ENTROPIE (C5, F3, G6)
(21 / 04 / 2020, © Monfort, Dicostat2005, 2005-2020)

Le concept d'**entropie** (R.E. CLAUSIUS, L.E. BOLTZMANN) caractérise l'**hétérogénéité** d'une **population statistique** ou la **dispersion** d'une **loi de probabilité**. On le présente dans des situations usuelles.

(i) **Cas scalaire discret**

Soit $(\Omega, \mathcal{T}, P)$ un **espace probabilisé**, $(\mathcal{X}, \mathcal{B})$ un **espace mesurable** (au plus) dénombrable (avec $\mathcal{B} = \mathcal{P}(\mathcal{X})$) et $\xi : \Omega \mapsto \mathcal{X}$ une **variable discrète** de **loi** P^{ξ}. On pose $f(x) = P^{\xi}(\{x\})$, $\forall x \in \mathcal{X}$ (**distribution des « masses »**).

On appelle **entropie**, ou **incertitude**, ou **degré d'indétermination**, de ξ (ou de P^{ξ}) la valeur de la série (si elle converge) :

(1) $$H(P^{\xi}) = \Sigma_{x \in \mathcal{X}} f(x) . \mathrm{Log}(1 / f(x)),$$

avec la convention de continuité $u(0) = 0$ pour la fonction $u : t \mapsto u(t) = t . \mathrm{Log}(1/t)$, définie sur $]0, 1]$.

En particulier, si $\mathcal{X}$ est **fini** (Card $\mathcal{X}$ = I) et si $P^{\xi} = \mathcal{U}(\mathcal{X})$ (**loi uniforme discrète** sur $\mathcal{X}$), on définit la notion d'entropie la plus simple, l'**entropie de R.V. HARTLEY** :

(2) $$H(\mathcal{U}(\mathcal{X})) = \mathrm{Log}(\mathrm{Card}\ \mathcal{X}) = \mathrm{Log}\ I.$$

On note l'entropie $H(\xi)$ (entropie de la va ξ), ou $H(P^{\xi})$ (entropie de la loi P^{ξ}), ou $H(f)$ (entropie de la **densité** f de P^{ξ}), ou encore $H(F)$ (entropie de la **fr** F associée à P^{ξ}).

(ii) Si $\mathcal{Y}$ est un autre ensemble tq $f : \mathcal{X} \mapsto \mathcal{Y}$ soit bijective (donc $f(\mathcal{X}) \subset \mathcal{Y}$ au plus dénombrable), on établit la **propriété d'invariance de l'entropie** (dans le cas discret) :

(3) $$H(f(\xi)) = H(\xi), \quad P\text{-p.s.}.$$

Cette propriété est parfois utilisée dans l'étude des variables qualitatives : f est alors un **codage**.

Dans le cas d'une **variable qualitative**, l'**entropie** se définit selon :

(4) $$H(p) = \Sigma_{i \in I}\, p_i\, \mathrm{Log}(1 / p_i) = - \Sigma_{i \in I}\, p_i\, \mathrm{Log}\, p_i ,$$

où les p_i $(i \in I)$ sont des **proportions** ou des **fréquences empiriques** (eg les **proportions empiriques** attachées à un **tableau de contingence** (multidimensionnel), i désignant alors le multi-**indice** associé aux **critères** définissant ce tableau (cf **loi qualitative**).

(iii) On montre que :

(a) $H(\xi) = 0 \Rightarrow \xi = c$ (P-p.s.), ie $P^{\xi} = \delta_c$ (**loi de DIRAC** chargeant le point c), où c est une **constante** ;

(b) si Card $\mathscr{X} = I < +\infty$, alors $H(\xi) \leq \text{Log } I$;

(c) si $\mathscr{Y}$ est dénombrable et $\varphi : \mathscr{X} \mapsto \mathscr{Y}$ mesurable, alors : $H(\varphi(\xi)) \leq H(\xi)$;

(d) si $\mathscr{Y}$ est dénombrable, si $\xi : \Omega \mapsto \mathscr{X}$, $\eta : \Omega \mapsto \mathscr{Y}$, et si $H(\xi) < +\infty$ et $H(\eta) < +\infty$, alors : $H((\xi, \eta)) \leq H(\xi) + H(\eta)$. De plus, la saturation à égalité équivaut à l'**indépendance** de ξ et de η.

