POSITIONNEMENT MULTIDIMENSIONNEL

(MDS)

On considère n individus.

Les données sont les $\frac{n(n-1)}{2}$ valeurs d'un indice (distance, similarité ou dissimilarité) observées ou construites pour chaque couple d'individus. L'objectif est de construire, à partir de cette matrice d'indices une représentation (euclidienne) des individus dans un espace de dimension réduite qui approche "au mieux" les indices observés.

Distances entre variables

Pratiquement, le MDS peut servir à comprendre la structure de liaison d'un grand ensemble de variables.

2 variables quantis : $d^2(X, Y) = 1 - \text{cor}(X, Y)$

(supposées centrées sans perte de généralité)

Rque : les distances associées à une dépendance

2 variables qualis : $d^2(P, Q) = 1 - T^2(X, Y)$, où [illegible]

$$T = \sqrt{\frac{\chi^2_{obs}}{n\sqrt{\nu}}}, \quad \nu = [\text{nb. modalités}(P) - 1][\text{nb. modalités}(Q) - 1] = (r-1)(s-1)$$

$$\text{et } \chi^2_{obs} = n\left[\sum_{\ell=1}^{r}\sum_{h=1}^{s} \frac{n_{\ell h}^2}{n_{\ell .}\, n_{. h}} - 1\right]$$

($n_{\ell h}$:= nb d'individus t.q. $p = \ell$ et $q = h$; $n_{\ell .}$ et $n_{. q}$: comptes marginaux)

donne une approximation empirique de $\frac{P(P=\ell, Q=h)}{P(P=\ell) \cdot P(Q=h)}$, rapport égal à 1 ssi P et Q sont indépendantes

<u>1 variable quali / 1 variable quanti</u> : $d^2(X, C) = 1 - R^2_v(X, C)$

où $R^2_v(X, C)$ est le rapport de variances expliquées sur totale :

$$R^2_v(X,C) = \frac{\text{Var expliquée (par les modalités de la variable quali } C)}{\text{Var totale}}$$

$$= \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{\sum_{k=1}^{K} (\mu_k - \mu)^2}.$$

Maintenant, à partir d'une matrice Δ de distances Δ_{ij}, $i = 1, \dots, n$, $j = 1, \dots, n$ entre points de $\mathbb{R}^n$, on cherche une configuration de ces points dans un espace de dimension q ($q = 2$ le plus souvent) (...suite) qui représente de façon unique et au mieux les Δ_{ij}.

Mis sous forme d'un problème d'optimisation, on cherche $\underset{X \in M_{n,q}(\mathbb{R})}{\arg\min} \| X X^T - \tilde{\Delta} \|_F$

n points dans un espace de dimension q

pour norme de Frobenius = racine carrée de la somme des carrés des éléments

$\mathrm{Id} - \frac{\mathbb{1} . \mathbb{1}^T}{n}$ vecteur de "1"

où $\tilde{\Delta} := -\frac{1}{2} J . \Delta_2 J$ et $(\Delta_2)_{ij} = \Delta_{ij}^2$, $\forall i,j$

<u>rque</u> : • si Δ est euclidienne, on trouve que le minimum ci-dessus vaut 0. Δ euclidienne ssi $\tilde{\Delta}$ a toutes ses valeurs propres ≥ 0

• J est la matrice de projection sur le sous-espace orthogonal à $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \mathbb{1}$

Si on diagonalise $\tilde{\Delta} = U . \Lambda . U^T$, la solution du problème est donnée par $X = U \Lambda^{1/2}$; on a les coordonnées des vecteurs de la représentation euclidienne. Si on ne garde que les 2 premières valeurs propres, X a n lignes et 2 colonnes.

<u>ex</u> : (R) library(MASS) ; loc <- cmdscale(eurodist) ;
x <- loc[,1] ; y <- loc[,2] ; plot(x,y,type='n') ; text(x,y,rownames(loc), cex=0.7