

Analyse factorielle des correspondances (AFC)

1) Introduction:

L'analyse factorielle des correspondances a pour objet le traitement de l'information contenue dans un tableau appelé de contingence ou de dépendance, relatif à deux ensembles de nature quelconque, en relation par moyen d'un processus naturel ou expérimental plus ou moins bien connu.

On peut ainsi étudier les correspondances entre les catégories socio-professionnelles et les principales maladies, entre les personnages d'œuvre de racine et les principaux mots prononcés par ces personnages, etc. Remarquons que les éléments des ensembles précédents sont pondérés : le nombre des malades est différent pour chaque maladie ; le nombre de mots utilisés par chaque personnage n'est pas identique, etc. Ces fréquences de répétition s'interprète facilement

en termes de probabilité.

Le tableau de dépendance peut être ainsi représenté dans un espace approprié par un nuage de points affectés de probabilités.

L'analyse des correspondances consiste en l'étudier la proximité et la dispersion des différents points du nuage entre eux. Pour cela, on projette le nuage transformé sur un sous-espace de dimensions petites (habituellement une, deux ou trois) de telle manière que le nuage projeté soit le moins affecté possible par la projection.

Historiquement, l'étude des nuages de points pondérés s'est fait dans le cadre de la mécanique physique, de-là que parfois dans la littérature scientifique, un vocabulaire identique (inertie, axe principal de l'inertie, tension d'inertie, ...) s'utilise pour l'analyse factorielle des correspondances.

2) Notation:

Considérons un tableau à double entrée, dans lequel les lignes représentent les n éléments (les individus) d'un ensemble I et où les colonnes représentent les m éléments (les paramètres) d'un ensemble J en correspondance avec I .

		Ensemble J (paramètres)						
		1	2	3	...	j	...	m
Ensemble I (individus)	1	x_{11}	x_{12}	x_{13}	...	x_{1j}	...	x_{1m}
	2	x_{21}	x_{22}	x_{23}	...	x_{2j}	...	x_{2m}
	3	:	:	:				
	i	x_{i1}	x_{i2}	x_{i3}	...	x_{ij}	...	x_{im}
	n	x_{n1}	x_{n2}	x_{n3}	...	x_{nj}	...	x_{nm}

les éléments x_{ij} du tableau représentent la mesure du paramètre j pour l'individu i .

Rappelons que tableau précédent se présente sous la forme d'un tableau des uns et des zéros (suivant si l'individu i possède ou non le paramètre j) dans le cas où les observations sont qualitatives.

la probabilité associée au terme x_{ij} est :

$$P_{ij} = \frac{x_{ij}}{\sum_{i=1}^n \sum_{j=1}^m x_{ij}}$$

les lois de probabilités marginales sont alors :

$$P_{i.} = \sum_{j=1}^m P_{ij} \quad \text{avec } i = 1, \dots, n$$

$$P_{.j} = \sum_{i=1}^n P_{ij} \quad \text{avec } j = 1, \dots, m$$

qui vérifient les propriétés : $\sum_{i=1}^n P_{i.} = 1$ et $\sum_{j=1}^m P_{.j} = 1$

le tableau des probabilités, associé au tableau des observations, et ainsi :

Ensemble J

	1	2	...	j	...	m	
1	P_{11}	P_{12}	...	P_{1j}	...	P_{1m}	$P_{1.0}$
2	P_{21}	P_{22}	...	P_{2j}	...	P_{2m}	$P_{2.0}$
i	P_{i1}	P_{i2}	...	P_{ij}	...	P_{im}	$P_{i.0}$
n	P_{n1}	P_{n2}	...	P_{nj}	...	P_{nm}	$P_{n.0}$
	$P_{.1}$	$P_{.2}$...	$P_{.j}$...	$P_{.m}$	1

le tableau des observations peut être aussi représenté par n points L_i (i variant de 1 à n).

