5 Partitionnement - Différenciation - Extremum

5.1 Partitionnement

Il est souvent utile de considérer (quand des lignes ou des colonnes ont des caractéristiques communes) qu'une matrice est constituée d'éléments qui sont eux-mêmes des matrices. De telles matrices s'écrivent :

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1p} \\ A_{21} & A_{22} & \cdots & A_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{np} \end{bmatrix}$$

A est une matrice *partitionnée*.

Les matrices A_{ij} à n_i lignes et p_j colonnes sont appelées *sous-matrices*.

Toutes les sous-matrices d'une ligne donnée de A ont le même nombre de lignes. De même, toutes les sous-matrices d'une colonne donnée de A ont le même nombre de colonnes.

Les opérations sur des matrices partitionnées ressemblent à celles déjà présentées, la seule différence réside dans la nature non scalaire des éléments de la matrice.

5.1.1 Opérations élémentaires

1 - *Transposition* : On note A' la matrice transposée de la matrice A.

$$A' = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1p} \\ A_{21} & A_{22} & \cdots & A_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{np} \end{bmatrix}' = \begin{bmatrix} A'_{11} & A'_{21} & \cdots & A'_{n1} \\ A'_{12} & A'_{22} & \cdots & A'_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ A'_{1p} & A'_{2p} & \cdots & A'_{np} \end{bmatrix}$$

2 - *Addition* : Soient A et B deux matrices partitionnées formées de sous-matrices de *dimensions similaires*. L'addition $A + B$ est :

$$A + B = \begin{bmatrix} A_{11} + B_{11} & A_{12} + B_{12} & \cdots & A_{1p} + B_{1p} \\ A_{21} + B_{21} & A_{22} + B_{22} & \cdots & A_{2p} + B_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ A_{n1} + B_{n1} & A_{n2} + B_{n2} & \cdots & A_{np} + B_{np} \end{bmatrix}$$

3 - *Multiplication* : Quand les dimensions des sous-matrices de A et B sont conformes à la règle du produit matriciel, le produit de A par B s'écrit :

$$AB = \begin{bmatrix} \sum_{i=1}^{p} A_{1i}B_{i1} & \cdots & \sum_{i=1}^{p} A_{1i}B_{im} \\ \cdots & \cdots & \cdots \\ \sum_{i=1}^{p} A_{ni}B_{i1} & \cdots & \sum_{i=1}^{p} A_{ni}B_{im} \end{bmatrix}$$

4 - *Inversion* : Il est possible d'exprimer l'inverse d'une matrice partitionnée non singulière en termes de sous-matrices. Un cas important en statistique est le suivant :

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \qquad (1)$$

où A_{11} et A_{22} sont toutes deux des sous-matrices carrées non singulières. Si on note :

$$B = A^{-1} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

on peut vérifier que :

$$\begin{aligned} B_{11} &= (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} \\ B_{22} &= A_{22}^{-1} + A_{22}^{-1}A_{21}B_{11}A_{12}A_{22}^{-1} \\ B_{12} &= -A_{11}^{-1}A_{12}B_{22} \\ B_{21} &= -A_{22}^{-1}A_{21}B_{11} \end{aligned}$$

Une autre expression de A^{-1} peut être obtenue en permutant les rôles de A_{11} et A_{22}. On déduit alors de ces deux expressions de A^{-1} une identité matricielle très utile :

$$(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} = A_{11}^{-1} + A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}A_{11}^{-1}$$

En particulier, si A est une matrice non singulière $(n \times n)$, B un vecteur $(n \times 1)$, α un scalaire, on a l'identité de Bartlett (1951) :

$$(A + \alpha BB')^{-1} = A^{-1} - \frac{\alpha}{1 + \alpha B' A^{-1} B} A^{-1} BB' A^{-1}$$

5 - *Déterminant* : Il est parfois nécessaire de calculer le déterminant de la matrice partitionnée (1). Si A_{11} est non singulière,

$$det(A) = det(A_{11}) \times det(A_{22} - A_{21} A_{11}^{-1} A_{12})$$

Si A_{22} est non singulière,

$$det(A) = det(A_{22}) \times det(A_{11} - A_{12} A_{22}^{-1} A_{21})$$

5.1.2 Exemple 1 : inverse et déterminant d'une matrice partitionnée

Calculer l'inverse et le déterminant de :

$$A = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 2 & 2 & 2 \\ 3 & -1 & 4 & 0 & 0 \\ 3 & -1 & 0 & 4 & 0 \\ 3 & -1 & 0 & 0 & 4 \end{bmatrix}$$

Réponse

$$A_{11}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad A_{22}^{-1} = \begin{bmatrix} \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{4} \end{bmatrix}$$

$$A_{11} - A_{12} A_{22}^{-1} A_{21} = \begin{bmatrix} -\frac{5}{4} & \frac{3}{4} \\ -\frac{9}{2} & \frac{5}{2} \end{bmatrix}$$

$$(A_{11} - A_{12} A_{22}^{-1} A_{21})^{-1} = \begin{bmatrix} 10 & -3 \\ 18 & -5 \end{bmatrix}$$

$$A_{12} A_{22}^{-1} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \qquad A_{22}^{-1} A_{21} = \begin{bmatrix} \frac{3}{4} & -\frac{1}{4} \\ \frac{3}{4} & -\frac{1}{4} \\ \frac{3}{4} & -\frac{1}{4} \end{bmatrix}$$

Ainsi,

$$A^{-1} = \begin{bmatrix} 10 & -3 & -1 & -1 & -1 \\ 18 & -5 & -2 & -2 & -2 \\ -3 & 1 & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ -3 & 1 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ -3 & 1 & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

$$det(A) = 6$$

5.2 Différenciation

On a parfois besoin de dériver des fonctions de vecteurs ou de matrices. Si on désigne par :

1 - $f(X)$ une fonction continue des éléments d'un vecteur

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

2 - $\frac{\partial f(X)}{\partial x_i}$, $\frac{\partial^2 f(X)}{\partial x_i \partial x_j}$ les dérivées partielles première et seconde.

