

Analyse Factorielle Discriminante (AFD)

Marie Chavent

2015-2016

On se place dans le cadre de la modélisation d'une variable Y **qualitative** à K modalités à partir de p variables explicatives X_1, \dots, X_p **quantitatives**. On suppose que l'on dispose d'un échantillon de taille n pour lequel les p variables explicatives et la variable à expliquer ont été mesurées simultanément. On se place donc dans un cadre dit **supervisé**, où chaque modalité de Y représente une classe (un groupe) d'individus que l'on cherche à discriminer. L'objectif de l'analyse factorielle discriminante (AFD) est alors essentiellement descriptif. Il s'agit de chercher quelles sont les combinaisons linéaires des variables quantitatives qui permettent de séparer le mieux possible les K modalités et de donner une représentation graphique (comme pour les méthodes factorielles telles que l'analyse en composantes principales) qui rende au mieux compte de cette opération. Cette visualisation sur le ou les plans factoriels permet aussi de décrire les liaisons entre la variable à expliquer Y et les p variables explicatives.

Exemple. On dispose ici de données où $n = 23$ poissons sont répartis en trois groupes selon leur site de pêche (site 1, site 2, site 3). La variable qualitative à expliquer Y est la variable *site* qui possède 3 trois modalités (site 1, site 2, site 3). Sur ces 23 poissons, on a mesuré les $p = 14$ variables quantitatives suivantes :

- $X_1 = \text{YEU}$ (radioactivité des yeux),
- $X_2 = \text{BR}$ (radioactivité des branchies),
- $X_3 = \text{OP}$ (radioactivité des opercules),
- $X_4 = \text{NAG}$ (radioactivité des nageoires),
- $X_5 = \text{FOI}$ (radioactivité du foie),
- $X_6 = \text{TUB}$ (radioactivité du tube digestif),
- $X_7 = \text{EC}$ (radioactivité des écailles),
- $X_8 = \text{MUS}$ (radioactivité des muscles),
- $X_9 = \text{POI}$ (poids),

- X_{10} = LON (longueur),
- X_{11} = LART (largeur de la tête),
- X_{12} = LAR (largeur),
- X_{13} = LARM (largeur du museau),
- X_{14} = DYEU (diamètre des yeux).

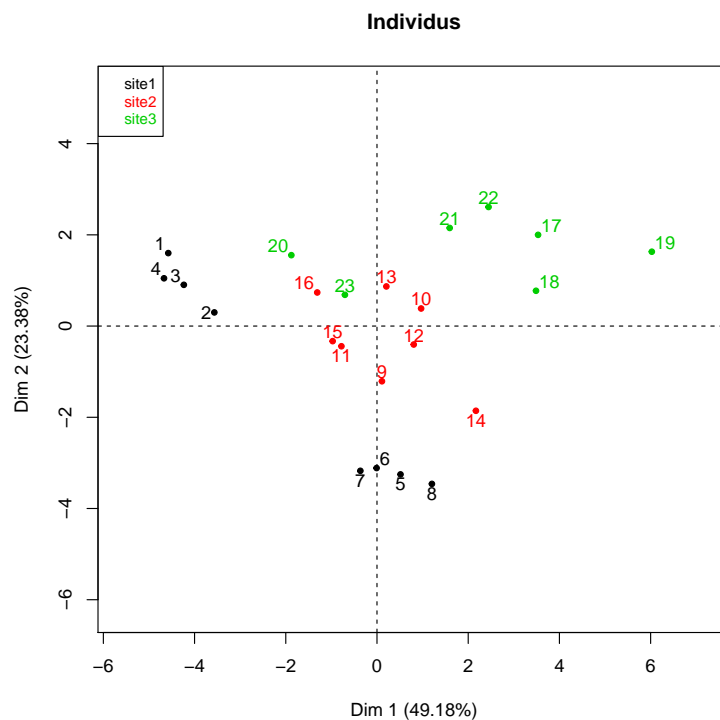
On veut pouvoir décrire Y ou encore expliquer l'appartenance \tilde{A} un site de pêche, en fonction de ces 14 variables explicatives X_1, \dots, X_{14}

