

L'ANALYSE DISCRIMINANTE

MÉTHODES DÉCISIONNELLES

MÉTHODES PROBABILISTES

1. La règle bayésienne

- k groupes en proportion $p_1, p_2 \dots p_k$
- La distribution de probabilité du vecteur observation $\underline{x} = (x_1, \dots, x_p)$ est donnée pour chaque groupe j par une densité (ou une loi discrète) $f_j(\underline{x})$.

Observation $(x_1, x_2 \dots x_p)$



probabilité qu'elle provienne du groupe j
formule de Bayes

$$P(G_j|\underline{x}) = \frac{P(\underline{x}|G_j)P(G_j)}{\sum_{j=1}^k P(\underline{x}|G_j)P(G_j)}$$

$$P(G_j|\underline{x}) = \frac{p_j f_j(\underline{x})}{\sum_{j=1}^k p_j f_j(\underline{x})}$$

Règle bayésienne

Affecter \underline{x} au groupe qui a la probabilité a posteriori maximale.

⇒ chercher le maximum de $p_j f_j(\underline{x})$

Il est nécessaire de connaître ou d'estimer $f_j(\underline{x})$

- méthodes non paramétriques
- méthodes paramétriques : cas gaussien p-dimensionnel, discrimination logistique.

2. Le modèle normal multidimensionnel

Hypothèse de travail

\underline{x} distribués selon $N_p(\underline{\mu}, \Sigma_j)$ pour chaque groupe

Densité:

$$f_j(\underline{x}) = \frac{1}{(2\pi)^{p/2} (\det \Sigma_j)^{1/2}} \exp \left[-\frac{1}{2} (\underline{x} - \underline{\mu}_j)' \Sigma_j^{-1} (\underline{x} - \underline{\mu}_j) \right]$$

2-1. Cas général

Règle bayésienne : $\max_j p_j f_j(\underline{x})$ devient par passage aux logarithmes

(de l'expression $-2 \text{Log } p_j f_j$) :

$$\min : (\underline{x} - \underline{\mu}_j)' \Sigma_j^{-1} (\underline{x} - \underline{\mu}_j) - 2 \text{Log } p_j + \text{Log} (\det \Sigma_j)$$

Lorsque les Σ_j sont différents, cette règle est donc **quadratique** \rightarrow il faut comparer k fonctions quadratiques de \underline{x}

$$\Sigma_j \text{ est en général estimé par } \frac{n}{n-1} V_j$$

$$\underline{\mu}_j \text{ par } \underline{g}_j$$

2-2. Cas d'égalité des matrices de variance covariance

Si $\Sigma_1 = \Sigma_2 = \dots = \Sigma$, la règle devient linéaire car :

$$\text{Alors } \left\{ \begin{array}{l} \log(\det \Sigma_j) = \text{constante} \\ (\underline{x} - \underline{\mu}_j)' \Sigma^{-1} (\underline{x} - \underline{\mu}_j) = \Delta^2(\underline{x}, \underline{\mu}_j) \end{array} \right.$$

= distance de Mahalanobis
de \underline{x} à $\underline{\mu}$

En développant, en éliminant $\underline{x}' \Sigma^{-1} \underline{x}$, on obtient :

d'où en divisant par -2 \rightarrow $\max \left[\underline{x}' \Sigma^{-1} \underline{\mu}_j - \frac{1}{2} \underline{\mu}'_j \Sigma^{-1} \underline{\mu}_j + \log p_j \right]$

Si Σ est estimé par $\frac{n}{n-k}W$:

règle Bayésienne \Leftrightarrow règle géométrique (si égalité des p_j)

La règle géométrique est alors optimale.

La probabilité a posteriori d'appartenance au groupe j est proportionnelle à :

$$p_j \exp \left[-\frac{1}{2} \Delta^2 (\underline{x}, \underline{\mu}_j) \right]$$

2-3. Cas de deux groupes avec égalité de Σ_1 et Σ_2

On affectera \underline{x} au groupe 1 si :

$$\left[\underline{x}' \Sigma^{-1} \underline{\mu}_1 - \frac{1}{2} \underline{\mu}'_1 \Sigma^{-1} \underline{\mu}_1 + \text{Log } p_1 \right] > \left[\underline{x}' \Sigma^{-1} \underline{\mu}'_2 - \frac{1}{2} \underline{\mu}'_2 \Sigma^{-1} \underline{\mu}_2 + \text{Log } p_2 \right]$$



$$\underline{x}' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2) > \frac{1}{2} (\underline{\mu}_1 + \underline{\mu}_2)' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2) + \text{Log } \frac{p_2}{p_1}$$

Si $p_1 = p_2 = 0,5$, on retrouve la règle de Fisher en estimant Σ par $\frac{n}{n-2}W$.

