2019

Algèbre linéaire pour les statistiques

Joseph Salmon
Université de Montpellier

Préface

Ce polycopié se veut être une introduction aux éléments d'algèbre linéaire utiles pour les statistiques. Noter que ce travail est en construction, les points notés TODO: blabla sont en cours de rédaction (clémence donc sur les commentaires à ces endroits).

Table des matières

Première partie

Rappels

1

Matrices classiques

1.1 Matrice orthogonale

Définition 1.1. *Une matrice $U \in \mathbb{R}^{n\times n}$ est dite orthogonale si elle vérifie la propriété suivante*

$$U^\top U = UU^\top = \mathrm{Id}_n \quad . \tag{1.1}$$

Une telle matrice U peut s'écrire en colonne $U = [u_1, \dots, u_n]$ et ses colonnes satisfont les relations

$$\forall i \in [\![1, n]\!],\ u_i^\top u_j = \delta_{i,j} \quad , \tag{1.2}$$

où le symbole de Kronecker δ est défini pour tout $(i,j) \in [\![1, n]\!]$ par

$$\delta_{i,j} = \begin{cases} 1, & \text{si } i = j, \\ 0, & \text{si } i \neq j. \end{cases} \tag{1.3}$$

Ainsi les vecteurs $u_1, \dots, u_n$ forment une base orthonormale.

1.2 Matrice J_n et recentrage

Soit $\mathbb{1}_n \in \mathbb{R}^n$ le vecteur de taille n ne contenant que des 1. On définit alors la matrice J_n par

$$J_n = \mathbb{1}_n \mathbb{1}_n^\top \in \mathbb{R}^n \quad . \tag{1.4}$$

Propriétés : La matrice J_n est symétrique définie positive :

$$J_n^2 = nJ_n \quad . \tag{1.5}$$

On associe souvent aussi le projecteur : $\bar{J}_n$ défini par $\bar{J}_n = \frac{1}{n}J_n$. Le projecteur orthogonal associé $\overline{C}_n = \mathrm{Id}_n - \bar{J}_n$ est la matrice que l'on appelle la *matrice de centrage* car pour tout vecteur $x \in \mathbb{R}^n$ on observe que

$$\overline{C}_n x = (x_1 - \overline{x}_n, \dots, x_n - \overline{x}_n) \ . \tag{1.6}$$

Décomposition spectrale :

$$J_n = H_n^\top \operatorname{diag}(n, 0, \dots, 0) H_n \tag{1.7}$$

$$\bar{J}_n = H_n^\top \operatorname{diag}(1, 0, \dots, 0) H_n \ . \tag{1.8}$$

où $H_n \in \mathbb{R}^{n \times n}$ est la matrice de Helmert, c'est-à-dire la matrice orthogonale définie par :

$$H_n = \begin{bmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{n}} & \frac{-1}{\sqrt{n}} & 0 & 0 & \cdots & 0 & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & 0 & \cdots & 0 & 0 \\ \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{-3}{\sqrt{12}} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \cdots & \frac{1}{\sqrt{n(n-1)}} & \frac{-(n-1)}{\sqrt{n(n-1)}} \end{bmatrix} .$$

Propriétés : Pour tout $(a, b, a', b') \in \mathbb{R}^4$, on a

$$(a \mathrm{Id}_n + bJ_n)(a' \mathrm{Id}_n + b'J_n) = aa' \mathrm{Id}_n + (ab' + a'b + nbb')J_n, \tag{1.9}$$

$$(a \mathrm{Id}_n + bJ_n)^{-1} = \frac{1}{a}\left(\mathrm{Id}_n - \frac{b}{a+nb}J_n\right), \quad \text{pour } a \neq 0 \text{ et } a \neq -nb, \tag{1.10}$$

$$a \mathrm{Id}_n + bJ_n = H_n^\top \operatorname{diag}(a + nb, \underbrace{a, \dots, a}_{(n-1)\text{fois}})H_n, \quad \text{(décomp. spectrale)} \tag{1.11}$$

$$\det(a \mathrm{Id}_n + bJ_n) = a^{n-1}(a + nb) \tag{1.12}$$

Il est aussi intéressant de considérer des matrices de la forme :

$$(\rho - 1) \mathrm{Id}_n + \rho J_n = \begin{bmatrix} 1 & \rho & \dots & \dots & \rho \\ \rho & 1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 1 & \rho \\ \rho & \dots & \dots & \rho & 1 \end{bmatrix}$$