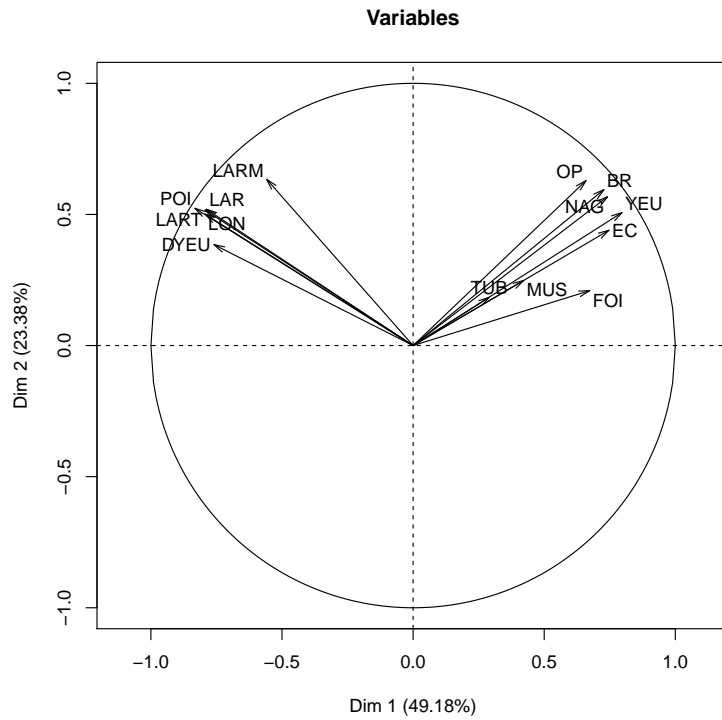
```
load("poissons.rda")
head(poissons)

##      site YEU BR OP NAG FOI TUB  EC MUS POI LON LART LAR LARM DYEU
## 1 site1 10 65 65 107  7  76 142  1 132 214  54  47  18  11
## 2 site1  9 43 39  67 29 113  99  2 122 220  49  44  16  10
## 3 site1  6 47 71  95 11 192 121  2 129 220  49  45  17  11
## 4 site1  7 70 40  66  8 310  90  2 133 225  52  48  15  11
## 5 site1  8 59 67 100 14 289 244  1  57 168  37  37  9  9
## 6 site1  8 46 55 112 17 115 153  1  59 178  38  35  11  9

X <- subset(poissons,select=-site)
y <-poissons[,1]
```

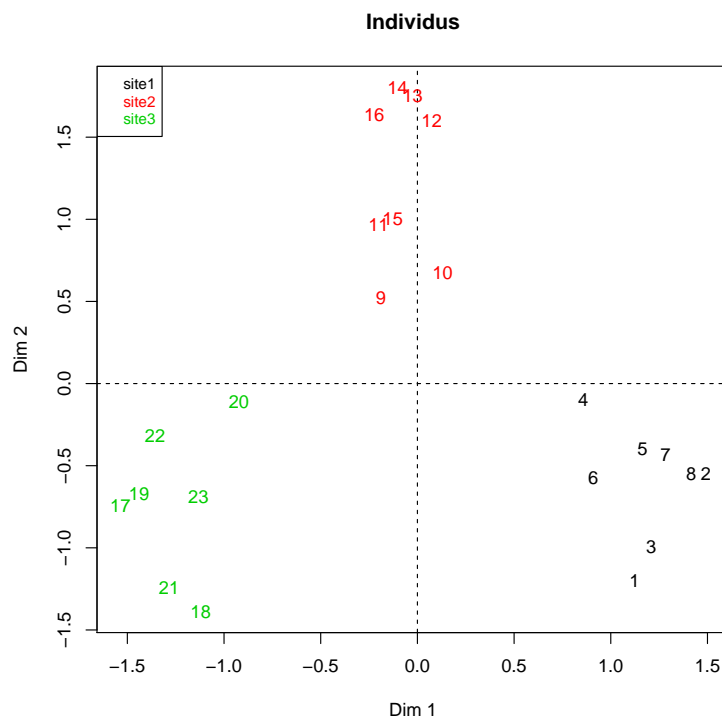
- Premier plan factorielle de l'ACP des 14 variables explicatives