Soit
$$s(\underline{x}) = \underline{x}' \Sigma^{-1}(\underline{\mu}_1 - \underline{\mu}_2) - \frac{1}{2}(\underline{\mu}_1 + \underline{\mu}_2)' \Sigma^{-1}(\underline{\mu}_1 - \underline{\mu}_2) \text{Log} \frac{p_2}{p_1}$$

On affectera \underline{x} au groupe 1 si $s(\underline{x}) > 0$

au groupe 2 si $s(\underline{x}) < 0$

$s(\underline{x})$ appelée « score » ou statistique d'Anderson.

Propriété

$S(\underline{x})$ est liée simplement à la probabilité a posteriori d'appartenance au groupe 1.

Démonstration :

$$P(G_1|\underline{x}) = \frac{p_1 f_1(\underline{x})}{p_1 f_1(\underline{x}) + p_2 f_2(\underline{x})} = P$$

$$\Rightarrow \frac{1}{P} = 1 + \frac{p_2 f_2(\underline{x})}{p_1 f_1(\underline{x})}$$

$$= 1 + \frac{p_2}{p_1} \exp \left[-\frac{1}{2} (\underline{x} - \underline{\mu}_2)' \Sigma^{-1} (\underline{x} - \underline{\mu}_2) + \frac{1}{2} (\underline{x} - \underline{\mu}_1)' \Sigma^{-1} (\underline{x} - \underline{\mu}_1) \right]$$

$$\Rightarrow \frac{1}{P} - 1 = \frac{p_2}{p_1} \exp \left[\frac{1}{2} \Delta^2(\underline{x}, \underline{\mu}_1) - \frac{1}{2} \Delta^2(\underline{x}, \underline{\mu}_2) \right]$$

$$\text{d'où } \text{Log} \left(\frac{1}{P} - 1 \right) = -S(\underline{x})$$

$$P = \frac{1}{1 + e^{-s(\underline{x})}} = \frac{e^{s(\underline{x})}}{1 + e^{s(\underline{x})}}$$

P « fonction logistique du score »

Remarque : Lorsque $p_1 = p_2 = \frac{1}{2}$

$$P = \frac{1}{1 + e^{-\frac{1}{2} [\Delta^2(\underline{x}, \underline{\mu}_1) - \Delta^2(\underline{x}, \underline{\mu}_2)]}}$$

2-4. A propos de certains tests :

Test d'égalité des matrices Σ_i : test de Box

Si l'hypothèse $\Sigma_1 = \Sigma_2 = \dots = \Sigma_k$ est vraie, la quantité :

$$1 - \frac{2p^2 + 3p - 1}{6(p+1)(k-1)} \left[\left(\sum \frac{1}{n_i - 1} - \frac{1}{n - k} \right) (n - k) \text{Log} \left| \frac{n}{n - k} W \right| - \sum (n_i - 1) \text{Log} \left| \frac{n_i}{n_i - 1} V_i \right| \right]$$

suit approximativement une loi $\chi^2_{\frac{p(p+1)(k-1)}{2}}$

Si on rejette l'hypothèse d'égalité, doit-on utiliser les règles quadratiques ?

Ce n'est pas sûr :

- Test de Box pas parfaitement fiable
- Règle quadratique \Rightarrow estimation de chaque Σ_j (donc de plus de paramètres).

Lorsque les échantillons sont de petite taille, les fonctions obtenues sont très peu robustes.

 il vaut mieux choisir une règle linéaire.