Le vecteur des dérivées partielles est :

$$\frac{\partial f(X)}{\partial X} = \begin{bmatrix} \frac{\partial f(X)}{\partial x_1} \\ \frac{\partial f(X)}{\partial x_2} \\ \vdots \\ \frac{\partial f(X)}{\partial x_n} \end{bmatrix}$$

1 - Si $f(X)$ est constante pour tout X :

$$\frac{\partial f(X)}{\partial X} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

2 - Soit C un vecteur $(n \times 1)$ constant et soit

$$f(X) = C'X = [\, c_1 \quad c_2 \quad \cdots \quad c_n \,] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$\frac{\partial f(X)}{\partial X} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$$

La forme colonne du vecteur dérivée est inchangée si on écrit $f(X) = X'C$.

3 - Vecteur des dérivées d'une forme quadratique $X'AX$:

$$\frac{\partial X'AX}{\partial X} = 2AX$$

Plus généralement, si $f(X) = (C - BX)'A(C - BX)$ avec A une matrice symétrique de dimension $(p \times p)$, $B = [B_1 \ \cdots \ B_n]$ une matrice de constantes de dimension $(p \times n)$ et C un vecteur de dimension $(p \times 1)$ alors :

$$\frac{\partial f(X)}{\partial X} = -2B'A(C - BX)$$

4 - La matrice des dérivées partielles du second ordre, d'une fonction de n variables, est appelée matrice *hessienne* :

$$H = \frac{\partial^2 f(X)}{\partial X' \partial X} = \begin{bmatrix} \frac{\partial^2 f(X)}{\partial x_1^2} & \cdots & \frac{\partial^2 f(X)}{\partial x_1 \partial x_n} \\ \cdots & \cdots & \cdots \\ \frac{\partial^2 f(X)}{\partial x_1 \partial x_n} & \cdots & \frac{\partial^2 f(X)}{\partial x_n^2} \end{bmatrix}$$

Par exemple, la matrice hessienne de $f(X) = X'AX$ est $2A$. La matrice hessienne est symétrique si les conditions de continuité et d'existence de toutes les dérivées première et seconde sont satisfaites par $f(X)$.

5.3 Extremum

La condition nécessaire pour obtenir un maximum ou un minimum de $f(X)$ au point $X = X_0$ est :

$\frac{\partial f(X)}{\partial X} = 0$ en ce point.

Une condition suffisante pour avoir en X_0 :

1 - Un *Maximum*, est que la matrice hessienne évaluée en X_0 soit *définie négative*.

2 - Un *Minimum*, est que la matrice hessienne évaluée en X_0 soit *définie positive*.

5.3.1 Maximisation sous contrainte

Comment maximiser (minimiser) une fonction $f(X)$ sous la contrainte $g(X) = c$?

On utilise la méthode des *multiplicateurs de Lagrange* qui consiste à écrire la nouvelle fonction :

$$h(X, \lambda) = f(X) - \lambda[g(X) - c]$$

On cherche l'extremum en résolvant :

$$\frac{\partial h(X, \lambda)}{\partial X} = \frac{\partial f(X)}{\partial X} - \lambda \frac{\partial g(X)}{\partial X} = 0$$

$$\frac{\partial h(X, \lambda)}{\partial \lambda} = -g(X) + c = 0$$

Cette seconde équation n'étant autre que la condition initiale.

5.3.2 Interprétation géométrique des valeurs et des vecteurs propres

Considérons dans l'espace des individus de dimension $p = 2$ l'ensemble des points M situés sur l'ellipse centrée en μ d'équation $(X - \mu)'\Sigma^{-1}(X - \mu) = 1$. Avec :

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

On veut déterminer les axes principaux de cette ellipse.

Le premier axe principal est la droite qui passe par le centre μ de l'ellipse et qui la traverse dans sa plus grande dimension. Le second axe principal de l'ellipse passe aussi par μ, c'est une droite perpendiculaire au premier axe principal. Donc, le demi-premier axe principal de l'ellipse est celui qui maximise la distance d'un point de l'ellipse au centre μ de cette ellipse. Soit :

$$(X-\mu)'(X-\mu) \text{ maximum sachant } (X-\mu)'\Sigma^{-1}(X-\mu) \;=\; 1$$

Pour maximiser la forme quadratique $(X-\mu)'(X-\mu)$ sous la contrainte $(X-\mu)'\Sigma^{-1}(X-\mu) \;=\; 1$ on utilise la méthode des multiplicateurs de Lagrange. La fonction à maximiser est :

$$h(X,\lambda) \;=\; (X-\mu)'(X-\mu) - \lambda[(X-\mu)'\Sigma^{-1}(X-\mu)-1]$$

on annule sa dérivée par rapport à X. soit :

$$\frac{\partial h(X,\lambda)}{\partial X} \;=\; 2(X-\mu) - 2\lambda\Sigma^{-1}(X-\mu) \;=\; 0 \qquad (1)$$

En prémultipliant par Σ on obtient le système d'équations homogène $(\Sigma-\lambda I)(X-\mu) \;=\; 0$ qui admet des solutions non triviales si et seulement si $det(\Sigma-\lambda I) \;=\; 0$. Cette dernière équation n'est autre que le polynôme caractéristique. Ainsi,

- le premier axe principal de l'ellipse est un vecteur propre de Σ.