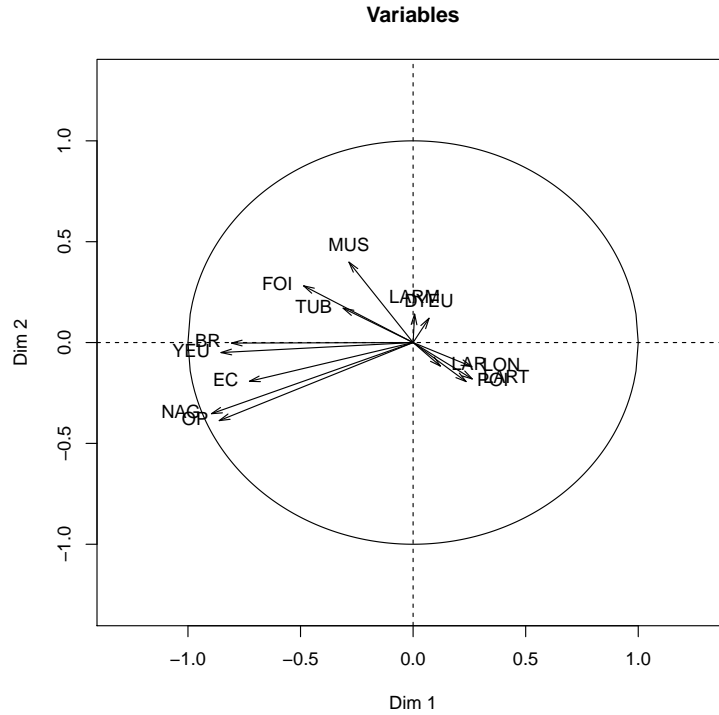




— Premier plan factoriel de l'AFD des 14 variables explicatives avec la variable *site* comme variable \tilde{A} expliquée.

```
source("AFD_procedures.R")
afd <- AFD(X,y)
plotAFD(afd)
```





1 Les données

On dispose d'un échantillon de n observations de Y et de $X = (X_1, \dots, X_p)$: sur les n individus de l'échantillon, on a mesuré une variable qualitative à K modalités et p variables quantitatives. En notant y_i la valeur de la variable à expliquer mesurée sur les i ème individu, on obtient le vecteur $\mathbf{y} = (y_1, \dots, y_n)' \in \{1, \dots, K\}^n$. En notant x_{ij} la valeur de la j ème variable explicative mesurée sur le i ème individu, on obtient ainsi la matrice de données de dimension $n \times p$ suivante :

$$\mathbf{X} = \begin{array}{c|ccc|} & 1 & \dots & j & \dots & p \\ \hline 1 & & & & & \\ \vdots & & & \vdots & & \\ i & \dots & & x_{ij} & \dots & \\ \vdots & & & \vdots & & \\ n & & & & & \\ \hline \end{array}$$

Notons :

- $x_i = (x_{i1}, \dots, x_{ip})' \in \mathbb{R}^p$ une ligne de \mathbf{X} décrivant le i ème individu.

- $x^j = (x_{1j}, \dots, x_{nj})' \in \mathbb{R}^n$ une colonne de \mathbf{X} décrivant la j^{e} me variable.
- G_k est le groupe des individus de l'échantillon qui possèdent la modalité k .
- $n_k = \text{card}(G_k)$ est le nombre d'individus qui possèdent la modalité k .

Si les n individus sont affectés des poids p_1, \dots, p_n , (tels que $\forall i = 1, \dots, n, p_i \geq 0$ et $\sum_{i=1}^n p_i = 1$) alors le poids de chaque groupe G_k est :

$$P_k = \sum_{i \in G_k} p_i$$

En général, on prend $p_i = \frac{1}{n}$ et donc $\mu_k = \frac{n_k}{n}$. On a alors les définitions suivantes :

- Le centre de gravité global est le vecteur de \mathbb{R}^p défini par :

$$\mathbf{g} = \sum_{i=1}^n p_i x_i = \frac{1}{n} \sum_{i=1}^n x_i.$$

- Le centre de gravité du groupe G_k est le vecteur de \mathbb{R}^p défini par :

$$\mathbf{g}_k = \frac{1}{\mu_k} \sum_{i \in G_k} p_i x_i = \frac{1}{n_k} \sum_{i \in G_k} x_i.$$

- La matrice $p \times p$ de variance-covariance globale est définie par :

$$\mathbf{V} = \sum_{i=1}^n p_i (x_i - \mathbf{g})(x_i - \mathbf{g})' = \frac{1}{n} \sum_{i=1}^n (x_i - \mathbf{g})(x_i - \mathbf{g})'.$$

