REDONDANCE (C5, F3, G6)
(11 / 10 / 2020, © Monfort, Dicostat2005, 2005-2020)

On appelle parfois **redondance** une **caractéristique légale** comparable à celles de **coefficient de concentration**, d'**entropie**, ou d'**information**.

(i) Soit $(\Omega, \mathcal{T}, P)$ un **espace probabilisé** et $\xi : \Omega \mapsto \mathbf{R}_+$ une **vars** positive (ie non négative) de **loi** P^ξ. On suppose que $\xi \,.\, \mathrm{Log}\, \xi \in \mathcal{L}_{\mathbf{R}+}{}^1 (\Omega, \mathcal{T}, P)$ et l'on pose $\mu = E\,\xi$.

On appelle **coefficient de redondance**, ou **indice de redondance**, théorique (au sens de H. THIEL) le scalaire :

(1) $\rho(\xi) = E\,\{(\xi / \mu) \,.\, \mathrm{Log}\,(\xi / \mu)\}$.

(ii) Soit $X = (X_1, ..., X_N)$ un **échantillon iid** constitué de **répliques** $X_n \sim P^\xi$ de ξ. On définit les **proportions empiriques** f_n selon (cf **fréquence empirique**) :

(2) $f_n = X_n / e_N' X, \quad \forall\, n \in N_N{}^*$,

ainsi que l'**entropie empirique** H_N (p) selon :

(3) $H_N (p) = \Sigma_{n=1}{}^N f_n \,.\, \mathrm{Log}\,(1 / f_n)$,

avec $f = (f_1, ..., f_N)$.

On établit que $\max_f H_N (f) = \mathrm{Log}\, N$. Par suite, on définit le **coefficient de redondance**, ou **indice de redondance**, empirique $R_N (X)$ (ou $R_N (f)$, ou simplement R_N) comme l'**écart** (en niveau) pr à ce maximum :

(4) $R_N = \max_f H_N (f) - H_N (f) = \mathrm{Log}\, N - H_N (f)$.

(iii) On établit les propriétés suivantes :

(a) **convergence presque sûre** :

(5) $R_N \to \rho(\xi)$ (P-p.s.) ;

(b) **convergence en loi**. Si $\xi\, \mathrm{Log}\, \xi \in L_{\mathbf{R}+}{}^2 (\Omega, \mathcal{T}, P)$, alors :

(6) $N^{1/2} \,.\, (R_N - \rho(\xi)) \to^{\mathcal{L}}{}_{N \to +\infty} \mathcal{N}_1 (0, \sigma^2)$ (**loi normale**),

avec $\sigma^2 = \mu^{-2}\, E\,\{\xi\, \mathrm{Log}\, \xi - \alpha\, (\xi / \mu) - \xi + \mu\}^2$ et $\alpha = E\,(\xi\, \mathrm{Log}\, \xi)$.