

Solution de l'exercice N°1 de la **Série N°2**

**Exercice 1** Reprenons l'exercice 1 du TD 1.

On interroge 1587 étudiants de M2 sur la catégorie socioprofessionnelle de leurs parents. Les étudiants suivent différents cursus: écoles d'ingénieurs, écoles de commerce, universités scientifiques. Les résultats sont les suivants :

	Ouvriers	Employés	Cadres	Professions libérales
Écoles d'ingénieurs	50	280	120	20
Écoles de commerce	8	29	210	350
Universités Scientifiques	150	230	100	40

On veut étudier l'influence du milieu socioprofessionnel des parents sur le type d'étude des enfants. On rappelle que, dans l'exercice 1 du TD 1, nous avons déjà vérifié, par le test du Khi-deux qu'il existe une dépendance entre le milieu socioprofessionnel des parents et le type d'étude des enfants. On peut alors effectuer une analyse factorielle des correspondances sur les données.

1-Donner les centre de gravités,  $g_r$  et  $g_c$ , associés aux profils-linges et profils-colonnes.

2-Donner les matrices diagonales des profils-linges  $D_r$  et des profils-colonnes  $D_c$ .

3-Donner les matrices des profils-linges  $X_r$  et profils-colonnes  $X_c$ .

4-Calculer les deux matrices  $A_r := X_r^t X_c^t$  et  $A_c := X_c^t X_r^t$ .

5-Calculer les valeurs propres et les vecteurs propres de  $A_r$  et de  $A_c$ . Que peut-on déduire?

6-Que représente  $g_r$  (resp.  $g_c$ ) pour la matrice  $A_r$  (resp.  $A_c$ )?

\*\*\*\*\*

## Solution

**Rappel:** la matrice des effectifs observés est

$$N^* = \begin{pmatrix} 50 & 280 & 120 & 20 \\ 8 & 29 & 210 & 350 \\ 150 & 230 & 100 & 40 \end{pmatrix}.$$

La matrice des fréquences observées est

$$N = \begin{pmatrix} 50/1587 & 280/1587 & 120/1587 & 20/1587 \\ 8/1587 & 29/1587 & 210/1587 & 350/1587 \\ 150/1587 & 230/1587 & 100/1587 & 40/1587 \end{pmatrix}.$$

Les fréquences marginales des lignes sont

$$\begin{aligned} f_{1.} &= 50/1587 + 280/1587 + 120/1587 + 20/1587 = 0.296\,16 \\ f_{2.} &= 8/1587 + 29/1587 + 210/1587 + 350/1587 = 0.376\,18 \, , \\ f_{3.} &= 150/1587 + 230/1587 + 100/1587 + 40/1587 = 0.327\,66 \end{aligned}$$

et les fréquences marginales des colonnes sont

$$\begin{aligned} f_{.1} &= 50/1587 + 8/1587 + 150/1587 = 0.131\,06 \\ f_{.2} &= 280/1587 + 29/1587 + 230/1587 = 0.339\,63 \\ f_{.3} &= 120/1587 + 210/1587 + 100/1587 = 0.270\,95 \, . \\ f_{.4} &= 20/1587 + 350/1587 + 40/1587 = 0.258\,35 \end{aligned}$$

1) Le centre de gravité des profils-lignes:

$$g_r = (0.131\,06, 0.339\,63, 0.270\,95, 0.258\,35)^t .$$

Le centre de gravité des profils-colonnes:

$$g_c = (0.296\,16, 0.376\,18, 0.327\,66)^t .$$

Matrice diagonale des profils-lignes

$$D_r = \begin{pmatrix} 0.296\,16 & 0 & 0 \\ 0 & 0.376\,18 & 0 \\ 0 & 0 & 0.327\,66 \end{pmatrix} .$$

2) Matrice diagonale des profils-colonnes

$$D_c = \begin{pmatrix} 0.131\,06 & 0 & 0 & 0 \\ 0 & 0.339\,63 & 0 & 0 \\ 0 & 0 & 0.270\,95 & 0 \\ 0 & 0 & 0 & 0.258\,35 \end{pmatrix} .$$