  En prémultipliant (1) par $(X-\mu)'$ on obtient :

  $$\lambda \;=\; \frac{(X-\mu)'(X-\mu)}{(X-\mu)'\Sigma^{-1}(X-\mu)} \qquad (2)$$

  La valeur propre obtenue apparaît donc comme un rapport de carrés de longueurs associées à deux métriques définies sur E (la métrique unité I et la métrique Σ^{-1}). D'autre part, en utilisant la contrainte, on obtient $(X-\mu)'(X-\mu) \;=\; \lambda$. On peut donc conclure que le premier axe principal n'est autre que le vecteur propre associé à la plus grande valeur propre de Σ.

- Le second axe principal est le vecteur propre associé à la plus petite valeur propre car il n'y a que deux valeurs propres distinctes et on a démontré que des vecteurs propres associés à des valeurs propres distinctes étaient orthogonaux. Pour les mêmes raisons que celles qui viennent d'être évoquées, la seconde valeur propre est aussi un rapport de carrés de longueurs définies par les métriques I et Σ^{-1}.

5.3.3 Relations entre deux ellipsoïdes

Les résultats qui précèdent se généralisent au cas de deux ellipsoïdes dans un espace vectoriel de dimension quelconque muni de métriques Σ_1 et Σ_2.

Soit E un espace vectoriel de dimension p, deux produits scalaires π_1 et π_2 peuvent être définis sur E en spécifiant leurs ellipsoïdes dans E. Soient E_1 et E_2 les ellipsoïdes centrés au point moyen μ :

$$(X-\mu)'\Sigma_1^{-1}(X-\mu) \;=\; 1 \text{ et } (X-\mu)'\Sigma_2^{-1}(X-\mu) \;=\; 1$$

Pour ces deux ellipsoïdes il existe un ensemble de p droites passant par μ qui définissent des ensembles d'axes conjugués communs à E_1 et E_2. Les directions de ces p droites sont déterminées par les segments de droite μB_1, $\cdots$, μB_p où B_1, $\cdots$, B_p est une base de vecteurs propres de π_1 relative à π_2.

La valeur propre λ_i $(i \;=\; 1,\; \cdots,\; p)$ est le rapport du carré de la longueur du demi-axe de E_2 dans la direction B_i au carré de la longueur du demi-axe de E_1 dans la même direction.

Pour illustrer ce qui vient d'être dit, nous considérons l'espace de dimension $p = 2$, la situation qui vient d'être décrite est représentée sur la figure 5.1.

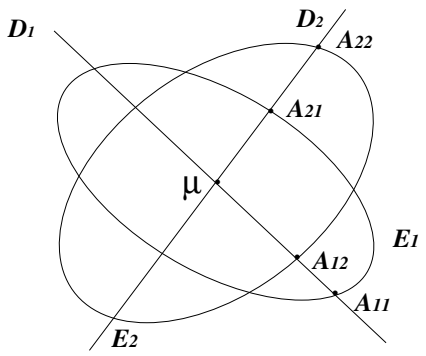

Figure 5.1 : Ellipses E_1 et E_2 centrées en μ. Les droites D_1 et D_2 définissent un couple d'axes conjugués. Les valeurs propres de E_1 relatives à E_2 sont $\lambda_i \;=\; [\mu A_{i2}/\mu A_{i1}]^2$ $(i \;=\; 1,\; 2)$.

Soient maintenant les ellipses obtenues en considérant que E est un espace Euclidien pour E_1 (figure 5.2). Dans cet espace Euclidien E_1 devient un cercle et les axes conjugués de E_2 correspondent aux axes principaux. Les valeurs propres λ_i, $(i = 1, 2)$ sont les inverses des carrés des longueurs des demi-axes principaux.

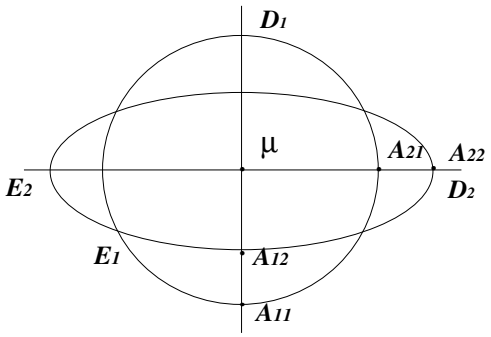

Figure 5.2 : Même chose que pour la Figure 5.1 avec μA_{21} et μA_{11} perpendiculaires, même unité de longueur sur ces axes de telle sorte que E_1 soit un cercle unité.

5.3.4 A propos de la maximisation

D'un point de vue plus général il est facile de montrer que les valeurs propres rangées dans l'ordre décroissant assurent les maxima du rapport de deux formes quadratiques $X'\Sigma_1 X$ et $X'\Sigma_2 X$ dans des directions orthogonales. Pour cela, il suffit de différencier, par rapport à X :

$$\frac{X'\Sigma_1 X}{X'\Sigma_2 X}$$

puis, d'annuler cette dérivée et enfin d'utiliser la propriété d'orthogonalité des vecteurs propres.

$$\frac{\partial}{\partial X}\left(\frac{X'\Sigma_1 X}{X'\Sigma_2 X}\right) = \frac{(X'\Sigma_2 X)(2\Sigma_1 X) - (X'\Sigma_1 X)(2\Sigma_2 X)}{(X'\Sigma_2 X)^2} = 0$$

$$\Updownarrow$$

$$\Sigma_1 X = \left(\frac{X'\Sigma_1 X}{X'\Sigma_2 X}\right)\Sigma_2 X$$

$$\Updownarrow$$

$$[\Sigma_2^{-1}\Sigma_1 - \left(\frac{X'\Sigma_1 X}{X'\Sigma_2 X}\right) I]X = 0$$

