STATISTIQUE DE HELLINGER (C4, C5, G8, I, K5)
(29 / 08 / 2020, © Monfort, Dicostat2005, 2005-2020)

Cette **statistique** exprime une **distance entre lois** (cf **distance de HELLINGER**, **analyse sphérique**).

(i) Dans le cadre du **test d'adéquation** du chi-deux (cf **test du chi-deux**), on appelle parfois **« statistique » de E. HELLINGER** la **statistique** :

(1) $D_N^2(p) = \text{Arc}\cos T_{NK}$,

dans laquelle $T_{NK} = \Sigma_{k=1}^{K} (N^{-1} . X_k . p_k)^{1/2}$.

(ii) Les **probabilités** « théoriques » p_k peuvent être estimées d'une façon analogue à celle du test du chi-deux.

L'expression $D_N^2(p)$ permet alors de fonder le **test de E. HELLINGER**, en remplaçant dans (1) le vecteur $p = (p_1 ,..., p_K)$ par l'estimateur $p (\theta_N\tilde{}) = (p_1 (\theta_N\tilde{}) ,..., p_K (\theta_N\tilde{}))$ correspondant.

(iii) Plus généralement, étant donné un modèle image $(\mathscr{X}, \mathscr{B}, \mathscr{P}^X)$, on appelle **distance de E. HELLINGER** sur $\mathscr{P}^X$ une application d_H définie par :

(2) $d_H (P_1^X, P_2^X) = \text{Arc}\cos \int \{f_1 (x) . f_2 (x)\}^{1/2} d\mu (x)$,

pour tout couple (P_1^X, P_2^X) d'éléments de $\mathscr{P}^X$, où μ est une **mesure positive** dominant la famille $\mathscr{P}^X$ (cf **famille de lois dominée**).

On peut aussi écrire :

(2)′ $d_H (P_1^X, P_2^X) = \text{Arc}\cos \int \{f_2 (x) / f_j (x)\}^{1/2} dP_j^X (x)$,

où $(i, j) \in N_2^* \times N_2^*$ et $j \neq i$, avec des densités $f_i = dP_i^X / d\mu$ $(i = 1, 2)$.

(iv) En particulier, si $\mathscr{P}^X$ est paramétrée selon $(P_\theta^X)_{\theta \in \Theta}$ et si P_N est la **probabilité empirique** associée à l'**échantillon** X, la distance $d_H (P_\theta^X, P_N)$ est parfois appelée **statistique de E. HELLINGER**.

Une méthode d'**estimation** de θ consiste à minimiser cette distance pr à $\theta \in \Theta$.