Ce type de matrice est obtenu en considérant la matrice de variance covariance d'un vecteur (gaussien) tel que décrit : ici `https://stats.stackexchange.com/questions/278227/representation-of-equicorrelated-normal-random-variables`

Exemple 1.2. *Prenons un modèle à effet aléatoire pour une catégorie à K modalités $C_1, \dots, C_K$ et n observations (avec $n = n_1 + \dots + n_K$) :*

$$y = \mu \mathbb{1}_n + \sum_{k=1}^{K} \mathbb{1}_{C_k} a_k + \varepsilon \ . \tag{1.13}$$

avec $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2 \operatorname{Id}_n)$ *et* $a \sim \mathcal{N}(0, \sigma_a^2 \operatorname{Id}_K)$*. Dans ce cas la matrice de covariance de* $y \in \mathbb{R}^n$ *s'écrit :*

$$\begin{aligned}
\mathbb{V}\mathrm{ar}(y) &= \mathbb{V}\mathrm{ar}(\varepsilon) + \mathbb{V}\mathrm{ar}(\begin{bmatrix}\mathbb{1}_{C_1} & \cdots & \mathbb{1}_{C_K}\end{bmatrix} a) && (1.14)\\
&= \sigma_\varepsilon^2 \operatorname{Id}_n + \begin{bmatrix}\mathbb{1}_{C_1} & \cdots & \mathbb{1}_{C_K}\end{bmatrix} \mathbb{V}\mathrm{ar}(aa^\top) \begin{bmatrix}\mathbb{1}_{C_1} & \cdots & \mathbb{1}_{C_K}\end{bmatrix}^\top && (1.15)\\
&= \sigma_\varepsilon^2 \operatorname{Id}_n + \begin{bmatrix}\mathbb{1}_{C_1} & \cdots & \mathbb{1}_{C_K}\end{bmatrix} \sigma_a^2 \operatorname{Id}_K \begin{bmatrix}\mathbb{1}_{C_1} & \cdots & \mathbb{1}_{C_K}\end{bmatrix}^\top && (1.16)
\end{aligned}$$

$$\mathbb{V}\mathrm{ar}(y) = \sigma_\varepsilon^2 \operatorname{Id}_n + \sigma_a^2 \begin{bmatrix} J_{n_1} & 0 & \dots & \dots & 0 \\ 0 & J_{n_2} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & J_{n_{K-1}} & 0 \\ 0 & \dots & \dots & 0 & J_{n_K} \end{bmatrix} \tag{1.17}$$

$$\mathbb{V}\mathrm{ar}(y) = \begin{bmatrix} \sigma_\varepsilon^2 \operatorname{Id}_{n_1} + \sigma_a^2 J_{n_1} & 0 & \dots & \dots & 0 \\ 0 & \sigma_\varepsilon^2 \operatorname{Id}_{n_2} + \sigma_a^2 J_{n_2} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \sigma_\varepsilon^2 \operatorname{Id}_{n_{K-1}} + \sigma_a^2 J_{n_{K-1}} & 0 \\ 0 & \dots & \dots & 0 & \sigma_\varepsilon^2 \operatorname{Id}_{n_K} + \sigma_a^2 J_{n_K} \end{bmatrix} \tag{1.18}$$

On peut donc en déduire avec Eq. (1.10)

$$(\mathbb{V}\mathrm{ar}(y))^{-1} = \begin{bmatrix} \frac{1}{\sigma_\varepsilon^2}\left(\operatorname{Id}_{n_1} - \frac{\sigma_a^2}{\sigma_\varepsilon^2 + n_1 \sigma_a^2} J_{n_1}\right) & 0 & \dots\dots & & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \dots\dots & & 0 & \frac{1}{\sigma_\varepsilon^2}\left(\operatorname{Id}_{n_K} - \frac{\sigma_a^2}{\sigma_\varepsilon^2 + n_K \sigma_a^2} J_{n_K}\right) \end{bmatrix} \tag{1.19}$$

1.3 Matrices stochastiques, doublement stochastiques

SENETA (*Non-negative matrices and Markov chains*)