- La matrice $p \times p$ de variance-covariance du groupe G_k est définie par :

$$\mathbf{V}_k = \frac{1}{\mu_k} \sum_{i \in G_k} p_i (x_i - \mathbf{g}_k)(x_i - \mathbf{g}_k)' = \frac{1}{n_k} \sum_{i \in G_k} (x_i - \mathbf{g}_k)(x_i - \mathbf{g}_k)'.$$

- La matrice $p \times p$ de variance-covariance intra-groupe est définie par :

$$\mathbf{W} = \sum_{k=1}^K \mu_k \mathbf{V}_k = \sum_{k=1}^K \frac{n_k}{n} \mathbf{V}_k.$$

- La matrice $p \times p$ de variance-covariance inter-groupe est définie par :

$$\mathbf{B} = \sum_{k=1}^K \mu_k (\mathbf{g}_k - \mathbf{g})(\mathbf{g}_k - \mathbf{g})' = \sum_{k=1}^K \frac{n_k}{n} (\mathbf{g}_k - \mathbf{g})(\mathbf{g}_k - \mathbf{g})'.$$

Cette matrice est la matrice de variance-covariance des p centres de gravités \mathbf{g}_k pondérés par μ_k (en général $\frac{n_k}{n}$).

Relations fondamentales. On a les relations classiques suivantes :

$$\mathbf{g} = \sum_{k=1}^K \mu_k \mathbf{g}_k = \sum_{k=1}^K \frac{n_k}{n} \mathbf{g}_k, \quad (1)$$

et

$$\mathbf{V} = \mathbf{W} + \mathbf{B}. \quad (2)$$

Les expressions données en (1) et (2) sont de “nature” assez usuelles en statistique :

- moyenne totale (globale) = moyenne (pondérée) des moyennes marginales,
- variance totale = moyenne (pondérée) des variances marginales + variance des moyennes marginales.

Exercice : démontrez (1) et (2).

Centrage des données. En AFD comme en analyse en composantes principales (ACP), on suppose que $\mathbf{g} = \mathbf{0}_p$, c’est à dire que les données sont centrées. Cela revient à translater le nuage de points de \mathbf{g} (ancien centre de gravité) vers l’origine. La “structure” du nuage de points n’est pas modifiée, mais les calculs ultérieurs seront simplifiés. En particulier, l’écriture des matrices de variance-covariance globale et inter-groupe est simplifiée :

$$\mathbf{V} = \sum_{i=1}^n \frac{1}{n} x_i x_i' \quad \text{et} \quad \mathbf{B} = \sum_{k=1}^K \frac{n_k}{n} \mathbf{g}_k \mathbf{g}_k',$$

et on obtient l’écriture matricielle $\mathbf{V} = \frac{1}{n} \mathbf{X}' \mathbf{X}$.

Exemple. Les matrices de variance-covariance totale, intra et inter-groupe pour les données sur les poissons sont trois matrices de dimension 14×14 .

```
# Centres de gravité's
data.frame(afd$g, afd$gk)[1:5,]

##      g site1 site2 site3
## YEU  15   8.2   16   24
## BR  105  57.0  108  156
## OP  109  52.4   80  208
## NAG 165  91.1  133  285
## FOI  27  15.2   34   34

# Matrice de variance covariance totale
afd$V[1:5,1:5]
```

```

##      YEU   BR   OP   NAG FOI
## YEU  54  327  455  560  83
## BR   327 2549 3021 3626 478
## OP   455 3021 5215 6028 430
## NAG  560 3626 6028 7577 696
## FOI  83  478  430  696 259

# Matrice de variance covariance intra-groupe
afd$W[1:5,1:5]

##      YEU   BR   OP   NAG FOI
## YEU  16   80   69   78   37
## BR   80  944  566  549  170
## OP   69  566  823  623   64
## NAG  78  549  623  911  220
## FOI  37  170   64  220  183

# Matrice de variance covariance inter-groupe
afd$B[1:5,1:5]

##      YEU   BR   OP   NAG FOI
## YEU  38  247  386  482  46
## BR  247 1606 2455 3077 308
## OP  386 2455 4392 5405 367
## NAG 482 3077 5405 6666 476
## FOI  46  308  367  476  76

```

2 Axes, facteurs et variables discriminantes

L'AFD consiste à rechercher 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'individus.

