ESTIMATEUR DE HUBER (F6, F8, H4, H5, J)
(14 / 01 / 2020, © Monfort, Dicostat2005, 2005-2020)

(i) Soit $(\mathcal{X}, \mathcal{B}, P_\theta^X)_{\theta \in \Theta}$ un **modèle statistique** dans lequel :

(a) Θ = **R** et θ représente un **paramètre de position** caractérisant la **lp** P_θ^X générant l'**échantillon** $X = (X_1, ..., X_N)$;

(b) $\mathcal{X}$ = $\mathbf{R}^N$ et $X : \Omega \mapsto \mathbf{R}^N$ est un **échantillon iid** constitué de N **copies** X_n dont la **variable parente** ξ admet l'une des lois P_θ^X ($\theta \in \Theta$) pour **loi de probabilité**. Autrement dit, $P_\theta^X = (P_\theta^\xi)^{\otimes N}$, $\forall\ \theta \in \Theta$;

(c) il existe une **densité** f_0 sur **R** tq (**dérivée de NIKODYM-RADON**) :

(1) $f(x, \theta) = (dP_\theta^\xi / d\lambda)(x) = f_0(x - \theta)$, $\forall\ (x, \theta) \in \mathcal{X} \times \Theta$.

où λ désigne la **mesure de LEBESGUE** sur **R**.

Soit $\rho : \mathbf{R} \mapsto \mathbf{R}$ une fonction (non constante) donnée.

On appelle **estimateur (de type M) de P.J. HUBER**, ou **M-estimateur de P.J. HUBER**, de θ une **statistique** $T_N = t_N(X)$ qui rend minimum pr à θ la **va** (cf **équation estimante**) :

(3) $s_\rho(\theta) = \Sigma_{n=1}^N \rho(X_n - \theta)$.

(ii) A titre d'exemples :

(a) si $\rho(u) = u^2$, on obtient la **moyenne empirique** $T_N = \bar{X}_N$;

(b) si $\rho(u) = |u|$, on obtient la **médiane empirique** $T_N = q_{1/2}\ X$;

(c) si $\rho(u) = -\text{Log}\ f_0(u)$, on obtient l'**estimateur du maximum de vraisemblance** de θ. Cet exemple relie ainsi l'estimateur de type M (« maximum » de vraisemblance) à celui du maximum de vraisemblance.

(iii) Sous certaines hypothèses, on montre que la solution T_N de (3) est unique et que :

(a) T_N converge presque-sûrement vers θ (cf **convergence presque sûre**) :

(4) $T_N \to_{N \to +\infty} \theta$ (P_θ-p.s.), $\forall\ \theta \in \Theta$;

(b) T_N est asymptotiquement gaussien :

(5) $N^{1/2}(T_N - \theta^\#) \to^{\mathcal{L}}_{N \to +\infty} \mathcal{N}(0, \sigma^2)$ (**loi normale réduite**),

où $\theta^{\#}$ est la solution en θ de $s_\rho(\theta) = 0$ et $\sigma^2 = (D\, s_\rho(\theta^{\#})^{-2}\, \sigma_0^2$, avec $\sigma_0^2 = \int \{\rho(x - \theta^{\#})\}^2 dF_N(x)$, F_N étant la **fonction de répartition empirique** associée à X.

(iv) La notion de M-estimateur généralise celle de **moyenne de WINSOR** (cf **transformation de WINSOR**). En effet, si :

$$(6) \qquad W_N(L,M) = N^{-1} \{L\, X^{(L+1)} + X^{(L+1)} + \ldots + X^{(N-M)} + M\, X^{(N-M)}\}$$

est la **moyenne (arithmétique) de WINSOR** d'ordre (L, M) et $X^{(.)}$ l'échantillon ordonné associé à X qui permet de la calculer (cf **statistique ordonnée**), alors l'estimateur de HUBER T_N est de la forme :

$$(7) \qquad T_N = N^{-1} \{L\, U + X^{(L+1)} + X^{(L+2)} + \ldots + X^{(N-M)} + M\, V\},$$

pour des couples (L, M) et (U, V) tq :

$$(8) \qquad X^{(L)} \leq U \leq X^{(L+1)} \text{ et } X^{(N-M)} \leq V \leq X^{(N-M+1)}.$$

(v) Un M-estimateur peut être utilisé pour détecter une **aberration** éventuelle, ie un « point aberrant » éventuel X_n , car il est robuste (eg pr à un **mélange de lois**) (cf **robustesse**).

