Analyse Factorielles des correspondances
A. F. C.

I/ Introduction

Soient X et Y deux variables observées sur une population I de taille n. On suppose que X et Y ont respectivement p et q modalités (p et q classes dans le cas de variables quantitatives)

On note par $M_1, \ldots, M_p$ les modalités de X et par $N_1, \ldots, N_q$ celles de Y.

Soit K le tableau de contingence associé.

$$\underset{(p,q)}{K} = \begin{array}{c} \\ M_1 \\ M_2 \\ \vdots \\ M_p \end{array} \begin{pmatrix} N_1 & N_2 & \cdots & N_q \\ n_{11} & & & \\ & & & \\ & & n_{ij} & \\ & & & \end{pmatrix}$$

n_{ij} est le nombre d'individus ayant la modalité M_i pour X et N_j pou[r Y]

$$\sum_{i=1}^{p} \sum_{j=1}^{q} n_{ij} = n.$$

Rq: dans le cas où X et Y ne sont indépendantes, on desir remonter les liens pouvant exister entre les modalités de X et de Y

II/ Problème posé

Rechercher la meilleure représentation entre les modalités de X et celles de Y. En d'autre terme, on desire représenter sur un même axe les points communs M_i et N_j $i=1,\ldots,p$ et $j=1,\ldots,q$.

L'A.F.C répond à ce type de problème. Ce qui revient à effectuer deux A.F.G, une sur le tableau profils lignes et l'autre sur le tableau profils colonnes. Où $\mathbb{R}^p$ et $\mathbb{R}^q$ sont munis resp des metrique Q et S qui on définira

Notation

$n_{i\cdot} = \sum_{j=1}^{q} n_{ij}$ effectif de la modalité M_i $i=1,\ldots,p$

$f_{i\cdot} = \frac{n_{i\cdot}}{n}$ fréquence de M_i

$n_{\cdot j} = \sum_{i=1}^{p} n_{ij}$ effectif de la modalité N_j $j=1,\ldots,q$

$f_{\cdot j} = \frac{n_{\cdot j}}{n}$ fréquence de N_j

On a $\sum_{i=1}^{p} n_{i\cdot} = \sum_{j=1}^{q} n_{\cdot j} = n$ taille de la population

$$f_{ij} = \frac{n_{ij}}{n}$$

$$\sum_{i=1}^{p} \sum_{j=1}^{q} f_{ij} = \sum_{i=1}^{p} f_{i\cdot} = \sum_{j=1}^{q} f_{ij} = 1$$

III / Tableaux profils lignes et profils colonnes.

III.1 / Tableau profils lignes

$$\underset{(p,q)}{X_I} = \begin{pmatrix} \frac{n_{11}}{n_{1\cdot}} & \frac{n_{12}}{n_{1\cdot}} & \cdots & \frac{n_{1q}}{n_{1\cdot}} \\ \vdots & & & \\ \frac{n_{p1}}{n_{p\cdot}} & \cdots & & \frac{n_{pq}}{n_{p\cdot}} \end{pmatrix}$$

$x_{i\cdot} = \left(\frac{n_{i1}}{n_{i\cdot}}, \cdots, \frac{n_{iq}}{n_{i\cdot}}\right)$ et le $i^{\text{ème}}$ profil ligne, ${}^t x_i \in \mathbb{R}^q$
$i = 1, \cdots, q$

III.2 / Tableau profils colonnes

$$\underset{(q,p)}{X_J} = \begin{pmatrix} \frac{n_{11}}{n_{\cdot 1}} & \cdots & \cdots & \frac{n_{p1}}{n_{\cdot 1}} \\ \frac{n_{12}}{n_{\cdot 2}} & & & \vdots \\ \vdots & & & \vdots \\ \frac{n_{1q}}{n_{\cdot q}} & & & \frac{n_{pq}}{n_{\cdot q}} \end{pmatrix}$$

$y_j = {}^t\left(\frac{n_{1j}}{n_{\cdot j}}, \cdots, \frac{n_{pj}}{n_{\cdot j}}\right)$ ${}^t y_j \in \mathbb{R}^p$ le $j^{\text{ème}}$ profil colonne
$j = 1 \cdots q$

Effectuer l'AF.G du tableau X revient à faire une
a. AFG du tableau. X_I
b. AF.G " " X_J

Dans le cas des metriques suivantes.

