STATISTIQUE DE BALAYAGE (C5, I2)
(07 / 10 / 2019, © Monfort, Dicostat2005, 2005-2019)

La **statistique de balayage** est une **statistique** permettant de tester si un échantillon comporte des **observations groupées** (**accumulation** autour de certaines valeurs). Dans ce cas, la loi qui les génère peut être une **loi multimodale** (cf **mélange de lois**, **test d'unimodalité**).

(i) Soit $\xi \sim \mathcal{U}(0, 1)$ une va uniforme (cf **loi uniforme continue**) et $X = (X_1 ,..., X_N)$ un **échantillon iid** comme ξ. On pose, $\forall$ h $\in [0 , 1]$:

(1) $\quad C_N (x, h) = \Sigma_{n=1}^{N} \mathbf{1}_{[In(x , h)]} = F_N (x + h) - F_N (x)$,

où $I_n (x, h) = [X_n \in]x, x+h]] = \{\omega \in \Omega : X_n (\omega) \in]x, x+h]\}$ et F_N est la **fonction de répartition empirique** associée à X. Autrement dit, $C_N (x, h)$ est le nombre d'observations X_n appartenant à l'intervalle]x, x+h].

(ii) On appelle alors **statistique de balayage** la statistique (ou **fonction aléatoire**) $C_N : [0 , 1] \mapsto N_N$ définie par :

(2) $\quad C_N (h) = \sup_{x \in [0 , 1-h]} C_N (x , h), \quad \forall h \in [0 , 1]$.

(iii) Cette statistique, de type fonctionnel, permet notamment de tester l'**hypothèse de base** (hypothèse d' « uniformité ») :

(3) $\quad H_0 : \xi \sim \mathcal{U}(0, 1)$

contre l'**alternative de groupement** H_a tq eg :

(4) $\quad H_a : \xi \sim \mathcal{G}(a, b)$,

où $\mathcal{G}(a, b)$ est une loi dont la densité pr à λ_1 est de la forme :

(5) $$g (x) = \begin{cases} \{1 + (b - 1) h\}^{-1}, & \forall x \in]0, a], \\ \{1 + (b - 1) h\}^{-1} . b, & \forall x \in]a, a + h], \\ \{1 + (b - 1) h\}^{-1}, & \forall x \in]a + h, 1], \end{cases}$$

où a et b sont connus et h inconnu.

(iv) C_N s'associe naturellement au **test du rapport des vraisemblances** de l'hypothèse (3) contre l'hypothèse (4). Ce test, fondé sur la loi de C_N lorsque H_0 est supposée vraie, admet pour **région critique** :

(6) $w = [C_N \geq q_{1-\alpha}]$

où $q_{1-\alpha}$ est le **quantile** d'ordre $1 - \alpha$ de la loi précédente. Il existe plusieurs approximations de (6) (formules complexes).

(v) L'approche précédente peut s'étendre aux cas où (a) ξ n'est plus uniforme (cf **uniformisation**) et où (b) ξ est à valeurs dans un ensemble plus général $\mathscr{X}$.