

REVUE DE STATISTIQUE APPLIQUÉE

H. CHAMLAL

S. SLAOUI CHAH

Proposition d'une nouvelle règle discriminante et comparaison avec les règles linéaire et quadratique

Revue de statistique appliquée, tome 49, n° 3 (2001), p. 61-72

http://www.numdam.org/item?id=RSA_2001__49_3_61_0

© Société française de statistique, 2001, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

PROPOSITION D'UNE NOUVELLE RÈGLE DISCRIMINANTE ET COMPARAISON AVEC LES RÈGLES LINÉAIRE ET QUADRATIQUE

H. CHAMLAL et SLAOUI CHAH S.

*Département de Mathématiques et Informatique, Faculté des Sciences,
B.P. 1014, Rue Ibn Batouta, Rabat (Maroc)*

RÉSUMÉ

L'objet de cet article est d'introduire une règle d'affectation en analyse discriminante combinant la notion d'atypicité et la règle quadratique bayésienne, sous les hypothèses d'équiprobabilité a priori et d'égalité des coûts de mauvaise classification. Après une présentation de l'indice d'atypicité, on montre que la présence des «régions critiques» lors de l'application de la règle quadratique bayésienne, sous certaines hypothèses, conduit à des erreurs d'affectations. La performance de la nouvelle règle est mise en évidence au moyen de simulations. Une étude comparative avec la règle quadratique bayésienne et l'analyse discriminante linéaire est effectuée.

Mots-clés : Indice d'atypicité, Régions critiques, Règle quadratique bayésienne, Affectation, Analyse discriminante linéaire.

ABSTRACT

We propose a method of discriminant analysis, based on the atypicity index and the density function. After a short survey of the atypicity index, we show that the presence of "critical regions", when we apply the bayesian quadratic discrimination, under some hypotheses, leads to misclassifications. The performance of the proposed method versus quadratic and linear discrimination is assessed via simulation. It's generally shown that the discrimination based on the ratio (atypicity index/density function) improves consistently the percentage of well classified individuals relative to the traditional methods. The method is illustrated with a numerical example and is compared to quadratic discrimination.

Keywords : Atypicity index, Critical regions, Bayesian quadratic rule, Affectation, Linear discriminant analysis.

1. Introduction

Dans les règles d'affectation en analyse discriminante, on s'intéresse au groupe d'individus ayant une «ressemblance» maximale avec l'individu à classer. Les méthodes géométriques telles que l'analyse discriminante linéaire, les méthodes de discrimination quadratique de Sebestyen, celle de Rohlf et de Rioux...définissent

chacune une distance entre un groupe et l'individu à classer, dans le but de déterminer le groupe le plus proche de l'individu à classer.

S'agissant du modèle probabiliste, et particulièrement de la règle quadratique bayésienne (RQB), on se place généralement sous les hypothèses d'équiprobabilité a priori et d'égalité des coûts de mauvaise classification. L'affectation se base sur la comparaison des fonctions de densité à l'intérieur de chaque groupe. Cependant, bien que cette règle donne pratiquement des résultats satisfaisants, la comparaison des fonctions de densité à l'intérieur de chaque groupe, peut conduire à des affectations incorrectes, à partir du moment où cette comparaison privilégie le groupe dans lequel les valeurs de la fonction de densité sont élevées.

Avant d'appliquer la règle de discrimination bayésienne, certains auteurs recommandent d'étudier l'atypicité de l'individu à classer au moyen de l'indice d'atypicité qui est significatif quand il atteint ses valeurs extrêmes.

Dans cet article, on introduit une règle d'affectation, permettant de remédier au problème généré par la différence entre fonctions de densité à l'intérieur de chaque groupe, lors de l'application de la règle quadratique bayésienne (RQB), sous les hypothèses d'égalité des coûts de mauvaise classification et d'équiprobabilité a priori.

