

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITÉ MOHAMED KHIDER, BISKRA
Faculté des Sciences Exactes et des Sciences de la Nature et de la Vie
DÉPARTEMENT DE MATHÉMATIQUES



Mémoire présenté en vue de l'obtention du Diplôme :

MASTER en Mathématiques

Option : STATISTIQUE

Par

MENZER Fahima

Titre :

ANALYSE DISCRIMINANTE

Membres du Comité d'Examen :

Pr.	BENATIA Fatah	UMKB	Président
Dr.	BENELMIR Imane	UMKB	Encadreur
Dr.	TOUBA Sonia	UMKB	Examinatrice

Juin 2024

Dédicace

Je dédie cet humble travail

À celle qui m'a

aidé par ses sincères prières

et douaa à la plus précieuse

personne de ma vie

ma mère.



À celui qui a bien

travaillé pour m'apprendre

c'est quoi le combat et qui

m'a fait ce que je suis

mon père



À mes sœur et mes frères qui m'ont soutenu et m'ont souhaité du succès

*Je prie **ALLAH** de protégé ma famille pour nous...♡*

À celle qui était parmi nous et a laissé un vide...

À celle qui ne nuisait en rien, et qui était une personne précieuse...

*Que **ALLAH** ait ton âme en Sa miséricorde...*

ma grand-mère...♡

À la fin, j'espère que ce travail sera très utile pour les autres candidats de cette spécialité ♡.

REMERCIEMENTS

Au nom d'**ALLAH** le plus clément et le plus miséricordieux.

Avant tout, je tiens à remercier "**ALLAH**" le Tout-Puissant, qui m'a accordé la santé, le courage, la patience et la volonté nécessaires pour accomplir ce travail.

Je tiens tout d'abord à remercier à mon encadreur **D.r. BENELMIR Imane** d'avoir accepté diriger de ce travail, pour ses remarques et ses conseils. Sans tout cela, ce mémoire n'aurait pas pu être réalisé.

Par la suite, j'aimerais remercier tous mes enseignants qui m'enseignent dans tout au long de mon parcours d'études.

Mes vifs remerciements vont également aux membres du jury, d'avoir accepté d'examiner et d'évaluer ce travail.

Table des matières

Remerciements	ii
Table des matières	iii
Table des figures	v
Liste des tables	vi
Introduction	1
1 Méthode géométrique	3
1.1 Données Caractéristiques	4
1.1.1 Matrice des observations	4
1.1.2 Centre de gravité	5
1.1.3 Poids	5
1.1.4 Nuage des points	5
1.1.5 Matrice de variance-covariance	7
1.1.6 Matrice de corrélation	9
1.1.7 Tableau des données	10
1.1.8 Fonctions, axes et variables discriminantes	10

1.1.9	Analyse factorielle discriminante et analyse en composantes	
	principales	15
1.2	Règles géométriques	17
1.2.1	Distances	17
1.2.2	Règle géométrique d'affectation	20
2	Méthode Probabiliste	22
2.1	Règle bayésienne	23
2.1.1	Modèle gaussien	24
2.1.2	Règle bayésienne avec estimation non paramétrique	29
	Conclusion	34
	Bibliographie	35
	Annexe A : Logiciel R	37
2.2	Qu'est-ce-que le langage R?	37
	Annexe B : Abréviations et Notations	38

Table des figures

1.1 Représentation des nuages des points.	6
1.2 Décomposition de corrélation.	9
1.3 Nuages concentriques.	13
1.4 Projection de deux groupes sur un axe discriminant.	14
1.5 Classes séparées.	14
1.6 Comparaison entre ACP et AFD.	15
1.7 Insectes : premier plan factoriel de l'ACP.	16
1.8 Insectes : premier plan factoriel de l'AFD.	16
1.9 affectation d'une observation à une classe	19

Liste des tableaux

1.1 Tableau des observations	18
--	----

Introduction

L'*analyse discriminante* a été développée pour la première fois par **Ronald A. Fisher** dans l'année 1930. Fisher, un statisticien britannique, a créé cette méthode dans le cadre de ses travaux sur la classification des plantes en fonction de leurs caractéristiques. Depuis lors, l'analyse discriminante a été largement utilisée dans divers domaines, notamment :

- **En biologie** : peut-être utiliser pour classer différentes espèces d'animaux ou de plantes en fonction de leurs caractéristiques morphologiques, génétiques ou physiques.
- **En psychologie** : peut-être utiliser pour identifier les traits ou les comportements qui distinguent différents groupes de personnes, tels que les individus atteints de différents troubles mentaux.
- **En apprentissage automatique et intelligence artificielle** : est utilisée pour la classification et la prédiction dans un large éventail d'applications, telles que la reconnaissance de motifs dans les données biomédicales, la détection de fraudes dans les transactions financières, etc.
- **En médecine** : par exemple pour détecter les groupes à haut risque cardiaque à partir de caractéristiques telles que l'alimentation, le fait de fumer ou pas, les antécédents familiaux...etc.

Analyse discriminante (AD) est une technique statistique qui vise à décrire, expliquer et prédire l'appartenance à des groupes prédéfinis d'un ensemble d'observations à

partir d'une série de variables prédictives. L'objectif d'AD est de prédire une variable qualitative à K classes à l'aide de p prédicteurs quantitatifs.

Mon mémoire est divisé en deux chapitres :

Dans le **premier chapitre** de titre : "**Méthodes géométriques**", on va aborder quelques notions de base nécessaires pour l'AD, puis nous explorerons les fonctions, axes et variables discriminantes. De plus, on va présenter l'analyse factorielle discriminante (AFD) et l'analyse en composantes principales (ACP), en soulignant que l'AFD peut être considérée comme ACP sous certaines conditions. Enfin, on va explorer des règles géométriques telles que les distances euclidienne et de Mahalanobis, ainsi que la règle géométrique d'affectation de Mahalanobis-Fisher.

Dans le **deuxième chapitre** de titre : "**Méthodes probabilistes**", on va commencer par aborder la règle bayésienne, en expliquant d'abord le modèle gaussien unidimensionnel, puis multidimensionnel. Dans ce contexte, nous traiterons de la règle bayésienne appliquée au modèle normal et de l'estimation des paramètres. Par la suite, nous examinerons la règle bayésienne avec estimation non paramétriques, en mettant en lumière la méthode du K plus proches voisins (K ppv).

Chapitre 1

Méthode géométrique

La **méthode géométrique** est essentiellement descriptive, elle ne repose que sur des notions de distances, les représentations graphiques des individus discriminant "au mieux" les K classes engendrées par la variable X .

L'idée de base est très simple. Elle consiste à classifier des observations en fonction de leurs distances aux centres de gravité des groupes prédéfinis. En calculant les distances entre une nouvelle observation et les centres de chaque groupe, l'observation est assignée au groupe dont le centre est le plus proche. Pour cela, on va mentionner dans ce chapitre les notions de base nécessaires pour l'analyse discriminante et des distances les plus utilisés et puis des règles d'affectation géométrique.

1.1 Données Caractéristiques

Nous allons mentionner les notions de base nécessaires comme les formes matricielles, et on va exposer le principe de l'analyse discriminante.

1.1.1 Matrice des observations

On rassemble les observations des p variables quantitatives sur n individus dans une matrice notée \mathcal{X} , à n lignes et p colonnes définie comme suit

$$\mathcal{X} = (x_{ij})_{n \times p} = \begin{pmatrix} x_{11} & \cdots & x_{1j} & \cdots & x_{1p} \\ & \ddots & & & \\ x_{i1} & \cdots & x_{ij} & \cdots & x_{ip} \\ & & & \ddots & \\ x_{n1} & \cdots & x_{nj} & \cdots & x_{np} \end{pmatrix} \in \mathcal{M}_{(n \times p)},$$

où x_{ij} indique la valeur de la variable j qui est mesurée sur l'individu i , avec $i = \overline{1, n}$ et $j = \overline{1, p}$.