Nombre de paramètres à estimer

- Exemple:

- Avec $p = 10$ variables
- Avec $k = 4$ groupes

L'analyse discriminante linéaire demande l'estimation de 95 paramètres et l'analyse discriminante quadratique l'estimation de 260 paramètres

2.5. Qualité de la discrimination

a. Cas de 2 groupes

Soit un ensemble de ℓ variables parmi les p composantes de \underline{x}

Supposons que $\Delta_p^2 = \Delta_1^2$: en d'autres termes les $(p - \ell)$ autres variables n'apportent aucune information pour séparer les deux populations, alors :

$$\frac{(n_1 + n_2 - p - 1)n_1n_2(D_p^2 - D_1^2)}{(p - 1)(n_1 + n_2)(n_1 + n_2 - 2) + n_1n_2D_1^2} = F(p - 1, n_1 + n_2 - p - 1)$$

On peut ainsi tester :

- l'apport d'une nouvelle variable en prenant $\ell = p - 1$
- l'apport de toutes les variables ($\Delta_p^2 = 0$)

b. Plus de 2 groupes :

On utilise le Λ de Wilks

Sous $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$

$$\Lambda = \frac{|W|}{|V|} = \frac{|W|}{|W + B|} = \frac{1}{|W^{-1}B + I|}$$

suit la loi de Wilks de paramètres $(p, n - k, k - 1)$

Justification : nV, nW, nB suivent des lois de Wishart à

$n - 1, n - k, k - 1$ d.d.l.

c. Remarque dans le cas de deux groupes

Le test de Wilks et le test de la distance de Mahalanobis $(H_0 \Delta_p^2 = 0)$ sont identiques :

B étant de rang **1**, on a :

$$\Lambda = \frac{1}{1 + D_p^2 \frac{n_1 n_2}{(n_1 + n_2)(n_1 + n_2 - 2)}} = \frac{1}{1 + \mu} = 1 - \lambda$$

$$\left\{ \begin{array}{l} \mu = \text{valeur propre de } W^{-1}B \\ \lambda = \text{valeur propre de } V^{-1}B \\ \mu = \frac{\lambda}{1 - \lambda} \end{array} \right.$$

d. Paramètres usuels fournis par les logiciels

➤ **Wilks:**
$$\Lambda = \prod_{i=1}^{k-1} (1 - \lambda_i)$$

λ_i = corrélation canonique au carré

Plus le Λ (Wilks) est faible, meilleure est la discrimination

➤ **Trace de Pillai** =
$$\text{Trace}(V^{-1}B) = \sum_{i=1}^k \lambda_i$$

➤ **Trace de Hotelling-Lawley**

$$\text{Trace}(W^{-1}B) = \sum_{i=1}^{k-1} \frac{\lambda_i}{1 - \lambda_i} = \sum_{i=1}^{k-1} \mu_i$$

➤ **Plus grande valeur propre de Roy** : μ_1

2.6. Sélection de variables pas à pas

En discriminante à k groupes, on utilise souvent le test de variation de Λ mesuré par :

$$\frac{n-k-p}{k-1} \left(\frac{\Lambda_p}{\Lambda_{p+1}} - 1 \right) \text{ que l'on compare à un } F_{(k-1, n-k-p)}$$

La plupart des logiciels présentent des techniques de **sélection ascendante, descendante ou mixte des variables**. SAS propose la procédure STEPDISC.

Sélection ascendante (option **Forward**)

- A l'étape initiale aucune variable n'est présente.
- A chaque étape on fait entrer la variable qui contribue le plus au pouvoir discriminant du modèle, mesuré par le lambda de Wilks.
- La sélection s'arrête quand aucune des variables non sélectionnées ne convient au sens du seuil de probabilité choisi pour le F de Fisher.

Sélection descendante (option **Backward**)

- On démarre avec le modèle complet (construit avec toutes les variables)
- A chaque étape, la variable contribuant le moins au pouvoir discriminant du modèle est éliminée.
- La sélection s'arrête quand on ne peut plus éliminer de variables étant donné le seuil de probabilité choisi pour le F de Fisher.

Sélection mixte (option **Stepwise**)

- On démarre comme dans la procédure ascendante.
- Dès qu'une variable entre dans le modèle, on vérifie compte tenu de cette entrée si l'une des variables déjà présentes est susceptible d'être éliminée.
- La sélection s'arrête quand on ne plus ajouter ou éliminer de variables.