Dans les sections 2 et 3, on rappelle la notion d'atypicité, la règle quadratique bayésienne et on étudie les «régions critiques». Dans la section 4, on introduit le critère d'affectation qui combine celui de la fonction de densité à l'intérieur du groupe (RQB) d'une part, et l'indice d'atypicité d'autre part. Le nouveau critère établit un équilibre entre fonctions de densité à l'intérieur de chaque groupe. Dans la section 5, on se place sous l'hypothèse d'un modèle de O'Neill [O'Neill, (1992)] et par simulation, on effectue une étude comparative de la méthode proposée avec la règle quadratique bayésienne (RQB) et l'analyse discriminante linéaire (ADL). Deux exemples numériques illustrent les performances de la règle proposée.

Données du problème

E : un n -échantillon.

X_1, \dots, X_p : p variables explicatives quantitatives.

Q : variable qualitative à expliquer, ayant m modalités.

G_1, \dots, G_m : groupes induits par Q sur E .

2. Indice d'atypicité

Soit $X = (X_1, \dots, X_p)^t$, un vecteur explicatif distribué à l'intérieur de chaque groupe G_r selon une loi de fonction de densité, $f(x/r)$ ($r = 1, \dots, m$).

DÉFINITION 1. – *Étant donné deux individus i et j , on dit que i est plus typique du groupe G_r que j si et seulement si $f(i/r) > f(j/r)$.*

DÉFINITION 2. – L'indice d'atypicité du groupe G_l affecté à l'individu i est donné par :

$$ind(i, l) = \sum_{j \in E, f(j/l) > f(i/l)} f(j/l)$$

Si $ind(i, l) = 0$, alors il résulte de la définition précédente que

$$\forall j \in E, f(j/l) \leq f(i/l), \quad i \in E, \quad l \in \{1, \dots, m\}$$

Par ailleurs, $ind(i, l)$ atteindra son maximum si

$$\forall j \in E, \quad j \neq i, \quad f(j/l) > f(i/l)$$

De façon générale, un individu sera affecté au groupe minimisant l'indice d'atypicité. Pourtant, si on décide de se baser sur cet indice seul pour construire une règle d'affectation, la comparaison des indices d'atypicité des différents groupes conduit à privilégier les groupes pour lesquels les valeurs de la fonction de densité sont faibles.

3. Règle bayésienne

Notons π_r la probabilité a priori d'appartenance au groupe G_r . $c(j/r)$ désigne le coût de mauvaise classification résultant d'une affectation d'un individu du groupe G_r au groupe G_j , $((j, r) \in \{1, 2, \dots, m\}^2, c(j/j) = 0)$.

Soit i_s une observation qu'on cherche à affecter. La règle bayésienne consiste à chercher une partition de la population en m régions $R_1^*, R_2^*, \dots, R_m^*$, i_s étant affectée au groupe G_l si i_s appartient à R_l^* . Cette partition est celle minimisant le risque global moyen, elle est donnée par le théorème suivant (voir par exemple [Nakache J. P. (1980)])

THÉORÈME 1. – La règle optimale bayésienne consiste à choisir la partition en m régions $R_1^*, R_2^*, \dots, R_m^*$, telles que :

$$R_s^* = \{x \in P / h_s(x) = \min(h_j(x), \quad j = 1 \dots m)\}$$

$$\text{avec } h_j(x) = \sum_{r=1}^m \pi_r c(j/r) f(x/r)$$

Si on suppose que les coûts de mauvaise classification sont égaux ($c(j/r) = c$ si $j \neq r, c(j/j) = 0$), compte tenu de ce que $h_l(x) = c \left[\sum_{j=1}^m \pi_j f(x/j) - \pi_l f(x/l) \right]$, on a :

$$(i_s \text{ est affecté au groupe } G_r) \iff \pi_r f(i_s/r) = \max_{l=1 \dots m} \pi_l f(i_s/l). \quad (3.1)$$

En général, on suppose que les probabilités a priori sont égales,

$$(3.1) \iff f(i_s/r) = \max_{l=1..m} f(i_s/l)$$

Le problème se ramène à une comparaison des fonctions de densité à l'intérieur de chaque groupe, ce qui privilégie le groupe dans lequel la fonction de densité prend des valeurs élevées.