1- On note par X_j le vecteur variable à n dimensions, ce qui signifie qu'elle peut prendre n valeurs pour chaque individu, tel que

$$X_j = (x_{1j}, \dots, x_{nj})^t \in \mathbb{R}^n.$$

2- On note par X_i le vecteur individu à p dimensions, tel que

$$X_i = (x_{i1}, \dots, x_{ip})^t \in \mathbb{R}^p.$$

1.1.2 Centre de gravité

Le centre de gravité noté g est le vecteur dont la j -ème coordonnée g_j correspond à la moyenne arithmétique $\overline{X_j}$ de la variable j sur les n individus, tel que

$$g = (g_1, \dots, g_p)^t = (\overline{X_1}, \dots, \overline{X_p})^t,$$

où $\overline{X_j} = \frac{1}{n} \sum_{i=1}^n x_{ij}$.

1.1.3 Poids

Si les n individus sont affectés des poids $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n$, ces poids peuvent être regroupées dans une matrice diagonale de taille n définie comme

$$\mathcal{D}_p = \begin{pmatrix} \mathbf{p}_1 & 0 & & 0 \\ 0 & \mathbf{p}_2 & & \\ & & \ddots & \\ 0 & & 0 & \mathbf{p}_n \end{pmatrix},$$

où $\sum_{i=1}^n \mathbf{p}_i = 1$, avec $0 < \mathbf{p}_i < 1$.

Remarque 1.1.1 *En général, tous les individus ont le même poids, et donc $\mathbf{p}_i = \frac{1}{n}$.*

1.1.4 Nuage des points

Les n individus X_i de l'échantillon constituent un nuage E , de \mathbb{R}^p partagé en K sous-nuages E_1, E_2, \dots, E_K prédéfinie des centres de gravité g_1, g_2, \dots, g_K de matrices de variances $\mathcal{V}_1, \mathcal{V}_2, \dots, \mathcal{V}_K$. Soit g le centre de gravité et \mathcal{V} la matrice de variance du nuage E .

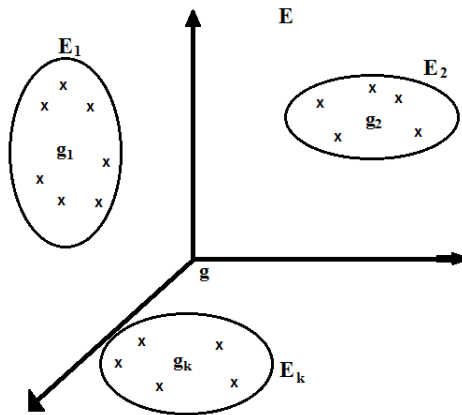


FIG. 1.1 – Représentation des nuages des points.

Si les n individus X_i sont affectés des poids $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n$, les poids $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_K$ de chaque sous-nuage sont alors

$$\mathbf{q}_k = \sum_{X_i \in E_K} \mathbf{p}_i.$$

Le centre de gravité de chaque sous-nuage est définie comme suit

$$g_k = \frac{1}{\mathbf{q}_k} \sum_{i=1}^n \mathbf{p}_i X_i, X_i \in E_K.$$

Le centre de gravité totale est définie comme suit

$$g = \sum_{k=1}^K \mathbf{q}_k g_k.$$

1.1.5 Matrice de variance-covariance

La matrice de variance-covariance notée \mathcal{V} est une matrice carrée symétrique de dimension $p \times p$ définie comme suit

$$\mathcal{V} = \mathbf{p}\mathcal{X}^t\mathcal{X} \in \mathcal{M}_{(p \times p)}.$$

Comme $\mathbf{p} = \frac{1}{n}$, alors

$$\mathcal{V} = \frac{1}{n}\mathcal{X}^t\mathcal{X} \in \mathcal{M}_{(p \times p)}.$$

On peut écrire

$$\begin{aligned} \mathcal{V} &= \begin{pmatrix} \text{var}(X_1) & \text{cov}(X_1, X_2) & \cdots & \text{cov}(X_1, X_p) \\ \text{cov}(X_2, X_1) & \text{var}(X_2) & \cdots & \text{cov}(X_2, X_p) \\ & & \ddots & \\ \text{cov}(X_p, X_1) & \text{cov}(X_p, X_2) & \cdots & \text{var}(X_p) \end{pmatrix} \\ &= \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{pmatrix}. \end{aligned}$$

Sachant que

- La covariance entre X_L et X_k notée $\text{cov}(X_L, X_k)$ ou σ_{Lk} est définie comme suit

$$\text{cov}(X_L, X_k) = \sum_{i=1}^n \mathbf{p}_i (x_{iL} - \bar{x}_L)(x_{ik} - \bar{x}_k).$$

- La variance de X_k notée $\text{var}(X_k)$ ou σ_k^2 est définie comme suit

$$\text{var}(X_k) = \text{cov}(X_k, X_k) = \sum_{i=1}^n \mathbf{p}_i (x_{ik} - \bar{x}_k)^2.$$

On analyse discriminante, il existe trois types de variance-covariance qui sont

Matrice de variance-covariance de la classe E_k :

La matrice de variance de la classe E_k , notée \mathcal{V}_k est définie ci-dessous

$$\mathcal{V}_k = \frac{1}{\mathbf{q}_k} \sum_{X_j \in E_k} \mathbf{p}_i (X_i - g_k)(X_i - g_k)^t.$$

Matrice de variance-covariance inter-classe :

La matrice de variance inter-classe, notée \mathcal{B} des k centres de gravité affectés des poids q_k définie ci-dessous

$$\mathcal{B} = \sum_{k=1}^K \mathbf{q}_k (g_k - g)(g_k - g)^t.$$

Matrice de variance-covariance intra-classe :

La matrice de variance intra-classe, notée \mathcal{W} est la moyenne des matrices \mathcal{V}_k définie comme ci-dessous

$$\mathcal{W} = \sum_{k=1}^K \mathbf{q}_k \mathcal{V}_k.$$

- En générale, \mathcal{W} est inversible et \mathcal{B} est non inversible.
- La variance totale est la somme de la variance inter-classe et la variance intra-classe.

$$\boxed{\underbrace{\mathcal{V}}_{\text{Variance totale}} = \underbrace{\mathcal{W}}_{\text{Variance intra-classe}} + \underbrace{\mathcal{B}}_{\text{Variance inter-classe}}.} \quad (1.1)$$

1.1.6 Matrice de corrélation

On note par \mathcal{R} la matrice de corrélation, c'est une matrice carrée symétrique de dimension $p \times p$ définie comme suit

$$\mathcal{R} = \mathbf{p}\mathcal{Z}^t\mathcal{Z} \in \mathcal{M}_{(p \times p)}.$$

Comme $\mathbf{p} = \frac{1}{n}$, alors

$$\mathcal{R} = \frac{1}{n}\mathcal{Z}^t\mathcal{Z} \in \mathcal{M}_{(p \times p)},$$

sachant que \mathcal{Z} est une matrice centrée réduite de la matrice \mathcal{X} .

Les termes en diagonal de cette matrice sont égaux à 1, on écrit alors

$$\mathcal{R} = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{12} & 1 & \cdots & r_{2p} \\ \vdots & & \ddots & \vdots \\ r_{1p} & r_{2p} & \cdots & 1 \end{pmatrix},$$

où $r_{Lf} = \text{cor}(X_L, X_f) = \frac{\text{cov}(X_L, X_f)}{\sigma_L\sigma_f}$.

Remarque 1.1.2 *On a déjà annoncé trois types de matrices de variances-covariances dans l'analyse discriminante, et donc il existe trois types de corrélations qui sont : la matrice de corrélation totale, la matrice de corrélation inter-classe et la matrice de corrélation intra-classe.*



FIG. 1.2 – Décomposition de corrélation.