3. Mesures d'efficacité des règles de classement

Critère usuel **Probabilité de bien classer une observation quelconque.**
Les diverses méthodes sont comparées en fonction de leurs taux d'erreur.

3.1 Taux d'erreur théorique pour deux groupes avec $\Sigma_1 = \Sigma_2$ et distribution normale

Quand $p_1 = p_2$, on affecte l'individu au groupe 1 si :

$$S(\underline{x}) = \left[\underline{x}' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2) - \frac{1}{2} (\underline{\mu}_1 + \underline{\mu}_2)' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2) \right] > 0$$

La probabilité d'erreur de classement est donc :

$$P[S(\underline{x}) > 0 | \underline{x} \in N_p(\underline{\mu}_2, \Sigma)]$$

La loi de $S(\underline{x})$ est une loi de Gauss à une dimension comme combinaison linéaire des composantes de \underline{x} .

$$\begin{aligned}
 E(S(\underline{x})) &= \underline{\mu}'_2 \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2) - \frac{1}{2} (\underline{\mu}_1 + \underline{\mu}_2)' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2) \\
 &= -\frac{1}{2} (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2) \\
 &= -\frac{1}{2} \Delta_p^2
 \end{aligned}$$

$$V(S(\underline{x})) = (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} \Sigma \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2) = \Delta_p^2$$

$$S(\underline{x}) \text{ suit } N \left(-\frac{1}{2} \Delta_p^2 ; \Delta_p \right) \text{ si } \underline{x} \in G_2$$

La probabilité de classer dans le groupe 1 une observation du groupe 2 est :

$$P(1|2) = P\left(\frac{S(\underline{x}) + \frac{1}{2}\Delta_p^2}{\Delta_p} > \frac{\mathbf{0} + \frac{1}{2}\Delta_p^2}{\Delta_p}\right)$$

$$P(1|2) = P\left(U > \frac{\Delta_p}{2}\right) \quad \text{où } U \text{ suit } N(\mathbf{0}; \mathbf{1})$$

Elle est égale à $P(2|1)$.

Cette relation donne une interprétation concrète à la distance de Mahalanobis.

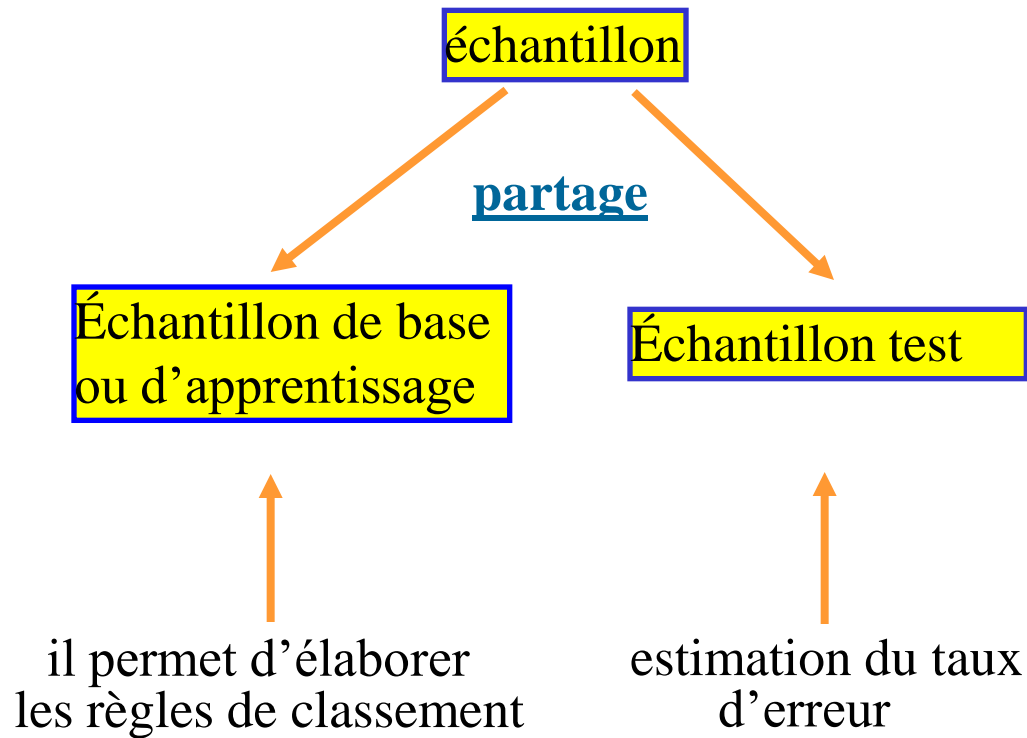
Remarque Estimations biaisées → sous-estimation du taux d'erreur.