III.3 / Choix de metriques

Pour l'analyse de X_I, on se place sur $\mathbb{R}^q$, et on muni $\mathbb{R}^q$ de la metrique Q_q et $\mathbb{R}^p$ de la metrique Δ_p

$$\underset{(q,q)}{Q_q} = \begin{pmatrix} 1/f_{.1} & & 0 \\ & \ddots & \\ 0 & & 1/f_{.q} \end{pmatrix} \qquad \underset{(p,p)}{D_p} = \begin{pmatrix} f_{1.} & & 0 \\ & \ddots & \\ 0 & & f_{p.} \end{pmatrix}$$

<u>Proposition</u>

On a $X_I = D_p^{-1} F$ F tableau des fréquences $F = \left(\frac{n_{ij}}{n}\right)_{\le}$

Pour l'analyse du tableau X_J, on muni $\mathbb{R}^p$ de la metrique Q'_p et $\mathbb{R}^q$ de la metrique D'_q et on se place sur $\mathbb{R}^p$

$$\underset{(p,p)}{Q'_p} = D_p^{-1} = \begin{pmatrix} 1/f_{1.} & & 0 \\ & \ddots & \\ 0 & & 1/f_{p.} \end{pmatrix} \qquad D'_q = Q_q^{-1} = \begin{pmatrix} f_{.1} & & 0 \\ & \ddots & \\ 0 & & f_{.q} \end{pmatrix}$$

IV/ <u>Analyse factorielle</u>.

IV.1/ <u>Analyse factorielle du tableau profils lignes</u>.

Soit X_I le tableau de profils lignes de dim (p,q), on se place sur $(\mathbb{R}^q, Q_q)$, et on muni $(\mathbb{R}^p, D_p)$

On suit les même étapes de l'A.F.G, on se ramène à la diagonalisation de la matrice $\underset{(q,p)}{{}^tX_I}\,\underset{(p,p)}{D_p}\,\underset{(p,q)}{X_I}\,\underset{(q,q)}{Q_q}$

Rq: Le tableau X_I n'est pas centré.

Defs: - La matrice ${}^tX_I D_p X_I$ est dite matrice d'inertie

- Les axes factoriels sont les droites passant par l'origine de $\mathbb{R}^q$ et engendrés par les vecteurs propres normés de ${}^tX_I D_p X_I Q_q$ associés au plus grandes valeurs propres.

Soit w_i la projection Q_q-orthogonale du $i^{ème}$ profil ligne sur Δu

$w_i = \langle {}^tx_i, u \rangle_{Q_q} = {}^tx_i Q_q u$

$w = \begin{pmatrix} w_1 \\ \vdots \\ w_p \end{pmatrix} = X_I Q_q u$ le vecteur des projections d's p profils lignes

Centre de gravité du nuage profils lignes.

$$N(I) = \left\{ {}^t\underline{x}_i , i = 1, \ldots, p \text{ et } {}^t\underline{x}_i \in \mathbb{R}^q \right\}$$

$$\underline{g}_I = \sum_{i=1}^{p} f_{i\cdot} {}^t\underline{x}_i = \sum_{i=1}^{p} f_{i\cdot} \begin{pmatrix} \frac{n_{i1}}{n_{i\cdot}} \\ \vdots \\ \frac{n_{iq}}{n_{i\cdot}} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{p} \frac{n_{i\cdot}}{n} \times \frac{n_{i1}}{n_{i\cdot}} \\ \vdots \\ \sum_{i=1}^{p} \frac{\cancel{n_{i\cdot}}}{n} \times \frac{n_{iq}}{\cancel{n_{i\cdot}}} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{p} \frac{n_{i1}}{n} \\ \vdots \\ \sum_{i=1}^{p} \frac{n_{iq}}{n} \end{pmatrix} = \begin{pmatrix} f_{\cdot 1} \\ \vdots \\ f_{\cdot q} \end{pmatrix}$$

~~Rq~~ Proposition

On pose $\underset{(p,q)}{F} = (f_{ij})_{\substack{1 \le i \le p \\ 1 \le j \le q}}$ tableau de fréquence.