1.1.7 Tableau des données

On considère que le tableau des données à étudier se met sous la forme suivante

$$\begin{array}{c}
 1 \\
 \vdots \\
 i \\
 \vdots \\
 n
 \end{array}
 \left(
 \begin{array}{cccc|cccc}
 1 & \cdots & j & \cdots & p & 1 & \cdots & k & \cdots & K \\
 x_{11} & & x_{1j} & & x_{1p} & 1 & & & & 0 \\
 \vdots & \ddots & & & & & 1 & & & \\
 x_{i1} & & x_{ij} & & x_{ip} & \vdots & & \ddots & & \\
 \vdots & & & & & & & & & \\
 x_{n1} & & x_{nj} & & x_{np} & 0 & & & & 1
 \end{array}
 \right)$$

$\underbrace{\hspace{10em}}_{\mathcal{X}} \quad \underbrace{\hspace{10em}}_{\mathcal{F}}$

1.1.8 Fonctions, axes et variables discriminantes

Pour faire l'étude de l'analyse discriminante sur nos observations, il faut chercher de nouvelles variables, qui seront appelées variables discriminantes correspondant à des directions de \mathbb{R}^p qui séparent le mieux possible en projection les K groupes d'observations.

Puisqu'on cherche une combinaison linéaire qui permet de séparer au mieux les différents groupes, il faut maximiser les variances inter-classe et minimiser les variances intra-classe.

On note par S la variable définie comme un vecteur de \mathbb{R}^p qui est une combinaison linéaire des vecteurs X_1, \dots, X_p définie comme suit

$$\boxed{S = Xu = u_1X_1 + u_2X_2 + \cdots + u_pX_p,}$$

où $u = (u_1, u_2, \dots, u_p)^t \in \mathbb{R}^p$ est un facteur discriminant.

Définition 1.1.1 On muni \mathbb{R}^p d'une métrique M et on projette les n points qui forment les lignes de X sur un axe Δ_u . On effectue des projections M -orthogonales sachant que l'axe Δ_u est M -normé de 1. La liste des coordonnées $S_i = X_i^t M \Delta_u$ des individus sur l'axe Δ_u forme la nouvelle variable S définie comme ci-dessous

$$S = (s_1, s_2, \dots, s_n)^t = XM\Delta_u = Xu.$$

Remarque 1.1.3

1. Si on dit un axe factoriel discriminant cela signifie une fonction discriminante.
2. Δ_u est un axe discriminant avec $\Delta_u^t M \Delta_u = 1$ où $\Delta_u \in \mathbb{R}^p$.
3. Le facteur discriminant u associé à la métrique M et à l'axe discriminant Δ_u avec $u^t M^{-1} u = 1$.
4. Xu est une variable discriminante.

On a déjà dit pour une mieux séparation des groupes, il faut maximiser les variances inter-classe (maximiser l'inertie inter-classe) et minimiser les variances intra-classe (minimiser l'inertie intra-classe).

D'après [1.1](#), on a

$$MVM = MWM + MBM.$$

On projette sur l'axe Δ_u , on obtient

$$\underbrace{\Delta_u^t MVM \Delta_u}_{\text{inertie totale}} = \underbrace{\Delta_u^t MWM \Delta_u}_{\text{inertie intra-classe}} + \underbrace{\Delta_u^t MBM \Delta_u}_{\text{inertie inter-classe}}.$$

Afin de choisir le meilleur axe, on doit maximiser l'inertie inter-classe sur l'inertie intra-classe, i.e.

$$\max_{\Delta_u \in \mathbb{R}^p} \frac{\Delta_u^t MBM \Delta_u}{\Delta_u^t MVM \Delta_u}.$$

Si Δ_u est un vecteur propre de $(M\mathcal{V}M)^{-1}(M\mathcal{B}M)$ associé à sa plus grande valeur propre λ (le pouvoir discriminant), alors

$$M^{-1}\mathcal{V}^{-1}\mathcal{B}M\Delta_u = \lambda\Delta_u. \quad (1.2)$$

Remarque 1.1.4 *L'axe discriminant Δ_u est alors associé au facteur discriminant u , tel que*

$$u = M\Delta_u. \quad (1.3)$$

D'après [1.2](#) et [1.3](#), on a $M^{-1}\mathcal{V}^{-1}\mathcal{B}u = \lambda M^{-1}u = M^{-1}\lambda u$.

Donc

$$\mathcal{V}^{-1}\mathcal{B}u = \lambda u. \quad (1.4)$$

Proposition 1.1.1 *"Les anglo-saxons" utilisent souvent comme critère pour mesurer la qualité de la discrimination d'une variable discriminantes i.e : $\max_{\Delta_u \in \mathbb{R}^p} \frac{\Delta_u^t M\mathcal{B}M\Delta_u}{\Delta_u^t M\mathcal{W}M\Delta_u}$. La solution est le vecteur propre u de $\mathcal{W}^{-1}\mathcal{B}$ associé à la plus grande valeur propre*

$$\eta = \frac{\lambda}{1 - \lambda}.$$

Ce critère varie entre zero et l'infini.

Proposition 1.1.2 *Les matrices $\mathcal{V}^{-1}\mathcal{B}$ et $\mathcal{W}^{-1}\mathcal{B}$ ont les même vecteurs propres u mais ce dernier a pour la valeur propre η .*

Preuve. En effet d'après 1.4, on a $\mathcal{V}^{-1}\mathcal{B}u = \lambda u \Leftrightarrow \mathcal{B}u = \lambda\mathcal{V}u$.

Si on remplace \mathcal{V} par $\mathcal{W} + \mathcal{B}$, on obtient

$$\begin{aligned} \mathcal{B}u &= \lambda(\mathcal{W} + \mathcal{B})u \\ &= \lambda\mathcal{W}u + \lambda\mathcal{B}u. \end{aligned} \quad \text{et}$$

$$\mathcal{B}u - \lambda\mathcal{B}u = \lambda\mathcal{W}u.$$

$$(1 - \lambda)\mathcal{B}u = \lambda\mathcal{W}u.$$

$$\mathcal{B}u = \lambda\mathcal{W}u / (1 - \lambda).$$

Donc

$$\mathcal{W}^{-1}\mathcal{B}u = \frac{\lambda}{1 - \lambda}u.$$

Puisque $\eta = \frac{\lambda}{1 - \lambda}$, alors

$$\mathcal{W}^{-1}\mathcal{B}u = \eta u.$$

■

Cas particuliers selon λ :

Puisque λ est le quantile à maximiser, alors $0 \leq \lambda \leq 1$.

- Pour $\lambda = 0$: Le meilleur axe ne permet pas de séparer les centres de gravité g_k , c'est le cas où ces centres de gravité sont confondus.

Donc, les nuages sont concentriques et il est impossible de faire une séparation linéaire. cependant, il existe une possibilité de discrimination non linéaire.

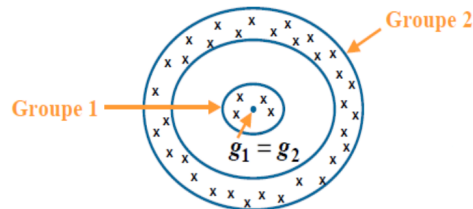


FIG. 1.3 – Nuages concentriques.

- Pour $\lambda = 1$: La dispersion intra-classe est nulle pour les projections sur l'axe Δ_u . Si les centres de gravité se projettent en des points différents, alors il y a une discrimination parfaite comme dans

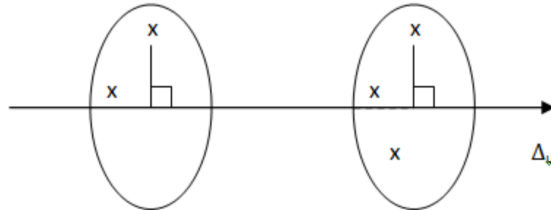


FIG. 1.4 – Projection de deux groupes sur un axe discriminant.

- Pour $0 < \lambda < 1$: La valeur propre λ est une mesure pessimiste du pouvoir discriminant d'un axe, car on peut avoir un cas où les classes sont parfaitement séparées et pourtant on a $\lambda < 1$, comme on le voit dans

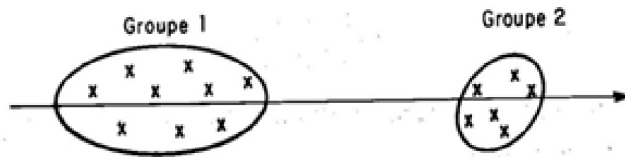


FIG. 1.5 – Classes séparées.