L'objectif. Il s'agit de trouver une nouvelle variable, combinaison linéaire des variables explicatives, qui “discrimine” au mieux les groupes définis par les modalités de la variable à expliquer. Cette variable notée \mathbf{s} est définie ici comme un vecteur de \mathbb{R}^n , combinaison linéaire des vecteurs x^1, \dots, x^p (colonnes de \mathbf{X}) :

$$\mathbf{s} = \mathbf{X}\mathbf{u} = u_1x^1 + \dots + u_px^p,$$

où $\mathbf{u} = (u_1, \dots, u_p)' \in \mathbb{R}^p$ est le vecteur des coefficients de cette combinaison linéaire. Il va donc falloir définir :

- comment mesurer que \mathbf{s} “discrimine” bien,
- comment trouver \mathbf{u} pour que $\mathbf{s} = \mathbf{X}\mathbf{u}$ “discrimine” au mieux.

Définitions. En AFD, on muni \mathbb{R}^p d’une métrique \mathbf{M} et on projette les n points de \mathbb{R}^p x_1, \dots, x_n (les n lignes de \mathbf{X}) sur un axe Δ de vecteur directeur \mathbf{a} . On effectue des projections \mathbf{M} -orthogonales et \mathbf{a} est \mathbf{M} -normé à 1. La liste des coordonnées $s_i = x_i' \mathbf{M} \mathbf{a}$ des individus sur Δ forme la nouvelle variable \mathbf{s} et on a donc :

$$\mathbf{s} = (s_1, \dots, s_n)' = \mathbf{X} \mathbf{M} \mathbf{a} = \mathbf{X} \mathbf{u},$$

où

- $\mathbf{a} \in \mathbb{R}^p$ est appelé l’axe discriminant ($\mathbf{a}' \mathbf{M} \mathbf{a} = 1$),
- $\mathbf{u} \in \mathbb{R}^p$ est appelé le facteur discriminant ($\mathbf{u}' \mathbf{M}^{-1} \mathbf{u} = 1$),
- \mathbf{s} est appelé la variable discriminante.

2.1 Le critère à optimiser

On définit maintenant le critère qui mesure la capacité d’un axe à discriminer les groupes. Un axe sera discriminant si les groupes sont bien séparés en projection. Pour cela, on utilise la variance intra-groupe et la variance inter-groupe de la variable discriminante \mathbf{s} constituée des coordonnées des projections des individus sur l’axe. On a les définitions suivantes :

- La variance de \mathbf{s} est définie par :

$$\begin{aligned} Var(\mathbf{s}) &= \frac{1}{n} \sum_{i=1}^n (s_i - \bar{s})^2 \\ &= \mathbf{u}' \mathbf{V} \mathbf{u} \end{aligned} \tag{3}$$

où \bar{s} est la moyenne de s .

- La variance intra-groupe de \mathbf{s} est définie par :

$$\begin{aligned} Intra(\mathbf{s}) &= \sum_{k=1}^K \frac{n_k}{n} \sum_{i \in E_k} \frac{1}{n_k} (s_i - \bar{s}_k)^2 \\ &= \mathbf{u}' \mathbf{W} \mathbf{u} \end{aligned} \tag{4}$$

où \bar{s}_k est la moyenne de s dans le groupe k .

— La variance inter-groupe de \mathbf{s} est définie par :

$$\begin{aligned} Inter(\mathbf{s}) &= \sum_{k=1}^K \frac{n_k}{n} (\bar{\mathbf{s}} - \bar{\mathbf{s}}_k)^2 \\ &= \mathbf{u}'\mathbf{B}\mathbf{u} \end{aligned} \quad (5)$$

Exercice : Démontrez (3) (4) et (5).

On a donc la relation classique $Var(\mathbf{s}) = Intra(\mathbf{s}) + Inter(\mathbf{s})$ qui s'écrit :

$$\mathbf{u}'\mathbf{V}\mathbf{u} = \mathbf{u}'\mathbf{W}\mathbf{u} + \mathbf{u}'\mathbf{B}\mathbf{u} \quad (6)$$

La quantité $\mathbf{u}'\mathbf{V}\mathbf{u}$ étant indépendante des groupes, minimiser $\mathbf{u}'\mathbf{W}\mathbf{u}$ est équivalent à maximiser $\mathbf{u}'\mathbf{B}\mathbf{u}$.