On a obtenu le système homogène d'équations relatif aux vecteurs propres X de la matrice $\Sigma_2^{-1}\Sigma_1$ et aux valeurs propres associées :

$$\lambda = \frac{X'\Sigma_1 X}{X'\Sigma_2 X}$$

5.4 Dérivée d'un déterminant

La dérivée du déterminant d'une matrice $A(n \times n)$ par rapport à ses éléments a_{ij} peut être obtenue à partir du développement de A en cofacteurs de la i*ème* ligne ou de la j*ème* colonne.

$$\frac{\partial det(A)}{\partial a_{ij}} = \frac{\partial}{\partial a_{ij}}(a_{i1}A_{i1} + \cdots + a_{ij}A_{ij} + \cdots + a_{in}A_{in})$$

Si A est symétrique :

$$\frac{\partial det(A)}{\partial a_{ii}} = A_{ii}, \frac{\partial det(A)}{\partial a_{ij}} = 2A_{ij}$$

5.5 Quelques formules utiles

1 - Soit $X(n \times p)$ une matrice d'éléments x_{ij}, sa dérivée par rapport à x_{ij} est :

$$\frac{\partial X}{\partial x_{ij}} = J_{ij}$$

où $J_{ij}(n \times p)$ est une matrice avec la valeur un en ij$ème$ position et des zéros partout ailleurs. Si X est symétrique, alors :

$$\frac{\partial X}{\partial x_{ij}} = J_{ij} + J_{ji} \quad i \neq j$$

2 - La règle pour différencier des produits de matrices est identique à celle utilisée pour les scalaires. Supposons que les éléments $x_{ij}(z)$ et $y_{ij}(z)$ des matrices $X(n \times p)$ et $Y(p \times q)$ sont fonctions d'une variable z. Alors,

$$\frac{\partial XY}{\partial z} = \frac{\partial X}{\partial z}Y + X\frac{\partial Y}{\partial z}$$

3 - Cette dernière formule permet de différencier l'inverse d'une matrice carrée non singulière. Soit $I = XX^{-1}$, alors :

$$\frac{\partial I}{\partial x_{ij}} = J_{ij}X^{-1} + X\frac{\partial X^{-1}}{\partial x_{ij}} = 0$$

On en déduit que :

$$\frac{\partial X^{-1}}{\partial x_{ij}} = -X^{-1}J_{ij}X^{-1}$$

Si X est symétrique :

$$\frac{\partial X^{-1}}{\partial x_{ij}} = -X^{-1}J_{ii}X^{-1} \quad i = j$$

$$\frac{\partial X^{-1}}{\partial x_{ij}} = -X^{-1}(J_{ij} + J_{ji})X^{-1} \quad i \neq j$$

5.6 Exemple 2 : description d'un ensemble d'observations

Soit le tableau 5.1 des données iris constitué par des longueurs (X) et des largeurs (Y) des sépales mesurées sur $n_1 = 50$ iris de la variété *Versicolor* et $n_2 = 50$ iris de la variété *Setosa* (R. A. Fisher (1936) ; Kendall et Stuart, vol 3, pp : 317-322).

iris *Versicolor*				iris *Setosa*			
X	Y	X	Y	X	Y	X	Y
7,0	3,2	6,6	3,0	5,1	3,5	5,0	3,0
6,4	3,2	6,8	2,8	4,9	3,0	5,0	3,4
6,9	3,1	6,7	3,0	4,7	3,2	5,2	3,5
5,5	2,3	6,0	2,9	4,6	3,1	5,2	3,4
6,5	2,8	5,7	2,6	5,0	3,6	4,7	3,2
5,7	2,8	5,5	2,4	5,4	3,9	4,8	3,1
6,3	3,3	5,5	2,4	4,6	3,4	5,4	3,4
4,9	2,4	5,8	2,7	5,0	3,4	5,2	4,1
6,6	2,9	6,0	2,7	4,4	2,9	5,5	4,2
5,2	2,7	5,4	3,0	4,9	3,1	4,9	3,1
5,0	2,0	6,0	3,4	5,4	3,7	5,0	3,2
5,9	3,0	6,7	3,1	4,8	3,4	5,5	3,5
6,0	2,2	6,3	2,3	4,8	3,0	4,9	3,6
6,1	2,9	5,6	3,0	4,3	3,0	4,4	3,0
5,6	2,9	5,5	2,5	5,8	4,0	5,1	3,4
6,7	3,1	5,5	2,6	5,7	4,4	5,0	3,5
5,6	3,0	6,1	3,0	5,4	3,9	4,5	2,3
5,8	2,7	5,8	2,6	5,1	3,5	4,4	3,2
6,2	2,2	5,0	2,3	5,7	3,8	5,0	3,5
5,6	2,5	5,6	2,7	5,1	3,8	5,1	3,8
5,9	3,2	5,7	3,0	5,4	3,4	4,8	3,0
6,1	2,8	5,7	2,9	5,1	3,7	5,1	3,8
6,3	2,5	6,2	2,9	4,6	3,6	4,6	3,2
6,1	2,8	5,1	2,5	5,1	3,3	5,3	3,7
6,4	2,9	5,7	2,8	4,8	3,4	5,0	3,3

Tableau 5.1 : longueur (X), largeur (Y) des sépales mesurées sur iris Setosa et iris Versicolor.

Sur la figure 5.3 on a représenté le nuage des points observés.