On se place dans le cas simple d'une distribution normale unidimensionnelle, (ce choix sera justifié dans la dernière section), et une variable à expliquer ayant deux modalités. La variable explicative est distribuée, à l'intérieur du premier groupe, suivant une loi normale centrée réduite, à l'intérieur du deuxième, suivant une loi normale de moyenne μ , d'écart type σ . Pour l'affectation d'un individu, on étudie le rapport

$$R = \frac{f(x/2)}{f(x/1)}$$

ou encore le signe de l'expression :

$$2\sigma^2 \ln\left(\frac{f(x/2)}{f(x/1)}\right) = (\sigma^2 - 1)x^2 + 2\mu x - 2\sigma^2 \ln(\sigma) - \mu^2$$

L'étude des variations de cette quantité, en supposant $\sigma^2 > 1$, est résumée dans le tableau 3.1, où

$$x_1^* = \frac{-\mu - \sqrt{\delta}}{\sigma^2 - 1}, \quad x_2^* = \frac{-\mu + \sqrt{\delta}}{\sigma^2 - 1}, \quad \delta = \mu^2 + (\sigma^2 - 1)(\mu^2 + 2\sigma^2 \ln(\sigma))$$

TABLEAU 3.1

Etude des variations de la quantité $\ln\left(\frac{f(x/2)}{f(x/1)}\right)$

x		x_1^*		x_2^*	
$\ln(R)$	+	0	-	0	+
R	$f(. / 2) > f(. / 1)$	$f(. / 2) = f(. / 1)$	$f(. / 2) < f(. / 1)$	$f(. / 2) = f(. / 1)$	$f(. / 2) > f(. / 1)$

Si on considère que les observations les plus probables à l'intérieur du premier groupe sont éléments de l'intervalle $[-2, 2]$, celles à l'intérieur du deuxième groupe sont éléments de $[\mu - 2\sigma, \mu + 2\sigma]$.

Le risque d'affecter incorrectement des observations du groupe G_2 , augmente avec la probabilité de l'événement : $([\mu - 2\sigma, \mu + 2\sigma] \cap [x_1^*, x_2^*])$.

Le risque d'affecter incorrectement des observations du groupe G_1 , augmente avec la probabilité de l'événement : $([-2, 2] \cap]-\infty, x_1^*]) \cup ([-2, 2] \cap [x_2^*, +\infty[)$.

La région $[\mu - 2\sigma, \mu + 2\sigma] \cap [x_1^*, x_2^*]$ est appelée région critique de G_2 .

La région $([-2, 2] \cap]-\infty, x_1^*]) \cup ([-2, 2] \cap [x_2^*, +\infty[)$ est appelée région critique de G_1 .

Exemple : supposons qu'à l'intérieur du deuxième groupe, X est distribuée suivant une loi normale $\mathcal{N}(3, 2)$: les valeurs les plus probables à l'intérieur du groupe G_2 sont éléments de $[-1, 7]$. $[-1, 7] \cap [-2, 2] \neq \emptyset$, $x_1^* = -3.418$, $x_2^* = 1.418$. La région critique du deuxième groupe est $[-1, 1.418]$ de probabilité 0.192. La région critique du premier groupe est $[1.418, 2]$ de probabilité 0.055. Si on suppose que $\mu = 3$ et $\sigma = 3$; les valeurs les plus probables à l'intérieur du groupe G_2 sont éléments de $[-3, 9] \supset [-2, 2]$, $x_1^* = -2.31$ et $x_2^* = 1.56$. La région critique du deuxième groupe est $[-2.31, 1.56]$ de probabilité 0.2772. Celle du groupe G_1 est $[1.56, 2]$ de probabilité 0.0366.

On voit qu'il est plus probable d'affecter incorrectement les observations provenant du deuxième groupe compte tenu des faibles valeurs de la fonction de densité. Ces deux cas seront étudiés ultérieurement.