On a
$$X_I = D_p^{-1} F$$

Preuve

$$D_p^{-1} F = \begin{pmatrix} \frac{1}{f_{1\cdot}} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{f_{p\cdot}} \end{pmatrix} \begin{pmatrix} f_{11} & \cdots & f_{1q} \\ \vdots & & \vdots \\ f_{p1} & \cdots & f_{pq} \end{pmatrix} = \begin{pmatrix} \frac{f_{11}}{f_{1\cdot}} & \cdots & \frac{f_{1q}}{f_{1\cdot}} \\ \vdots & & \vdots \\ \frac{f_{p1}}{f_{p\cdot}} & \cdots & \frac{f_{pq}}{f_{p\cdot}} \end{pmatrix} = X_I$$

on a $= \frac{f_{ij}}{f_{i\cdot}} = \frac{\frac{n_{ij}}{n}}{\frac{n_{i\cdot}}{n}} = \frac{n_{ij}}{n_{i\cdot}} = j^{\text{ème}}$ composante de $\underline{x}_i$

Rq : la matrice à diagonaliser devient :
$${}^t(D_p^{-1} F)\, D_p\, (D_p^{-1} F)\, Q_q = {}^tF\, D_p^{-1}\, F\, Q_q .$$

IV.2) Analyse factorielle du tableau profils colonnes

Soit X_J le tableau profil colonne de dimension (q,p), on se place sur $(\mathbb{R}^p, Q'_p) = (\mathbb{R}^p, D_p^{-1})$ et on munie $(\mathbb{R}^q, D'_q) = (\mathbb{R}^q, Q_q^{-1})$

On effectue l'A.F.G du tableau X_J, on se ramène à la diagonalisation de la matrice $\cancel{{}^tX_J D'_q X_J Q'_p}$ $\underset{(p,q)}{{}^tX_J}\ \underset{(q,q)}{D'_q}\ X_J\ Q'_p$

De même X_J n'est pas forcément centré

et ${}^tX_J\, D'_q\, X_J\, Q'_p$ ~~n'est~~ est dite matrice d'inertie.

Les axes factoriels sont les droites engendrées par les vecteurs propres normés de ${}^tX_J D'_q X_J Q'_p$ associés aux plus grandes valeurs propres.

On note par ϑ^j la projection du $j^{ème}$ profil colonne sur Δ_s.

$$\vartheta^j = \langle {}^t y_j, s \rangle_{Q'_p} = {}^t y_j \cdot Q'_p \, s$$

$\vartheta = \begin{pmatrix} \vartheta^1 \\ \vdots \\ \vartheta^q \end{pmatrix}$ le vecteur des projections des q profils colonnes.

$$\boxed{\vartheta = {}^t X_J \, Q'_p \, . s.}$$

Centre de gravité du nuage profils colonnes

$$N(J) = \{ y_j \,,\ j = 1, \dots, q \text{ et } y_j \in \mathbb{R}^p \}$$

$$g_J = \sum_{j=1}^{q} f_{\cdot j} \, y_j = \sum_{j=1}^{q} f_{\cdot j} \begin{pmatrix} \frac{n_{1j}}{n_{\cdot j}} \\ \vdots \\ \frac{n_{pj}}{n_{\cdot j}} \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{q} f_{\cdot j} \frac{n_{1j}}{n_{\cdot j}} \\ \vdots \\ \sum_{j=1}^{q} f_{\cdot j} \frac{n_{pj}}{n_{\cdot j}} \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{q} \frac{n_{\cdot j}}{n} \cdot \frac{n_{1j}}{n_{\cdot j}} \\ \vdots \\ \sum_{j=1}^{q} \frac{n_{\cdot j}}{n} \cdot \frac{n_{pj}}{n_{\cdot j}} \end{pmatrix} = \begin{pmatrix} f_{1\cdot} \\ \vdots \\ f_{p\cdot} \end{pmatrix}$$

