STATISTIQUE DE HOEFFDING (G, H, I)
(12 / 11 / 2019, © Monfort, Dicostat2005, 2005-2019)

(i) Soit ($\mathscr{X}$, $\mathscr{B}$, $\mathscr{P}^X$) un **modèle image** d'un modèle de base par un N-**échantillon** X, avec $\mathscr{X} = \mathscr{X}_0^N$, et ($\mathscr{U}$, $\mathscr{E}$) un **espace mesurable** auxiliaire doté d'une **structure** de **groupe mesurable**.

On appelle **statistique de W. HOEFFDING**, ou **U-statistique**, associée au modèle précédent une **statistique** U_N définie par une **application mesurable** $u_N : \mathscr{X} \mapsto \mathscr{U}$ de la forme suivante (**moyenne** calculée sur N_0 coordonnées de X) :

(1) $U_N = u_N(X) = (C_N^{N(0)})^{-1} \Sigma_{A(0)} \; \phi (X_{\alpha(1)} ,..., X_{\alpha(N(0))})$,

où $X = (X_1 ,..., X_N)$, l'entier N_0 (aussi noté N(0)) est généralement donné et tq $1 \leq N_0 \leq N$, $C_N^{N(0)}$ est nombre de **combinaisons** de N objets pris N_0 à N_0 , $A(0) = \{(\alpha_1 ,..., \alpha_{N(0)}) : 1 \leq \alpha_1 \leq ... \leq \alpha_{N(0)} \leq N\}$ et $\phi : \mathscr{X}_0^{N(0)} \mapsto \mathscr{U}$ est une **application** symétrique. On a noté $\alpha(n)$ pour désigner α_n (n = 1 ,..., N(0)).

L'entier N_0 est appelé **degré** de ϕ et l'application ϕ est appelée **noyau** de U_N .

En pratique, on a généralement ($\mathscr{U}$, $\mathscr{E}$) = (**R**, $\mathscr{B}_{\mathbf{R}}$).

(ii) Sous des **conditions de régularité** générales, dont :

(a) les va X_n sont à valeurs dans $\mathscr{X}_0$ = **R** et forment une **suite iid** ;

(b) ($\mathscr{U}$, $\mathscr{E}$) = (**R**, $\mathscr{B}_{\mathbf{R}}$),

on montre que U_N est asymptotiquement gaussienne (cf **normalité asymptotique**).

Si U_N est de carré intégrable et si l'on pose $E\, \phi (X_1 ,..., X_{N0}) = \alpha$, alors $E\, U_N = \alpha$ et :

(4) $V\, U_N = (N_0^2 / N) \,.\, \beta + O\,(N^{-2})$ (grand zéro),

où $\beta = V\, \phi_1 (X_1)$, en notant $\phi_n (x_1 ,..., x_n) = E\, \phi (x_1 ,..., x_n , X_{n+1} ,..., X_{N(0)})$, $\forall$ n $\in$ {1 ,..., N_0} et $\forall$ $(x_1 ,..., x_n) \in \mathbf{R}^n$.

(ii) Les **U-statistiques de W. HOEFFDING** interviennent dans divers contextes, eg :

(a) soit ϕ le **noyau** (supposé symétrique) d'une **fonctionnelle** linéaire h : $\mathscr{F}^X \mapsto$ **R** et N_0 le **degré** de h. La **va** réelle U_N définie par l'**application mesurable** u_N selon (1), où l'entier $N_0 \leq N$ est généralement donné et où les va X_n (n = 1 ,..., N) sont indépendantes, est une U-statistique. Cette statistique est un **estimateur sans biais** de la valeur de la fonctionnelle h au point $F \in \mathscr{F}^X$:

(2) $E\, U_N = h(F), \qquad \forall\, F \in \mathscr{F}^X$ et $\forall\, N \geq N_0$.

Elle possède une **variance** minimale dans la classe des **estimateurs sans biais** de h (F). Enfin, sous des hypothèses générales de régularité, la loi asymptotique de U_N est gaussienne ;

(b) dans le cadre du problème à deux échantillons $X = (X_1 ,..., X_M)$ et $Y = (Y_1 ,..., Y_N)$ (cf **problème à plusieurs échantillons**), on utilise souvent (cf **test de WILCOXON**, **test de MANN-WHITNEY**) une statistique de la forme :

(3) $U_{MN} = \Sigma_{m=1}^{M} \Sigma_{n=1}^{N} \mathbf{1}_{A(m,n)}$,

où $A(m,n) = [X^{(m)} \geq Y^{(n)}] = \{\omega \in \Omega : X^{(m)}(\omega) \geq Y^{(n)}(\omega)\}$ et où $X^{(.)} = (X^{(1)} ,..., X^{(N)})$ et $Y^{(.)} = (Y^{(1)} ,..., Y^{(N)})$ sont les **statistiques d'ordre** (ou échantillons ordonnés) resp associées à X et Y.

U_{MN} est une U-statistique de HOEFFDING qui indique le nombre de fois où une coordonnée de X dépasse une coordonnée de Y. Elle peut être utilisée pour un **test de décentrement** (cf **test de déplacement**).