Remarque 1.1.5 *Si le nombre des valeurs propres est non nulles, donc l'axe discriminant est égal à $K - 1$, dans le cas habituel où $n > p > K$ et où les variables ne sont pas liées par des relations linéaires.*

1.1.9 Analyse factorielle discriminante et analyse en composantes principales

L'analyse factorielle discriminante notée (AFD) est une analyse en composantes principales notée (ACP) du nuage de K centre de gravité avec la métrique \mathcal{V}^{-1} .

1. Si $K > 2$; on peut utiliser les représentations graphiques de l'ACP (cercle de corrélation, ect).
2. L'ACP maximise la variance des projections sur le sous-espace, par contre L'AFD maximise la différentiation entre les classes dans le sous-espace.

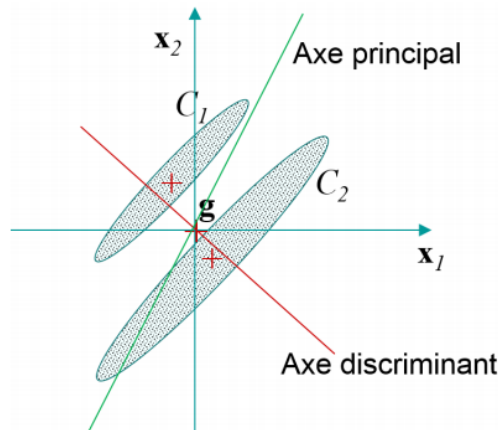


FIG. 1.6 – Comparaison entre ACP et AFD.

Exemple 1.1.1 *Les insectes de Lubitsc :*

Cette méthode est illustrée par une comparaison des sorties graphiques issues d'une ACP et d'une AFD. Les données décrivent 3 classes d'insectes (a, b et c) sur lesquels ont été réalisées 6 mesures anatomiques. On cherche à savoir si ces mesures permettent de retrouver la typologie de ces insectes. Ce jeu de données "scolaire" conduit à une bien meilleure discrimination que ce que l'on peut obtenir dans une situation concrète. [7]

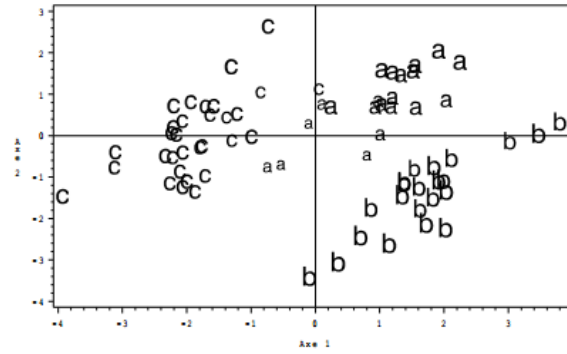


FIG. 1.7 – Insectes : premier plan factoriel de l'ACP.

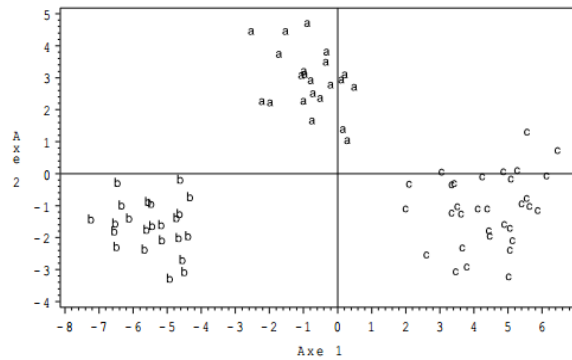


FIG. 1.8 – Insectes : premier plan factoriel de l'AFD.

****Commentaire :***

On note qu'en ACP, il y a un certain chevauchement entre la classe a et la classe c.

Alors qu'en AFD, les classes sont éloignées et distinctes les unes des autres.

1.2 Règles géométriques

La **règle géométrique** consiste à calculer les distances de la nouvelle observation à chacun des K centres de gravité et à affecter cette observation au groupe le plus proche. Pour cela, il faut préciser la métrique à utiliser dans le calcul des distances. La règle la plus utilisée est celle de **MahalanobisFisher** qui consiste à prendre la métrique \mathcal{W}^{-1} (ou bien \mathcal{V}^{-1} ce qui est équivalent).

1.2.1 Distances

Il existe plusieurs distances ou méthodes utilisées pour la classification comme la distance euclidienne, distance de mahalanobis, distance de Minkowski, ect.

On s'intéresse sur les deux types de distances utilisés à savoir la distance euclidienne et la distance de Mahalanobis.

1.2.1.1 Distance euclidienne

C'est une mesure de la ressemblance, c'est probablement le type de distance le plus couramment utilisé. Elle est donnée par la formule ci-dessous

$$\forall i_1, i_2 = 1, \dots, n : d(X_{i_1}, X_{i_2}) = \sqrt{\sum_{k=1}^K (x_{i_1k} - x_{i_2k})^2}.$$

1.2.1.2 Distance de Mahalanobis

La formule est donnée par

$$\forall i_1, i_2 = 1, \dots, n : d(X_{i_1}, X_{i_2}) = \sqrt{(x_{i_1} - x_{i_2})^t \mathcal{V}^{-1} (x_{i_1} - x_{i_2})},$$

où \mathcal{V}^{-1} est l'inverse de la matrice de variance-covariance.

On peut aussi utiliser la matrice de variance intra-classe \mathcal{W}^{-1} au lieu de \mathcal{V}^{-1} .

Exemple 1.2.1 Soit le tableau suivant des observations (notre distribution) :

A	1	2	4	2	5
B	100	300	200	600	100
C	10	15	20	10	30

TAB. 1.1 – Tableau des observations

calculer la distance entre X et notre distribution tel que $X = (4, 500, 40)^t$:

$$\mathcal{X} = \begin{pmatrix} 1 & 2 & 4 & 2 & 5 \\ 100 & 300 & 200 & 600 & 100 \\ 10 & 15 & 20 & 10 & 30 \end{pmatrix},$$

– Pour la distance "euclidienne" :

$$\text{La moyenne de } \mathcal{X} : m = \begin{bmatrix} 2.8 \\ 260 \\ 17 \end{bmatrix}.$$

$$\begin{aligned} d(X, m) &= \sqrt{\sum_{i=1}^3 (x_i - m_i)^2} \\ &= \sqrt{(4 - 2.8)^2 + (500 - 260)^2 + (40 - 17)^2} \\ &= \sqrt{1.44 + 57600 + 529} \\ &= \sqrt{58130.44} \\ &= 241.103 \end{aligned}$$

– Pour la distance "Mahalanobis" :

$$\text{La moyenne de } \mathcal{X} : m = (2.8, 260, 17)^t.$$

$$d^2(X, m) = (x_i - m_i)^t \mathcal{V}^{-1} (x_i - m_i)$$

$$* (x_i - m_i) = \begin{bmatrix} 1.2 \\ 240 \\ 23 \end{bmatrix}, (x_i - m_i)^t = \begin{bmatrix} 1.2 & 240 & 23 \end{bmatrix}$$

on va calculer V^{-1} :

$$\mathcal{V} = \begin{pmatrix} 2.7 & -110 & 13 \\ -110 & 4300 & -900 \\ 13 & -900 & 70 \end{pmatrix}, \mathcal{V}^{-1} = \begin{pmatrix} 5.5 & -0.01 & -1.15 \\ -0.01 & 0.00005 & 0.0025 \\ -1.15 & 0.0025 & 5.2 \end{pmatrix}$$

Alors,

$$d^2(X, m) = \begin{bmatrix} 1.2 & 240 & 23 \end{bmatrix} \begin{pmatrix} 5.5 & -0.01 & -1.15 \\ -0.01 & 0.00005 & 0.0025 \\ -1.15 & 0.0025 & 5.2 \end{pmatrix} \begin{bmatrix} 1.2 \\ 240 \\ 23 \end{bmatrix}$$

$$= 106.7$$

$$d(X, m) = \sqrt{106.7}$$

$$= 10.33$$

***Commentaire :**

On note que la distance "euclidienne" est liée aux valeurs des observation ; plus les valeurs sont grands, plus la distance est grande ; i.e. elle est affectée par l'unité de mesure utilisée pour mesurer les variables. Ainsi, elle ne prend pas en sa compte la relation entre les variables, qu'elles soient fortes ou faibles. Les valeurs du carré de distance euclidienne ne suivent aucune distribution statistique et la signification de cette valeur ne peut donc pas être jugée.