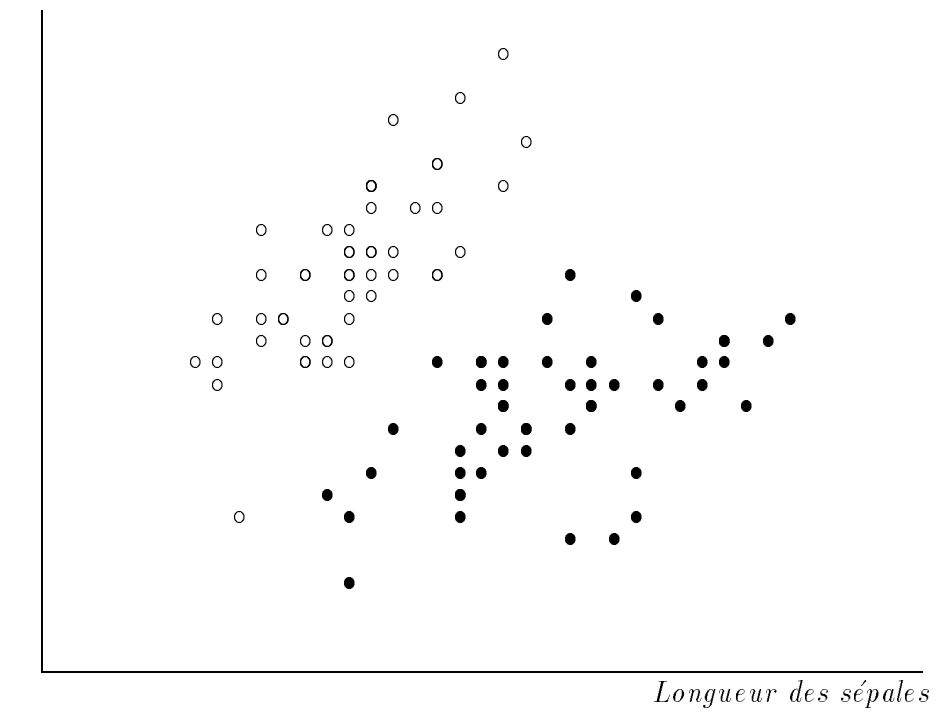

Figure 5.3 : Représentation des données : iris Versicolor (•), iris Setosa (∘).

Pour décrire ces données, on peut estimer des paramètres de position (moyenne générale, moyenne de chacune des populations d'iris), des paramètres de dispersion, par exemple des matrices d'inertie telles que :

1 - <u>*Les matrices d'inertie des populations*</u> :

$$W_i,\ (i = 1 : iris\ Versicolor,\ i = 2 : iris\ Setosa)$$

$$W_i = \begin{bmatrix} \sum_{j=1}^{n_i}(x_{ij} - x_{i.})^2 & \sum_{j=1}^{n_i}(x_{ij} - x_{i.})(y_{ij} - y_{i.}) \\ \sum_{j=1}^{n_i}(x_{ij} - x_{i.})(y_{ij} - y_{i.}) & \sum_{j=1}^{n_i}(y_{ij} - y_{i.})^2 \end{bmatrix}$$

2 - *La matrice somme des inerties des populations* :

$$W = W_1 + W_2$$

$$W = \begin{bmatrix} \sum_{i=1}^{2}\sum_{j=1}^{n_i}(x_{ij} - x_{i.})^2 & \sum_{i=1}^{2}\sum_{j=1}^{n_i}(x_{ij} - x_{i.})(y_{ij} - y_{i.}) \\ \sum_{i=1}^{2}\sum_{j=1}^{n_i}(x_{ij} - x_{i.})(y_{ij} - y_{i.}) & \sum_{i=1}^{2}\sum_{j=1}^{n_i}(y_{ij} - y_{i.})^2 \end{bmatrix}$$

3 - *La matrice d'inertie des points moyens des populations* : B

$$B = \begin{bmatrix} \sum_{i=1}^{2} n_i(x_{i.} - x_{..})^2 & \sum_{i=1}^{2} n_i(x_{i.} - x_{..})(y_{i.} - y_{..}) \\ \sum_{i=1}^{2} n_i(x_{i.} - x_{..})(y_{i.} - y_{..}) & \sum_{i=1}^{2} n_i(y_{i.} - y_{..})^2 \end{bmatrix}$$

$$B = \frac{n_1 \times n_2}{n_1 + n_2} \begin{bmatrix} (x_{1.} - x_{2.})^2 & (x_{1.} - x_{2.})(y_{1.} - y_{2.}) \\ (x_{1.} - x_{2.})(y_{1.} - y_{2.}) & (y_{1.} - y_{2.})^2 \end{bmatrix}$$

4 - *La matrice d'inertie totale* : $T = B + W$

A chacune de ces matrices d'inertie est associé un nombre de *degrés de liberté (ddl)* et on estime les matrices de variance-covariance en divisant ces matrices d'inertie par les degrés de liberté correspondants *(Tableau 5.2)*.

		MATRICES	
	ddl	*Inertie*	*Var-cov*
Populations ($i = 1, 2$)	$n_i - 1$	W_i	$\frac{W_i}{n_i - 1}$
Somme populations	$n_1 + n_2 - 2$	$W = W_1 + W_2$	$\frac{W}{n_1 + n_2 - 2}$
Points moyens populations	1	B	B
Totale	$n - 1$	T	$\frac{T}{n - 1}$

Tableau 5.2 : Degrés de liberté (ddl), matrices d'inertie et de variance-covariancœ (Var-cov).

Estimations des paramètres

1 - *Paramètres de position*

Les estimations des moyennes de chaque variété d'iris et de la moyenne générale sont données dans le tableau 5.3.

	Moyennes		
	Versicolor	*Setosa*	*Générale*
Longueur sépales	5,936	5,006	5,471
Largeur sépales	2,770	3,428	3,099

Tableau 5.3 : Estimations des moyennes iris Versicolor, iris Setosa et de la moyenne générale.

