TEST DE RANG (F6, I2)
(27 / 07 / 2020, © Monfort, Dicostat2005, 2005-2020)

En **Statistique non paramétrique**, de nombreux **tests d'hypothèses** font intervenir la statistique de rang associée à un échantillon aléatoire.

(i) Soit $(\Omega, \mathcal{T}, (P_\theta)_{\theta \in \Theta})$ un **modèle statistique** dans lequel $\Theta = \mathbf{R}$, $\xi : \Omega \mapsto \mathbf{R}$ une **vars** et $X = (X_1 ,..., X_N)$ un **échantillon aléatoire** constitué de **copies** indépendantes et équidistribuées comme ξ (ie dont chacune suit, indépendamment des autres, la loi P_θ^ξ) (**échantillon iid**).

Si la **fonction de répartition** F_θ associée à P_θ^ξ est de la forme :

(1) $\quad F_\theta (x) = F (x - \theta), \quad \forall x \in E,$

où F est une fonction indépendante de θ, on peut chercher à tester l'**hypothèse nulle** :

(2) $\quad H_0 : \theta = 0,$

contre une **alternative** tq (cas unilatéral) :

(3) $\quad H_1 : \theta > 0.$

Soit $s_N : \mathcal{X} \mapsto \mathbf{R}$ une **statistique de test** associée à un **test** φ de l'hypothèse H_0 . La **statistique** $S_N = s_N (X)$ ainsi définie est tq, pour un seuil $\alpha \in]0, 1[$ donné :

(4) $\quad S_N \geq q_{1-\alpha} \Rightarrow \theta > 0,$

où désigne le **quantile** associé à ce seuil et calculé selon la **loi** de S_N .

On dit alors que φ est un **test de rang** ssi il existe une une **application mesurable** $s : N_N^* \times \{-1, 0, +1\} \mapsto \mathbf{R}$ tq :

(5) $\quad S_N = s_N (X) = s (R^{|.|}, \text{sgn } X)),$

expression dans laquelle :

(a) R est la **statistique de rang** associée à l'échantillon ordonné $(|X_1|^{(1)} ,..., |X_N|^{(N)})$, lequel se déduit de l'échantillon noté symboliquement $|X| = (|X_1| ,..., |X_N|)$, constitué des valeurs absolues des coordonnées de X (cf **statistique d'ordre**) ;

(b) $\text{sgn } X = (\text{sgn } X_1 ,..., \text{sgn } X_N)$, où sgn X = - 1 ssi X < 0, sgn X = 0 ssi X = 0 et sgn X = + 1 ssi X > 0, pour tout n = 1 ,..., N (cf **fonction signe**).

(ii) Un test de rang classique est défini par :

(7) $S_N = s_N(X) = \Sigma_{n=1}^N a_N(R_n) \operatorname{sgn} X_n$,

où les $(a_N(n))_{n=1,\dots,N}$ sont des **scores**.

Cette **statistique** admet les cas particuliers suivants :

(a) **statistique de J.L. HODGES - E.L. LEHMANN** avec les scores $a_N(n) = n$ (pour tout n) :

(8) $S_N = 2 . \Sigma_D R_n - (1/2) N (N+1)$,

où $D = \{X_n : X_n > 0\}$.

Le test associé à cette statistique est asymptotiquement optimal pour des lois de type logistique (cf **loi logistique**). En effet, lorsque $N \to +\infty$, ce test est adapté au cas où, avec les notations de (1), $F(x) = (1 + e^{-x})^{-1}$.

L'**estimateur** de $\theta \in$ **R** associé à s est ici la **médiane empirique** des moyennes deux à deux. On l'appelle **estimateur de J.L. HODGES - E.L. LEHMANN** de θ, ie :

(9) $T_N = q_{1/2}(M_{\alpha\beta})$, calculé sur les **indices** $1 \leq \alpha \leq \beta \leq N$,

où $M_{\alpha\beta} = (X^{(\alpha)} + X^{(\beta)}\} / 2$ (moyennes simples 2 à 2).

(b) **test des signes** avec les scores $a_N(n) = n$ (pour tout n). Ce test est optimal lorsque F est de type exponentiel (cf **loi exponentielle**), ie $F(x) = (1/2)\, e^{-|x|}$. L'estimateur correspondant est encore la médiane empirique associée à X, ie $T_N = q_{1/2} X$;

(c) test associé à l'**estimateur de P.J. BICKEL - J.L. HODGES** de θ, où (9) est remplacé par :

(10) $T_N = q_{1/2}(L_N)$,

avec $L_N = (X^{(n)} + X^{(N-n+1)}) / 2$;

(d) **test de B.L. van der WAERDEN** associé à la **statistique de B.L. van der WAERDEN**, où la fonction de répartition F est celle Φ de la **loi normale réduite** $\mathcal{N}(0, 1)$ et où les scores sont $a_N(n) = E\,|U|^{(n)}$, avec $U = (U_1, \dots, U_N) \sim \mathcal{N}_N(0, I_N)$, $|U| = (|U_1|, \dots, |U_N|)$, et $|U|^{(.)} = (|U|^{(1)}, \dots, |U|^{(N)})$ (statistique ordonnée associée à $|U|$).