3-2. Méthode de resubstitution

Réaffectation des observations selon les fonctions discriminantes trouvées.

Inconvénient On sous-estime le taux d'erreur.

3-3. Échantillon d'apprentissage; Échantillon test



3.4 Validation croisée

- Pour $i = 1$ à n , on construit la règle de décision sur la base privée de son $i^{\text{ème}}$ élément et on affecte ce dernier à l'un des groupes suivant cette règle.
- **Le taux d'erreur estimé est alors la fréquence de points mal classés de la sorte. L'estimation du taux d'erreur ainsi obtenu est pratiquement sans biais**
- La variance de l'estimation est d'autant plus importante que n est grand, puisque dans ce cas, les différentes règles de décision construites à partir de $n-2$ observations communes ont tendance à se ressembler.

4. Méthodes non paramétriques

Les méthodes non paramétriques consistent à estimer la densité de probabilité en chaque point de l'échantillon.

Deux méthodes sont souvent utilisées :

- méthode du noyau
- méthode des k plus proches voisins.

4.1 La méthode des noyaux

La méthode des noyaux généralise la notion d'histogramme. Dans le cas unidimensionnel, pour estimer la densité en un point \mathbf{x} , on centre l'intervalle de longueur \mathbf{R} de l'histogramme en ce point. La densité est alors le rapport de la probabilité de l'intervalle sur la longueur de l'intervalle.

Dans le cas multidimensionnel, considérons l'ellipsoïde centré sur $\underline{\mathbf{x}}$:

$$\Omega_{r,j}(\underline{\mathbf{X}}) = \left\{ \mathbf{y} / (\mathbf{y} - \mathbf{x})' \mathbf{V}_j^{-1} (\mathbf{y} - \mathbf{x}) \leq r^2 \right\}$$

Notons $I_j(\mathbf{z})$ la variable indicatrice de l'ellipsoïde $\{\mathbf{z} / \mathbf{z}'\mathbf{V}_j^{-1}\mathbf{z} \leq r^2\}$

La densité de probabilité estimée s'écrit :
$$f_j(\mathbf{x}) = \frac{\sum_{y \in G_j} I_j(\mathbf{y} - \mathbf{x})}{n_j v_r(\mathbf{j})}$$

avec $n_j =$ nombre d'éléments du groupe \mathbf{j}

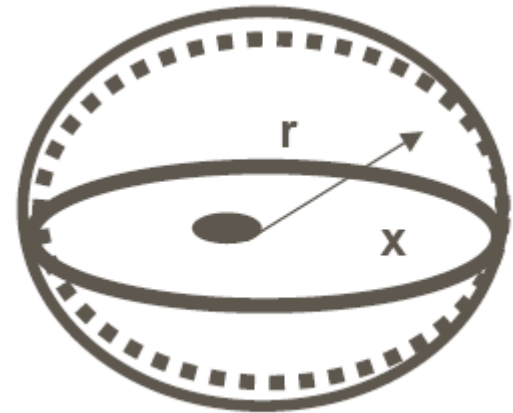
$v_r(\mathbf{j}) =$ volume de l'ellipsoïde

La méthode du noyau consiste à utiliser une fonction (le noyau) plus « lisse » que la variable indicatrice $I_j(\mathbf{z})$.

On trouve dans la littérature (et les logiciels) **différents types de noyaux** :

- uniforme: On compte le nombre d'observations appartenant à la boule de rayon R . Ce nombre est aléatoire.

- normal
- Epanechnikov
- biweight kernel
- triweight kernel



La difficulté d'utilisation de ces méthodes réside dans le choix du noyau et le choix de r .

4.2. Méthode des k plus proches voisins

On cherche les k points les plus proches de l'individu \underline{x} et on classe \underline{x} dans le groupe le plus représenté : la probabilité a posteriori d'appartenir au groupe j est égale au quotient entre le nombre d'individus du groupe j parmi les k points, et le nombre de voisins (k).