Mais, la distance de "Mahalanobis" est venue éliminer ou réduire tous ces défauts afin de donner des résultats plus précises.

Exemple 1.2.2 affectation d'une observation à une classe

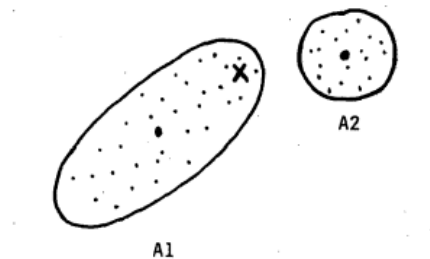


FIG. 1.9 – affectation d'une observation à une classe

Avec la distance de Mahalanobis, x sera affecté à sa classe naturelle ($A1$), avec la distance euclidienne il sera affecté à $A2$. [2]

1.2.2 Règle géométrique d'affectation

Une règle simple et géométrique d'affectation est de choisir la classe dont le centre de gravité est le plus proche du point-individu. La métrique généralement utilisée dans les applications les plus courantes est celle de **Mahalanobis-Fisher** qui consiste à prendre la métrique \mathcal{W}^{-1} (ou bien \mathcal{V}^{-1}).

Règle de Mahalanobis-Fisher

Soit $b = (b_1, \dots, b_p)$ le nouvel individu à classer.

On va calculer la distance pour la métrique \mathcal{W}^{-1} entre b et le centre de gravité g_k comme suit

$$d^2(b, g_k) = (b - g_k)^t \mathcal{W}^{-1} (b - g_k).$$

On obtient

$$d^2(b, g_k) = b^t \mathcal{W}^{-1} b + g_k^t \mathcal{W}^{-1} g_k - 2b^t \mathcal{W}^{-1} g_k.$$

Elles sont des fonctions quadratiques de b , mais elles ont toutes en commun le terme carré $b^t \mathcal{W}^{-1} b$ qui ne dépend pas de k , on pourra donc comparer les fonctions linéaires discriminantes f_{E_k} de b , définies ci-dessous, relatives à chaque groupe E_k :

$$f_{E_k}(b) = d^2(b, g_k) - b^t \mathcal{W}^{-1} b = g_k^t \mathcal{W}^{-1} g_k - 2b^t \mathcal{W}^{-1} g_k.$$

On décide d'affecter b à la classe E telle que

$$f_E(b) = \min_{k=1, K} f_{E_k}(b).$$

Remarque 1.2.1 *Cette méthode n'est pas très efficace car elle donne des affectations incorrectes lorsque les dispersions des groupes sont très différentes entre elles, ou bien, il n'y a pas d'un garanti d'utilisation la même métrique pour les différentes groupes.*

Chapitre 2

Méthode Probabiliste

La **méthode probabiliste** dans l'analyse discriminante repose sur la modélisation des distributions probabilistes des classes à partir des données disponibles.

L'idée de base est de classer une nouvelle observation dans le groupe pour lequel la probabilité conditionnelle d'appartenance à ce groupe, étant donné les valeurs observées, est maximale. En utilisant des concepts statistiques comme la distribution normale multivariée, cette méthode estime la probabilité qu'un individu appartienne à une classe donnée. Ces probabilités ne peuvent être calculées que si les observations suivent une loi multi-normale ; sinon, les données doivent être transformées pour s'en rapprocher.

Dans ce chapitre, nous définirons et étudierons des règles de décision probabilistes.

2.1 Règle bayésienne

L'objectif est de trouver une règle d'affectation qui permet de prédire, pour un individu b donné, sa valeur associée de Y à partir des valeurs prises par les p variables quantitatives X .

On suppose maintenant que l'échantillon d'apprentissage est issu d'une population en K groupes, Y est une variable aléatoire (v.a.) qui prend ses valeurs dans $\{1, \dots, K\}$ et $X = (X_1, \dots, X_p)^t$ est un vecteur de variable aléatoire réelle (v.a.r). Soit $(P_1, \dots, P_k, \dots, P_K)^t$ le vecteur des distributions de Y , où $P_k = \mathbb{P}(Y = k)$ est la proportion théorique de E_k (probabilité a priori de E_k), $f_k : \mathbb{R}^p \rightarrow [0, 1]$ la densité de probabilité du vecteur X dans le groupe E_k (X est absolument continue), $f_k(b) = \mathbb{P}(b | E_k)$ la probabilité de l'événement $X = b$ dans le groupe E_k (X est discret).

La densité de X est comme suit

$$X \sim \sum_{k=1}^K P_k f_k(b).$$

La probabilité que b provienne du groupe E_k est donnée par la formule de Bayes

$$\mathbb{P}(E_k | b) = \frac{P_k f_k(b)}{\sum_{k=1}^K P_k f_k(b)}, \quad (2.1)$$

c'est la probabilité a posteriori de E_k .

Remarque 2.1.1 La formule [2.1](#) est la base de la règle de décision bayésienne.

Proposition 2.1.1 *On affecte une nouvelle observation b dans le groupe E_{k^*} ; on obtient*

$$E_{k^*} = \arg \max_{k=1, \dots, K} \mathbb{P}(E_k | b),$$

elle peut s'écrire comme suit

$$E_{k^*} = \arg \max_{k=1, \dots, K} P_k f_k(b). \quad (2.2)$$

1. La règle bayésienne consiste à affecter b à la classe pour laquelle P_k est maximal.
2. Si P_k sont égales, la formule de la règle de décision bayésienne [2.2](#) devient

$$E_{k^*} = \arg \max_{k=1, \dots, K} f_k(b).$$

2.1.1 Modèle gaussien

La règle de décision la plus utilisée en pratique est basée sur le modèle gaussien.

2.1.1.1 Modèle gaussien unidimensionnel

Soit X une v.a. d'espérance mathématique μ et de variance σ^2 , sa densité est

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right\},$$

où $X \sim \mathcal{N}(\mu, \sigma^2)$.

Si on fait un changement $U = \frac{X - \mu}{\sigma}$, on trouve que la densité de la v.a. U est définie comme suit

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

où $U \sim \mathcal{N}(0, 1)$.

Donc, toute v.a. X se ramène simplement à la variable U par $X = \mu + \sigma U$.

2.1.1.2 Modèle gaussien multidimensionnel

Soit $X = (X_1, \dots, X_p)^t$ avec $p > 1$ un vecteur gaussien, de distribution multinormal d'espérance $\mu = (\mu_1, \dots, \mu_p)^t$ et de matrice de variance ξ de dimension $(p \times p)$, sa densité s'écrit comme

$$f(x) = \frac{1}{(2\pi)^{p/2} (\det \xi)^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu)^t \xi^{-1} (x - \mu) \right\},$$

avec $X \sim \mathcal{N}_p(\mu, \xi)$.