BHATIA (*Matrix analysis*)

2

Inversion de matrices

2.1 Inversion des matrices par blocs

Pour toutes matrices $A \in \mathbb{R}^{n_1 \times n_1}$, $B \in \mathbb{R}^{n_1 \times n_2}$, $C \in \mathbb{R}^{n_2 \times n_1}$, $D \in \mathbb{R}^{n_2 \times n_2}$ telles que les matrices $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$, A et D sont inversibles on a les relations suivantes :

$$\begin{aligned}\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} &= \begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix} \\ &= \begin{bmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1} + D^{-1} \end{bmatrix} \ .\end{aligned}$$

<u>Conséquence</u> : les déterminants suivants sont égaux :

$$\left|\begin{bmatrix} A & B \\ C & D \end{bmatrix}\right| = |A||D - CA^{-1}B| = |D||A - BD^{-1}C| \ .$$

De plus si $AC = CA$ monter qu'alors $\left|\begin{bmatrix} A & B \\ C & D \end{bmatrix}\right| = |AD - CB|$.

Identité de Woodbury sous les mêmes hypothèses :

$$(A + BDC)^{-1} = A^{-1} - A^{-1}B(D^{-1} + CA^{-1}B)^{-1}CA^{-1} \ . \tag{2.1}$$

La formule précédente donne la relation suivante pour des scalaires $x \neq 0$ et $\delta \neq 0$:

$$\frac{1}{x + \delta} = \frac{1}{x} - \left(\frac{1}{x}\right)^2 \cdot \left(\frac{1}{\delta} + \frac{1}{x}\right)^{-1} \ . \tag{2.2}$$

Enfin relation suivante pour des vecteurs $u \in \mathbb{R}^{n_1}$ et $v \in \mathbb{R}^{n_1}$:

$$\left(A + uv^\top\right)^{-1} = A^{-1} - \frac{A^{-1}uv^\top A^{-1}}{1 + v^\top A^{-1}u} \ . \tag{2.3}$$

Rem: On appelle complément de Schur du bloc D de la matrice M, la matrice de dimension $n_1 \times n_2$ suivante : $A - BD^{-1}C$, qu'on note parfois M/D.

Démonstration. Premier point : commençons par appliquer un premier pivot (par bloc) sur notre matrice, en éliminant la matrice "C" en bas à gauche :

$$\begin{bmatrix} I_{n_1} & 0 \\ -CA^{-1} & I_{n_2} \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A & B \\ 0 & D - CA^{-1}B \end{bmatrix} .$$

On élimine de la même manière le bloc en haut à droite :

$$\begin{bmatrix} I_{n_1} & 0 \\ -CA^{-1} & I_{n_2} \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I_{n_1} & -A^{-1}B \\ 0 & I_{n_2} \end{bmatrix} = \begin{bmatrix} A & B \\ 0 & D - CA^{-1}B \end{bmatrix} \begin{bmatrix} I_{n_1} & -A^{-1}B \\ 0 & I_{n_2} \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix} .$$

Remarquons alors que l'on peut facilement inverser les matrices triangulaires par blocs :

$$\begin{aligned} \begin{bmatrix} I_{n_1} & A^{-1}B \\ 0 & I_{n_2} \end{bmatrix} \begin{bmatrix} I_{n_1} & -A^{-1}B \\ 0 & I_{n_2} \end{bmatrix} &= I_{n_1+n_2} \\ \begin{bmatrix} I_{n_1} & 0 \\ CA^{-1} & I_{n_2} \end{bmatrix} \begin{bmatrix} I_{n_1} & 0 \\ -CA^{-1} & I_{n_2} \end{bmatrix} &= I_{n_1+n_2} . \end{aligned}$$

Ainsi en inversant les matrices diagonales, on obtient la factorisation LDU (Low triangular, Diagonal, Upper triangular)

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I_{n_1} & 0 \\ CA^{-1} & I_{n_2} \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix} \begin{bmatrix} I_{n_1} & A^{-1}B \\ 0 & I_{n_2} \end{bmatrix} . \tag{2.4}$$

