STATISTIQUE LINÉAIRE DE RANG (F6, G1, I2)
(24 / 09 / 2020, © Monfort, Dicostat2005, 2005-2020)

(i) Soit $(\mathscr{X}, \mathscr{B}, \mathscr{P}^X)$ un **modèle statistique** dans lequel $(\mathscr{X}, \mathscr{B}) = (\mathbf{R}^N, \mathscr{B}(\mathbf{R}^N))$ et soit $R = (R_1, \dots, R_N) = r(X)$ la **statistique de rang** associée à X.

On appelle aussi, plus généralement, **statistique de rang** une **statistique** (application mesurable) qui dépend de R, ie une application $s : \sigma_N \mapsto \mathbf{R}$ définissant la statistique $S = s(R)$, où σ_N représente le groupe des **permutations** de $N_N^* = \{1, \dots, N\}$.

(ii) On définit alors une statistique linéaire de rang selon l'un des deux procédés suivants.

(a) soit $c = (c_1, \dots, c_N)' \in \mathbf{R}^N$ un vecteur donné et $(a_n)_{n \in \mathbf{N}^*}$ une **suite** finie, ou **score**, sur **R** (ie $a_N : \{1, \dots, N\} \mapsto \mathbf{R}$).

On appelle **statistique linéaire de rang** une statistique $S = s(R)$ tq :

(1) $$S = s(R) = \Sigma_{n=1}^N c_n a_N(R_n).$$

Sous l'hypothèse H_0 de l'**échantillon aléatoire**, les premiers **moments algébriques** de S sont :

(2) $$\begin{aligned} E_0 S &= N \,.\, \bar{c} \,.\, \bar{a}, \\ V_0 S &= (N-1)^{-1} \,.\, \{\Sigma_{n=1}^N (c_n - \bar{c})^2\} \,.\, \{\Sigma_{n=1}^N (a_N(n) - \bar{a})^2\}, \end{aligned}$$

avec $\bar{c} = N^{-1} \Sigma_{n=1}^N c_n$ (moyenne des c_n) et $\bar{a} = N^{-1} \Sigma_{n=1}^N a_N(n)$ (moyenne des scores).

Si $\phi :]0, 1[\mapsto \mathbf{R}$ est une fonction λ_1-intégrable (**fonction score**), on établit, sous certaines conditions, la **convergence en loi** :

$(3)_a$ $$\mathscr{L}\{(S - E_0 S) / \sigma_0\} \rightarrow_{N \to +\infty} \mathscr{N}_1(0, 1),$$

avec :

$(3)_b$ $$\sigma_0^2 = \{\Sigma_{n=1}^N (c_n - \bar{c})^2\} \,.\, \{\int_{[0, 1]} (\phi(u) - \bar{\phi})\, du\}, \quad \text{où } \bar{\phi} = \int_{[0, 1]} \phi(u)\, du \,;$$

(b) soit $a_N(n)_{n=1,\dots,N}$ une suite de **scores** et :

(4) $$R_N^+ = \Sigma_{\alpha=1}^N H(|X_n| - |X_\alpha|), \qquad \forall\, n \in \{1, \dots, N\},$$

où H désigne la **fonction de HEAVYSIDE**.

On appelle **statistique linéaire de rang** la statistique :

(5) $S = s(R)$ ou $s \circ r(X) = \Sigma_{n=1}^{N} a_N(R_n^+) \,.\, H(X_n)$.

Si X est un **échantillon iid** selon la loi P^ξ et si la **densité de probabilité** f de P^ξ pr à λ_1 est symétrique pr à $\alpha \in E$, ie :

(6) $f(x - \alpha) = f(x + \alpha), \qquad \forall x \in \mathbf{R}$,

alors, sous l'hypothèse $H_0 : \alpha = 0$, les premiers **moments algébriques** s'écrivent :

(7)
$$E_0 S = (1/2) \,.\, \Sigma_{\alpha=1}^{N} a_N(n),$$
$$V_0 S = (1/4) \,.\, \Sigma_{\alpha=1}^{N} \{a_N(n)\}^2 = \sigma_0^2.$$

Par suite, sous des conditions générales portant sur les scores, on établit la **convergence en loi** suivante :

(8) $\mathscr{L}((S - E_0 S) / \sigma_0) \rightarrow_{N \rightarrow +\infty} \mathscr{N}_1(0, 1)$,

ce qui permet d'effectuer un test de **centralité** relatif à α.