

STATISTIQUE DES RANGS SIGNES (F6, I2)

(23 / 10 / 2019, © Monfort, Dicostat2005, 2005-2019)

(i) Soit $(\Omega, \mathcal{F}, P)$ un **espace probabilisé**, $(\mathcal{X}_0^N, \mathcal{B}_0^{\otimes N})$ un **espace mesurable** (eg un **espace d'observation**) et $X = (X_1, \dots, X_N) : \Omega \mapsto \mathcal{X}_0^N$ une **va** (ou **échantillon aléatoire**) de taille N . On suppose que $(\mathcal{X}_0, \leq)$ est un ensemble totalement ordonné (cf **relation d'ordre**) et muni d'une **valeur absolue** $|\cdot|$.

On note symboliquement $|X| = (|X_1|, \dots, |X_N|)$ l'échantillon associé à X , $|R_n|$ le **rang** de $|X_n|$ dans la **suite** $|X|$ et $R = (R_1, \dots, R_N)$ la suite de ces rangs.

On appelle alors **statistique des rangs signés** toute **statistique** de la forme :

$$(1) \quad S = s(X, |R|).$$

où $s : \mathcal{X}_0^N \times (N_N^*)^N \mapsto \mathcal{S}$, $N_N^* = \{1, \dots, N\}$ et $\mathcal{S}$ représente un ensemble de valeurs prises par S .

(ii) Un exemple de statistique des rangs signés est la **statistique linéaire des rangs signés**, définie selon :

$$(2) \quad S = \sum_{n=1}^N u_n \cdot a(|R_n|),$$

dans laquelle $0 \leq a(1) \leq \dots \leq a(N)$ est une suite de **scores**, $u_n = u(X_n)$, $\forall n \in N_N^*$, et u est la **fonction de HEAVYSIDE** (ie $u(x) = 1$ si $x > 0$, et $u(x) = 0$ sinon).

En particulier, une **statistique des rangs signés** classique s'écrit :

$$(3) \quad S_N = \sum_{n=1}^N |R_n| - (1/4) \cdot N(N+1),$$

où N est le nombre de coordonnées $X_n \neq 0$.

S'il existe des valeurs multiples pour $|X_n|$, on utilise souvent la **méthode des rangs moyens**. Sous l'hypothèse H_0 d'indépendance des coordonnées de X (cf **échantillon indépendant**), on établit la **propriété asymptotique** suivante :

$$(4) \quad \mathcal{L}\{\sigma_0^{-1}(S_N - E_0 S_N)\} \rightarrow_{N \rightarrow \infty} \mathcal{N}(0, 1),$$

où les deux premiers **moments**, calculés sous H_0 , valent :

$$(5) \quad \begin{aligned} E_0 S_N &= 1/2, \\ \sigma_0^2 = V_0 S_N &= (1/24) \{(N(N+1)(2N+1) - (1/2) \sum_{g=1}^G (N_g - 1) N_g (N_g + 1))\}, \end{aligned}$$

où N_g est le nombre de coordonnées multiples $|X_n|$ de $|X|$.

Ces propriétés permettent ainsi de tester H_0 .

(iii) Une statistique tq (2) est très utilisée en **Statistique non paramétrique** (**théorie des tests**) en raison de sa simplicité (cf **test des rangs signés**).