En inversant les deux côtés :

$$\begin{aligned} \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} &= \begin{bmatrix} I_{n_1} & A^{-1}B \\ 0 & I_{n_2} \end{bmatrix}^{-1} \begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix}^{-1} \begin{bmatrix} I_{n_1} & 0 \\ CA^{-1} & I_{n_2} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} I_{n_1} & -A^{-1}B \\ 0 & I_{n_2} \end{bmatrix} \begin{bmatrix} A^{-1} & 0 \\ 0 & (D - CA^{-1}B)^{-1} \end{bmatrix} \begin{bmatrix} I_{n_1} & 0 \\ -CA^{-1} & I_{n_2} \end{bmatrix} \\ &= \begin{bmatrix} A^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ 0 & (D - CA^{-1}B)^{-1} \end{bmatrix} \begin{bmatrix} I_{n_1} & 0 \\ -CA^{-1} & I_{n_2} \end{bmatrix} \\ &= \begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix} . \end{aligned} \tag{2.5}$$

On procède de même pour la seconde relation et on obtient la factorisation UDL (Upper triangular, Diagonal, Lower Triangular) :

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I_{n_1} & BD^{-1} \\ 0 & I_{n_2} \end{bmatrix} \begin{bmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} I_{n_1} & 0 \\ D^{-1}C & I_{n_2} \end{bmatrix} . \tag{2.6}$$

De nouveau on peut inverser les deux membres :

$$\begin{aligned} \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} &= \begin{bmatrix} I_{n_1} & 0 \\ D^{-1}C & I_{n_2} \end{bmatrix}^{-1} \begin{bmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{bmatrix}^{-1} \begin{bmatrix} I_{n_1} & BD^{-1} \\ 0 & I_{n_2} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} I_{n_1} & 0 \\ -D^{-1}C & I_{n_2} \end{bmatrix} \begin{bmatrix} (A - BD^{-1}C)^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix} \begin{bmatrix} I_{n_1} & -BD^{-1} \\ 0 & I_{n_2} \end{bmatrix} \\ &= \begin{bmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1} + D^{-1} \end{bmatrix} . \end{aligned} \tag{2.7}$$

Second point : il suffit de calculer les déterminants dans les relations données par les Équations (2.4) et (2.6). □

3

Décompositions classiques

3.1 Décomposition spectrale

Horn et Johnson (*Topics in matrix analysis*)

Golub et van Loan (*Matrix computations*)

Strang (*Introduction to linear algebra*)

3.1.1 Valeurs propres du laplacien sur graphe

TODO: `http://www.math.ucsd.edu/~fan/research/cb/ch1.pdf`

3.2 Décomposition en valeurs singulières (SVD)

TODO: SVD et ses variantes

Deuxième partie

Outils pour l’ANOVA

4

Matrice pour l'analyse de la variance

Plan d'expériences : on observe K classes $C_1, \dots, C_K$ (*e.g.*, les variétés de blé sur une parcelle) et n observations d'un phénomène (*e.g.*, le rendement de la variété) sont consignées. On fait l'hypothèse que les classes C_k sont disjointes et forment une partition des observations : $\cup_{k=1}^{K} C_k = [\![1, n]\!]$ et $\forall (k, k') \in [\![1, K]\!], C_k \cap C_{k'} = \varnothing$. On appelle cet encodage l'encodage "un-chaud" (: *one-hot*) ou par variable indicatrice, ou encore par variable factice (: *dummy variable*). Enfin on suppose que la cardinalité de chaque classe C_k est n_k, et donc que $n = \sum_{k=1}^{K} n_k$.

Le modèle de l'ANOVA peut alors s'écrire, sous l'hypothèse gaussienne que pour tout $i \in [\![1, n]\!]$, $\varepsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$ et pour tout $(i, k) \in [\![1, n]\!] \times [\![1, K]\!]$,

$$y_i = \mu + a_k \delta_{i,C_k} + \varepsilon_i \tag{4.1}$$

où $\delta_{i,C_k} = \begin{cases} 1, & \text{si } i \in C_k \\ 0, & \text{si } i \notin C_k \end{cases}$.