- Règle bayésienne avec modèle normal

Soit $X \sim \mathcal{N}(\mu_k, \xi_k)$ dans chaque groupe E_k

$$f_k(x) = \frac{1}{(2\pi)^{p/2} (\det \xi_k)^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu_k)^t \xi_k^{-1} (x - \mu_k) \right\},$$

Dans ce cas, la règle de Bayes devient

$$E_{k^*} = \arg \min_{k=1, \dots, K} \mathfrak{h}_k^2(b),$$

où \mathfrak{h}^2 est le carré de la distance de mahalanobis théorique généralisée; tel que

$$\mathfrak{h}_k^2(b) = (b - \mu_k)^t \xi_k^{-1} (b - \mu_k) - 2 \ln P_k + \ln (\det \xi_k). \quad (2.3)$$

Preuve. maximiser $P_k f_k(x)$ est équivalent à maximiser $\ln(P_k f_k(x))$; car la fonction \ln est croissante.

$$\begin{aligned} \ln(P_k f_k(b)) &= \ln P_k + \ln f_k(b) \\ &= \ln P_k + \ln \left(\frac{1}{(2\pi)^{p/2} (\det \xi_k)^{1/2}} \exp \left\{ -\frac{1}{2} (b - \mu_k)^t \xi_k^{-1} (b - \mu_k) \right\} \right) \\ &= \ln P_k - \frac{p}{2} \ln 2\pi - \frac{1}{2} [\ln(\det \xi_k) + (b - \mu_k)^t \xi_k^{-1} (b - \mu_k)]. \end{aligned}$$

Comme $-\frac{p}{2} \ln 2\pi$ est indépendant de k , alors maximiser $\ln(P_k f_k(b))$ est équivalent à minimiser

$$-2 \left(\ln P_k - \frac{1}{2} [\ln(\det \xi_k) + (b - \mu_k)^t \xi_k^{-1} (b - \mu_k)] \right) = \mathfrak{h}_k^2(b).$$

■

- Estimation des paramètres

La règle bayésienne consiste à trouver une estimation de la probabilité a posteriori d'affectation car en pratique, on ne connaît ni P_k , ni les ξ_k et ni les μ_k .

A partir de l'échantillon d'apprentissage, on veut estimer le paramètre θ où

$$\theta = (P_1, \dots, P_k, \mu_1, \dots, \mu_k, \xi_1, \dots, \xi_k).$$

on cherche à l'estimer en utilisant la méthode du maximum de vraisemblance (M.V) qui s'écrit comme suit

$$\mathcal{L}(\theta) = \prod_{i=1}^n f_X(b_i) = \prod_{k=1}^K \prod_{i \in E_k} P_k f_k(b_i).$$

La fonction log-vraisemblance s'écrit de la manière suivante

$$\ln(\mathcal{L}(\theta)) = \sum_{k=1}^K \sum_{b_i \in E_k} \left(\ln P_k - \frac{p}{2} \ln 2\pi - \frac{1}{2} [\ln(\det \xi_k) + (b - \mu_k)^t \xi_k^{-1} (b - \mu_k)] \right)$$

on obtient alors les estimateurs du (M.V) suivants

$$\begin{aligned} \widehat{P}_k &= \frac{n_k}{n} \\ \widehat{\mu}_k &= \frac{1}{n_k} \sum_{b_i \in E_k} b_i \\ \widehat{\xi}_k &= \begin{cases} \widehat{\xi} = \frac{1}{n} \sum_{k=1}^K \sum_{b_i \in E_k} (b_i - \mu_k)^t (b_i - \mu_k), \text{ cas homoscédastique.} \\ \widehat{\xi}_k = \frac{1}{n_k} \sum_{b_i \in E_k} (b_i - \mu_k)^t (b_i - \mu_k), \text{ cas hétéroscédastique.} \end{cases} \end{aligned}$$

Ces estimateurs de ξ_k étant biaisé, on a les estimateurs sans biais suivants :

$$\begin{cases} \widehat{\xi} = \frac{1}{n - K} \sum_{k=1}^K \sum_{b_i \in E_k} (b_i - \mu_k)^t (b_i - \mu_k). \\ \widehat{\xi}_k = \frac{1}{n_k - 1} \sum_{b_i \in E_k} (b_i - \mu_k)^t (b_i - \mu_k). \end{cases}$$

1. **Cas hétéroscédastique** (matrices des variances différentes) i.e. $\xi_k \neq \xi_{k^*}$ avec $k \neq k^*$.

On note

$$\begin{aligned} - \widehat{P}_k &= q_k = \frac{n_k}{n} \\ - \widehat{\mu}_k &= a_k = \frac{1}{n_k} \sum_{b_i \in E_k} b_i \\ - \widehat{\xi}_k &= V_k = \frac{1}{n_k} \sum_{b_i \in E_k} (b_i - a_k)^t (b_i - a_k), \end{aligned}$$

ou bien $\widehat{\xi}_k = V_k = \frac{1}{n_k - 1} \sum_{b_i \in E_k} (b_i - a_k)^t (b_i - a_k)$ la version sans biais.

Définition 2.1.1 *Analyse discriminante quadratique est basée sur la règle bayésienne en fonction des estimateurs dans le cas hétéroscédastique comme suit*

$$E_{k^*} = \arg \min_{k=1, \dots, K} \mathfrak{Q}_k(b),$$

où $\mathfrak{Q}_k(b) = (b - a_k)^t V_k^{-1} (b - a_k)^t - 2 \ln q_k - \ln(\det V_k)$ est appelée la fonction quadratique discriminante du groupe E_k .

Remarque 2.1.2 *Une nouvelle observation sera affectée au groupe pour lequel la fonction quadratique discriminante sera la plus petite.*

2. Cas homoscdastique (matrices des variances identiques)

i.e. $\xi_1 = \dots = \xi_K = \xi$.

La règle de Bayes s'écrit comme suit

$$E_{k^*} = \arg \max_{k=1, \dots, K} b^t \xi^{-1} \mu_k - \frac{1}{2} \mu_k^t \xi^{-1} \mu_k + \ln P_k.$$

Preuve. $\forall k = 1, \dots, K : \xi_k = \xi$,

d'après [2.3](#) on a

$$\mathfrak{h}_k^2(b) = b^t \xi^{-1} b - 2b^t \xi^{-1} \mu_k + \mu_k^t \xi^{-1} \mu_k - 2 \ln P_k + \ln(\det \xi),$$

où $b^t \xi^{-1} b$ et $\ln(\det(\xi))$ sont indépendants de k et ξ^{-1} est symétrique.

Donc, minimiser $\mathfrak{h}_k^2(b)$ équivalent à maximiser $-\frac{1}{2} (-2b^t \xi^{-1} \mu_k + \mu_k^t \xi^{-1} \mu_k - 2 \ln P_k)$.

■

Remarque 2.1.3 - La matrice de variances ξ commune des différents groupes est

estimée par
$$W = \frac{1}{n} \sum_{k=1}^K \sum_{b_i \in E_k} (b_i - a_k)^t (b_i - a_k)$$

ou par sa version sans biais
$$W = \frac{1}{n - K} \sum_{k=1}^K \sum_{b_i \in E_k} (b_i - a_k)^t (b_i - a_k).$$

- Lorsqu'on suppose que P_k sont égaux, la règle de l'analyse discriminante linéaire est équivalent à la règle de classement géométrique.

Définition 2.1.2 Analyse discriminante linéaire est basée sur la règle bayésienne en fonction des estimateurs dans le cas homoscédastique comme suit

$$E_{k^*} = \arg \max_{k=1, \dots, K} \mathfrak{L}_k(b),$$

où $\mathfrak{L}_k(b) = b^t W^{-1} a_k - \frac{1}{2} a_k^t W^{-1} a_k + 2 \ln(q_k)$ est appelée la fonction linéaire discriminante du groupe E_k .

2.1.2 Règle bayésienne avec estimation non paramétrique

En statistique, on utilise le terme "estimation non paramétrique" lorsqu'on doit estimer un nombre infini de paramètres. Au lieu de supposer qu'on a affaire à une densité de type connu dont on estime les paramètres, on cherche plutôt une estimation \hat{f} de la fonction de densité f .

Cette méthode très flexible présente l'avantage de ne pas exiger d'hypothèse spécifiques sur la loi. Cependant, elle n'est utilisable qu'avec des échantillons de grande taille, surtout lorsque le nombre de dimensions p est grand.

Dans le domaine de l'analyse discriminante, ces méthodes permettent une estimation directe des densités f_k . On examine ici la méthode des K plus proches voisins (ppv).

