MÉTHODE DES NUÉES DYNAMIQUES (K9)
(10 / 05 / 2020, © Monfort, Dicostat2005, 2005-2020)

La **méthode des « nuées » dynamiques** est une méthode de **classification** formelle à caractère itératif (cf aussi **classification automatique**).

(i) Soit (E, d) un **espace métrique** tq E soit fini (Card E = N $< +\infty$).

On définit les **pseudo-distances** (notées avec le même symbole δ) :

(a) entre tout élément de E et toute **partie** de E :

$(1)_a$ $\quad \delta(x, A'') = \Sigma_{y \in A''} d(x, y), \qquad \forall x \in E, \forall A'' \in \mathscr{P}(E)$;

(b) entre parties de E :

$(1)_b$ $\quad \delta(A', A'') = \Sigma_{x \in A'} \delta(x, A''), \qquad \forall (A', A'') \in \mathscr{P}(E) \times \mathscr{P}(E)$.

Si Q $\in$ **N***, on appelle **noyau de taille Q**, ou **Q-noyau**, de E toute partie A $\in \mathscr{P}(E)$ tq Card A = Q. On note $\mathscr{N}_Q$ l'**ensemble** des Q-noyaux de E et $\mathscr{U}$ la **mesure de probabilité** uniforme sur $\mathscr{N}_Q$ (cf **loi uniforme discrète**). Un entier R $\in$ **N*** étant donné, on tire au **hasard** uniforme (ie selon $\mathscr{U}$) R noyaux $A_r^{(0)}$ de taille donnée Q (avec r $\in N_R$*), d'où une **suite** initiale (finie) $A^{(0)} = (A_r^{(0)})_{r=1,\ldots,R}$.

(ii) La **méthode des nuées dynamiques de E. DIDAY** procède par étapes :

(a) on pose, $\forall$ r $\in N_R$* :

(2) $\quad B_r^{(0)} = \{x \in E : \delta(x, A_r^{(0)}) \leq \delta(x, A_s^{(0)}), \ \forall s \neq r\}$.

Moyennant une convention d'affectation simple lorsque $\delta(x, A_r^{(0)}) = \delta(x, A_s^{(0)})$, la suite $(B_r^{(0)})_{r=1,\ldots,R}$ constitue une **partition** $\Pi_E(0)$ de E.

On définit alors une **application** $f_0 : A^{(0)} \mapsto \Pi_E^{(0)}$ selon l'affectation $f_0(A_r^{(0)}) = B_r^{(0)}$, $\forall$ r $\in N_R$* ;

(b) on associe à $\Pi_E^{(0)}$ une deuxième suite $A^{(1)} = (A_r^{(1)})_{r=1,\ldots R}$ tq, $\forall$ r $\in N_R$* :

$$(3) \quad \begin{cases} \text{Card } A_r^{(1)} = Q \qquad (\text{ie } A_r^{(1)} \in \mathscr{N}_Q), \\ \delta(A_r^{(1)}, B_r^{(0)}) = \min \{\delta(A, B_r^{(0)}) : A \in \mathscr{N}_Q\}, \end{cases}$$

et l'on pose alors $f_1 : \Pi_E^{(0)} \mapsto A^{(1)}$, application tq $f_1(B_r^{(0)}) = A_r^{(1)}$, $\forall$ r $\in N_R$* ;

(c) on réitère les étapes (a) et (b).

(ii) On établit que la suite double $(A^{(n)}, \Pi_E^{(n)})_{n \in \mathbf{N}}$ ainsi définie converge, au bout d'un nombre fini N d'itérations, vers un couple limite $(A^{(N)}, \Pi_E^{(N)})$, où $\Pi_E^{(N)}$ représente la partition finale de E obtenue.

La méthode dépend de la suite initiale $A^{(0)}$ (tirée au **hasard**) (cf **nombre au hasard**).

Si l'on procède à $S \in \mathbf{N}^*$ tirages au hasard tq le précédent, on obtient S partitions finales tq la partition $\Pi_E^{(N)}$ précédente, soit $\Pi_E^{(N(s), s)}$, où $s \in N_S^*$ et où N(s) désigne (par commodité) N_s .

En posant, $\forall (x, y) \in E^2$:

(4) $\nu (x, y) = \{$nombre de partitions $\Pi_E^{(N(s), s)}$ tq $x \neq y\}$

(où $x \neq y$ signifie que x et y n'appartiennent pas à la même **classe d'équivalence**), on définit une **relation d'équivalence** ~ sur E.

On appelle alors **forme forte** tout élément (ie classe) de l'ensemble quotient E / ~. Le cardinal d'une forme forte diffère, en général, de R.

(iii) En pratique, la méthode est assise sur un **tableau statistique** $T \in M_{NK} (\mathbf{R})$ constitué de N **observations** relatives à K **variables**. Une éventuelle transformation préalable de T conduit à un tableau $X \in M_{NK} (\mathbf{R})$. On choisit alors comme ensemble de départ E l'ensemble (indicié par N_N^*) des vecteurs lignes X_n (observations ou **unités statistiques**) de X.

La **distance** initiale δ est parfois remplacée par une pseudo-distance (ou **semi-distance**, ou quasi-distance) (cf aussi **indice de dissimilarité**, **indice de similarité**).