

Chapitre 25

Analyse canonique

Pour conclure le cours sur les analyses statistiques multivariées, ce chapitre propose une petite synthèse sur ce que l'on appelle l'**analyse canonique**. Elle opère la synthèse entre les méthodes exploratoires et les méthodes explicatives. Ce chapitre est large inspiré par [Bouroche et Saporta, 2002, p. 63-82].

L'analyse canonique fut inventée par Harold Hotelling en 1969 [Hotelling, 1936]. Elle constitue la base théorique de la plupart des méthodes d'analyse de données :

1. la régression multiple ;
2. l'analyse de la variance ;
3. l'analyse des correspondances ;
4. l'analyse discriminante.

La méthode est difficilement interprétable.

« Le but de l'analyse canonique est d'étudier les relations linéaires existant entre deux groupes de caractères qualitatifs observés sur un même ensemble d'individus. De façon plus précis, on cherche une combinaison linéaire des caractères du premier ensemble et une combinaison linéaire des caractères du deuxième qui soient les plus corrélées possibles » [Bouroche et Saporta, 2002, p. 64].

On pose n le nombre d'observations dans \mathbb{R}^n .

— Soient p caractères du groupe 1 : $x^1, \dots, x^j, \dots, x^p$.

— Soient q caractères du groupe 2 : $y^1, \dots, y^k, \dots, y^q$.

On établit des combinaisons linéaires des caractères des deux groupes :

$$\xi = a_1x^1 + \dots + a_jx^j + \dots + a_px^p \quad (25.1)$$

$$\eta = b_1y^1 + \dots + b_ky^k + \dots + b_qy^q \quad (25.2)$$

Attention ! x^1, y^1, \dots , utilisent des **indices** en exposant.

On isole les coefficients.

$${}^t a = (a_1 \quad \dots \quad a_j \quad \dots \quad a_p) \quad (25.3)$$

$${}^t b = (b_1 \quad \dots \quad b_k \quad \dots \quad b_q) \quad (25.4)$$

Ils maximisent le carré de corrélation entre ξ et η .

On appelle **caractères canoniques** les vecteurs ξ et $\eta \in \mathbb{R}^n$, **facteurs canoniques** les vecteurs de coefficient $a \in \mathbb{R}^p$ et $b \in \mathbb{R}^q$, et **corrélation canonique** le coefficient de corrélation entre ξ et η .

L'ensemble des caractères ξ , combinaisons linéaires des $x^1, \dots, x^j, \dots, x^p$, forme un sous-espace vectoriel W_1 de \mathbb{R}^n que l'on appelle **potentiel de précision** du premier groupe.

L'ensemble des caractères η , combinaisons linéaires des $y^1, \dots, y^k, \dots, y^q$, forme un sous-espace vectoriel W_2 de \mathbb{R}^n que l'on appelle **potentiel de précision** du deuxième groupe.

La question revient à déterminer deux vecteurs $\xi \in W_1$ et $\eta \in W_2$ faisant un **angle minimal**.

N.B. 1. Si les caractères sont centrées, cosinus et corrélation sont liés. Il existe une solution triviale : η^1 et ξ^1 tels que $\cos^2(\eta^1, \xi^1) = 1$. Lorsqu'un premier couple de variables canoniques a été obtenu, on recherche un autre couple de caractères η^2 et ξ^2 tels que $\rho(\eta^2, \xi^2)$ soit maximal et tel que η^1 et η^2 , et ξ^1 et ξ^2 , aient chacun une **corrélation nulle**. On répète le processus jusqu'à p ou q caractères.

N.B. 2. Le problème de l'analyse canonique peut être rapproché de celui de la régression multiple.

25.1 Formulation géométrique de l'analyse canonique

25.1.1 Projection orthogonale sur un sous-espace vectoriel

Avant de résoudre le problème de l'analyse canonique, on pose celui de la régression multiple. Soit un caractère « à expliquer » y et de p caractères « explicatifs » $x^1, \dots, x^j, \dots, x^p$, on suppose que les $p + 1$ caractères sont observés sur le même ensemble de n individus, chaque individu étant muni du poids $p_i > 0$ avec $\sum_{i=1}^n p_i = 1$. à partir de là, il s'agit de déterminer une combinaison linéaire des p caractères explicatifs :

$$\xi = a_1 x^1 + \dots + a_j x^j + \dots + a_p x^p \quad (25.5)$$

telle que ξ soit le plus proche possible de y au sens de la distance dans l'espace des caractères, c'est-à-dire le **critère des moindres carrés**.