4. Méthode d'affectation proposée

L'approche que nous proposons consiste à intégrer, la notion d'atypicité dans la méthode basée sur la fonction de densité. Le critère qui en résulte, présente alors les qualités des deux méthodes (i.e. la (RQB) sous les hypothèses d'équiprobabilité a priori et d'égalité des coûts de mauvaise classification, et la règle basée sur l'indice d'atypicité); il permet en outre d'obtenir des résultats meilleurs que ceux obtenus, par chacune des approches appliquées séparément.

PROPOSITION 1. – La règle proposée consiste à affecter i_s au groupe G_{l_0} , vérifiant :

$$\frac{ind(i_s, l_0)}{f(i_s/l_0)} = \min_{l=1, \dots, m} \frac{ind(i_s, l)}{f(i_s/l)}$$

Convention : La règle discriminante proposée est basée sur le Rapport (Atypicité/Densité), on convient de l'intituler (RAD).

PROPOSITION 2. – Supposons qu'à l'intérieur de chaque groupe G_l , $f(\cdot/l)$ est la fonction de densité d'une loi multinormale de moyenne $\mu_l \in \mathbb{R}^p$, de matrice de variance covariance Σ_l alors :

$$\frac{ind(i_s, l)}{f(i_s/l)} = \frac{\sum_{f_0(i^l) > f_0(i_s^l)} f_0(i^l)}{f_0(i_s^l)}$$

où f_0 est la fonction de densité d'une loi multinormale centrée réduite, i^l est l'observation i normalisée, pour le groupe G_l ($i^l = \Sigma_l^{-1/2}(i - \mu_l)$).

En effet, considérons la quantité :

$$\frac{ind(i_s, l)}{f(i_s/l)}$$

(i_s individu à classer)

$$\begin{aligned} \frac{ind(i_s, l)}{f(i_s/l)} &= \frac{(2\pi)^{(-p/2)} (det(\Sigma_l))^{(-1/2)} \sum_{f(i/l) > f(i_s/l)} exp(-\frac{1}{2}(i - \mu_l)^t \Sigma_l^{-1} (i - \mu_l))}{(2\pi)^{(-p/2)} (det(\Sigma_l))^{(-1/2)} exp(-\frac{1}{2}(i_s - \mu_l)^t \Sigma_l^{-1} (i_s - \mu_l))} \\ &= \frac{\sum_{f(i/l) > f(i_s/l)} exp(-\frac{1}{2}(i - \mu_l)^t \Sigma_l^{-1} (i - \mu_l))}{exp(-\frac{1}{2}(i_s - \mu_l)^t \Sigma_l^{-1} (i_s - \mu_l))} \\ &= \frac{\sum_{\sqrt{(det(\Sigma_l))} f(i/l) > \sqrt{(det(\Sigma_l))} f(i_s/l)} exp(-\frac{1}{2}(i - \mu_l)^t \Sigma_l^{-1} (i - \mu_l))}{exp(-\frac{1}{2}(i_s - \mu_l)^t \Sigma_l^{-1} (i_s - \mu_l))} \\ &= \frac{\sum_{f_0(i^l) > f_0(i_s^l)} f_0(i^l)}{f_0(i_s^l)} \end{aligned}$$

À ce niveau, on se base sur un critère qui ne privilégie ni le groupe dans lequel les variables sont les plus dispersées, ni les moins dispersées, puisque cela revient à travailler avec des observations normalisées.

Dans le cas normal, construire une règle d'affectation sur la base du critère

$$\frac{ind(i_s, l)}{f(i_s/l)}$$

permet d'obtenir des résultats meilleurs que ceux obtenus par application de la règle quadratique bayésienne sous l'hypothèse d'équiprobabilité a priori, avec normalisation. En effet, d'après la proposition 2 :

$$\frac{ind(i_s, l)}{f(i_s/l)} = \frac{\sum_{f_0(i^l) > f_0(i_s^l)} f_0(i^l)}{f_0(i_s^l)}$$

en minimisant ce rapport, on minimise la dissemblance à un groupe (après normalisation) et on maximise le critère de la règle quadratique bayésienne sous l'hypothèse d'équiprobabilité a priori, avec normalisation.

