Analyse canonique généralisée

Jacques Bailhache (jacques.bailhache@gmail.com)

July 31, 2020

1 Présentation

L'analyse canonique généralisée est une méthode d'analyse de tableaux de données de n lignes (individus) et p colonnes (variables) groupées en q groupes, que l'on peut écrire sous la forme :

$$X = (X_1, X_2, \dots, X_k, \dots, X_q)$$

Si q = 2, on retrouve l'analyse canonique classique.
Si chaque bloc X_k est un tableau disjonctif complet, on retrouve l'analyse des correspondances multiples.
Si chaque bloc n'est formé que d'une seule colonne, on retrouve l'analyse en composantes principales.

2 Formulation mathématique

Il s'agit de maximiser $\sum_{k=1}^{q} {}^tyX_k({}^tX_kX_k)^{-1}\,{}^tX_ky$ avec la contrainte ${}^tyy = 1$.
Le vecteur y est le vecteur propre correspondant à la plus grande valeur propre de la matrice :

$$S = \sum_{k=1}^{q} X_k({}^tX_kX_k)^{-1}\,{}^tX_k$$

3 Cas q = 2 : Analyse canonique classique

Pour q = 2, on a :

$$X_1({}^tX_1X_1)^{-1}\,{}^tX_1y + X_2({}^tX_2X_2)^{-1}\,{}^tX_2y = \lambda y$$

Posons $({}^tX_1X_1)^{-1}\,{}^tX_1y = a$ et $({}^tX_2X_2)^{-1}\,{}^tX_2y = b$.
On a alors

$$X_1a + X_2b = \lambda y$$

En multipliant à gauche par $({}^tX_1X_1)^{-1}\,{}^tX_1$, on obtient :

$$({}^tX_1X_1)^{-1}\,{}^tX_1X_2b = (\lambda - 1)a$$

De même, on a aussi

$$({}^tX_2X_2)^{-1}\,{}^tX_2X_1a = (\lambda - 1)b$$

et par substitution

$$({}^tX_2X_2)^{-1}\,{}^tX_2X_1({}^tX_1X_1)^{-1}\,{}^tX_1X_2b = (\lambda - 1)^2b$$

On retrouve la matrice à diagonaliser de l'analyse canonique classique.

4 Cas où chaque bloc comprend une seule colonne : Analyse en composantes principales

Dans ce cas on a :

$$S = \sum_{k=1}^{q} x_k ({}^t x_k x_k)^{-1} \, {}^t x_k = \sum_{k=1}^{q} \frac{1}{n s_k^2} x_k \, {}^t x_k$$

avec

$$s_k^2 = \frac{1}{n} \, {}^t x_k x_k$$

Soit la matrice T dont la k-ième colonne vaut $t_k = \frac{1}{s_k} x_k$.
Alors on a $S = \frac{1}{n} T \, {}^t T$.
La relation $Sy = \lambda y$ s'écrit alors :

$$\frac{1}{n} T {}^t T y = \lambda y$$

En multipliant à gauche par ${}^t T$ on obtient :

$$\frac{1}{n} \, {}^t T T {}^t T y = \lambda {}^t T y$$

et en posant ${}^t T y = u$:

$$\frac{1}{n} \, {}^t T T u = \lambda u$$

ou

$$Cu = \lambda u$$

5 Référence

Statistique exploratoire multidimensionnelle, éditions DUNOD, 8.3.5 Analyse canonique généralisée