Chacun des $p + 1$ caractères peut être représenté par un vecteur de \mathbb{R}^n :

$$y = \begin{pmatrix} y_1 \\ \dots \\ y_i \\ \dots \\ y_n \end{pmatrix} \in \mathbb{R}^n \quad (25.6)$$

et

$$x^j = \begin{pmatrix} x_1^j \\ \dots \\ y_i^j \\ \dots \\ y_n^j \end{pmatrix} \in \mathbb{R}^n \quad (25.7)$$

avec $j = 1, 2, \dots, p$.

On suppose que les $p + 1$ caractères sont centrés :

$$\sum_{i=1}^n p_i y_i = 0 \quad (25.8)$$

et

$$\sum_{i=1}^n p_i x_i^j = 0 \quad (25.9)$$

avec $j = 1, 2, \dots, p$.

Soit le sous-espace vectoriel W de \mathbb{R}^n engendré par les combinaisons linéaires des caractères x^j :

$$\xi \in W \Leftrightarrow \xi = a_1 x^1 + \dots + a_j x^j + \dots + a_p x^p \quad (25.10)$$

On suppose que la dimension de W est égale à p , ce qui revient à dire que les p caractères x^j forme une base de W , ou encore que le rang de la matrice :

$$X_{np} = \begin{pmatrix} x_1^1 & \dots & x_1^j & \dots & x_1^p \\ \dots & \dots & \dots & \dots & \dots \\ x_i^1 & \dots & x_i^j & \dots & x_i^p \\ \dots & \dots & \dots & \dots & \dots \\ x_n^1 & \dots & x_n^j & \dots & x_n^p \end{pmatrix} \quad (25.11)$$

est égal à p .

N.B. En notation abrégée, on pose :

$$W = \{\xi \in \mathbb{R}^n / \xi = Xa \text{ avec } a \in \mathbb{R}^p\} \quad (25.12)$$

On suppose que l'espace des caractères est muni du produit scalaire associé à la matrice diagonale des poids :

$$D = \begin{pmatrix} p_1 & \dots & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & p_i & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & \dots & p_n \end{pmatrix} \quad (25.13)$$

Sur l'espace des caractères centrés, le produit scalaire et la covariance sont **identiques** :

$${}^t x^i . D . x^k = s_{jk} \quad (25.14)$$

De même, la norme et la variance sont **identiques** :

$$||x^j||^2 = s_j^2 \quad (25.15)$$

La distance entre deux caractères est donnée par :

$$d^2(x^i, x^k) = ||x^i - x^k||^2 = (x^j - x^k) . D . (x^i - x^k) \quad (25.16)$$

$y \in \mathbb{R}^n$ est donné. On cherche $\xi \in W$ tel que la distance entre y et ξ soit minimale. En appliquant le critère des moindres carrés, on peut écrire :

$$\min_{\xi \in W} ||y - \xi||^2 \quad (25.17)$$

Cela permet de déterminer \hat{y} , le point de W le plus proche de y . \hat{y} est la **projection orthogonale de y sur W** .

25.1.2 Recherche du projecteur orthogonal sur W

Le **projecteur orthogonal** sur W est l'application linéaire de \mathbb{R}^n dans \mathbb{R}^n faisant correspondre à tout vecteur de \mathbb{R}^n sa projection orthogonale sur W . On note A la matrice de cette application :

$$y \rightarrow A.y = \hat{y} \quad (25.18)$$

avec la condition d'orthogonalité vérifiée ${}^t(y - \hat{y}) . D . \hat{y} = 0$.

Comment construire A à partir des vecteurs $x^1, \dots, x^j, \dots, x^p$, base de W ? Tout vecteur $\xi \in W$ peut s'écrire sous la forme : $\xi = X.a$, en particulier $\hat{y} \in W$, pour lequel on pose : $\hat{y} = X.\hat{a}$. $(y - \hat{y})$ est orthogonal à tout vecteur de W , donc, en particulier, aux vecteurs de base. Par conséquent, il existe p équations : ${}^t x^j . D . (y - \hat{y}) = 0$ avec $j = 1, 2, \dots, p$, ou, puisque $\hat{y} = X.\hat{a}$:

$${}^t x^j . D . y = {}^t x^j . D . X . \hat{a} \quad (25.19)$$

avec $j = 1, 2, \dots, p$. Cela peut s'écrire sous la forme d'une seule équation matricielle :

$${}^tX.D.X.\hat{a} = {}^tX.D.y \quad (25.20)$$

Puisque le rang de X vaut p , la matrice ${}^tX.D.X$ est **inversible**, ce qui permet de déterminer \hat{a} en application des systèmes linéaires :

$$\hat{a} = ({}^tX.D.X)^{-1} \cdot {}^tX.D.y \quad (25.21)$$

Le vecteur \hat{a} contient les p coefficients de la combinaison linéaire $\hat{y} = \hat{a}_1x^1 + \dots + \hat{a}_jx^j + \dots + \hat{a}_px^p \in W$ **la plus proche** de y .