2.1.2.1 K plus proches voisins[\[9\]](#)

La méthode des K ppv est très utilisée en pratique. Elle est basée sur les considérations suivantes : La probabilité qu'une v.a. X de densité $f(x)$ prenne sa valeur dans une région $D(x)$ de volume D centrée sur un point b est donnée par

$$F = \int_{D(x)} f(u) du.$$

Lorsque $D(x)$ est petit, on peut écrire

$$F \simeq f(x) \times D.$$

Lorsqu'on a un échantillon indépendant et identiquement distribuées (i.i.d.) avec une densité $f(x)$, on peut estimer cette probabilité par la proportion d'observation appartenant à $D(x)$.

Si on note c le nombre d'observations (parmi n) appartenant à $D(x)$, alors on a

$$F \simeq \frac{c}{n} \simeq f(x) \times D.$$

par conséquent, on en déduit une estimation de la densité

$$\hat{f}(x) = \frac{c}{nD}.$$

La méthode des K ppv implique de définir $\frac{c}{n}$ et de calculer D de la plus petite hypersphère $D(x)$ centrée sur b contenant c observations.

Par exemple, si on note b_c le $c^{\text{ième}}$ ppv de b dans \mathbb{R}^p pour la norme euclidienne $\|\cdot\|$, alors on prend pour $D(x)$ l'hypersphère centrée sur b , de rayon $\|b - b_c\|$.

Le volume de cette hypersphère est donné par

$$D = 2 \|b - b_c\|^p \pi^{p/2} \Gamma\left(\frac{p}{2}\right).$$

□

Le rapport $\frac{c}{nD}$ donne l'estimation souhaitée.

Le paramètre c . Si celle-ci est trop élevée, l'estimation sera trop régulière et les détails seront effacés. En revanche, si elle est trop faible, l'estimation présentera de nombreux pics.

On peut montrer que pour que $\hat{f}(b)$ converge en probabilité vers $f(b)$ sachant que f est continue en b , il suffit que

$$\begin{cases} c(n) \rightarrow \infty & n \rightarrow \infty \\ \frac{c(n)}{n} \rightarrow 0 & n \rightarrow \infty \end{cases}$$

2.1.2.2 Application des K plus proches voisins à l'analyse discriminante [\[9\]](#)

Disposant d'une estimation de la densité, on peut l'utiliser pour construire une règle de décision.

On suppose que parmi les c ppv de b , c_k appartiennent à la classe E_k . On a donc

$$\sum_{k=1}^K c_k = c.$$

On note n_k le nombre d'observations issues de la classe E_k . On a donc

$$\sum_{k=1}^K n_k = n.$$

${}^1\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$
 $\Gamma(n) = (n-1)\Gamma(n-1) = (n-1)!$
 $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

On note D le volume de la plus petite hypersphère centrée sur b contenant c observations.

Alors, en combinant l'approche par ppv, on peut estimer la densité conditionnelle $\mathbb{P}(E_k | b)$ par

$$\widehat{\mathbb{P}}(E_k | b) = \frac{c_k}{n_k D}.$$

et la probabilité a priori P_k par

$$\widehat{P}_k = \frac{n_k}{n}.$$

La règle de décision consiste à affecter b à E_k si

$$\forall j : \widehat{\mathbb{P}}(E_k | b) \widehat{P}_k \geq \widehat{P}_j \widehat{\mathbb{P}}(E_j | b).$$

ce qui s'écrit, en appliquant le théorème de Bayes,

$$\forall j : \frac{c_k}{n_k D} \times \frac{n_k}{n} \geq \frac{c_j}{n_j D} \times \frac{n_j}{n}.$$

Ainsi, on affecte b à la classe E_k si

$$\forall j : c_k \geq c_j.$$

Autrement dit, on affecte b à la classe la plus représentée parmi ses c ppv.

En cas de présence d'ex-aequo, on procède en général de l'une des façons suivantes :

1. On choisit une classe arbitrairement parmi les ex-aequo.
2. On affecte b à la classe E_k , parmi les ex-aequo, dont la moyenne (calculée sur c_k observations) est la plus proche de b .
3. On affecte b à la classe la plus proche parmi les ex-aequo.

On peut démontrer que l'erreur de classification asymptotique ε de la règle du ppv ($k = 1$) vérifie

$$\varepsilon_B \leq \varepsilon \leq \varepsilon_B \left(2 - \frac{K\varepsilon_B}{K-1} \right),$$

où ε_B désigne l'erreur de la règle de Bayes, donc de la règle optimale.

Ainsi, pour de grands échantillons, l'erreur de classification par ppv est bornée supérieurement par deux fois l'erreur de la règle de Bayes.

Elle fournit également une borne inférieure pour celle-ci puisqu'on déduit de l'inégalité précédente que

$$\frac{K-1}{K} - \sqrt{\frac{K-1}{K}} \sqrt{\frac{K-1}{K} - \varepsilon} \leq \varepsilon_B \leq \varepsilon.$$

Conclusion

Ce mémoire donne une aide générale sur l'analyse discriminante.

En conclusion, ce mémoire a exploré deux approches principales de l'analyse discriminante : les méthodes géométriques et les méthodes probabilistes. Nous avons examiné les notions fondamentales, les techniques avancées et les différentes règles appliquées dans chaque approche. Ces méthodes offrent des perspectives distinctes et complémentaires pour l'analyse discriminante.

Les méthodes géométriques se concentrent sur la structure spatiale des données, en utilisant des distances et des règles d'affectation pour classifier les individus. En revanche, les méthodes probabilistes s'appuient sur des modèles statistiques et des probabilités pour estimer la distribution des données et effectuer des classifications basées sur la vraisemblance ou l'estimation non paramétrique.

Ainsi, le choix entre ces deux approches dépend des caractéristiques spécifiques des données et des objectifs de l'analyse. Les méthodes géométriques sont particulièrement efficaces lorsque les données sont bien séparées dans l'espace, tandis que les méthodes probabilistes sont mieux adaptées aux situations où il est important de modéliser la distribution sous-jacente des données.

Bibliographie

- [1] Baey, C. (2019). Analyse de données. *M2 Ingénierie Statistique et Numérique. Université de Lille.*
- [2] Bonneau, F., & Porth, J. M. (1985). *Analyse discriminante : Methode du type plus proches voisins utilisant un pretraitement des donnees* (Doctoral dissertation, INRIA).
- [3] Casin, P. (2015). L'analyse factorielle discriminante de tableaux multiples. *Journal de la Société Française de Statistique*, 156(4), 1-20.
- [4] Crucianu, M., (2021). Apprentissage statistique : modélisation descriptive et introduction aux réseaux de neurones (RCP208) Méthodes d'analyse factorielle. *Conservatoire National des Arts & Métiers, Paris, France.*
- [5] Décarie, Y. (2012). *Analyse en composantes principales et analyse discriminante fonctionnelles appliquées à des données de prises alimentaires animales.* Library and Archives Canada= Bibliothèque et Archives Canada, Ottawa.
- [6] Duby, C., & Robin, S. (2006). Analyse en composantes principales. *Institut National Agronomique, Paris-Grignon*, 80, 53.
- [7] Gadat, S. Analyse Factorielle Discriminante. *Laboratoire de Statistique et Probabilités, UMR 5583 CNRS-UPS*, 10-11.
- [8] GONZALEZ, P. L., Analyse discriminante. https://maths.cnam.fr/IMG/pdf/Analyse_discriminante_-_Methodes_geometriques_cle8b864f.pdf

- [9] Lardjane, S., (2022). Analyse discriminante et Régression logistique. *Université Bretagne Sud*, 22-24
- [10] Necir, A. (2022.a). Analyse en Composantes Principales (Modèle linéaire), Cours de 1^{ère} Année Master, Université de Mohamed Khider Biskra.
- [11] Piron, M., Lebart, L., & Morineau, A. (1995). Statistique exploratoire multidimensionnelle. *Paris*.
- [12] Saporta, G. (1997). Méthodes statistiques de discrimination. *Statistique et méthodes neuronales*, 20-30.
- [13] Saporta, G. (2006). *Probabilités Analyse des données et Statistique*, 2^{ème} édition, Editions technip.
- [14] Saporta, G. (2012). Analyse discriminante, classification supervisée, scoring. *Conservatoire National des Arts et Métiers*. <http://cedric.cnam.fr/~saporta>.

