TEST DE HOEFFDING (D1, F6, I2)
(25 / 11 / 2019, © Monfort, Dicostat2005, 2005-2019)

Le test de HOEFFDING est un **test d'indépendance** relatif à un un **couple aléatoire**.

(i) Soit $(\Omega, \mathcal{T}, P)$ un **espace probabilisé** et $(\xi, \eta) : \Omega \mapsto \mathbf{R}^2$ un **vecteur aléatoire** réel dont la loi $P^{(\xi, \eta)}$ est supposée être une **loi absolument continue** pr à la **mesure de LEBESGUE** λ_2 . On observe un **échantillon** $((X_1 , Y_1) ,..., (X_N , Y_N))$ constitué de N couples indépendants (X_n , Y_n) de même loi que (ξ, η).

On considère le test de l'**hypothèse** d'**indépendance** suivante :

(1) $H_0 : P^{(\xi , \eta)} = P^\xi \otimes P^\eta$,

où l'on note P^ξ (resp P^η) la **loi propre** (ie la **loi marginale**) de ξ (resp de η).

On définit les **statistiques** suivantes :

(a) $X = (X_1 ,,..., X_N)$ et $Y = (Y_1 ,..., Y_N)$ (échantillons « marginaux ») ;

(b) $X^{(.)} = (X^{(1)} ,..., X^{(N)})$ et $Y^{(.)} = (Y^{(1)} ,..., Y^{(N)})$ (**statistiques d'ordre** resp associées à X et Y) ;

(c) $R = (R_1 ,..., R_N)$ et $S = (S_1 ,..., S_N)$ (**statistiques de rang** resp associées à X et Y) ;

(d) $M_n (X, Y) = \Sigma_{\alpha=1}^N u (X_\alpha - X_n) u (Y_\alpha - Y_n)$, pour tout n = 1 ,..., N, où u est la **fonction de HEAVYSIDE** ($u (h) = 1$ ssi $h > 0$, $u (h) = 0$ ssi $h \leq 0$).

Le **test de W. HOEFFDING** est fondé sur la **statistique de test** suivante, appelée **statistique de W. HOEFFDING** :

(2) $H_N = \{A - 2 (N - 2) B + (N - 2) (N - 3) C\} / \Pi_{n=0}^4 (N - n)$,

expression dans laquelle :

$A = \Sigma_{n=1}^N (R_n - 1) (R_n - 2) (S_n - 1) (S_n - 2)$,

(3) $B = \Sigma_{n=1}^N (R_n - 2) (S_n - 2) M_n (X , Y)$,

$C = \Sigma_{n=1}^N M_n (X , Y) \{M_N (X , Y) - 1\}$.

La statistique H_N est souvent notée D_N ou D.

(ii) Lorsque l'hypothèse H_0 est vraie, la loi de H_N est tabulée. Le test possède aussi des **propriétés asymptotiques**.

Lorsque des valeurs multiples apparaissent dans l'échantillon, on utilise la **méthode des rangs moyens** pour R et S, et l'on remplace M_n (X, Y) par l'expression modifiée (cf **modification**) :

(4) $M_n' (X, Y) = \Sigma_{\alpha=1,\ldots,n-1,n+1,\ldots N}\ w (X_\alpha - X_n) . w (Y_\alpha - Y_n)$,

où la fonction w est tq :

$$
(5)\quad w(h) = \begin{cases} 1 & \text{ssi } h > 0, \\ 1/2 & \text{ssi } h = 0, \\ 0 & \text{ssi } h < 0. \end{cases}
$$