Interprétation : les a_k sont les coefficients qui correspondent au niveau d'influence de la k^{e} classe

$$y = \underbrace{\begin{bmatrix} \mathbb{1}_n & \mathbb{1}_{C_1} & \dots & \mathbb{1}_{C_K} \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} \mu \\ a_1 \\ \vdots \\ a_K \end{bmatrix}}_{\beta} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_K \end{bmatrix} . \tag{4.2}$$

Remarque 4.1. *La matrice* $X \in \mathbb{R}^{n \times K+1}$ *est de rang* K *car* $\mathbb{1}_n = \sum_{k=1}^{K} \mathbb{1}_{C_k}$ *et les vecteurs* $\mathbb{1}_{C_1}, \dots, \mathbb{1}_{C_K}$ *sont linéairement indépendants et générateurs (cela est vrai car les classes* C_k *sont disjointes et forment une partition de l'espace :* $\cup_{k=1}^{K} C_k = [\![1, n]\!]$*). Ainsi on ne peut pas appliquer la formule* $(X^\top X)^{-1} X^\top y$ *pour obtenir une solution des moindres carrés : en effet* $\mathrm{rg}(X) = \mathrm{rg}(X^\top X) = K < K + 1$ *mais* $X^\top X \in \mathbb{R}^{(K+1)\times(K+1)}$.

On peut expliciter la matrice de Gram $X^\top X$ dans ce contexte :

$$X^\top X = \begin{bmatrix} n & n_1 & \cdots & \cdots & n_K \\ n_1 & n_1 & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & n_{K-1} & 0 \\ n_K & 0 & \cdots & 0 & n_K \end{bmatrix} .$$

4.1 Moindres carrées et contraintes

On cherche pour l'estimation des moindres carrés à résoudre le problème suivant, avec des contraintes sur les a_k, ce qui correspond à

$$\min_{\beta \in \mathbb{R}^{K+1}} \frac{1}{2} \|y - X\beta\|^2 \quad , \tag{4.3}$$

$$\text{t.q. } \beta = (\mu, a_1, \ldots, a_K)^\top \text{ et } \sum_{k=1}^{K} c_k a_k = 0 \quad , \tag{4.4}$$

où le vecteur $c = (c_1, \ldots, c_K)^\top \in \mathbb{R}^K$ est un vecteur encodant les contraintes choisies telles que $\sum_{k=1}^K c_k \neq 0$. On distingue trois types de contraintes :

1. Le cas où l'on choisit une classe C_{k_0} comme référence. Cela revient à prendre $c = (0, \ldots, \underbrace{1}_{\text{en } k_0^{\text{e}} \text{ position}}, \ldots, 0)$, ce qui revient à la contrainte "$a_{k_0} = 0$".
2. Le cas où l'on choisit des contributions des classes centrées autour de la moyenne $c = (1, \ldots, 1)$, ce qui revient à la contrainte $\sum_{k=1}^K a_k = 0$.
3. Le cas où l'on choisit des contributions pondérées des classes centrées autour de la moyenne pondérée $c = (n_1, \ldots, n_K)$, ce qui revient à la contrainte $\sum_{k=1}^K n_k a_k = 0$.

Formation du Lagrangien :

$$\mathcal{L}(\beta, \gamma) = \frac{1}{2} \|y - X\beta\|^2 + \gamma a^\top c \ . \tag{4.5}$$

Condition nécessaire du premier ordre :

$$\frac{\partial \mathcal{L}(\beta, \gamma)}{\partial \beta} = 0 \iff X^\top (X\beta - y) + \gamma c = 0 \tag{4.6}$$

$$\iff X^\top X\beta + \gamma c = X^\top y \tag{4.7}$$

$$\iff \begin{bmatrix} n & n_1 & \cdots & \cdots & n_K & 0 \\ n_1 & n_1 & 0 & \cdots & 0 & c_1 \\ \vdots & 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & n_{K-1} & 0 & c_{K-1} \\ n_K & 0 & \cdots & 0 & n_K & c_K \\ 0 & c_1 & \cdots & c_{K-1} & c_K & 0 \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} X^\top y \\ 0 \end{bmatrix} . \tag{4.8}$$

Tout d'abord il est facile de vérifier que $X^\top y = \begin{bmatrix} n\bar{y}_n \\ n_1\bar{y}_{C_1} \\ \vdots \\ n_K\bar{y}_{C_K} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i\in C_1} y_i \\ \vdots \\ \sum_{i\in C_K} y_i \end{bmatrix}$ avec la notation $\bar{y}_{C_k} = \frac{1}{n_k}\sum_{i\in C_k}$ pour tout $k \in [\![1, K]\!]$.

