ESTIMATEUR DE HODGES-LEHMANN (G, H)
(10 / 06 / 2020, © Monfort, Dicostat2005, 2005-2020)

L'**estimateur de HODGES - LEHMANN** est une **statistique** qui sert notamment à estimer le **paramètre** de **centralité** d'une **loi de probabilité** (cf **paramètre de position**) et à réaliser les **tests** associés. On le rencontre notamment dans le **problème à un échantillon** ou dans le problème à deux échantillons (cf **problème à plusieurs échantillons**).

(i) Dans le problème à un **échantillon**, $X = (X_1, ..., X_N)$ est un **échantillon iid**, équidistribué comme la **vars** $\xi : \Omega \mapsto \mathbf{R}$ dont la **loi** est P^ξ.

On peut estimer la **médiane** $Q_{1/2}\ \xi$ de ξ (ou de P^ξ) à l'aide de l'**estimateur de J.L. HODGES - E.L. LEHMANN**, défini comme la **médiane empirique** suivante :

(1) $A_N = q_{1/2}(M)$,

où $M = (M_{\alpha\beta})_{(\alpha,\beta)}$ est la suite des **moyennes de J.E. WALSH**, avec :

$(2)_a$ $M_{\alpha\beta} = (X_\alpha + X_\beta) / 2, \quad \forall (\alpha, \beta) \in (\mathbf{N}_N{}^*)^2$,

ou aussi la suite des $C_N{}^2$ moyennes partielles :

$(2)_b$ $M_{\alpha\beta}' = (X^{(\alpha)} + X^{(\beta)}) / 2, \quad \forall (\alpha, \beta) \in (\mathbf{N}_N{}^*)^2_{\leq}$,

où $X^{(.)} = (X^{(1)}, ..., X^{(N)})$ est la **statistique ordonnée** associée à X.

Sous certaines hypothèses, A_N est un **estimateur** symétrique, invariant, sans biais (au sens de la médiane théorique $Q_{1/2}\ \xi$) et asymptotiquement gaussien (cf **symétrie**, **invariance** et **estimateur invariant**, **estimateur sans biais**, **normalité asymptotique**).

L'**efficacité asymptotique** de A_N pr à la **moyenne empirique** $\overline{X}_N = e_N' X / N$ vaut :

(3) $e_\infty (A_N / \overline{X}_N) = 3 / \pi$.

(ii) Dans le problème à deux échantillons, on considère un **espace mesurable** $(\Omega, \mathcal{T})$ et deux vars $\xi : \Omega \mapsto \mathbf{R}$ et $\eta : \Omega \mapsto \mathbf{R}$, de lp resp P^ξ et P^η. On note F (resp G) la **fr** associée à P^ξ (resp à P^η) et l'on suppose ces deux fr liées par la relation :

(4) $G(x) = F(x - \alpha), \quad \forall x \in \mathbf{R}$,

où α est le **paramètre de position** de la **va** ξ. Ce **paramètre** α s'interprète aussi comme un **paramètre d'écart**, ou **paramètre de déplacement**, ou encore **paramètre de glissement**, entre les deux « **populations** » d'où sont tirés les échantillons.

Soit $X = (X_1 ,..., X_{N(1)})$ et $Y = (Y_1 ,..., Y_{N(2)})$ deux **échantillons iid** indépendants entre eux (où l'on note N(i) pour désigner N_i , i = 1, 2) : ie X est iid comme ξ et Y comme η. On définit $N_1 . N_2$ va (**écarts** deux à deux) :

(5) $$D_{\alpha\beta} = Y_\alpha - X_\beta , \qquad \forall (\alpha , \beta) \in \{1,..., N_1\} \times \{1,..., N_2\}.$$

On appelle alors **estimateur de J.L. HODGES - E.L. LEHMANN** du paramètre α la **médiane empirique** du N-échantillon $D = (D_{\alpha\beta})_{(\alpha,\beta)}$ ainsi constitué, ie :

(6) $$A_N = q_{1/2} (D).$$

Sous certaines hypothèses, on montre que A_N admet une **loi asymptotique** gaussienne de moyenne nulle, ie (cf **normalité asymptotique**) :

(7) $$(N_1 + N_2)^{-1/2} . (N_1 N_2)^{1/2} . (A_N - \alpha) \to^{\mathscr{L}}_{\min (N(1),N(2)) \to +\infty} \mathcal{N}_1(0, (\sigma^*)^2)$$

(**loi gaussienne**), où $\sigma^* > 0$ est une **constante** dépendant de ces hypothèses.