

ANALYSE SPHÉRIQUE (K5)

(25 / 08 / 2020, © Monfort, Dicostat2005, 2005-2020)

L'**analyse sphérique** est une méthode d'**analyse des données**, voisine de l'**analyse des correspondances**, qui se ramène aussi à l'**analyse générale des données** par transformation du tableau initial.

(i) Soit $T \in M_{IJ}(\mathbf{N})$ un **tableau de contingence** et $F \in M_{IJ}(\mathbf{R}_+)$ un tableau (à éléments positifs) donné.

On appelle **analyse factorielle sphérique**, ou simplement **analyse sphérique** (M. VOLLE) de T l'application de l'**agd** au tableau X obtenu à partir de T par les transformations suivantes :

$$(a) T \mapsto F = (f_{ij})_{(i,j)}, \text{ avec } f_{ij} = t_{ij} / t_{..},$$

$$(b) (F, \Phi) \mapsto X = (x_{ij})_{(i,j)}, \text{ avec } x_{ij} = f_{ij}^{1/2} - \varphi_{ij}^{1/2},$$

où $t_{..} = e_i' T e_j$ (total général) et $\Phi = (\varphi_{ij})_{(i,j)}$.

L'**agd** appliquée à X consiste en la **décomposition spectrale** (**diagonalisation**) de $X'X$ ou de XX' .

(ii) La méthode est fondée sur la **distance de HELLINGER** entre les tableaux F et Φ , ie :

$$(1) \quad d^2(F, \Phi) = \sum_{i=1}^I \sum_{j=1}^J (f_{ij}^{1/2} - \varphi_{ij}^{1/2})^2.$$

Cette distance s'associe à la **distance géodésique** sur la sphère :

$$(2) \quad d^2(F, \Phi) = 2 \cdot \text{Arc cos } (\alpha' \beta),$$

dans laquelle $\alpha_k^2 = F_k^v$, $\beta_k^2 = \Phi_k^v$ (resp k -ième coordonnée des vectorialisés F^v de F et Φ^v de Φ) (cf **vectorialisation d'une matrice**).

(iii) Le tableau Φ est appelé **tableau de référence**. On peut choisir pour Φ le produit des marges de F (ie $\Phi = F e_j e_i' F$) (cf **indépendance stochastique**) ou le tableau nul (ie $\Phi = 0_{IJ}$).

Si T est un **tableau statistique** comportant des éléments négatifs, on peut remplacer F par le tableau F^+ , avec $f_{ij}^+ = |t_{ij}| / \sum_{i=1}^I \sum_{j=1}^J |t_{ij}|$, $\forall (i, j) \in N_I^* \times N_J^*$, et analyser le tableau noté X^+ , d'élément général $x_{ij}^+ = (f_{ij}^+)^{1/2} w (f_{ij}^+)$, où w est la **fonction signe** définie par :

$$(3) \quad w(u) = \begin{cases} +1 & \text{si } u > 0, \\ 0 & \text{si } u = 0, \\ -1 & \text{si } u < 0. \end{cases}$$

Il en va de même pour Φ .

L'inertie du nuage X par rapport à l'origine est :

$$(4) \quad \text{tr } X'X = \sum_{i=1}^I \sum_{j=1}^J x_{ij}^2 = d^2(F, \Phi).$$

(iv) L'analyse sphérique ne vérifie pas, en général, le principe d'**équivalence distributionnelle**, sauf (notamment) lorsque $\Phi = F e_J e_I' F$ (produit des marges de F).