Annexe A : Logiciel R

2.2 Qu'est-ce-que le langage R ?

- Le langage R est un langage de programmation et un environnement mathématique utilisés pour le traitement de données. Il permet de faire des analyses statistiques aussi bien simples que complexes comme des modèles linéaires ou non-linéaires, des tests d'hypothèse, de la modélisation de séries chronologiques, de la classification, etc. Il dispose également de nombreuses fonctions graphiques très utiles et de qualité professionnelle.

- R a été créé par Ross Ihaka et Robert Gentleman en 1993 à l'Université d'Auckland, Nouvelle Zélande, et est maintenant développé par la R Development Core Team. L'origine du nom du langage provient, d'une part, des initiales des prénoms des deux auteurs (Ross Ihaka et Robert Gentleman) et, d'autre part, d'un jeu de mots sur le nom du langage S auquel il est apparenté.

Annexe B : Abréviations et Notations

Les différentes abréviations et notations utilisées tout au long de ce mémoire sont expliquées ci-dessous :

- n : Nombre d'individu.
- p : Nombre de variable quantitative.
- \mathcal{X} : Matrice des observations.
- x_{ij} : Valeur de la variable j mesurée sur l'individu i (une observation).
- $\mathcal{M}_{(\cdot \times \cdot)}$: Espace des matrices de dimension $(\cdot \times \cdot)$.
- X_j : Vecteur variable de \mathbb{R}^p .
- X_i : Vecteur individu de \mathbb{R}^n .
- \mathbb{R}^\cdot : Espace réel de dimension \cdot .
- g : Centre de gravité totale.
- \overline{X}_j : Moyenne arithmétique des variables.
- \mathbf{p} : Poids des individus.
- D_p : Matrice des poids.
- E : Nuage de points.
- K : Nombre des classes (sous-nuages).
- E_k : Sous-nuages.
- g_k : Centre de gravité des sous-nuages.

\mathcal{V}	Matrice de variance totale.
\mathcal{V}_k	Matrice de variance de chaque classe (sous-nuage).
\mathbf{q}	Poids de chaque classe (sous-nuage).
\mathcal{X}^t	Matrice transposé de la matrice des observations.
$var(\cdot), \sigma^2$	variance.
$cov(\cdot)$	covariance.
\mathcal{B}	Matrice variance des moyennes (inter-classe).
\mathcal{W}	Matrice moyenne des variances (intra-classe).
\mathcal{R}	Matrice de corrélation.
\mathcal{Z}	Matrice centrée réduite de la matrice \mathcal{X} .
\mathcal{Z}^t	Matrice transposé de la matrice \mathcal{Z} .
$cor(\cdot), r$	Corrélation.
\mathcal{F}	Tableau logique associé à la variable qualitative.
S	Variable discriminante.
u	facteur discriminante.
X	Variable aléatoire quantitative.
M	Métrique.
Δ_u	Axe discriminante.
i.e.	C'est à dire.
λ, η	Valeur propre (pouvoir discriminante d'un axe)
$M^{-1}, \mathcal{V}^{-1}, \mathcal{W}^{-1}$	Matrice inverse des M, \mathcal{V} et \mathcal{W} .
AFD	A nalyse F actorielle D iscriminante.
ACP	A nalyse en C omposantes P incipales.
$d(X_{i_1}, X_{i_2})$	Distance entre X_{i_1} et X_{i_2} .
b	Nouvelle observation à classé $\in \mathbb{R}^p$.
f_{E_k}	Fonction linéaire discriminante relatives à le groupe E_k .
Y	Variable aléatoire qualitative.

v.a	V ariable aléatoire .
v.a.r	V ariable aléatoire réelle .
$\mathbb{P}(Y = k), P_k$	Proportion théorique (probabilité a priori) de E_k .
f_k	Densité de probabilité/probabilité de X dans E_k
f	Densité de probabilité.
U	Variable aléatoire.
μ	Espérance.
ξ	Matrice de variance.
\mathfrak{h}^2	Le carré de la distance de Mahalanobis théorique généralisée.
θ	Paramètre.
M.V	M aximum de V raisemblance.
\hat{P}_k, q_k	Estimation de la probabilité a priori P_k .
$\hat{\mu}_k, a_k$	Estimation de μ_k .
$\hat{\xi}_k, V_k$	Estimation de ξ_k .
$\hat{\xi}, W$	Estimation de ξ .
$\xi^{-1}, \xi_k^{-1}, V_k^{-1}, W^{-1}$	Inverse de ξ_k, ξ, V_k et W .
\mathfrak{Q}_k	Fonction quadratique discriminante du groupe E_k .
\mathfrak{L}_k	Fonction linéaire discriminante du groupe E_k .
ppv	P lus p roches v oisins.
$D(x)$	Région (hypersphère).
D	Volume.
F	Fonction de répartition de la densité f sur la région $D(x)$.
i.i.d.	i ndépendant et i dentiquement d istribuées.
c	Nombre d'observation appartenant à $D(x)$.
\hat{f}	Estimation de la fonction de densité f .
b_c	observation fixée.
$\ \cdot\ $	Norme euclidienne.

Γ	Fonction Gamma.
c_k	Observation parmi les c observations $\in E_k$.
n_k	Nombre d'observations issues de la classe E_k .
$\hat{\mathbb{P}}(E_k b)$	Estimation de la densité conditionnelle $\mathbb{P}(E_k b)$.
ε	Erreur de classification asymptotique de la règle du plus proches voisins.
ε_B	Erreur de la règle de Bayes.

ملخص :

التحليل التمييزي هو تقنية إحصائية تُستخدم لتصنيف البيانات وتحديد المجموعة التي ينتمي إليها عنصر معين بناءً على المتغيرات التوضيحية. تعتمد هذه التقنية على منهجين رئيسيين: المنهج الاحتمالي والمنهج الهندسي. يستخدم المنهج الاحتمالي لتقدير الاحتمالات الشرطية لانتماء العناصر إلى مجموعات، بينما يعتمد المنهج الهندسي على تحليل المسافات والفروق الهندسية بين المجموعات. تُستخدم هذه الأساليب في مجالات متعددة مثل التسويق، التمويل، والطب لتحسين دقة التنبؤات واتخاذ القرارات.

الكلمات المفتاحية: التحليل التمييزي، الأساليب الاحتمالية، الأساليب الهندسية، التصنيف، تحليل البيانات، مسافة مهالانوبيس، التحليل متعدد المتغيرات.

Résumé :

L'analyse discriminante est une technique statistique utilisée pour classifier les données et déterminer à quel groupe appartient un élément donné en fonction des variables explicatives. Cette technique repose sur deux approches principales : l'approche probabiliste et l'approche géométrique. L'approche probabiliste estime les probabilités conditionnelles d'appartenance, tandis que l'approche géométrique analyse les distances et différences géométriques entre les groupes. Ces méthodes sont utilisées dans des domaines tels que le marketing, la finance et la médecine pour améliorer la précision des prédictions et la prise de décision.

Mots clés : Analyse discriminante, méthodes probabilistes, méthodes géométriques, classification, Analyse des données, la distance de Mahalanobis, analyse multivariée.

Abstract:

Discriminant analysis is a statistical technique used to classify data and determine the group to which an element belongs based on explanatory variables. This technique relies on two main approaches: the probabilistic and geometric approaches. The probabilistic approach estimates conditional probabilities of group membership, while the geometric approach analyzes distances and geometric differences between groups. These methods are used in fields such as marketing, finance, and medicine to improve prediction accuracy and decision-making.

Key words: Discriminant analysis, probabilistic methods, geometric methods, classification, data analysis, Mahalanobis distance, multivariate analysis.