Lors de l'application de la règle quadratique bayésienne, on se place généralement sous l'hypothèse d'homoscédasticité (*hypothèse rarement testée*). Dans le cas normal, et sous cette hypothèse, l'application de la règle quadratique bayésienne sur les données initiales est équivalente à l'application sur les données normalisées. Le critère (RAD) ne suppose pas l'homoscédasticité, il donne des résultats meilleurs que ceux obtenus par application de la règle quadratique bayésienne sous cette hypothèse. En outre, se baser sur le critère (RAD) revient à travailler avec des observations normalisées (cf. Proposition 2), même si l'hypothèse d'homoscédasticité est violée.

La règle d'affectation (RAD) est une généralisation de la règle «(RQB) normalisée» au cas d'une famille quelconque de fonctions de densité.

Le critère (RAD) établit un «équilibre» entre fonctions de densité à l'intérieur des différents groupes sans privilégier les groupes selon la valeur plus ou moins élevée de la fonction de densité. En effet l'approche consiste à minimiser la dissemblance à un groupe et à maximiser la probabilité d'appartenance à un groupe.

5. Comparaison des méthodes

5.1. Résultats de simulations

Dans tout ce qui suit, on suppose que la variable à expliquer possède deux modalités. La population P est divisée en deux sous populations P_1 , P_2 . Le modèle de O'Neill a été proposé par O'Neill (1992) pour comparer la performance des règles de discrimination linéaire et quadratique. Les hypothèses du modèle de O'Neill sont les suivantes :

X est distribuée suivant une loi multivariée $\mathcal{N}_p(\omega_i, \Omega_i)$ ($i = 1, 2$) dans les deux sous populations.

Dans la population P_1 : $\omega_1 = 0$ et $\Omega_1 = I_p$.

Dans la population P_2 : $\omega_2 = \delta e_1$ et $\Omega_2 = I_p + \tau e_1 e_1^t$.

e_1 étant le premier vecteur de la base canonique de \mathbb{R}^p , $\tau > -1$ et $\delta \in \mathbb{R}$ sont des paramètres.

DÉFINITION 3. – Soit X un vecteur distribué suivant une loi $\mathcal{N}_p(\omega_i, \Omega_i)$ dans les populations P_i ($i = 1, 2$), on dit que X satisfait au modèle de O'Neill de dimension q ($q < p$), si

$$\omega_1 = 0 \text{ et } \Omega_1 = I_p, \quad \omega_2 = \sum_{j=1}^q \delta_j e_j \text{ et } \Omega_2 = I_p + \sum_{j=1}^q \tau_j e_j e_j^t$$

e_j étant le j^{em} vecteur de la base canonique de \mathbb{R}^p ,

Pour étudier la performance de la méthode (RAD), on se place dans le cas d'un modèle de O'Neill de dimension un, ce qui justifie l'étude unidimensionnelle faite dans la troisième section, lors de la détermination des régions critiques (RQB). Le choix de $q = 1$ est motivé par la pratique; il est pratiquement le cas le plus important [Flury B. et al (1996)].

Nous avons effectué des simulations dans les conditions suivantes :

$$\pi_1 = \pi_2, \quad \omega_1 = 0 \in \mathbb{R}^p, \quad \Omega_1 = I_p, \quad \omega_2 = \delta e_1 \in \mathbb{R}^p, \quad \Omega_2 = \Omega_1 + \tau e_1 e_1^t$$

p étant fixé égal à 4, $\delta \in \{3, 4, 5, 6\}$ et $\tau \in \{3, 8, 15, 24, 35\}$. Quatre cas résultant des combinaisons de ces paramètres sont étudiés ici. Pour chaque cas, nous considérons que les deux groupes induits par la variable à expliquer sont de taille égale