2 - *Paramètres de dispersion*

Inerties populations

$$W_1 = \begin{bmatrix} 13,055 & 4,175 \\ 4,175 & 4,825 \end{bmatrix} \qquad W_2 = \begin{bmatrix} 6,088 & 4,865 \\ 4,865 & 7,046 \end{bmatrix} \qquad W = \begin{bmatrix} 19,143 & 9,036 \\ 9,036 & 11,866 \end{bmatrix}$$

Inertie des points moyens des populations

$$B = \begin{bmatrix} 21,623 & -15,299 \\ -15,299 & 10,824 \end{bmatrix}$$

Inertie totale

$$T = \begin{bmatrix} 40,766 & -6,263 \\ -6,263 & 22,690 \end{bmatrix}$$

Pour décrire les populations d'iris, on utilise l'information véhiculée par les paramètres de position et de dispersion (résumés des données). Pour cela, on cherche les directions de plus grande variabilité des nuages de points c'est-à-dire les vecteurs propres des matrices d'inertie W_1, W_2 et W. Les résultats sont donnés dans le tableau 5.4 :

	W_1		W_2		W	
Valeurs propres	14,802	3,078	11,455	1,679	25,246	5,764
Vecteurs propres	1	1	1	1	1	1
	0,418	-2,390	1,103	-0,906	0,675	-1,481

Tableau 5.4 : Valeurs propres et vecteurs propres des matrices d'inertie W_1, W_2 et W.

On peut vérifier que les deux vecteurs propres associés à chaque matrice W_1, W_2 et W sont orthogonaux entre eux (le produit scalaire de ces deux vecteurs propres est égal à zéro). On a représenté les premières directions propres de W_1, W_2 et W sur la figure 5.4.

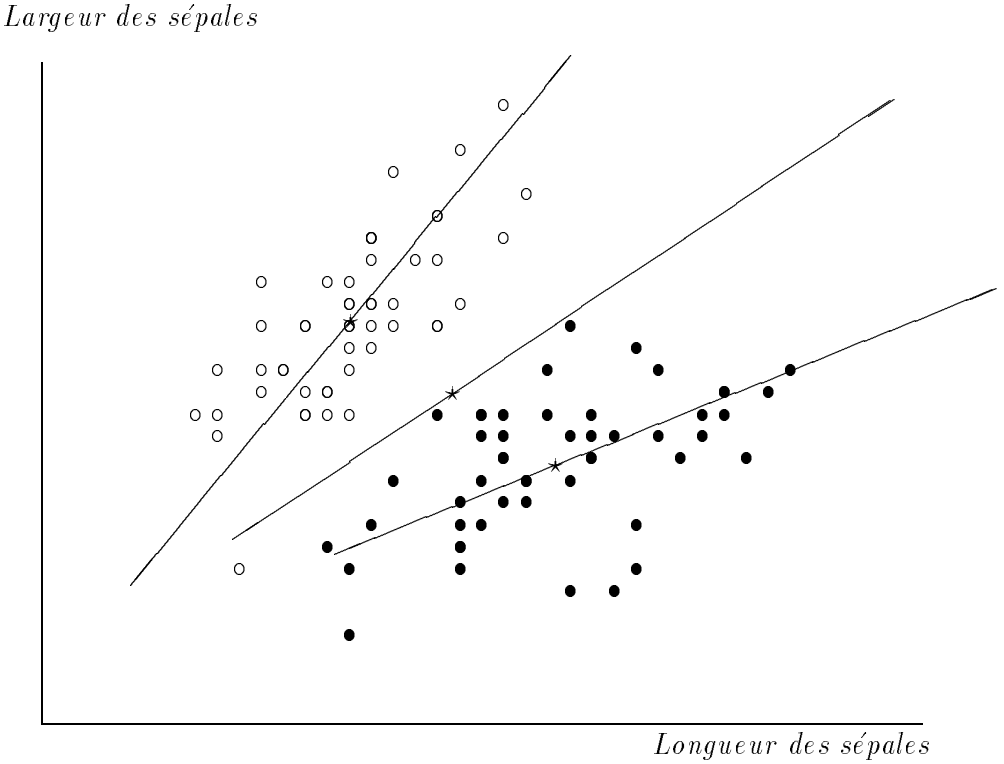

Figure 5.4 : Premier vecteur propre de W_1 : iris Versicolor (•), W_2 : iris Setosa (○) et W. ⋆ représente les points moyens des populations d'iris et le point moyen calculé sur la totalité des observations.

De même, la direction d'inertie maximum des points moyens des populations est obtenue en calculant les valeurs et vecteurs propres de la matrice B. Comme par deux points, il ne passe qu'une droite et qu'une seule, il faut s'attendre à ne trouver qu'une valeur propre non nulle et donc une seule direction propre. On obtient :

$$\lambda = 32,447 \qquad \begin{bmatrix} 1 \\ -0,707 \end{bmatrix}$$

Les résultats obtenus montrent que les deux populations d'iris sont représentées par des nuages de points qui ont des orientations et des dispersions comparables. On peut donc admettre que $W = W_1 + W_2$ représente la matrice d'inertie commune aux deux populations. Par ailleurs, si on trace la direction propre de B et si on projette orthogonalement sur cette direction les points individuels de chacune des populations d'iris, on constate que les iris *Versicolor* sont positionnés à droite et les iris *Setosa* à gauche du point moyen. Le vecteur propre de B est donc une direction de l'espace sur laquelle les projections orthogonales des observations relatives à chacune des deux populations d'iris sont séparées mais on peut mieux faire. En effet, cette direction ne prend pas en compte l'inertie W commune aux deux populations, ce n'est donc pas la direction de l'espace qui permet de séparer au mieux les deux populations d'iris. La question que l'on se pose donc est la suivante :

Existe-t-il dans le plan défini par les variables X et Y une direction optimale pour séparer les populations d'iris ?