On aura une bonne discrimination des groupes si :

- les centres de gravité projetés sont bien éloignés i.e. $Inter(\mathbf{s}) = \mathbf{u}'\mathbf{W}\mathbf{u}$ est maximum,
- les groupes projetés ne sont pas trop dispersés i.e. $Intra(\mathbf{s}) = \mathbf{u}'\mathbf{B}\mathbf{u}$ est minimum.

Le critère à maximiser,

$$\frac{\mathbf{u}'\mathbf{B}\mathbf{u}}{\mathbf{u}'\mathbf{V}\mathbf{u}} \in [0, 1], \quad (7)$$

est la proportion de la variance de \mathbf{s} expliquée par \mathbf{y} encore appelé le rapport de corrélation entre \mathbf{s} et \mathbf{y} . Ce critère désigne également le pouvoir discriminant de l'axe Δ . Dans le cadre de la discrimination parfaite, ce rapport serait égal à 1. Plus généralement, la discrimination est d'autant meilleure que le ratio est proche de 1.

2.2 Solution du problème d'optimisation

On veut maximiser le rapport entre la variance inter-groupe et la variance totale. La première variable discriminante $\mathbf{s} = \mathbf{X}\mathbf{u}$ est telle que le facteur discriminant \mathbf{u} vérifie :

$$\max_{\mathbf{u} \in \mathbb{R}^p} \frac{\mathbf{u}'\mathbf{B}\mathbf{u}}{\mathbf{u}'\mathbf{V}\mathbf{u}}. \quad (8)$$

Rappels sur la maximisation du quotient de deux formes quadratiques. Soient \mathbf{A} et \mathbf{B} deux matrices symétriques de même dimension, \mathbf{B} étant supposée inversible. Alors, le rapport $\frac{\mathbf{u}'\mathbf{A}\mathbf{u}}{\mathbf{u}'\mathbf{B}\mathbf{u}}$ est maximal pour u_1 vecteur propre de $\mathbf{B}^{-1}\mathbf{A}$ associé à la plus grande valeur propre notée λ_1 , la valeur du maximum étant λ_1 .

Exercice : Démontrez le résultat ci-dessus.

On déduit du rappel ci-dessus que, pour notre problème d'optimisation (8), le vecteur \mathbf{u} solution est le vecteur propre de $\mathbf{V}^{-1}\mathbf{B}$ associé à la plus grande valeur propre λ_1 .

En pratique, si on utilise la métrique $\mathbf{M} = \mathbf{V}^{-1}$ par exemple, on peut choisir parmi tous les vecteurs propres colinéaires, celui qui est \mathbf{V} -normé à 1 ($\mathbf{u}'\mathbf{V}\mathbf{u} = Var(\mathbf{s}) = 1$). Si on utilise une autre métrique \mathbf{M} , l'axe discriminant $\mathbf{a} = \mathbf{M}^{-1}\mathbf{u}$ est modifié mais le facteur discriminant \mathbf{u} est le même (à une constante près) et la variable discriminante $\mathbf{s} = \mathbf{X}\mathbf{M}\mathbf{a} = \mathbf{X}\mathbf{u}$ (coordonnées des projections des individus sur la droite engendrée par l'axe discriminant) est donc également la même (à une constante près). La valeur du maximum, c'est à dire le pouvoir discriminant, quand à lui est inchangé. Le choix de la métrique \mathbf{M} est donc indifférente pour la construction de variables et des facteurs discriminants, seule la norme change. Avec la métrique $\mathbf{M} = \mathbf{W}^{-1}$ par exemple, si \mathbf{u} est \mathbf{W} -normé à 1, alors $\mathbf{u}'\mathbf{W}\mathbf{u} = Intra(\mathbf{s}) = 1$.

On a alors les définitions suivantes :

- Le premier facteur discriminant est \mathbf{u}_1 , le premier vecteur propre de $\mathbf{V}^{-1}\mathbf{B}$.
- La première variable discriminante est $\mathbf{s}_1 = \mathbf{X}\mathbf{u}_1$.
- Le pouvoir discriminant est $\lambda_1 = \frac{\mathbf{u}_1'\mathbf{B}\mathbf{u}_1}{\mathbf{u}_1'\mathbf{V}\mathbf{u}_1}$.

