Eléments complémentaires pour l'AFC

A partir : réduction dimension

- *Proximités entre profils : analyse par rapport l'origine mais possible à partir des centres de gravité*
- *Dans espace R^p avec analyse par rapport à l'origine : réduction de dimension*
- *Critère de projection orthogonale selon un axe où inertie maximale (variance) passant par O et engendré par un vecteur unitaire u et de métrique* $\boldsymbol{D_p}$

Avec les :

- des matrices F contenant (k_{ij}/k) avec k_{ij} l'effectif d'individus de modalité ij simultanément
- D_n^{-1} : matrice diagonale des 1/ marges lignes
- D_p^{-1} : matrice diagonales de 1/marges colonnes (métrique M)

Sur le Profil ligne

$$distance\ du\ \chi 2 \quad : \quad d^2(i,i') = \sum_{j=1}^{p} \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{i'j}}{f_{i'.}} \right)^2$$

n points i du nuage des profils lignes dans R^p chaque point i a pour coordonnées :

$$x_{ij} = \frac{1}{\sqrt{f_{.j}}} \frac{f_{ij}}{f_{i.}}$$

Pour la moyenne pour une variable j = moyenne arithmétique pondérée :

$$\overline{x_j} = \sum_i f_{i.} x_{ij} = \sum_i f_{i.} \frac{1}{\sqrt{f_{.j}}} \frac{f_{ij}}{f_{i.}} = \frac{1}{\sqrt{f_{.j}}} f_{.j}$$

D'où

$$\overline{x_j} = \sqrt{f_{.j}}$$

1° La covariance entre deux var Xj et Xj' (centrage par rapport a la moyenne)

$$cov(x_j, x_{j'}) = V_{jj'} = \sum_i f_{i.} \left[\left(\frac{1}{\sqrt{f_{.j}}}\frac{f_{ij}}{f_{i.}} - \sqrt{f_{.j}}\right)\left(\frac{1}{\sqrt{f_{.j'}}}\frac{f_{ij'}}{f_{i.}} - \sqrt{f_{.j'}}\right)\right]$$

$$V_{jj'} = \sum_i \frac{f_{ij}f_{ij'}}{\sqrt{f_{.j}}\sqrt{f_{.j'}}f_{i.}} - \sqrt{f_{.j}}\sqrt{f_{.j'}}$$

On peut montrer que pour le *p* ième vecteur propre , avec V la matrice de cov a diagonaliser et S la matrice d'inertie :

On a la relation : $\lambda_p u_p = V u_p = S u_p$

Ou S est la matrice (p,p) pour l'espace des Rp (dim p des colonnes)

Ou chaque terme de S est :$s_{jj'} = \sum_i \frac{f_{ij}f_{ij'}}{\sqrt{f_{.j}}\sqrt{f_{.j'}}f_{i.}}$

Les vecteurs propres de la matrice V sont identiques à ceux de la matrice S : on peut donc diagonaliser l'un ou l'autre des matrices (cas ou on a centré les données avec la matrice de cov)

2° Dans le cadre vu en cours : analyse des proximités entre individus par rapport à l'origine : on recherche l'axe d'inertie max du nuages profils lignes passant par origine O et engendré par un vecteur unitaire u pour une métrique D_p^{-1} .

$$Max\left\{\sum_i f_{i.}\, d^2(i,O)\right\}$$

Soit maximiser la somme des distances au carré de Om_i

Soit rendre max la quantité

$$\mathbf{u' D^{-1}_p F' D^{-1}_n F D^{-1}_p u}$$

Avec la contrainte $\boldsymbol{u'} \mathbf{D^{-1}_p u = 1}$

Alors u et vecteur propre la matrice

$\mathbf{S} = \mathbf{F' D^{-1}_n \; F \, D^{-1}_p}$ de terme générale : $s_{jj'} = \sum_{i=1}^{n} \frac{f_{ij} f_{ij'}}{f_{i.} f_{.j'}}$

Or **S** n'est pas symétrique en générale

Alors :

Si on considère la matrice $\mathbf{\hat{A} = F' D^{-1}_n \; F}$ elle est symétrique et la matrice diagonale $\mathbf{D^{-1}_p}$

Alors on pose : la matrice $\mathbf{S = \hat{A} \, D^{-1/2}_p \, D^{-1/2}_p}$

Pour la solution : partant de la **relation** $\mathbf{S \, u = \lambda \, u}$

$$\mathbf{\hat{A} \, D^{-1/2}_p \, D^{-1/2}_p \, u = \lambda \, u}$$

En pré multipliant par $\mathbf{D^{-1/2}_p}$ et en posant que $\mathbf{D^{-1/2}_p \; u = w}$

$$\mathbf{D^{-1/2}_p \, \hat{A} \, D^{-1/2}_p \, w = \lambda \, w}$$

Puis la matrice **A** est symétrique :

$$\mathbf{A = D^{-1/2}_p \, \hat{A} \, D^{-1/2}_p = D^{-1/2}_p \, F' \, D^{-1}_n \; F \, D^{-1/2}_p}$$