Pour répondre à cette interrogation, on cherche une combinaison linéaire des X et Y ($Z = v_1 X + v_2 Y$) telle que, la distance entre les points moyens des populations d'iris soit maximum compte tenu des caractéristiques de chacune de ces populations. Autrement dit, on cherche à maximiser l'inertie des points moyens des populations (B) par rapport à l'inertie des populations (W).

Soient :

$$W_Z = \sum_{j=1}^{n_1} (Z_{1j} - Z_{1.})^2 + \sum_{j=1}^{n_2} (Z_{2j} - Z_{2.})^2 = V'WV$$

$$B_Z = \frac{n_1 n_2}{n_1 + n_2} (Z_{1.} - Z_{2.})^2 = V'BV$$

$$V = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

On résout ce problème soit en utilisant les résultats présentés au paragraphe 5.3.4 soit en maximisant :

$$\phi \;=\; \frac{V'BV}{V'WV} \text{ sous la contrainte } V'WV \;=\; 1$$

Dans ce cas, on utilise la méthode des *multiplicateurs de Lagrange* présentée au paragraphe 5.3.1. Nous avons choisi d'utiliser la méthode des multiplicateurs de Lagrange :

1 - On écrit :

$$h(V,\lambda) \;=\; V'BV - \lambda(V'WV - 1) \;=\; V'[B - \lambda W]V - \lambda$$

2 - On dérive $h(V,\lambda)$ par rapport à V :

$$\frac{\partial h(V,\lambda)}{V} \;=\; 2[B - \lambda W]V$$

3 - On égale la dérivée à zéro

$$[B - \lambda W]V \;=\; 0 \Longleftrightarrow [W^{-1}B - \lambda I]V \;=\; 0$$

On retrouve un problème de *valeur et de vecteur propre*.

Pour l'exemple des iris, on a le résultat suivant :

$$W^{-1} \;=\; \begin{bmatrix} 0,082 & -0,062 \\ -0,062 & 0,132 \end{bmatrix} \qquad\qquad W^{-1}B \;=\; \begin{bmatrix} 2,713 & -1,920 \\ -3,356 & 2,374 \end{bmatrix}$$

Equation caractéristique : $\lambda^2 - 5,088\lambda \;=\; 0$

Une seule valeur propre non nulle : $\lambda \;=\; 5,088$

Un seul vecteur propre : $V \;=\; \begin{bmatrix} 1 \\ -1,237 \end{bmatrix}$

Cette direction est représentée sur la figure 5.5.

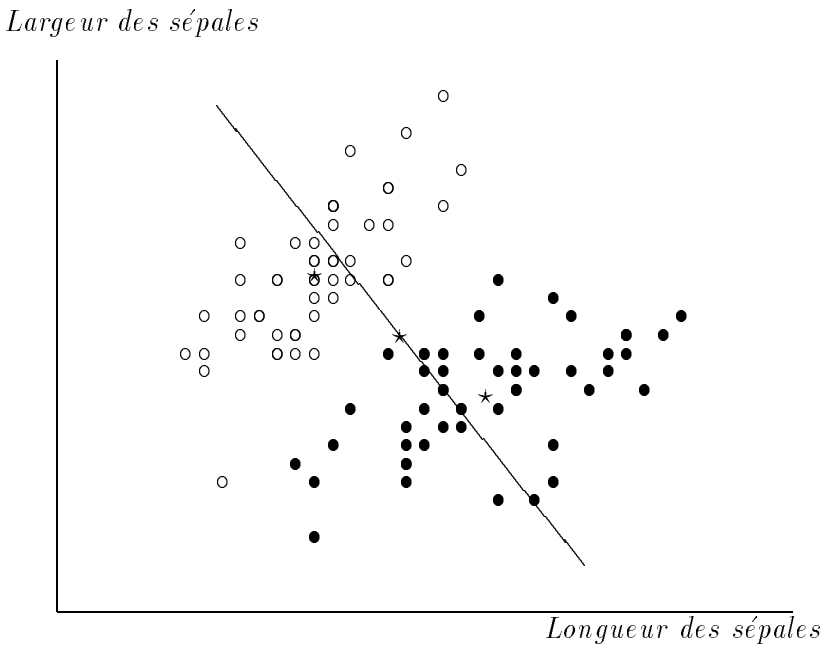

Figure 5.5 : Premier vecteur propre de $W^{-1}B$.

5.6.1 Illustration géométrique

On veut décrire par la géométrie les différentes étapes qui ont été suivies dans l'exemple des iris. Nous avons considéré deux familles d'iris, elles peuvent être représentées par deux ellipses centrées en μ_1 et μ_2. On passe de la première à la seconde par une translation $\mu_1\mu_2$. Si on projette ces ellipses sur un axe parallèlement à une direction donnée D_1, les centres des ellipses projetées sont les projections de μ_1 et μ_2 parallèlement à cette direction. Les écarts types s'obtiennent à partir des tangentes aux ellipses parallèles à cette direction. Dans l'ensemble des directions du plan, il en existe une telle que le rapport de la distance $\mu_1\mu_2$ à l'écart type commun soit maximum, c'est la direction D_2 conjuguée de $\mu_1\mu_2$ par rapport aux ellipses (Cyffers, 1965). La situation qui vient d'être décrite est représentée par la figure 5.6.