($n_1 = n_2$). On a généré 1000 fois $n_1 = n_2$ réalisations de la loi $\mathcal{N}(\omega_i, \Omega_i)$ ($i = 1, 2$). Nous avons fait varier n_1 de 15 à 120 et le pourcentage moyen de bien classés est la moyenne arithmétique des 1000 pourcentages de bien classés calculés par application de la technique de validation croisé due à Lachenbruch et Mickey [Saporta G., 1990] sur les 1000 échantillons simulés. Pour les quatre cas, les régions critiques sont de probabilité non négligeable (cf. section 3). D'où l'intérêt de présenter ces cas. L'étude effectuée est comparative; nous avons calculé les pourcentages de bien classés pour la nouvelle méthode (RAD), la règle quadratique bayésienne (RQB) et l'analyse discriminante linéaire (ADL).

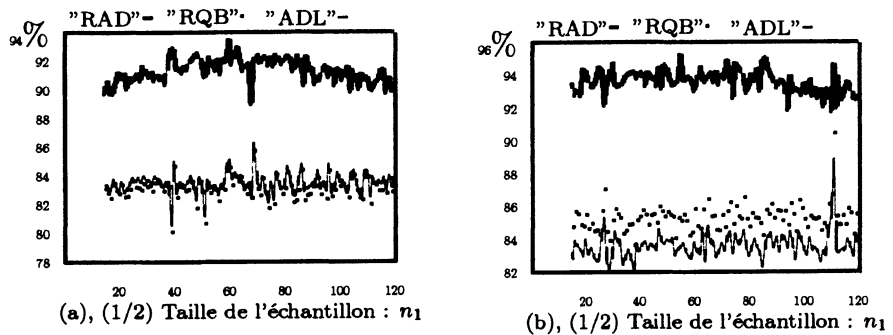


FIGURE 5.1

Pourcentage moyen de bien classés, $p = 4$, (a) $\delta = 3$, $\tau = 8$; (b) $\delta = 4$, $\tau = 15$

Pour les quatre cas, les meilleurs pourcentages de bien classés sont atteints par application de la règle d'affectation (RAD). Les plus bas pourcentages de bien classés pour la (RQB) sont obtenus dans les cas où les régions critiques sont de probabilité non négligeable (cf. Tableau 5.1). Il s'agit des cas :

$$\delta = 3 \tau = 8 ; \delta = 4 \tau = 15 ; \delta = 6 \tau = 35 ; \delta = 5 \tau = 24$$

TABLEAU 5.1
Régions critiques

Cas	Probabilité de la région critique de G_2	Probabilité de la région critique de G_1
$\delta = 3 \tau = 8$	0.277	0.0366
$\delta = 4 \tau = 15$	0.2306	0.02
$\delta = 6 \tau = 35$	0.178	0
$\delta = 5 \tau = 24$	0.195	0.006

Premier cas : (cf. fig 5.1(a)) Par application de l'(ADL), on obtient des pourcentages de bien classés légèrement meilleurs que ceux calculés par application de la (RQB). Par application de la règle (RAD), on obtient des pourcentages de bien classés très écartés des pourcentages calculés par application de l'(ADL) et de la (RQB) .

Deuxième cas : (cf. fig 5.1(b)) L'écart, d'une part, entre les pourcentages de bien classés calculés par application de la règle (RAD) et les pourcentages calculés par application de la (RQB) et de l'(ADL), d'autre part, est important. L'écart entre les pourcentages de bien classés, calculés par application de l'(ADL) et de la (RQB), est plus significatif que dans le premier cas. Les pourcentages de bien classés calculés par application de la (RQB) sont meilleurs que ceux calculés par application de l'(ADL).

Troisième cas : (cf. fig 5.2(b)) Les pourcentages de bien classés calculés par application de la règle (RAD) s'écartent nettement de ceux calculés par application de la (RQB) et de l'(ADL).

Quatrième cas : (cf. fig 5.2(a)) L' écart entre les pourcentages de bien classés calculés par application de la règle (RAD) et ceux calculés par application de la (RQB) et de l'(ADL) est important.