En multipliant à gauche par la matrice $\mathrm{diag}(\frac{1}{n}, \frac{1}{n_1}, \ldots, \frac{1}{n_K}, 1)$ le système précédent on obtient :

$$\begin{bmatrix} 1 & \frac{n_1}{n} & \cdots & \cdots & \frac{n_K}{n} & 0 \\ 1 & 1 & 0 & \cdots & 0 & \frac{c_1}{n_1} \\ \vdots & 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & 1 & 0 & \frac{c_{K-1}}{n_{K-1}} \\ 1 & 0 & \cdots & 0 & 1 & \frac{c_K}{n_K} \\ 0 & c_1 & \cdots & c_{K-1} & c_K & 0 \end{bmatrix} \begin{bmatrix} \mu \\ a \\ \gamma \end{bmatrix} = \begin{bmatrix} \bar{y}_n \\ \bar{y}_{C_1} \\ \vdots \\ \bar{y}_{C_K} \\ 0 \end{bmatrix} . \tag{4.9}$$

En prenant les équations de la ligne 2 à $K+1$, on en déduit les équations suivantes :

$$\begin{cases} \hat{a}_K & = \bar{y}_{C_K} - \hat{\mu} - \frac{c_K}{n_K}\gamma \\ \vdots & \\ \hat{a}_k & = \bar{y}_{C_k} - \hat{\mu} - \frac{c_k}{n_k}\gamma \\ \vdots & \\ \hat{a}_1 & = \bar{y}_{C_1} - \hat{\mu} - \frac{c_1}{n_1}\gamma \end{cases} \tag{4.10}$$

Avec la première équation du système précédent, *i.e.*,

$$\hat{\mu} + \sum_{k=1}^K \frac{n_k}{n} a_k = \bar{y}_n \tag{4.11}$$

et en sommant avec les poids du système d'équation précédent, on obtient :

$$\sum_{k=1}^K \frac{n_k}{n}\hat{a}_k = \sum_{k=1}^K \frac{n_k}{n}\bar{y}_{C_k} - \hat{\mu} + \frac{\gamma}{n}\sum_{k=1}^K c_k \tag{4.12}$$

ce qui implique que

$$\frac{\gamma}{n}\sum_{k=1}^K c_k = 0 \ . \tag{4.13}$$

Rappelant la condition $\sum_{k=1}^K c_k \neq 0$, on en déduit que $\gamma = 0$.

Enfin, en sommant avec les poids $c_K, \ldots, c_k, \ldots, c_1$ les sysème d'équations dans (4.10), on obtient :

$$0 = \sum_{k=1}^K c_k\hat{a}_k = \sum_{k=1}^K c_k\bar{y}_{C_k} - \sum_{k=1}^K c_k\hat{\mu} - \sum_{k=1}^K \frac{c_K^2}{n_K}\gamma = \sum_{k=1}^K c_k\bar{y}_{C_k} - \sum_{k=1}^K c_k\hat{\mu}$$
$$\iff \hat{\mu} = \frac{\sum_{k=1}^K c_k\bar{y}_{C_K}}{\sum_{k=1}^K c_k}$$

On obtient donc comme solution du système initial

$$\begin{cases} \hat{\mu} & = \frac{\sum_{k=1}^{K} c_k \bar{y}_{C_K}}{\sum_{k=1}^{K} c_k} \\ \hat{a}_1 & = \bar{y}_{C_1} - \hat{\mu} \\ \vdots & \\ \hat{a}_k & = \bar{y}_{C_k} - \hat{\mu} \\ \vdots & \\ \hat{a}_K & = \bar{y}_{C_K} - \hat{\mu} \end{cases} \tag{4.14}$$

On peut maintenant retourner sur les trois cas possibles :

1. $c = (0, \ldots, \underbrace{1}_{\text{en } k_0^{\text{e}} \text{ position}}, \ldots, 0)$:

$$\begin{cases} \hat{\mu} & = \bar{y}_{C_{k_0}} \quad \text{(moyenne de la modalité de référence)} \\ \hat{a}_1 & = \bar{y}_{C_1} - \bar{y}_{C_{k_0}} \\ \vdots & \\ \hat{a}_k & = 0 \\ \vdots & \\ \hat{a}_K & = \bar{y}_{C_K} - \bar{y}_{C_{k_0}} \end{cases} \tag{4.15}$$