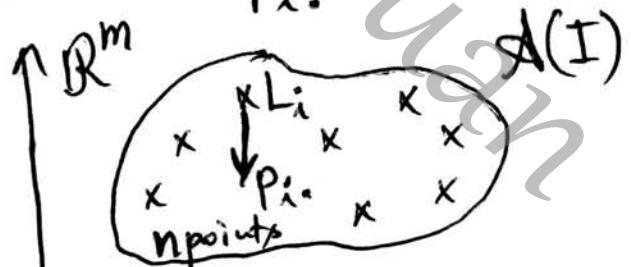
$$L_i = \left(\frac{P_{i1}}{P_{i.0}}, \frac{P_{i2}}{P_{i.0}}, \dots, \frac{P_{ij}}{P_{i.0}}, \dots, \frac{P_{im}}{P_{i.0}} \right)$$

dans un espace de dimension m (\mathbb{R}^m), chaque point est pondéré par la probabilité $P_{i.0}$.

Ensemble I

L_1	$P_{1.0}$
L_2	$P_{2.0}$
:	:
L_i	$P_{i.0}$
:	
L_n	$P_{n.0}$

$$P_{i.0}$$



On obtient alors le nuage de points pondérés associés à l'ensemble I :

$$A(I) = \{L_i; P_{i.0}\}$$

3) Un indice de similitude : la χ^2 du tableau de contingence :

Deux individus i et i' de l'ensemble I se sont « proches » si le pas de i à i' (ou de i' à i) a peu d'influence sur la structure de l'ensemble des paramètres. Cette proximité se traduit par une distance, notée $d^2(i, i')$.

La distance retenue pour définir la proximité ^{de la χ^2} entre i et i' est la distance appelée χ^2 associée à P.j

$$d^2(i, i') = \sum_{j=1}^m \frac{1}{P_{i,j}} \left(\frac{P_{ij}}{P_{i \cdot}} - \frac{P_{i'j}}{P_{i' \cdot}} \right)^2$$

Posons

$$\frac{P_{ij}}{P_{i \cdot}} = \alpha_{ij} \quad \text{et} \quad \frac{P_{i'j}}{P_{i' \cdot}} = \alpha_{i'j},$$

cette distance s'écrit sous forme matricielle :

$$d^2(i, i') = (\alpha_{ij} - \alpha_{i'j}) \begin{pmatrix} \frac{1}{P_{i \cdot}} & & & \\ & \frac{1}{P_{i' \cdot}} & & \\ & & \ddots & \\ & & & \frac{1}{P_{m \cdot}} \end{pmatrix} (\alpha_{ij} - \alpha_{i'j})$$

et aussi

$$d^2(i, i') = (\alpha_{ij} - \alpha_{i'j}) \cdot D \cdot (\alpha_{ij} - \alpha_{i'j})$$

la distance $d^2(i, i')$ apparaît ainsi comme la distance engendrée par une forme quadratique définie positive de matrice D , dans un espace vectoriel euclidien.

Par normalisation des axes, il est possible de rendre unitaire la matrice de la forme quadratique.