3) Matrices des profils-lignes

$$\begin{aligned} X_r &= D_r^{-1}N = \begin{pmatrix} 0.296\,16 & 0 & 0 \\ 0 & 0.376\,18 & 0 \\ 0 & 0 & 0.327\,66 \end{pmatrix}^{-1} \\ &\times \begin{pmatrix} 50/1587 & 280/1587 & 120/1587 & 20/1587 \\ 8/1587 & 29/1587 & 210/1587 & 350/1587 \\ 150/1587 & 230/1587 & 100/1587 & 40/1587 \end{pmatrix} \end{aligned}$$

Donc

$$X_r = \begin{pmatrix} 0.106\,38 & 0.595\,74 & 0.255\,32 & 4.255\,3 \times 10^{-2} \\ 0.013\,4 & 4.857\,6 \times 10^{-2} & 0.351\,76 & 0.586\,27 \\ 0.288\,46 & 0.442\,31 & 0.192\,31 & 7.692\,4 \times 10^{-2} \end{pmatrix} .$$

Matrices des profils-colonnes

$$\begin{aligned} X_c &= D_c^{-1}N^t = \begin{pmatrix} 0.131\,06 & 0 & 0 & 0 \\ 0 & 0.339\,63 & 0 & 0 \\ 0 & 0 & 0.270\,95 & 0 \\ 0 & 0 & 0 & 0.258\,35 \end{pmatrix}^{-1} \\ &\times \begin{pmatrix} 50/1587 & 280/1587 & 120/1587 & 20/1587 \\ 8/1587 & 29/1587 & 210/1587 & 350/1587 \\ 150/1587 & 230/1587 & 100/1587 & 40/1587 \end{pmatrix}^T . \end{aligned}$$

Donc

$$X_c = \begin{pmatrix} 0.240\,39 & 3.846\,3 \times 10^{-2} & 0.721\,18 \\ 0.519\,49 & 5.380\,4 \times 10^{-2} & 0.426\,72 \\ 0.279\,07 & 0.488\,37 & 0.232\,56 \\ 0.048\,78 & 0.853\,66 & 9.756\,1 \times 10^{-2} \end{pmatrix}.$$

4) Calcul de la matrice  $A_r := X_r^t X_c^t$  :

$$A_r = \begin{pmatrix} 0.106\,38 & 0.595\,74 & 0.255\,32 & 4.255\,3 \times 10^{-2} \\ 0.013\,4 & 4.857\,6 \times 10^{-2} & 0.351\,76 & 0.586\,27 \\ 0.288\,46 & 0.442\,31 & 0.192\,31 & 7.692\,4 \times 10^{-2} \end{pmatrix}^T \\ \times \begin{pmatrix} 0.240\,39 & 3.846\,3 \times 10^{-2} & 0.721\,18 \\ 0.519\,49 & 5.380\,4 \times 10^{-2} & 0.426\,72 \\ 0.279\,07 & 0.488\,37 & 0.232\,56 \\ 0.048\,78 & 0.853\,66 & 9.756\,1 \times 10^{-2} \end{pmatrix}^T.$$

Donc:

$$A_r = \begin{pmatrix} 0.234\,12 & 0.179\,08 & 0.103\,32 & 4.477\,1 \times 10^{-2} \\ 0.464\,06 & 0.500\,84 & 0.292\,84 & 0.113\,68 \\ 0.213\,60 & 0.233\,62 & 0.287\,76 & 0.331\,50 \\ 8.825\,5 \times 10^{-2} & 8.647\,5 \times 10^{-2} & 0.316\,08 & 0.510\,06 \end{pmatrix}.$$

Calcul de la matrice  $A_c := X_c^t X_r^t$  :

$$A_c = \begin{pmatrix} 0.240\,39 & 3.846\,3 \times 10^{-2} & 0.721\,18 \\ 0.519\,49 & 5.380\,4 \times 10^{-2} & 0.426\,72 \\ 0.279\,07 & 0.488\,37 & 0.232\,56 \\ 0.048\,78 & 0.853\,66 & 9.756\,1 \times 10^{-2} \end{pmatrix}^T \\ \times \begin{pmatrix} 0.106\,38 & 0.595\,74 & 0.255\,32 & 4.255\,3 \times 10^{-2} \\ 0.013\,4 & 4.857\,6 \times 10^{-2} & 0.351\,76 & 0.586\,27 \\ 0.288\,46 & 0.442\,31 & 0.192\,31 & 7.692\,4 \times 10^{-2} \end{pmatrix}^T.$$