2. Le cas où $c = (1, \ldots, 1)$:

$$\begin{cases} \hat{\mu} & = \frac{1}{K} \sum_{k=1}^{K} \bar{y}_{C_K} \quad \text{(moyenne des moyennes par classes)} \\ \hat{a}_1 & = \bar{y}_{C_1} - \hat{\mu} \\ \vdots & \\ \hat{a}_k & = \bar{y}_{C_k} - \hat{\mu} \\ \vdots & \\ \hat{a}_K & = \bar{y}_{C_K} - \hat{\mu} \end{cases} \tag{4.16}$$

3. Le cas où $c = (n_1, \ldots, n_K)$:

$$\begin{cases} \hat{\mu} & = \bar{y}_n \quad \text{(moyenne des observations)} \\ \hat{a}_1 & = \bar{y}_{C_1} - \hat{\mu} \\ \vdots & \\ \hat{a}_k & = \bar{y}_{C_k} - \hat{\mu} \\ \vdots & \\ \hat{a}_K & = \bar{y}_{C_K} - \hat{\mu} \end{cases} \tag{4.17}$$

Exercice 4.1. *Calculer les conditionnements et choisir numériquement la meilleure contrainte possible.*

5

Produit de Kronecker

5.1 Introduction et propriétés

Définition 5.1. *Soient $A \in \mathbb{R}^{m\times n}$ et $B \in \mathbb{R}^{m\times n}$. Leur produit tensoriel ou produit de Kronecker est la matrice $A \otimes B \in \mathbb{R}mp \times nq$, définie par blocs successifs de taille $p \times q$, le bloc d'indice i, j valant $a_{i,j}B$, ou de manière équivalente :*

$$A \otimes B = \begin{bmatrix} a_{1,1}B & \cdots & a_{1,n}B \\ \vdots & \ddots & \vdots \\ a_{m,1}B & \cdots & a_{m,n}B \end{bmatrix} \tag{5.1}$$

ou encore

$$A \otimes B = \begin{bmatrix} a_{1,1}b_{1,1} & a_{1,1}b_{1,2} & \cdots & a_{1,1}b_{1,q} & \cdots & \cdots & a_{1,n}b_{1,1} & a_{1,n}b_{1,2} & \cdots & a_{1,n}b_{1,q} \\ a_{1,1}b_{2,1} & a_{1,1}b_{2,2} & \cdots & a_{1,1}b_{2,q} & \cdots & \cdots & a_{1,n}b_{2,1} & a_{1,n}b_{2,2} & \cdots & a_{1,n}b_{2,q} \\ \vdots & \vdots & \ddots & \vdots & & & \vdots & \vdots & \ddots & \vdots \\ a_{1,1}b_{p1} & a_{1,1}b_{p,2} & \cdots & a_{1,1}b_{p,q} & \cdots & \cdots & a_{1,n}b_{p1} & a_{1,n}b_{p,2} & \cdots & a_{1,n}b_{p,q} \\ \vdots & \vdots & & \vdots & \ddots & & \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & & \ddots & \vdots & \vdots & & \vdots \\ a_{m,1}b_{1,1} & a_{m,1}b_{1,2} & \cdots & a_{m,1}b_{1,q} & \cdots & \cdots & a_{m,n}b_{1,1} & a_{m,n}b_{1,2} & \cdots & a_{m,n}b_{1,q} \\ a_{m,1}b_{2,1} & a_{m,1}b_{2,2} & \cdots & a_{m,1}b_{2,q} & \cdots & \cdots & a_{m,n}b_{2,1} & a_{m,n}b_{2,2} & \cdots & a_{m,n}b_{2,q} \\ \vdots & \vdots & \ddots & \vdots & & & \vdots & \vdots & \ddots & \vdots \\ a_{m,1}b_{p1} & a_{m,1}b_{p,2} & \cdots & a_{m,1}b_{p,q} & \cdots & \cdots & a_{m,n}b_{p1} & a_{m,n}b_{p,2} & \cdots & a_{m,n}b_{p,q} \end{bmatrix}. \tag{5.2}$$