En général, on note que l'écart entre les pourcentages moyens de bien classés calculés par application de la règle (RAD), et ceux calculés par application de la (RQB), augmente avec la probabilité des régions critiques.

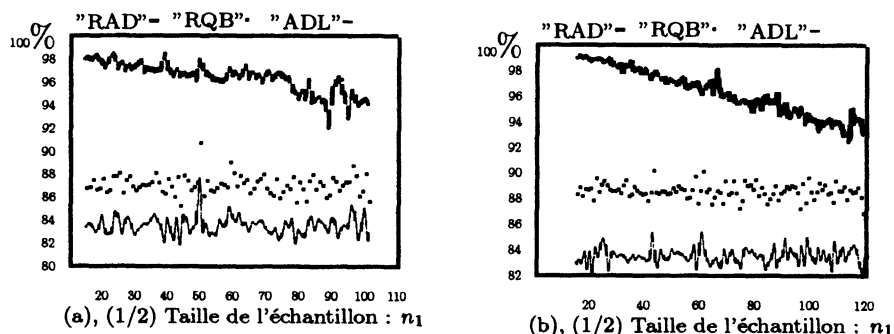


FIGURE 5.2

Pourcentage moyen de bien classés, $p = 4$, (a) $\delta = 5$, $\tau = 24$; (b) $\delta = 6$, $\tau = 35$

5.2. Exemples

Nous optons pour deux ensembles de données, pour lesquels, l'hypothèse d'homoscédasticité est «presque» réalisée. L'application de la (RQB) sur les données initiales est équivalente à l'application sur les données normalisées.

Le premier ensemble est appelé **Normal4** [Nikhil R. Pal & James C. Bezdek, 1995], le second **Iris** [Anderson E. (1935)].

Normal4. est un échantillon de 800 observations réparties en 4 groupes de 200 observations chacun. Les individus sont décrits par 4 variables explicatives. La moyenne et la matrice des variances covariances sont :

$$\mu_i = 3e_i, \text{ et } \Sigma_i = I_4, \quad i = 1, 2, 3, 4, \quad e_i = (0, \dots, \underbrace{1}_i, \dots, 0)^t, \quad e_i \in \mathbb{R}^4$$

Iris. il s'agit des «Iris d'Anderson». 100 iris répartis à égalité en deux classes,

1 : Iris Virginia,

2 : Iris Vesicolor.

Les fleurs sont décrites par quatre variables explicatives :

X_1, X_2 , mesurent la longueur et la largeur en (cm) des sépales ,

X_3, X_4 , mesurent la longueur et la largeur en (cm) des pétales.

Sur chaque ensemble de données, nous appliquons la règle (RAD) et la (RQB).

Normal4.

Méthode appliquée	Pourcentage de bien classés
(RAD)	96.125%
(RQB)	96%

Les deux pourcentages de bien classés, calculés par application de la technique de validation croisé due à Lachenbruch et Mickey [Saporta G., 1990], sont presque identiques.

Iris. le $\frac{1}{5}$ de l'échantillon total joue le rôle d'échantillon test. Dans le tableau (cf. Tableau 5.2), les 20 iris à réaffecter sont décrits par les quatre variables explicatives et la variable à expliquer Q . On donne pour chaque groupe la valeur du critère à minimiser noté $ct(l), l = 1, 2$. La variable Q' représente les affectations des individus par application de la règle (RAD).

On se place sous les hypothèses de normalité et d'homoscédasticité. La matrice des variances covariances est estimée par la matrice de variance intraclasse.

Les résultats sont les suivants : un seul individu de chaque groupe est mal classé; soit un pourcentage de bien classés égal à 90%. Les résultats par application de la (RQB) sont analogues; aussi bien au niveau des affectations qu'au niveau des pourcentages de bien classés.