Proposition

On a $\underset{(q,p)}{X_J} = \underset{(q,q)}{D_q'^{-1}} \, \underset{(q,p)}{{}^t F} = Q_q \, {}^t F$

Preuve

$$\underset{(q,q)}{Q_q} \cdot {}^t F = \begin{pmatrix} \frac{1}{f_{\cdot 1}} & & \\ & \ddots & \\ & & \frac{1}{f_{\cdot q}} \end{pmatrix} \begin{pmatrix} f_{11} & \dots & f_{p1} \\ \vdots & & \vdots \\ f_{1q} & \dots & f_{pq} \end{pmatrix} = \begin{pmatrix} \frac{f_{11}}{f_{\cdot 1}} & \dots & \frac{f_{p1}}{f_{\cdot 1}} \\ \vdots & & \\ \frac{f_{1q}}{f_{\cdot q}} & \dots & \frac{f_{pq}}{f_{\cdot q}} \end{pmatrix}$$

et $\frac{f_{ij}}{f_{\cdot j}} = \frac{\frac{n_{ij}}{n}}{\frac{n_{\cdot j}}{n}} = \frac{n_{ij}}{n_{\cdot j}}$

$$= \begin{pmatrix} \frac{n_{11}}{n_{\cdot 1}} & \dots & \frac{n_{p1}}{n_{\cdot 1}} \\ \vdots & & \\ \frac{n_{1q}}{n_{\cdot q}} & \dots & \frac{n_{pq}}{n_{\cdot q}} \end{pmatrix} = X_J$$

La matrice à diagonaliser devient

$$ {}^t(Q_q \, {}^t F) \, D'_q \, (Q_q \, {}^t F) \, Q'_p = F \, D_q'^{-1} \, {}^t F \, Q'_p$$

Rq : $\|g_I\|^2_{Q_q} = \|g_J\|^2_{Q'_p} = 1$. En effet

$$\|g_I\|_{Q_q} = {}^t g_I \, Q_q \, g_I = (f_{\cdot 1}, \dots, f_{\cdot q}) \begin{pmatrix} 1/f_{\cdot 1} & & 0 \\ & \ddots & \\ 0 & & 1/f_{\cdot q} \end{pmatrix} \begin{pmatrix} f_{\cdot 1} \\ \vdots \\ f_{\cdot q} \end{pmatrix}$$

$$= (1, \dots, 1) \begin{pmatrix} f_{\cdot 1} \\ \vdots \\ f_{\cdot q} \end{pmatrix} = \sum_{j=1}^{q} f_{\cdot j} = 1$$

$$\|g_J\|^2_{Q'_p} = {}^t g_J\, Q'_p\, g_J = (f_{1\cdot}, \dots, f_{p\cdot}) \begin{pmatrix} 1/f_{1\cdot} & & 0 \\ & \ddots & \\ 0 & & 1/f_{p\cdot} \end{pmatrix} \begin{pmatrix} f_{1\cdot} \\ \vdots \\ f_{p\cdot} \end{pmatrix}$$

$$= (1, \dots, 1) \begin{pmatrix} f_{1\cdot} \\ \vdots \\ f_{p\cdot} \end{pmatrix} = \sum_{i=1}^{p} f_{i\cdot} = 1$$

V/ Relation entre les espaces propres.

<u>Proposition</u>

Les matrices $\underset{(q,q)}{{}^tF\, D_p^{-1}\, F\, Q_q}$ et $\underset{(p,p)}{F\,(D'_q)^{-1}\, {}^tF\, Q'_p}$ ont les mêmes valeurs propres.

Preuve : voir l'A.F.G.