Propriétés et liens avec les opérations usuelles : *Prenons A, B, C et D quatre matrices quelconques. Alors les relations suivantes sont satisfaites :*

$$
\begin{aligned}
A \otimes (B + C) &= A \otimes B + A \otimes C, && (5.3)\\
(B + C) \otimes A &= B \otimes A + C \otimes A, && (5.4)\\
(kA) \otimes B = A \otimes (kB) &= k(A \otimes B) \quad (\textit{pour } k \in \mathbb{R}), && (5.5)\\
(A \otimes B) \otimes C &= A \otimes (B \otimes C), && (5.6)\\
A \otimes 0 = 0 \otimes A &= 0, && (5.7)\\
(A \otimes B)(C \otimes D) &= (AC) \otimes (BD) \quad (\textit{quand } AC \textit{ et } BD \textit{ existent}), && (5.8)\\
(A \otimes B)^\top &= (A^\top \otimes B^\top), && (5.9)\\
\begin{bmatrix} A_1 & A_2 \end{bmatrix} \otimes B &= \begin{bmatrix} A_1 \otimes B & A_2 \otimes B \end{bmatrix} \quad (\textit{pour des matrices concaténées}), && (5.10)\\
(A \otimes B)^{-1} &= A^{-1} \otimes B^{-1} \quad (\textit{pour } A \textit{ et } B \textit{ inversibles}). && (5.11)
\end{aligned}
$$

Propriétés spectrales *Prenons des éléments spectraux des matrices* $A \in \mathbb{R}^{n \times n}$ *et* $B \in \mathbb{R}^{m \times m}$*, i.e.,* $Ax = \alpha x$ *et* $By = \beta y$*, alors*

$$(A \otimes B)(x \otimes y) = \alpha\beta(x \otimes y) \ . \tag{5.12}$$

et par conséquent si l'on prend $\{\alpha_1, \ldots, \alpha_n\}, \{x_1, \ldots, x_n\}$ *et* $\{\beta_1, \ldots, \beta_m\}, \{y_1, \ldots, y_m\}$ *les couples valeurs/vecteurs propres des matrices* A *et* B*, alors la matrice* $A \otimes B$ *a pour éléments propres* $\{\alpha_i \beta_j, i = 1, \ldots n, j = 1, \ldots, m\}$ *et* $\{x_i y_j, i = 1, \ldots n, j = 1, \ldots, m\}$*. Ainsi*

$$
\begin{aligned}
\operatorname{tr}(A \otimes B) &= \operatorname{tr}(A) \operatorname{tr}(B), && (5.13)\\
\det(A \otimes B) &= \det(A)^m \det(B)^n, && (5.14)\\
\operatorname{rg}(A \otimes B) &= \operatorname{rg}(A) \operatorname{rg}(B). && (5.15)
\end{aligned}
$$

References

CLARKE *(Linear models : the theory and application of analysis of variance)*

SEARLE, CASELLA *et* MCCULLOCH *(Variance components)*

Bibliographie

BHATIA, R. *Matrix analysis*. T. 169. Graduate Texts in Mathematics. New York : Springer-Verlag, 1997 (p. 7).

CLARKE, B. R. *Linear models : the theory and application of analysis of variance*. T. 634. John Wiley & Sons, 2008 (p. 18).

GOLUB, G. H. et C. F. VAN LOAN. *Matrix computations*. Fourth. Johns Hopkins University Press, Baltimore, MD, 2013, p. xiv+756 (p. 11).

HORN, R. A. et C. R. JOHNSON. *Topics in matrix analysis*. Corrected reprint of the 1991 original. Cambridge : Cambridge University Press, 1994, p. viii+607 (p. 11).

SEARLE, S. R., G. CASELLA et C. E. MCCULLOCH. *Variance components*. T. 391. John Wiley & Sons, 2009 (p. 18).

SENETA, E. *Non-negative matrices and Markov chains*. Springer Series in Statistics. New York : Springer, 2006, p. xvi+287 (p. 7).

STRANG, G. *Introduction to linear algebra*. 5th edition. Wellesley, MA : Wellesley-Cambridge Press, 2016, p. x + 574 (p. 11).