De l'expression de \hat{a} , on trouve l'expression recherchée de $\hat{y} = X.\hat{a}$

$$\hat{y} = X.({}^tX.D.X)^{-1} \cdot {}^tX.D.y \quad (25.22)$$

La matrice $X.({}^tX.D.X)^{-1} \cdot {}^tX.D$ fait correspondre à y sa projection orthogonale sur W . On déduit la matrice recherchée A :

$$A = X.({}^tX.D.X)^{-1} \cdot {}^tX.D \quad (25.23)$$

25.1.3 Recherche de la droite W faisant un angle minimal

Comment prouver que \hat{y} est un vecteur de W faisant un angle minimal avec y ? D'après le théorème de Pythagore, $\|y\|^2 = \|y - \hat{y}\|^2 + \|\hat{y}\|^2$, donc minimiser $\|y - \hat{y}\|^2$ revient à maximiser $\|\hat{y}\|^2$ puisque $\|\hat{y}\|^2$ est une constante. \hat{y} est le vecteur W maximisant :

$$^2(y, \hat{y}) = \frac{\|\hat{y}\|^2}{\|y\|^2} \quad (25.24)$$

ce qui est par conséquent l'angle minimal entre y et \hat{y} .

N.B. Comme y et x^i sont centrés, le cosinus entre y et \hat{y} est le **coefficient de corrélation** entre les caractères y et \hat{y} .

25.2 Recherche des caractères canoniques

25.2.1 Présentation géométrique

Soient deux ensemble de caractères : $x^1, \dots, x^j, \dots, x^p$ et $y^1, \dots, y^k, \dots, y^q$, on suppose que les $p + q$ caractères sont observés sur le même ensemble de n individus munis de poids $p_i > 0$ avec $i = 1, 2, \dots, n$ et $\sum_{i=1}^n p_i = 1$. On

suppose que les $p + q$ caractères sont centrés. Chacun des $p + q$ caractères peut être représenté par un vecteur de \mathbb{R}^n :

$$x^j = \begin{pmatrix} x_1^j \\ \vdots \\ y_i^j \\ \vdots \\ y_n^j \end{pmatrix} \quad (25.25)$$

avec $j = 1, \dots, p$, et

$$y = \begin{pmatrix} y_1^k \\ \vdots \\ y_i^k \\ \vdots \\ y_n^k \end{pmatrix} \quad (25.26)$$

avec $k = 1, \dots, q$.

Aux vecteurs, x^j et y^k , on associe respectivement les sous-espaces vectoriels de \mathbb{R}^n , W_1 et W_2 :

$$W_1 = \{\xi \in \mathbb{R}^n / \xi = X.a, a \in \mathbb{R}^p\} \quad (25.27)$$

$$W_2 = \{\eta \in \mathbb{R}^n / \eta = X.b, b \in \mathbb{R}^q\} \quad (25.28)$$

avec X_{np} et Y_{nq} les matrices contenant respectivement en colonnes les vecteurs x^i et y^k étant centrés, leurs sous-espaces vectoriels respectifs W_1 et W_2 contiennent des vecteurs centrés.

On suppose que les X^j et les y^k forment respectivement une base de W_1 et une base de W_2 , donc :

$$\dim(W_1) = p \quad (25.29)$$

et

$$\dim(W_2) = q \quad (25.30)$$

et

$$\text{rg}(X) = p \quad (25.31)$$

et

$$\text{rg}(Y) = q \quad (25.32)$$

Cela permet de formuler **géométriquement** le problème de l'analyse canonique. Il s'agit de déterminer $\xi \in W_1$ et $\eta \in W_2$ tels que :

$$\cos^2(\eta, \xi) = \rho(\xi, \eta) \quad (25.33)$$

soit un **maximum**.