En effet, si F_θ est la **fr** associée à la lp P_θ^{ξ} ($\forall\, \theta \in \Theta$), et si cette fonction correspond à une loi normale contaminée (cf **contamination des lois**), ie :

$$(9) \qquad F_\theta(x) = (1 - \varepsilon)\, \Phi(x) + \varepsilon\, G(x), \qquad \forall\, x \in \mathbf{R},$$

où Φ est la fr de la loi $\mathcal{N}(0, 1)$, $\varepsilon \in [0, 1[$ est donné ($\varepsilon \leq 1$) et G est une fr inconnue, on montre que le choix de la fonction :

$$(10) \qquad \rho(u) = \begin{cases} u^2 / 2 & \text{si } |u| < c, \\ c\,|u| - (c^2 / 2) & \text{si } |u| \geq c, \end{cases}$$

dans laquelle $c > 0$ dépend de ε selon la relation :

$$(11) \qquad 1 - \varepsilon = \{\Phi(c) - \Phi(-c)\} + c^{-1} (2 / \pi)^{1/2} \exp(- c^2 / 2),$$

fournit l'**estimateur** de **centralité** le plus robuste (cf **robustesse**) au sens où sa **variance** asymptotique est minimum lorsque G parcourt la classe des fr dont les lp associées sont des **lois symétriques**.

(vi) Les hypothèses portant sur ρ sont diverses. Parmi les plus usuelles, on peut noter les suivantes :

(a) ρ est paire et convexe sur **R** (avec $\lim_{|u| \to +\infty} \rho(u) = +\infty$) ;

(b) ρ est à valeurs dans $\mathbf{R}_+$ et non décroissante sur **R** ;

(c) ρ est symétrique pr à 0 et tq ρ (0) = 0 ;

(d) ρ est différentiable et ρ' = D ρ est continue (sauf en un nombre fini de points de **R**).

(vii) Il existe des notions de **M-estimateur généralisé**, notamment au cas où ρ est définie sur $\mathbf{R}^K$ et à valeurs dans **R**, et où la somme s_ρ (θ) est remplacée par :

(12) $\Sigma_{(n(1),\ldots,n(K))}\ \rho\ (X_{n(1)} - \theta, \ldots, X_{n(K)} - \theta)$,

avec $\{n_1, \ldots, n_K\} \subset \{1, \ldots, N\}$ et $\beta \neq \alpha \Rightarrow n_\beta \neq n_\alpha$ (où les n(1) ,..., n(K) désignent, alternativement, les n_1 ,..., n_K).

(viii) Le modèle initial s'étend directement au cas d'un **paramètre d'échelle et de position**. On considère alors un modèle dans lequel :

(a) Θ = **R** x $\mathbf{R}_+^*$ et $\theta = (\alpha, \beta)$ représente un couple constitué d'un **paramètre de position** α et d'un **paramètre d'échelle** β (cf **paramètre d'échelle et de position**);

(b) $\mathcal{X} = \mathbf{R}^N$ et X est défini comme précédemment ;

(c) il existe une densité f_0 sur **R** tq la **dérivée de NIKODYM-RADON** de P_θ^ξ pr à la mesure de LEBESGUE λ est de la forme :

(13) $f(x, \theta) = \beta^{-1} f_0 \{(x - \alpha) / \beta\}, \qquad \forall (x, \theta) \in \mathcal{X} \times \Theta.$

On appelle alors **estimateur de type M de HUBER**, ou **M-estimateur de HUBER**, de θ toute solution $T_N = (A_N , B_N)$ du système en (α, β) :

(14)
$$\Sigma_{n=1}^N\ \varphi\ \{(X_n - \alpha) / \beta\} = 0,$$
$$\Sigma_{n=1}^N\ \psi\ \{(X_n - \alpha) / \beta\} = 0,$$

dans lequel le choix des fonctions φ et ψ dépend des propriétés recherchées pour T_N (notamment sa robustesse).

Un estimateur de HUBER particulier est tq :

(15) $$\varphi(u) = \begin{cases} -c & \text{si } u \leq -c, \\ u & \text{si } -c \leq u \leq +c, \\ +c & \text{si } +c \leq u, \end{cases}$$

et :

(16) $\psi(u) = \{\varphi(u)\}^2 - E\,\varphi^2(U)$,

où $E\,\varphi^2(U) = (2\pi)^{-1/2} \int \{\varphi(u)\}^2 \exp(-u^2/2)\, du$.

(ix) Si X est un **échantillon iid** comme ξ, si P_θ^ξ est une **loi symétrique** ($\forall\ \theta$), si φ est impaire et ψ paire, on montre que l'estimateur de HUBER (ie la solution $T_N = (A_N , B_N)$ de (14)) vérifie les propriétés suivantes :

(a) $R_N\ (A_N - \alpha) \rightarrow^{\mathcal{L}}_{N \rightarrow +\infty} \mathcal{N}(0, \beta^2 . \Sigma_A)$ (**loi gaussienne** centrée), avec $\Sigma_A = E\ \varphi^2\ (\zeta)\ E\ \{\varphi\ '\ (\zeta)\}^{\ 2}$, en notant $\zeta = (\xi - \alpha) / \beta$;

(b) $R_N\ (B_N - \beta) \rightarrow^{\mathcal{L}}_{N \rightarrow +\infty} \mathcal{N}(0, \beta^2\ \Sigma_B)$, avec $\Sigma_B = E\ \psi^2\ (\zeta)\ .\ E\ (\zeta\ .\ \psi\ '\ (\zeta))^2$ et $\zeta = (\xi - \alpha) / \beta$;

(c) A_N et B_N sont asymptotiquement indépendants.