On note par u un $\vec{vp}$ normé de ${}^tF\, D_q^{-1}\, F\, Q_q$ et par s un $\vec{vp}$ normé de $F(D'_q)^{-1}\, {}^tF\, Q'_p$ associés à la même vp λ

Les vecteurs $\underset{(q,1)}{\alpha} = Q_q\, u$ et $\underset{(p,1)}{\beta} = Q'_p\, s$ sont dits facteurs

<u>Proposition</u>

On a

$$\beta = \frac{1}{\sqrt{\lambda}}\, X_I\, \alpha$$

et

$$\alpha = \frac{1}{\sqrt{\lambda}}\, X_J\, \beta$$

Ce qui donne les formules de transition suivants

$$\begin{cases} \underset{(p,1)}{s} = \dfrac{1}{\sqrt{\lambda}}\, F\, Q_q\, u \\ \underset{(q,1)}{u} = \dfrac{1}{\sqrt{\lambda}}\, {}^tF\, Q'_p\, s \end{cases}$$

Preuve : voir A.F.G.

VI/ <u>Projection simultannée</u>.

- On se place sur $(\mathbb{R}^q, Q_q)$, la matrice à diagonaliser ${}^tF\, D_p^{-1}\, F\, Q_q$, on note par u ses $\vec{vp}$ et λ vp

. Projection des profils colonnes.

$w_i = \underline{x}_i Q_q u$ et $\boxed{w = X_I Q_q u}$

Projection des profils colonnes.

$v_j = {}^t y_j Q'_p . s = {}^t y_j Q'_p \left(\frac{1}{\sqrt{\lambda}} F Q_q u\right) ({}^t D'^{-1}_q {}^t F)$

$v = \begin{pmatrix} v_1 \\ \vdots \\ v_q \end{pmatrix} = {}^t X_J Q'_p \left(\frac{1}{\sqrt{\lambda}} {}^t F Q_q u\right) = \left(\cancel{Q'_p {}^t F}\right) D_p \left(\frac{1}{\sqrt{\lambda}} F Q_q u\right) = \frac{1}{\sqrt{\lambda}} {}^t F Q_p F Q_q u$

$\boxed{v = \sqrt{\lambda}\, Q_q u}$ $\quad Q_q = (D'^{-1}_q)$

. On se place sur $(\mathbb{R}^p, Q'_p)$ ou $(\mathbb{R}^p, D^{-1}_p)$

Projection des profils lignes.

$w_i = \underline{x}_i Q_q u = \underline{x}_i Q_q \left(\frac{1}{\sqrt{\lambda}} {}^t F Q'_p s\right)$

$w = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = X_I Q_q \left(\frac{1}{\sqrt{\lambda}} {}^t F Q'_p . s\right) = \left(D^{-1}_p F\right) Q_q \left(\frac{1}{\sqrt{\lambda}} {}^t F Q'_p . s\right)$

$= \frac{1}{\sqrt{\lambda}} D^{-1}_p \left(F Q_q {}^t F Q'_p s\right) = \frac{1}{\sqrt{\lambda}} D^{-1}_p (\lambda s) = \boxed{\sqrt{\lambda}\, D^{-1}_p s = \sqrt{\lambda}\, Q'_p s}$

Projection des profils colonnes

$v_j = {}^t y_j Q'_p . s \qquad v = \begin{pmatrix} {}^t y_1 \\ \vdots \\ {}^t y_q \end{pmatrix} Q'_p s = \boxed{{}^t X_J Q'_p . s}$

VI / Aide à l'interpretation

1) Contribution absolue

- Profils lignes $\quad Ca(i) = f_{i.} \frac{|w_i|^2}{\lambda}$
- Profils colonnes $\quad Ca(j) = f_{.j} \frac{|v_j|^2}{\lambda}$

2) Contribution relative

- Profils lignes $\quad Cr(i) = \lambda \frac{|w_i|^2}{d^2_{Q_q}(\underline{x}_i, g_I)}$
- Profils colonnes $\quad Cr(j) = \lambda \frac{|v_j|^2}{d^2_{Q'_p}(y_j, g_J)}$

R_q

$$d^2_{Q_q}(\underline{x}_i, g_I) = \|\underline{x}_i - g_I\|^2_{Q_q} = {}^t(\underline{x}_i - g_I)\, Q_q\, (\underline{x}_i - g_I)$$

$$d^2_{Q'_q}(y_j, g_J) = \|y_j - g_J\|^2_{Q'_q} = {}^t(y_j - g_J)\, Q'_q\, (y_j - g_J)$$