Et $\mathbf{A \, w = \lambda \, w}$

La matrice A est diagonalisable et la matrice A et la matrice S ont même valeur propre et le passage de **w** à **u** est donné par :

Le passage du vecteur propre : est donné par $\mathbf{D^{-1/2}_p \; u = w}$

Il est plus facile de diagonaliser la matrice **A avec** $\boldsymbol{\lambda}$ **et** w (valeurs propres et vecteurs propres) on peut reconstruire le vecteur **u.**

A est de terme général :

$$a_{jj'} = \sum_{i=1}^{n} \frac{f_{ij} f_{ij'}}{f_{i.} \sqrt{f_{.j} f_{.j'}}}$$

Si on a choisi comme coordonnées des points i, les p quantités :

$$x_{ij} = \frac{1}{\sqrt{f_{.j}}} \frac{f_{ij}}{f_{i.}} \; (j=1 ,..p)$$

Ce qui permet de diagonaliser une matrice symétrique. La distance du $\chi 2$ devient alors une distance euclidienne classique

$$d^2(i,i') = \sum_{j=1}^{p} \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{i.}\sqrt{f_{.j}}} - \frac{f_{i'j}}{f_{i'.}\sqrt{f_{.j}}} \right)^2$$

De centre de gravité Gj= $\sqrt{f_{.j}}$

Quelques éléments sous R : à tester ou à retrouvé

```
FF<-profil_ij
#D_n<-diag(nrow(x),nrow(x))/nrow(x) suelement la fréquence f_ij
D_n<-diag(f_i.)
D_n_1<-diag(1/f_i.)
Fprim<-t(FF)
D_p<-diag(f_.j)
D_p_1<-diag(1/f_.j)
FD_n_1<-D_n_1%*%FF #
A_at<-Fprim%*%D_n_1%*%FF
L1<-diag(f_.j^(-1/2))
A<-L1%*%A_at%*%L1
de<- eigen(A)
svd(A)
```

Puis à vous de *jouer* avec les packages ade4 ou FactoMineR

Compléments sur les formulations de passage nuage des profils ligne et des profils colonnes (octobre 2019)

Dans R^p profil ligne	Table 1	Dans R^p profil colonne
$\mathbf{S = F' D^{-1}_n\ F D^{-1}_p}$	Matrice à diagonaliser	$\mathbf{T = F D^{-1}_p\ F' D^{-1}_n}$
$\mathbf{S\ u_\alpha = \lambda_\alpha\ u_\alpha}$	Axe factoriel correspondant à la valeur propre λ_α et vecteur propre $\mathbf{u_\alpha}$	$\mathbf{T\ v_\alpha = \lambda_\alpha\ v_\alpha}$
$\mathbf{\psi_\alpha = D^{-1}_n\ F D^{-1}_p\ u_\alpha}$	Coordonnées factorielles	$\mathbf{\varphi_\alpha = D^{-1}_p\ F' D^{-1}_n\ v_\alpha}$
$\psi_{\alpha i} = \sum_{j=1}^{p} \frac{f_{ij}}{f_{i.} f_{.j}} u_{\alpha j}$	Coordonnées factorielles *A noter : coordonnées sont centrées et de variance = λ_α*	$\varphi_{\alpha j} = \sum_{i=1}^{n} \frac{f_{ij}}{f_{i.} f_{.j}} v_{\alpha i}$

Représentation simultanée par les formules de passage suivantes

Pour passer de l'espace profil ligne à espace profil colonne : vecteurs propres

$$\mathbf{v}_\alpha = \frac{1}{\sqrt{\lambda_\alpha}}\ \mathbf{F}\ \mathbf{D}_{\boldsymbol{p}}^{-1}\ \mathbf{u}_\alpha$$

$$\mathbf{u}_\alpha = \frac{1}{\sqrt{\lambda_\alpha}}\ \mathbf{F'}\ \mathbf{D}_{\boldsymbol{n}}^{-1}\ \mathbf{v}_\alpha$$

Avec le table(1) et les deux formules ci-dessus :

$$\boldsymbol{\psi}_\alpha = \sqrt{\lambda_\alpha}\mathbf{D}_{\boldsymbol{n}}^{-1}\ \mathbf{v}_\alpha$$

$$\boldsymbol{\varphi}_\alpha = \sqrt{\lambda_\alpha}\mathbf{D}_{\boldsymbol{p}}^{-1}\,\mathbf{u}_\alpha$$

Soit explicitement

$$\psi_{\alpha i} = \frac{\sqrt{\lambda_\alpha}}{f_{i.}} \quad v_{\alpha i}$$

$$\varphi_{\alpha j} = \frac{\sqrt{\lambda_\alpha}}{f_{.j}} \quad u_{\alpha j}$$

Remarque :

Nuage des profils-colonne (R^n) initial: chaque point *j* de coordonnes $f_{ij}/f_{.j}$ pour $i = 1,..n$; est donc affecté d'une masse $f_{.j}$ les *p* points sont situés dans un sous espace à *n-1* dimension puisque $\sum_{i:1}^{n} \frac{f_{ij}}{f_{.j}} = 1$.