1936.

[14] GOVAERT G. - "Classification automatique et distances adaptatives"
Thèse de 3è cycle - Université Paris VI 1975.

[15] LAFAYE DE MICHEAUX D. - "Approximations d'analyses canoniques non linéaires de variables aléatoires et analyses factorielles privilégiantes"
Thèse de Docteur Ingénieur - Université de Nice 1978.

[16] LEBART L. et FENELON J.P. - "Statistiques et informatique appliquées"
DUNOD Ed. Paris 1975.

[17] LEBART L., MORINEAU A. et TABARD N. - "Techniques de la description statistique "
DUNOD Ed. Paris 1977.

[18] J. de MONTGOLFIER et BERTIER P. - "Approche multicritère des problèmes de décision"
Editions Hommes et Techniques - Paris 1979.

[19] MUTEL B. - "Etude de la commande optimale du processus d'usinage dans un ilot de fabrication mécanique"
Thèse de Docteur Es Sciences - Université de Metz 1978.

[20] NAKACHE J.P. et DUSSERRE L. - "Etude de problème posés par l'analyse linéaire discriminante en pas a pas"
Revue de Statistique Appliquée, Vol. XXIII, n° 3, 1975.

[21] NAKACHE J.P. - "Comparaisons des règles d'affectation associées à différentes méthodes de discrimination"
Rapport de recherche INSERM 1976.

[22] OK-SAKUN Y. - "Analyse factorielle typologique et lissage typologique"
Thèse de 3° cycle - Université Paris VI 1975.

[23] PERENNOU G. - "Contribution à l'étude des discriminateurs, Calcul et optimisation"
Thèse d'Etat - Université Paul Sabatier, Toulouse 1968.

[24] RAO C.R. - "Advanced statistical methods in biometric research"
JOHN WILEY and sons - 1952.

[25] ROMEDER J.M. - "Méthodes et programmes d'analyse discriminante"
DUNOD - Paris 1973.

[26] SAPORTA G. - "Liaisons entre plusieurs ensembles de variables et codage de données qualitatives"
Thèse de 3° cycle - Université Paris VI 1975.

[27] SEBESTYEN G.S. - "Decision making process in pattern recognition"
The Macmillan Company 1962.

[28] SIMON J.C. - "Some current topics in clustering in relation with pattern recognition"
Proc. of the fourth international joint conference on pattern recognition
Kyoto 1978.

[29] ULMO J. - "Differents aspects de l'analyse discriminante"
Revue de Statistiques Appliquées, Vol. XXI, n° 2, 1973.

[30] WATANABE S. - "Knowing and guessing"
John Wiley & sons 1969.