N.B. Les vecteurs centrés ou **centrés réduits** ont un angle $\xi = X.a$ et $\eta = X.b$ **identique**. En pratique, on effectue généralement les calculs sur des caractères centrés **et** réduits.

En analyse canonique, on recherche **simultanément** une combinaison linéaire des variables du premier groupe, notée ξ , et une combinaison linéaire des variables du second groupe, notée η , telles que le coefficient de corrélation entre ξ et η soit maximal. Ce couple étant trouvé, on en recherche un deuxième orthogonal au premier qui satisfait le même critère, *etc.*

25.2.2 Recherche des caractères canoniques

Si les caractères ξ^1 et η^1 sont solutions du problème de l'analyse canonique, puisque l'angle entre ξ et η ne dépend pas de leur norme, on suppose que $\|\xi\| = \|\eta\| = 1$.

η^1 doit être colinéaire avec la projection orthogonale de ξ^1 sur W_2 qui est le vecteur de W_2 faisant un angle minimal avec ξ^1 . Cette condition s'écrit :

$$A_2.\xi^1 = \rho_1.\eta^1 \quad (25.34)$$

avec $\rho_1 = \cos(\xi^1, \eta^1)$ et A_2 l'opérateur de projection orthogonale sur W_2 . De même, on a :

$$A_1.\eta^1 = \rho_1.\xi^1 \quad (25.35)$$

On en déduit le système linéaire :

$$A_1.A_2.\xi^1 = \lambda_1.\xi^1 \quad (25.36)$$

$$A_2.A_1.\eta^1 = \lambda_1.\eta^1 \quad (25.37)$$

avec $\lambda_1 = \rho_1^2 = \cos^2(\xi^1, \eta^1)$. On en déduit que ξ^1 et η^1 sont respectivement **vecteurs propres** des opérateurs $A_1.A_2$ et $A_2.A_1$ associés à la même plus grande **valeur propre** λ_1 , égale à leur cosinus carré, c'est-à-dire à leur corrélation au carré.

Les caractères ξ^1 et η^1 se déduisent l'un de l'autre par une simple application linéaire :

$$\eta^1 = \frac{1}{\sqrt{\lambda_1}}.A_2.\xi^1 \quad (25.38)$$

$$\xi^1 = \frac{1}{\sqrt{\lambda_1}}.A_1.\eta^1 \quad (25.39)$$

Les caractères canoniques suivants sont les vecteurs propres de $A_1.A_2$ et $A_2.A_1$ associés avec des valeurs propres rangées par ordre décroissant. On peut prouver que les vecteurs propres de $A_1.A_2$ sont orthogonaux pour D , et que, par conséquent,

$\cos^2(\xi^i, \xi^j) = \cos^2(\eta^i, \eta^j) = 0$ lorsque $i \neq j$. À chaque étape, on choisit le couple de caractères canoniques ξ^i, η^i associé à la plus grande valeur propre λ_i non encore sélectionnée.

N.B. Le nombre maximal de caractères canoniques est égal à $\min(p, q)$. En supposant $p < q$, les ξ^i avec $i = 1, \dots, p$, forment une base W dans laquelle il est impossible d'obtenir d'autres vecteurs appartenant à W_1 et orthogonaux aux ξ^i .

25.2.3 Recherche des facteurs canoniques

Puisque $\xi \in W_1$, ξ peut s'écrire comme une combinaison linéaire des caractères $x^1, \dots, x^j, \dots, x^p$:

$$\xi = a_1 x^1 + \dots + a_j x^j + \dots + a_p x^p \quad (25.40)$$

Si on pose ${}^t a = (a_1 \dots a_j \dots a_p)$, on peut l'écrire plus simplement :

$$\xi = X.a \quad (25.41)$$

et

$$\eta = X.b \quad (25.42)$$

Les facteurs canoniques a et b peuvent être calculés **directement**. En posant :

$$A_1 = X.({}^t X.D.X)^{-1}.{}^t X.D \quad (25.43)$$

et

$$A_2 = Y.({}^t Y.D.Y)^{-1}.{}^t Y.D \quad (25.44)$$

et en remplaçant dans les équations donnant ξ et η , on obtient :

$$X.({}^t X.D.X)^{-1}.{}^t X.D.Y.({}^t Y.D.Y)^{-1}.{}^t Y.D.a = \lambda.X.a \quad (25.45)$$

$$Y.({}^t Y.D.Y)^{-1}.{}^t Y.D.X.({}^t X.D.X)^{-1}.{}^t X.D.b = \lambda.X.b \quad (25.46)$$