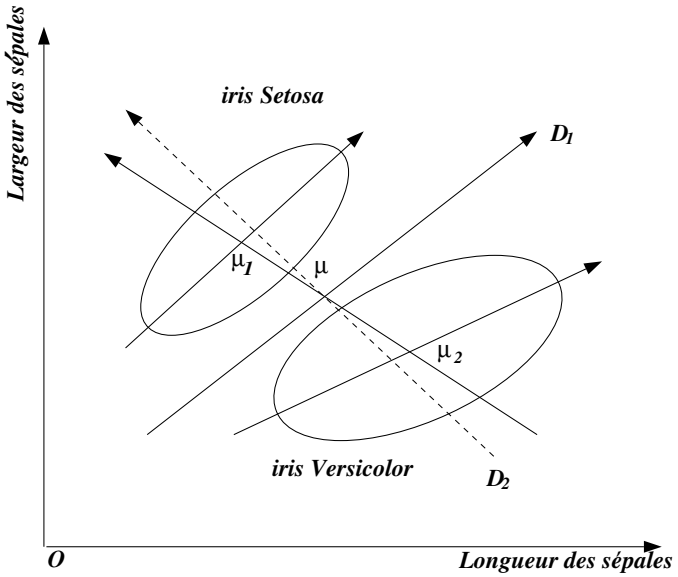

Figure 5.6 : Ellipses d'inertie centrées aux points moyens μ_1 (iris Versicolor) et μ_2 (iris Setosa), direction de projection (D_1) et axe conjugué (D_2) de $\mu_1\mu_2$ par rapport aux ellipses.

Considérons maintenant un cas moins pathologique en nous plaçant dans un cadre un peu plus général tel que : l'espace des individus est toujours défini par deux variables X_1 et X_2, mais dans lequel nous avons trois populations P_1, P_2, P_3 au lieu de deux. On fait cette supposition pour pouvoir représenter l'ellipse d'inertie des points moyens μ_1, μ_2, μ_3 des populations et on suppose que ces points moyens ne sont pas alignés.

1- Représentation des données

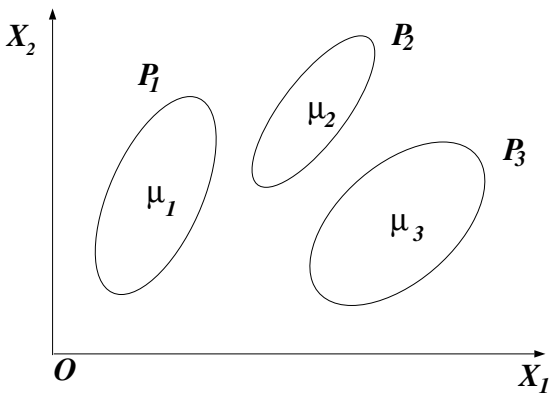

Figure 5.7 : Ellipses d'inertie des populations P_1, P_2 et P_3 centrées respectivement aux points moyens μ_1, μ_2 et μ_3.

On a représenté les données par les contours des ellipses d'inertie associées à chacune des populations P_1, P_2 et P_3. Ces ellipses sont centrées aux points moyens des populations (figure 5.7).

2- Représentation de l'ellipse d'inertie commune aux trois populations et de l'ellipse d'inertie des points points moyens μ_1, μ_2 *et* μ_3

Les trois populations étant représentées par des ellipses d'inertie, approximativement orientées dans la même direction et de tailles comparables, on représente l'inertie commune aux trois populations par une ellipse E_1 centrée au point moyen μ et l'inertie des points moyens par une ellipse E_2 elle aussi centrée au pont moyen μ (figure 5.8).

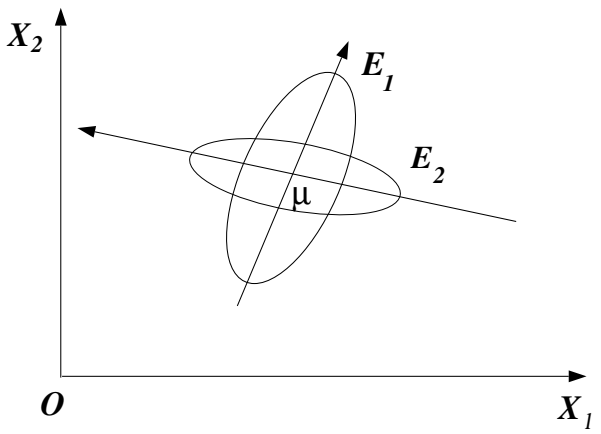

Figure 5.8 : Représentation de l'ellipse d'inertie E_1 commune aux trois populations et de l'ellipse d'inertieE_2 des points moyens .

3- Que cherche-t-on à faire ?

On veut trouver, en tenant compte de l'inertie intra-population, la direction de l'espace telle que les projections de μ_1, μ_2 et μ_3 sur cette direction soient les plus éloignées possible. Autrement dit, M_1 étant un point de l'ellipse E_1 et M_2 le point de l'ellipse E_2 à l'intersection du demi-axe μM_1 et de E_2, on cherche à maximiser la longueur du segment μM_2 par rapport à la longueur du segment μM_1 (figure 5.9).

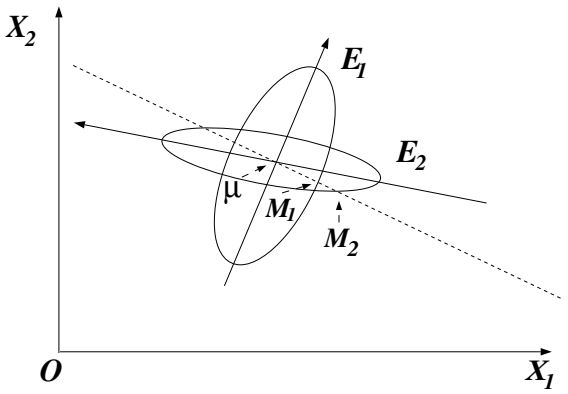

Figure 5.9 : Direction qui maximise le rapport des longueurs de deux segments $\mu M_2/\mu M_1$.

Cette dernière illustration, qui généralise l'exemple des iris, suggère que la géométrie est un outil intéressant pour explorer le champ sémantique de la statistique.