Le choix de k est moins crucial que le choix de r dans la méthode des noyaux. On peut choisir k optimisant une proportion de bien classés en validation croisée.

5. La régression logistique

Lorsqu'il n'y a pas que deux groupes, sous l'hypothèse de normalité et d'égalité des matrices de variance, la probabilité a posteriori est une fonction logistique du score, lui-même fonction linéaire des variables explicatives.

$$\text{Log} \left(\frac{1}{P} - 1 \right) = \text{Log} \frac{p_2}{p_1} \frac{f_2(\underline{x})}{f_1(\underline{x})} = \underbrace{-S(\underline{x})}$$

Donc :

$$\text{Log} \frac{f_2(\underline{x})}{f_1(\underline{x})} = \text{Log} \frac{p_1}{p_2} + \underbrace{\alpha + \underline{\beta}' \underline{x}}$$

$$(*) \quad \text{Log} \frac{f_2(\underline{x})}{f_1(\underline{x})} = \beta_0 + \underline{\beta}' \underline{x} \quad \text{où} \quad \underline{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}$$

Ceci amène à définir la régression logistique à partir de l'expression (*).

Hypothèse de travail

$$\text{Log} \quad \frac{f_2(\underline{x})}{f_1(\underline{x})} = \beta_0 + \beta' \underline{x} \quad \text{où} \quad \underline{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}$$

Le modèle de la régression logistique consiste à estimer les $(p+1)$ paramètres selon le maximum de vraisemblance.

$$P(G_1|\underline{x}) = \frac{p_1 f_1(\underline{x})}{p_1 f_1(\underline{x}) + p_2 f_2(\underline{x})} = \frac{\frac{p_1 f_1(\underline{x})}{p_2 f_2(\underline{x})}}{1 + \frac{p_1 f_1(\underline{x})}{p_2 f_2(\underline{x})}}$$

$$P(G_1|\underline{x}) = \frac{\exp\left(\text{Log} \frac{p_1}{p_2} + \beta_0 + \underline{\beta}' \underline{x}\right)}{1 + \exp\left(\text{Log} \frac{p_1}{p_2} + \beta_0 + \underline{\beta}' \underline{x}\right)} \quad (1)$$

$$P(G_2|\underline{x}) = \frac{1}{1 + \exp\left(\text{Log} \frac{p_1}{p_2} + \beta_0 + \underline{\beta}' \underline{x}\right)} \quad (2)$$

On montre que les expressions (1) et (2) sont conservées pour la famille de distributions :

$$f_i(\underline{x}) = c_i \exp\left[-\frac{1}{2}(\underline{x} - \underline{\mu}_i)' \Sigma^{-1}(\underline{x} - \underline{\mu}_i)\right] h(\underline{x})$$

où h est une fonction arbitraire de \underline{x} intégrable non négative et c_i une constante telle que f_i soit une densité de probabilité.

En effet, h n'intervient pas dans le calcul de (1) et (2) :

- si $h(x) \equiv 1$ on retombe sur la loi multinormale
- **on peut faire intervenir des variables binaires dans le modèle**
- **on peut appliquer le modèle au cas où un groupe de la population est dissymétrique ($h(x)$ constante dans la population normale, croissante ailleurs)**



[méthode générale](#)

Expression de la vraisemblance des β (n_1 et n_2 fixés)

$$L = \prod_{i \in G_1} f_1(\underline{x}_i) \prod_{i \in G_2} f_2(\underline{x}_i) \quad \text{avec : } f(\underline{x}) = p_1 f_1(\underline{x}) + p_2 f_2(\underline{x})$$

$$\text{On a } \begin{cases} f_1(\underline{x}) = \frac{P(G_1|\underline{x})f(\underline{x})}{p_1} \\ f_2(\underline{x}) = \frac{P(G_2|\underline{x})f(\underline{x})}{p_2} \end{cases}$$

$$\text{D'où } L = \frac{1}{p_1^{n_1} p_2^{n_2}} \prod_{i \in G_1} P(G_1|\underline{x}_i) \prod_{i \in G_2} P(G_2|\underline{x}_i) \prod_{i=1}^{n_1+n_2} f(\underline{x}_i)$$

