RÉGRESSION ROBUSTE (J1, J6)
(07 / 06 / 2020, © Monfort, Dicostat2005, 2005-2020)

L'expression abrégée **régression robuste** désigne (improprement) une méthode d'**estimation** robuste relative à un **modèle de régression** (multiple) (cf **régression multiple**, **robustesse**).

Plusieurs méthodes de ce type ont été définies, dont l'une des plus classiques est la suivante (cf aussi **estimateur de HUBER**).

(i) On considère le **modèle linéaire** (« vrai ») écrit dans l'**espace des variables** (ξ, η) selon $\eta = \xi' b^* + \varepsilon$, observé dans l'**espace des observations** selon $y = X b^* + u$. On suppose que la **perturbation aléatoire** ε suit une loi P^ε et que u est un **vecteur aléatoire** iid selon ε (cf **suite iid**). La « vraie valeur » du **paramètre d'intérêt** est notée b^*.

On appelle **estimateur (par régression) robuste**, ou **M-estimateur**, (au sens de P.J. HUBER) du paramètre $b \in \mathbf{R}^K$ la solution en b de l'équation vectorielle (cf **équation estimante**) :

(1) $\quad \Sigma_{n=1}^{N} \; \Psi \{(y_n - X_n b) / \beta_N\} X_n' = 0,$

dans laquelle :

(a) $\Psi : \mathbf{R} \mapsto \mathbf{R}$ est une fonction numérique tq :

(2) $\quad \int \Psi \, S (x / \beta) \, dP^\varepsilon (x) = 0, \quad \forall \beta > 0 ;$

(b) $\beta_N : \mathbf{R}^N \mapsto \mathbf{R}^*_+$ est un estimateur du **paramètre d'échelle** β de la loi P^ε tq :

(3) $\quad \beta_N (a y + X b) = |a| \, . \, \beta_N (y), \qquad \forall (a, b) \in \mathbf{R} \times \mathbf{R}^K.$

On note $b_\Psi\tilde{}$, ou simplement b_Ψ , l'estimateur robuste de b ainsi défini.

(ii) A titre d'exemple, si P^ε est une **loi symétrique** et si Ψ est une fonction impaire P^ε-intégrable, les conditions (2) et (3) sont vérifiées.

Ψ est souvent supposée monotone.

(iii) Lorsqu'on veut se prémunir contre l'influence de valeurs aberrantes u_n de ε sur l'estimateur de b (cas où P^ε est une **loi à queue épaisse** ou un **mélange légal**) (cf **aberration**), diverses formes analytiques ont été utilisées pour Ψ, eg :

(a) la fonction :

(4) $\quad \Psi (x) = A_{ab} (x) + B_{ab} (x) + C_{ab} (x) + D_{ab} (x),$

avec :

$$A_{ab}(x) = \mathbf{1}_{[|x| \le a]}(x) \cdot x,$$

$$B_{ab}(x) = \mathbf{1}_{[a < |x| \le b]}(x) \cdot a \cdot \operatorname{sgn} x,$$

(4)'

$$C_{ab}(x) = \mathbf{1}_{[b < |x| \le c]}(x) \cdot a\,(c - b)^{-1}\,(c - |x|) \cdot \operatorname{sgn} x,$$

$$D_{ab}(x) = \mathbf{1}_{[c < |x|]}(x),$$

où sgn désigne la **fonction signe** et $(a, b, c) \in \mathbf{R}^3$;

(b) la fonction :

(5) $$\Psi(x) = \mathbf{1}_{[|x| \le k\pi]}(x) \cdot \sin(x / k),$$

où $k > 0$ est un nombre déterminé ;

(c) ou encore la fonction :

(6) $$\Psi(x) = x \cdot \{1 - (x / \gamma)^2)^2 \cdot \mathbf{1}_{[|x| \le k]}(x),$$

où k et γ sont des nombres déterminés.

(iv) La notion s'étend directement à la classe des **modèles non linéaires**. Elle consiste essentiellement à remplacer les équations normales de la méthode des moindres carrés (ordinaires) par l'équation (1) et à choisir une fonction Ψ qui atténue l'effet sur b_Ψ des « grandes » valeurs u_n prises par ε, eg :

(a) si $\Psi(x) = x^2$, b_Ψ n'est autre que l'**estimateur des moindres carrés ordinaires** (b_Ψ se note souvent $\hat{b}$ et β se note aussi σ) ;

(b) si $\Psi(x) = |x|$, b_Ψ est appelé **estimateur des moindres écarts absolus**.

Si le paramètre d'échelle β n'est pas connu (cas général), on peut l'estimer, par « balayage », en même temps que le paramètre d'intérêt b.