RÉGRESSOGRAMME DE TUKEY (D2, H2, J1)
(12 / 11 / 2019, © Monfort, Dicostat2005, 2005-2019)

(i) Soit $(\Omega, \mathcal{T}, P)$ un **espace probabilisé** et $(\xi, \eta) : \Omega \mapsto \mathbf{R}^2$ un **couple aléatoire** réel dont la **loi** est $P^{(\xi, \eta)}$. On note $x \mapsto \rho(x) = E(\eta / \xi = x)$ la fonction de **régression** (**espérance conditionnelle**) de η sachant que $\xi = x$, ie :

(1) $\rho(x) = (f_x(x))^{-1} . \int y . f(x, y)\, dy, \qquad \forall x \in \mathbf{R},$

où $f = dP^{(\xi, \eta)} / d\lambda_2$ est la **densité** de $P^{(\xi, \eta)}$ pr à λ_2 et où f_x désigne la densité de la loi P^ξ de ξ pr à λ_1 (ie la **densité marginale** en x). Si le **support** $S_x = \mathrm{Supp}\, f_x$ de f_x est un segment $[a, b] \subset \mathbf{R}$, on le subdivise en k intervalles égaux :

(2) $S_i = \{a + k^{-1}(b - a) . (i - 1) , a + k^{-1}(b - a) . i\}, \qquad \forall i \in N_k^* = \{1 ,..., k\}.$

Par ailleurs, on note $((X_1 , Y_1),..., (X_N , Y_N))$ un **échantillon aléatoire** constitué de **copies**, indépendantes entre elles, du couple (ξ, η).

On appelle alors **régressogramme de J.W. TUKEY** le **graphe** de la fonction en escalier $r_N : [a,b] \mapsto \mathbf{R}$ définie selon :

(3) $$r_N(x) = \begin{cases} \mathbf{1}(S_i)(x) . \{B(X_n , Y_n)\}^{-1} . A(X_n , Y_n), \\ 0 \text{ sinon,} \end{cases}$$

expression dans laquelle, $\forall x \in \mathbf{R}$ et $\forall i \in N_k^*$:

(a) $\mathbf{1}(S_i)$ désigne la **fonction indicatrice** de la subdivision S_i pour la variable x (ie $x \in S_i$) ;

(b) $\mathbf{1}(S_i \times \mathbf{R})$ désigne la fonction indicatrice du produit cartésien $S_i \times \mathbf{R}$, pour la va X_n ;

(c) $A(X_n , Y_n) = \Sigma_{n=1}^N \mathbf{1}(S_i \times \mathbf{R})(X_n) . Y_n$,

(d) $B(X_n , Y_n) = \Sigma_{i=1}^k \mathbf{1}(S_i \times \mathbf{R})(X_n)$ (grandeur supposée non nulle).

(ii) Cette fonction (aléatoire) $r_N(.)$ dépend de l'échantillon et se note aussi $r_N(., X, Y)$. Elle constitue un estimateur de ρ et se généralise directement eg à des **vecteurs aléatoires** $\xi \in \mathbf{R}^K$ et $\eta \in \mathbf{R}^G$.