Remarques :

- Cas où $\lambda_1 = 1$: en projection sur la droite discriminante Δ , les dispersions intra-groupe sont nulles ; cela correspond au fait que les K sous-nuages sont donc chacun dans des hyperplans orthogonaux à Δ . Il y a alors évidemment discrimination parfaite si les centres de gravités se projettent en des points différents sur la droite Δ .
- Cas où $\lambda_1 = 0$: ici le meilleur axe discriminant ne permet pas de séparer les K centres de gravité g_k . C'est par exemple le cas si ces derniers sont confondus ; cela peut correspondre au cas de figure où les K sous-nuages sont concentriques et donc aucune séparation linéaire n'est possible. Remarquons qu'il peut cependant exister une possibilité de discrimination non linéaire (du type quadratique par exemple, si l'on pense à la distance au centre de gravité qui permet de séparer les groupes dans le cas de sous-nuages concentriques).
- Notons enfin que la valeur propre λ est une mesure pessimiste du pouvoir discriminant. En effet, même s'il est possible de discriminer parfaitement les groupes, on peut avoir $\lambda < 1$.

La décomposition spectrale de $\mathbf{V}^{-1}\mathbf{B}$ donne aussi les autres axes discriminants de l'AFD. En effet, on cherche ensuite $\mathbf{s}_2 = \mathbf{X}\mathbf{u}_2$, non corrélée à \mathbf{s}_1 , telle que $\frac{\mathbf{u}_2'\mathbf{B}\mathbf{u}_2}{\mathbf{u}_2'\mathbf{V}\mathbf{u}_2}$ soit maximum et ainsi de suite. On peut montrer que les vecteurs propres de $\mathbf{V}^{-1}\mathbf{B}$ sont les facteurs discriminants et les valeurs propres $\lambda_1, \dots, \lambda_{K-1}$ sont les pouvoirs discriminants.

On pourra construire $K-1$ axes discriminants car le rang de $\mathbf{V}^{-1}\mathbf{B}$ est au plus $\min(p, K-1)$ et on suppose que $K-1 < p$.

Remarques :

- Il faut que la matrice de variance-covariance \mathbf{V} soit inversible et que donc la matrice de données \mathbf{X} soit de plein rang. Il faudra donc faire attention en AFD, comme en régression linéaire multiple, au problème de la multicollinéarité des colonnes de \mathbf{X} .
- L'AFD peut être vue comme une ACP (analyse en composante principale) des centres de gravités g_k avec la métrique \mathbf{V}^{-1} .
- L'AFD est aussi un cas particulier de l'analyse canonique. Elle correspond en effet à l'analyse canonique des tableaux \mathbf{A} et \mathbf{X} , où \mathbf{A} (de dimension $n \times K$) représente l'ensemble des variables indicatrices associées à la variable qualitative à expliquer.

2.3 Les anglo-saxons

En AFD, d'autres critères ont été utilisés pour définir les facteurs discriminants et donc les variables discriminantes. Les anglo-saxons utilisent souvent comme critère pour mesurer la qualité de la discrimination d'une variable discriminante $\mathbf{s} = \mathbf{X}\mathbf{M}\mathbf{a} = \mathbf{X}\mathbf{u}$:

$$\frac{\mathbf{u}'\mathbf{B}\mathbf{u}}{\mathbf{u}'\mathbf{W}\mathbf{u}} = \frac{Inter(\mathbf{s})}{Intra(\mathbf{s})}.$$

Le problème d'optimisation s'écrit :

$$\max_{\mathbf{u} \in \mathbb{R}^p} \frac{\mathbf{u}'\mathbf{B}\mathbf{u}}{\mathbf{u}'\mathbf{W}\mathbf{u}}. \quad (9)$$

La solution est le vecteur propre \mathbf{u} de $\mathbf{W}^{-1}\mathbf{B}$ associé à la plus grande valeur propre μ .

Ce problème d'optimisation est équivalent en terme de facteurs discriminants au problème (8) mais les maximums sont différents. En d'autres termes, les matrices $\mathbf{W}^{-1}\mathbf{B}$ et $\mathbf{V}^{-1}\mathbf{B}$ ont les mêmes vecteurs propres et $\mu = \frac{\lambda}{1-\lambda}$.

