ESTIMATEUR DE L'ENTROPIE (C5, H)
(22 / 01 / 2020, © Monfort, Dicostat2005, 2005-2020)

Comme toute **caractéristique légale**, l'**entropie** peut théoriquement être estimée à l'aide de la **loi empirique** ou de la **fonction de répartition empirique**.

(i) Ainsi, $(\Omega, \mathcal{T}, P)$ étant un **espace probabilisé** et $\xi : \Omega \mapsto \mathbf{R}$ une **vars** de **loi** P^{ξ}, l'**entropie** associée est :

(1) $\quad H(P^{\xi}) = -\int f(x) \text{ Log } f(x)\, dx = \int \text{Log } \{(d / dp) F^{-1}(p)\}\, dp,$

où F est la **fonction de répartition** de P^{ξ}, f sa **densité** pr à λ et F^{-1} la **fonction quantile**.

On observe un **échantillon iid** $X = (X_1, ..., X_N)$, distribué comme la **variable parente** ξ (ie selon P^{ξ}) et l'on note $X^{(.)} = (X^{(1)}, ..., X^{(N)})$ l'échantillon ordonné (cf **statistique d'ordre**) correspondant à X (avec $X^{(1)} \leq ... \leq X^{(N)}$).

Un estimateur naturel de $H(P^{\xi})$ s'obtient en remplaçant F par F_N (fr empirique définie par X) dans (1) (cf **statistique naturelle**).

(ii) La **dérivée** de F^{-1}, ie la fonction $p \mapsto \psi(p) = (d / dp) F^{-1}(p)$, est estimée par :

(2) $\quad \psi_{Nm}(p) = (2m)^{-1} N \{X^{(n+m)} - X^{(n-m)}\}, \quad \forall p \in I_n =]N^{-1}(n-1), N^{-1} n[,$

pour tout $m \in \{1, ..., N-1\}$, avec $X^{(n)} = X^{(1)}$ si $n \leq 1$ et $X^{(n)} = X^{(N)}$ si $n \geq N$.

On en déduit (par **différence finie**) l'**estimateur de l'entropie** de $H(P^{\xi})$:

(3) $\quad H_{NM}(P^{\xi}) = N^{-1} \Sigma_{n=1}^{N} \text{ Log } \{\mathbf{1}_{I(n)}(p).\, \psi_{NM}(p)\},$

où I(n) désigne l'intervalle I_n défini en (2).

(iii) On montre que, si $\xi \in \mathcal{L}_{\mathbf{R}}^2(\Omega, \mathcal{T}, P)$, alors :

(4) $\quad H_{NM}(P^{\xi}) \rightarrow^{m.q.}{}_{\min(N,M) \rightarrow +\infty\,,\ M = o(N)}\ H(P^{\xi}).$

Par ailleurs, si $P^{\xi} = \mathcal{N}(\mu, \sigma^2)$ (**loi normale**), $H(P^{\xi})$ ne dépend pas de μ. Par suite, sous l'hypothèse de **normalité** suivant :

(5) $\quad H_0 : P^{\xi} = \mathcal{N}(\mu, \sigma^2),$

on montre que (cf **test de VASICEK**) :

(6) $\quad \text{P-lim}_{\min(N,M) \rightarrow +\infty\,,\ M/N \rightarrow 0}\ H_{NM}(P^{\xi}) = \text{Log } (2\, \pi\, .\, e\, .\, \sigma^2).$