6. Conclusion

Les résultats des simulations et ceux des exemples numériques montrent que l'approche (RAD) donne de meilleurs résultats que les méthodes classiques. Dans le

Tableau 5.2
 Résultats de la méthode (RAD) appliquée sur les données Iris

X_1	X_2	X_3	X_4	$ct(1)$	$ct(2)$	Q	Q'
5.5	2.4	3.7	1	34.553	7308603.322	1	1
6	2.7	5.1	1.6	779.841	99.439	1	2
6	3.4	4.5	1.6	120.161	8424.083	1	1
6.3	2.3	4.4	1.3	89.948	82342.583	1	1
5.5	2.5	4	1.3	5.633	16832.735	1	1
6.1	3	4.6	1.4	7.009	3521.558	1	1
5	2.3	3.3	1	98.742	29291229.515	1	1
5.7	3	4.2	1.2	33.872	222090.929	1	1
6.2	2.9	4.3	1.3	0	79949.503	1	1
5.7	2.8	4.1	1.3	0	28658.63	1	1
7.9	3.8	6.4	2	319161.645	2288.082	2	2
6.3	2.8	5.1	1.5	120.342	243.144	2	1
7.7	3	6.1	2.3	4221269.578	498.101	2	2
6.4	3.1	5.5	1.8	3503.24	31.994	2	2
6.9	3.1	5.4	2.1	5955.666	35.933	2	2
6.9	3.1	5.1	2.3	599162.718	4371.9	2	2
6.8	3.2	5.9	2.3	883351.593	4.613	2	2
6.7	3	5.2	2.3	265579.526	300.514	2	2
6.5	3	5.2	2	1364.973	9.071	2	2
5.9	3	5.1	1.8	2420.095	38.984	2	2

cas où les régions critiques sont de probabilité non négligeable, les résultats, obtenus par simulation, sont nettement meilleurs que ceux de la règle quadratique bayésienne et de l'analyse discriminante linéaire. Dans tout ce qui précède, on s'est placé lors de l'application de la (RQB), sous l'hypothèse d'équiprobabilité a priori. Dans le cas le plus fréquent, on se place sous cette hypothèse (voir par exemple [J.P.Nakache, 1980]). Le nouveau critère «**équilibre**» les fonctions de densité à l'intérieur des différents groupes. Il permet de ne privilégier ni le groupe dans lequel les valeurs de la fonction de densité sont faibles ni le groupe dans lequel les valeurs de la fonction de densité sont élevées. Pour les simulations, on s'est placé sous l'hypothèse d'un modèle de O'Neil, et sous l'hypothèse $q = 1$. C'est pratiquement le modèle le plus important. Pour les exemples numériques, les résultats par application de la règle (RAD) et de la (RQB) sont analogues.

Références bibliographiques

- ANDERSON E., The IRISes of the Gaspe peninsula : *Bull.Ame.IRIS SOC.*, vol. 59, pp. 2–5, (1935).
- FLURY B., NEL D. G. and PIENAAR I., Simultaneous Detection of Shift in Means and Variances : *Journal of the American Statistical Association* 91, pp. 1743–1748 (1996).
- NAKACHE J. P., Méthodes de discrimination pour variables de nature quelconque, théorie et pratique : Thèse présentée pour obtenir le grade de Docteur-es sciences, Université Pierre et Marie Curie (1980).
- NIKHIL R. Pal & JAMES C. Bezdek, On cluster validity for the fuzzy C-means model : *IEEE Transaction on fuzzy system*, vol 3, n 3, p 377, August (1995).
- O'NEILL T. J., Error Rates of Non-Bayesian Classification Rules and Robustness of Fisher's Linear Discriminant Function : *Biometrika*, 79, pp. 177–184 (1992).
- RIOUX P., Quadratic discriminant analysis : *EDV in Medizin and Biologie*, 6,112 (1975).
- ROHLFF J., Adaptative Hierarchical clustring shemes : *Syst. Zool*, 19, p. 58–82 (1970).
- SAPORTA G., Probabilités, analyse des données et statistique, Editions technip-France (1990).
- SEBESTYEN G. S., Decision Making Process in Pattern Recognition : *Vol 19, Mac Millan Company*.