Prevue : On suppose \mathbf{W} inversible.

\mathbf{u} vecteur propre de $\mathbf{V}^{-1}\mathbf{B} \Rightarrow \mathbf{V}^{-1}\mathbf{B}\mathbf{u} = \lambda\mathbf{u}$

$$\Rightarrow \mathbf{B}\mathbf{u} = \lambda\mathbf{V}\mathbf{u} = \lambda(\mathbf{B} + \mathbf{W})\mathbf{u}$$

$$\Rightarrow (1 - \lambda)\mathbf{B}\mathbf{u} = \lambda\mathbf{W}\mathbf{u}$$

$$\Rightarrow \mathbf{B}\mathbf{u} = \frac{\lambda}{1 - \lambda}\mathbf{W}\mathbf{u}$$

$$\Rightarrow \mathbf{W}^{-1}\mathbf{B}\mathbf{u} = \frac{\lambda}{1 - \lambda}\mathbf{u}$$

$$\Rightarrow \mathbf{u} \text{ vecteur propre de } \mathbf{W}^{-1}\mathbf{B} \text{ et } \frac{\lambda}{1 - \lambda} \text{ est la valeur propre } \mu$$

■

On retrouve donc les mêmes variables discriminantes qu'avec l'autre critère (au signe et à une constante près). En revanche le critère λ varie entre 0 et 1 et s'interprète facilement tandis que le critère μ varie entre zero et l'infini.

Exemple. On reprend l'exemple des poissons pour vérifier les liens entre les deux approches.

- AFD à la française.

```
res1 <- AFD(X,y)
#pouvoirs discriminants
res1$eig
## [1] 0.97 0.83
#variables discriminantes
head(res1$S)
##      s1      s2
## 1 1.12 -1.201
## 2 1.49 -0.546
## 3 1.21 -0.995
## 4 0.86 -0.095
## 5 1.17 -0.397
## 6 0.91 -0.576
#facteurs discriminants
res1$U[1:5,]
##          u1          u2
## YEU -0.1135  0.1450
## BR  -0.0029  0.0093
## OP   0.0114 -0.0124
## NAG -0.0122 -0.0120
## FOI  0.0170  0.0012
```

- AFD à l'anglo-saxonne des poissons.

```

res2 <- AFD(X,y,type="GB")
#pouvoirs discriminants
res2$eig
## [1] 30.4 4.8
#variables discriminantes
head(res2$S)
##   s1   s2
## 1 5.9 -2.71
## 2 7.8 -1.23
## 3 6.3 -2.24
## 4 4.5 -0.21
## 5 6.1 -0.89
## 6 4.7 -1.30
#facteurs discriminants
res2$U[1:5,]
##      u1    u2
## YEU -0.593 0.3267
## BR  -0.015 0.0210
## OP   0.060 -0.0280
## NAG -0.064 -0.0271
## FOI  0.089 0.0027

```

- Comparaison des deux.

```

res1$eig/(1-res1$eig)
## [1] 30.4 4.8
#correlations entre les variables et facteurs discriminants des deux approches
comp <- function(X,Y)
{
  corcol <- rep(NA,ncol(X))
  names(corcol) <- colnames(X)
  for (j in 1:ncol(X))
    corcol[j] <- cor(X[,j],Y[,j])
  return(corcol)
}
comp(res1$U,res2$U)
## u1 u2
## 1 1
comp(res1$S,res2$S)
## s1 s2
## 1 1

```

- Avec la fonction lda.

```

library(MASS)
res3 <- lda(X,y)
#facteurs discriminants
head(res3$scaling)
##      LD1      LD2
## YEU  0.59318 -0.32671
## BR   0.01527 -0.02102
## OP  -0.05950 0.02798

```

```
## NAG 0.06363 0.02709
## FOI -0.08887 -0.00267
## TUB -0.00086 -0.00084
head(predict(res3)$x)
## LD1 LD2
## 1 -5.9 2.71
## 2 -7.8 1.23
## 3 -6.3 2.24
## 4 -4.5 0.21
## 5 -6.1 0.89
## 6 -4.7 1.30
```

On retrouve bien directement en sortie de cette fonction les variables et les facteurs discriminants mais pas les pouvoirs discriminants ni les représentations graphiques des observations et des corrélations entre les variables et ces variables. Mais tout se retrouve...