Ensuite, suivant ces nouveaux axes, l'individu i a pour coordonnées

$$M_i = (\beta_{i1}, \beta_{i2}, \dots, \beta_{ij}, \dots, \beta_{im})$$

avec

$$\beta_{ij} = \frac{1}{\sqrt{P_j}} \alpha_{ij}$$

ou encore

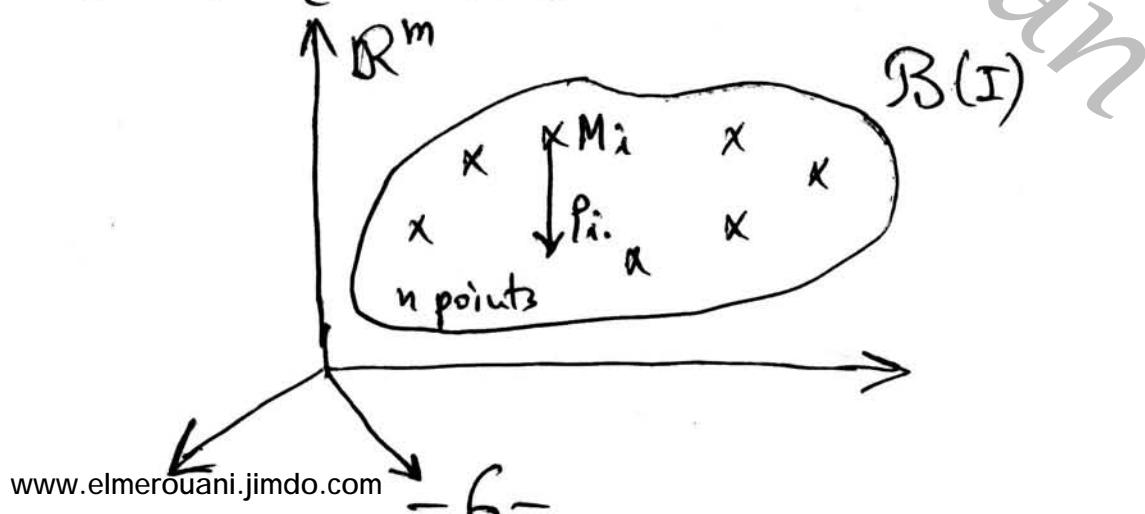
$$\beta_{ij} = \frac{1}{\sqrt{P_j}} \frac{\alpha_{ij}}{P_i}$$

étant toujours la pondération du point i , P_i .

Le nuage des points associés à l'ensemble I apparaît maintenant comme celui de n points M_i , chaque point M_i étant pondéré par la probabilité P_i .

On note par $B(I)$ ce nouveau nuage de points.

$$B(I) = \{M_i; P_i\}$$



La distance entre deux points M_i et $M_{i'}$ est maintenant:

$$d^2(M_i, M_{i'}) = \sum_{j=1}^m (\beta_{ij} - \beta_{i'j})^2$$

soit encore

$$d^2(M_i, M_{i'}) = (\beta_{ij} - \beta_{i'j})' (\beta_{ij} - \beta_{i'j}).$$

le nuage $B(I)$ est aussi un hyperplan d'équation:

$$\sum_{j=1}^m \sqrt{p_{.j}} \beta_{ij} = 1$$

En effet, d'une part :

$$\sum_{j=1}^m \alpha_{ij} = \sum_{j=1}^m \frac{p_{ij}}{p_{i.}} = \frac{1}{p_{i.}} \sum_{j=1}^m p_{ij} = \frac{1}{p_{i.}} p_{i.} = 1$$

d'autre part

$$\sum_{j=1}^m \alpha_{ij} = \sum_{j=1}^m \sqrt{p_{.j}} \beta_{ij};$$

et finalement

$$\sum_{j=1}^m \sqrt{p_{.j}} \beta_{ij} = 1.$$

4) la matrice des variances-covariances W

la matrice des variances-covariances W du nuage $B(I)$ relativement à un paramètre j est :

$$W = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1m} \\ v_{21} & v_{22} & \cdots & v_{2m} \\ \vdots & \ddots & \ddots & \vdots \\ v_{m1} & v_{m2} & \cdots & v_{mm} \end{pmatrix}$$

Où :

- la variance v_{jj} qui caractérise la dispersion du mage tout au long de l'axe j , et définie par :

$$v_{jj} = \sum_i p_i \cdot (\beta_{ij} - \sqrt{p_j})^2 ;$$

- la covariance v_{jk} est définie par :

$$v_{jk} = \sum_i p_i \cdot (\beta_{ij} - \sqrt{p_j})(\beta_{ik} - \sqrt{p_k})'$$

ou bien, en remplaçant, on aura :

$$v_{jk} = \sum_i \frac{p_{ij} p_{ik}}{p_i \cdot \sqrt{p_j p_k}} - \sqrt{p_j p_k} .$$

On appelle la variabilité totale du mage $B(I)$, la trace de la matrice W :

$$\sqrt{B} = \text{tr}(W) = \sum_j v_{jj}$$

5) Projection du mage $B(I)$ sur un axe :

Nous nous proposons de projeter le mage $B(I)$ orthogonalement sur un espace d'une dimension, de telle manière que l'information perdue par la projection soit minimale.

Soit un vecteur unitaire u de direction quelconque sur lequel on projette le mage $B(I)$.

Soit $C(I)$ le mage projeté.

Soit \sqrt{C} la variabilité totale du mage $C(I)$.