$$\text{soit : } L = \frac{L_1 L_2}{p_1^{n_1} p_2^{n_2}} \quad \begin{array}{l} L_1 = \text{vraisemblance conditionnelle des} \\ \text{paramètres connaissant les } \underline{x}_i \\ L_2 = \text{densité (inconditionnelle) des } \underline{x}_i \end{array}$$

f non connue, on estime $\beta_0, \beta_1 \dots \beta_p$ par une méthode de maximum de vraisemblance conditionnelle :

$$\max_{\beta} \prod_{i \in G_1} \frac{\exp\left(\text{Log} \frac{p_1}{p_2} + \beta_0 + \underline{\beta}' \underline{x}_i\right)}{1 + \exp\left(\text{Log} \frac{p_1}{p_2} + \beta_0 + \underline{\beta}' \underline{x}_i\right)} \prod_{i \in G_2} \frac{1}{1 + \exp\left(\text{Log} \frac{p_1}{p_2} + \beta_0 + \underline{\beta}' \underline{x}_i\right)}$$

Nécessité d'utiliser une méthode numérique.

(Pas de solution analytique à l'équation de vraisemblance).

Les β étant estimés, la règle Bayésienne peut être appliquée pour les classements.

$$\text{Log} \frac{P(G_1|\underline{x})}{P(G_2|\underline{x})} = \text{Log} \frac{p_1}{p_2} + \beta_0 + \underline{\beta}' \underline{x}$$

On affectera au groupe 1 si $\text{Log} \frac{p_1}{p_2} + \beta_0 + \underline{\beta}' \underline{x} > 0$

Avantages - Inconvénients de la régression logistique

Résultats meilleurs que la règle géométrique, pour :

- des populations non gaussiennes
- des populations où Σ_1 très différent de Σ_2
mais procédure de calcul plus complexe.

Lorsque les données proviennent de deux populations normales avec $\Sigma_1 = \Sigma_2$ la régression logistique est moins performante que l'analyse discriminante. Seul ($\frac{f_1}{f_2}$ supposé connu).

BIBLIOGRAPHIE CONCERNANT LES METHODES D'ANALYSE DISCRIMINANTE ET DE SEGMENTATION

Références générales en statistique

G. GOVAERT (Editeur) “ **Analyse des données** ” Hermés Lavoisier (2003)

L. LEBART, A. MORINEAU, M. PIRON
“ **Statistique exploratoire multidimensionnelle** ” 3^{ème} édition Dunod (2000)

G. SAPORTA “**Probabilités, analyse des données et statistique**” 2^{ème} édition Technip (2006).

S. TUFFERY “**Data mining et statistique décisionnelle**” Technip 2012

**S.TUFFERY « Étude de cas en statistique décisionnelle » Technip
2009**

**M. TENENHAUS "Statistique: Méthode pour décrire, expliquer et
prévoir ". Dunod (2006).**

Analyse discriminante et Segmentation

**BARDOS M. « Analyse discriminante : Application au risque et
scoring financier » Dunod (2001)**

**Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.J. “ Classification
And Regression Trees. ” Monterey, California, Wadsworth & Brooks
(1984)**

**CELEUX G. (Editeur scientifique) “ Analyse discriminante sur
variables continues ” Collection didactique INRIA (1990)**

**CELEUX G ;, NAKACHE J.P.“ Analyse discriminante sur variables
qualitatives ” Polytechnica (1994)**

DROESBEKE J-J., LEJEUNE M., SAPORTA G. (Editeurs) “ **Modèles statistiques explicative pour données qualitatives** ” Technip (2005)

HUBERTY C. “**Applied discriminant analysis**” Wiley (1994)

NAKACHE J-P., CONFAIS J. “ **Statistique explicative appliquée** ” Technip (2003)

TOMASSONE R., DANZART M., DAUDIN J.J., MASSON J.P. “ **Discrimination et classement** ” Masson (1988)

ZIGHED D.A., RAKOTOMALALA R. “ **Graphes d’induction** ” Hermès (2000)

Sites INTERNET

Le site de la Société Française de Statistique : www.sfds.asso.fr

L’aide en ligne du logiciel SAS : <http://support.sas.com/documentation/online.doc>

Le site de Statsoft sur la statistique et le data mining : www.statsoft.com