Pour rendre plus lisible les équations, on pose :

— les variances :

$$V_{11} = {}^t X.D.X \quad (25.47)$$

$$V_{22} = {}^t Y.D.Y \quad (25.48)$$

— les covariances :

$$V_{12} = {}^t X.D.Y \quad (25.49)$$

$$V_{21} = {}^t X.D.Y \quad (25.50)$$

ce qui devient :

$$X.V_{11}^{-1}.V_{12}.V_{22}^{-1}.V_{21}.a = \lambda.X.a \quad (25.51)$$

$$Y.V_{22}^{-1}.V_{21}.V_{11}^{-1}.V_{12}.b = \lambda.X.b \quad (25.52)$$

Les facteurs canoniques correspondent aux vecteurs propres de produits de matrices de covariance.

Les conditions de normalisation $\|\xi\|^2 = \|\eta\|^2 = 1$ deviennent :

$${}^t\xi.D.\xi = {}^t a.{}^t X.D.X.a = {}^t a.V_{11}.a = 1 \quad (25.53)$$

$${}^t\eta.D.\eta = {}^t b.{}^t X.D.X.b = {}^t b.V_{22}.b = 1 \quad (25.54)$$

Pour finir, a et b se déduisent l'un de l'autre par transformation linéaire. $\eta = \frac{1}{\sqrt{\lambda}}A_2.\xi$ devient $Y.b = \frac{1}{\sqrt{\lambda}}Y.({}^tY.D.Y)^{-1}.{}^tY.D.X.a$, et en simplifiant, on peut écrire :

$$b = \frac{1}{\sqrt{\lambda}}V_{22}^{-1}.V_{21}.a \quad (25.55)$$

et, de même, pour a , on peut écrire :

$$a = \frac{1}{\sqrt{\lambda}}V_{11}^{-1}.V_{12}.b \quad (25.56)$$

En pratique, on recherche a si $p < q$ afin de travailler sur la matrice de plus petite taille, puis on déduira b .

Conclusion générale

L'intérêt de l'analyse canonique réside essentiellement dans ses aspects méthodologiques. La régression multiple n'en est qu'un cas particulier. De même, pour l'A.C.P., l'A.F.C., l'A.C.M. ou l'A.D.F.

J. D. Carroll a proposé une généralisation de l'analyse canonique à l'analyse de plus de deux groupes de variables en 1968 [Carroll, 1968]. Le principe de la généralisation est simple. On dispose de m ensembles de caractères numériques centrés représentés par les tableaux $X_1, X_2, \dots, X_j, \dots, X_m$, soit W_i le potentiel de prévision associé à X_i . On recherche un nouveau caractère $z \in \mathbb{R}^n$ maximisant la somme des corrélations :

$$\sum_{i=1}^m \text{cor}^2(z, \xi_i) \quad (25.57)$$

avec $\xi_i \in W_i$.

N.B. 1. z est solution de $(\sum_{i=1}^m A_i).z = \mu.z$.

N.B. 2. Si $m = 2$, z est colinéaire à la bissectrice de ξ^1 et η^2 .

L'analyse canonique généralisée (ou analyse multicanonique) présente quatre particularités intéressantes :

1. l'analyse canonique simple dans le cas où $m = 2$;
2. l'A.C.P. dans le cas où il n'existe qu'un seul caractère par groupe. L'A.C.P. est une analyse canonique généralisée de groupes de variables réduits chacun à un élément. Les composantes principales sont les variables générales et les variables canoniques sont confondues avec les variables initiales ;
3. l'A.C.M. dans le cas où les tableaux X_i sont des tableaux de variables indicatrices. L'A.C.M. est une analyse canonique généralisée dans laquelle les groupes de variables sont composés chacun par les indicatrices des classes d'une même variable qualitative ;
4. l'A.F.M. dans le cas où les tableaux sont pondérés.

Bibliographie

- [Bouroche et Saporta, 2002] BOUROCHE, J.-M. et SAPORTA, G. (2002). L'analyse de données. Que sais-je ? P.U.F., Paris.
- [Carroll, 1968] CARROLL, J. D. (1968). A generalisation of canonical correlation analysis to three ou more sets of variables. Proceedings of the 76th Convention American Psychological Analysis, pages 227–228.
- [Hotelling, 1936] HOTELLING, H. (1936). Relation between two sets of variates. Biometrika, 28(3/4):321–377.