Alors l'information perdue par la projection est minime lorsque la variabilité V_c du mage projeté se réduit le moins possible par les projections orthogonales des points de $B(I)$ sur le vecteur u .

On peut démontrer que : $V_c = u' W u$, où u' est la transposée de u .

Ainsi, il s'agit de déterminer le vecteur u qui maximise $V_c = u' W u$ sous la restriction $u'u=1$.

Cette question est analogue à celle déjà résolue dans le chapitre de l'ACP.

Le maximum de V_c correspond à la plus grande valeur propre λ_{\max} de W , matrice des variances-covariances de $B(I)$, et alors $V_c = \lambda_{\max}$.

la partie de variabilité expliquée par la projection de $B(I)$ sur u est alors

$$S = \frac{V_c}{\sqrt{B}}, \text{ soit encore, en remplaçant :}$$

$$S = \frac{\lambda_{\max}}{\text{tr}(W)}.$$

6) Projection de $B(I)$ sur un plan, un hyperplan :

De manière analogue au chapitre de l'ACP, les vecteurs propres de la matrice W s'appellent axes factoriels du mage. La détermination des axes factoriels se fait en diagonalisant la matrice symétrique W . En pratique, on se contente des valeurs propres les plus grands.

Nous avons vu que le terme $v_{j,k}$ de la matrice W s'écrit : $v_{j,k} = \sum_i p_i (\beta_{ij} - \sqrt{p_{i,j}})(\beta_{ik} - \sqrt{p_{i,k}})$

Notre encore

$$v_{j,k} = \sum_i \left(\frac{\beta_{ij} - p_i \cdot p_{i,j}}{\sqrt{p_i \cdot p_{i,k}}} \right) \left(\frac{\beta_{ik} - p_i \cdot p_{i,k}}{\sqrt{p_i \cdot p_{i,k}}} \right)$$

Posons

$$\frac{\beta_{ij} - p_i \cdot p_{i,j}}{\sqrt{p_i \cdot p_{i,k}}} = r_{ij} \quad (i=1, \dots, n; j=1, \dots, m)$$

le terme r_{ij} est le terme général d'une matrice R de type (n, m) , et on a : $W = R'R$
où R' est la transposée de R .

Maximiser $u' W u$ revient à maximiser $u' R'R u$ sous la restriction $u'u = 1$, c'est-à-dire à déterminer les vecteurs propres associés aux valeurs propres de la matrice $R'R$.

7) Recherche des facteurs :

les points du nuage $\mathcal{G}(I)$ possèdent un nombre réduit de coordonnées dans le référentiel formé par les axes factoriels. Ces coordonnées sont les valeurs de nouvelles variables qui s'appellent : Facteurs.

Le premier facteur correspond aux coordonnées sur le premier axe factoriel.

On peut démontrer que les facteurs sont non-correlés entre eux et s'expriment comme combinaisons linéaires des données.

Réiproquement, les données ont des coefficients qui sont des combinaisons linéaires des facteurs.

Ainsi, à partir des facteurs, il est possible de reconstruire un tableau de données avec une minimale perte d'information, c'est-à-dire obtenir un tableau plus facilement accessible à l'analyse descriptive.

8) Proximité en \mathbb{R}^m et en \mathbb{R}^n :

les développements précédentes étudient les proximités entre n points de \mathbb{R}^m .

Par les calculs symétriques, on peut étudier les proximités de m points en \mathbb{R}^n .

Néanmoins, il existe des relations entre les facteurs de \mathbb{R}^m et les facteurs de \mathbb{R}^n .

Par conséquent, il est possible de représenter, sur le même graphique, dans le plan des deux premiers axes factoriels, les proximités entre les éléments de l'ensemble I et les proximités entre les éléments de l'ensemble J.

Cette simultanéité de représentation suggère parfois une signification économique, sociale, politique, ... des axes factoriels.

9) Avantages et inconvénients de l'analyse factorielle des correspondances :

les inconvénients sont les défauts de toute analyse factorielle : déformation inévitable du nuage durant la projection et détermination de la signification (interprétation) des axes factoriels.

l'avantage essentiel réside dans l'étude des caractères qualitatifs.