Donc

$$A_c = \begin{pmatrix} 0.408\,38 & 0.155\,22 & 0.356\,54 \\ 0.197\,16 & 0.675\,39 & 0.194\,48 \\ 0.394\,46 & 0.169\,39 & 0.449 \end{pmatrix}.$$

5) Calcul des valeurs propres et les vecteurs propres de  $A_r$  :

$$A_r = \begin{pmatrix} 0.234\,12 & 0.179\,08 & 0.103\,32 & 4.477\,1 \times 10^{-2} \\ 0.464\,06 & 0.500\,84 & 0.292\,84 & 0.113\,68 \\ 0.213\,60 & 0.233\,62 & 0.287\,76 & 0.331\,50 \\ 8.825\,5 \times 10^{-2} & 8.647\,5 \times 10^{-2} & 0.316\,08 & 0.510\,06 \end{pmatrix}.$$

Les valeurs propres

$$\lambda_1 = 1.0, \lambda_2 = 0.479\,66, \lambda_3 = 5.310\,9 \times 10^{-2}, \lambda_4 = 5.997 \times 10^{-7}.$$

Les vecteurs propres

$$u_1 = \begin{pmatrix} 0.250\,99 \\ 0.650\,40 \\ 0.518\,87 \\ 0.494\,74 \end{pmatrix}, \quad u_2 = \begin{pmatrix} 0.256\,30 \\ 0.630\,57 \\ -0.175\,67 \\ -0.711\,22 \end{pmatrix}$$

$$u_3 = \begin{pmatrix} 0.723\,78 \\ -0.625\,73 \\ -0.248\,77 \\ 0.150\,7 \end{pmatrix}, \quad u_4 = \begin{pmatrix} 4.257\,9 \times 10^{-2} \\ -0.409\,29 \\ 0.801\,19 \\ -0.434\,47 \end{pmatrix}.$$

Calcul des valeurs propres et les vecteurs propres de  $A_c$  :

$$A_c = \begin{pmatrix} 0.408\,38 & 0.155\,22 & 0.356\,54 \\ 0.197\,16 & 0.675\,39 & 0.194\,48 \\ 0.394\,46 & 0.169\,39 & 0.449 \end{pmatrix}.$$

Les valeurs propres

$$\lambda_1 = 1.0, \lambda_2 = 0.479\,66, \lambda_3 = 5.310\,9 \times 10^{-2}.$$

Les vecteurs propres

$$u_1 = \begin{pmatrix} 0.510\,48 \\ 0.648\,41 \\ 0.564\,78 \end{pmatrix}, u_2 = \begin{pmatrix} 0.383\,99 \\ -0.816\,03 \\ 0.432\,02 \end{pmatrix}, u_3 = \begin{pmatrix} 0.709\,33 \\ -4.450\,4 \times 10^{-3} \\ -0.704\,86 \end{pmatrix}.$$

6) On remarque que l'ensemble des valeurs propres de  $A_c$  est inclus dans l'ensemble des valeurs propres de  $A_r$ .

7) Nous avons

$$\begin{aligned} g_r &= (0.131\,06, 0.339\,63, 0.270\,95, 0.258\,35)^T \\ &= \begin{pmatrix} 0.131\,06 \\ 0.339\,63 \\ 0.270\,95 \\ 0.258\,35 \end{pmatrix}. \end{aligned}$$

Remarquons que

$$1.915\,1 g_r = 1.915\,1 \begin{pmatrix} 0.131\,06 \\ 0.339\,63 \\ 0.270\,95 \\ 0.258\,35 \end{pmatrix} = \begin{pmatrix} 0.250\,99 \\ 0.650\,40 \\ 0.518\,87 \\ 0.494\,74 \end{pmatrix} = u_1.$$

Comme  $u_1$  est un vecteur propre de  $A_r$  associé à la valeur propre  $\lambda_1 = 1$ , donc  $g_r$  est aussi un vecteur propre de  $A_r$  associé à la même valeur propre  $\lambda_1 = 1$ . Avec le même raisonnement, on en déduit que  $g_c$  est aussi un vecteur propre de  $A_c$  associé à la même valeur propre  $\lambda_1 = 1$ .