(iv) **Cas vectoriel « continu »**

Soit $(\Omega, \mathscr{T}, P)$ un **espace probabilisé** et $\xi : \Omega \mapsto \mathbf{R}^K$ un **vecteur aléatoire** réel de loi $P^{\xi} = \xi(P)$. On suppose que P^{ξ} admet une **densité** f pr à la **mesure de LEBESGUE** λ_K définie sur $\mathscr{B}(\mathbf{R}^K)$ et l'on note F sa **fr**.

On appelle **entropie de (la loi de)** ξ le nombre réel (s'il existe) :

(5) $$H(\xi) = E \text{ Log } (1 / f(\xi)) = \int f(x) \text{ Log } (1 / f(x)) \, d\lambda_K(x),$$

qui est l'**espérance** de la va $\eta : \Omega \mapsto \mathbf{R}$ définie par $\eta = \text{Log } (1 / f(\xi))$. On a donc $H(\xi) = - E \text{ Log } f(\xi)$.

On peut aussi écrire :

(6) $$H(\xi) = \int \text{Log } (1 / f(x)) \, dF(x) = \int (1 / f(x))^{-1} \text{ Log } (1 / f(x)) \, d\lambda_K(x)$$

(entropie de la fr ou de la densité).

Comme précédemment, on note aussi, selon le cas, $H(P^{\xi})$, ou $H(f)$ ou encore $H(F)$.

(v) Soit $\varphi : \mathbf{R} \mapsto \mathbf{R}$ une fonction définissant un **changement de variable aléatoire** $\xi^* = \varphi(\xi)$ et f^* la densité de la **lp** de ξ^* pr à λ_1 . Alors, en notant $H(f) = - E \text{ Log } f(\xi)$ l'entropie de ξ, celle de ξ^* s'écrit (par définition) :

(7) $$H(f^*) = - E \text{ Log } f^*(\xi^*) = - \int f^*(x^*) \text{ Log } f^*(x^*) \, dx^*,$$

soit explicitement :

(8) $$H(f^*) = - \int f^*(\varphi(x)) \text{ Log } f^*(\varphi(x)) \, \varphi'(x) \, dx.$$

(vi) Si $\xi \in L_{RK}^2 (\Omega, \mathscr{T}, P)$, si la densité f de la loi P^ξ de ξ pr à λ_K est continue, et si $\mu \in \mathbf{R}^K$ est un vecteur donné et $\Sigma \in M_K(\mathbf{R})$ une **matrice** donnée, alors le **problème d'optimisation** suivant :

$$
(9) \quad \begin{cases} \max H(P^\xi) \\ \text{sous } E\,\xi = \mu \text{ et } V\,\theta = \Sigma \end{cases}
$$

admet pour solution la **loi normale multidimensionnelle** $P^\xi = \mathcal{N}_K(\mu, \Sigma)$. Par suite :

(10) $H^* (\mathcal{N}_K(\mu, \Sigma)) = (1/2) . \{K \text{ Log } (2\,\pi) + K + \text{Log } |\Sigma|\}$.

En particulier, si K = 1, $E\,\xi = 0$ et $V\,\xi = 1$, on obtient la solution $\mathcal{N}(0, 1)$ (**loi normale réduite**) dont l'entropie vaut $H(\mathcal{N}(0,1)) = \text{Log } (2\,\pi\,e)$.

(vii) On appelle **neg-entropie**, ou **information de S. KULLBACK - R.A. LEIBLER**, de ξ, ou de sa loi, la quantité, si elle existe :

(11) $K(\xi) = - H(\xi)$.

(viii) Si K = 1 et si $p \mapsto F^{-1}(p) = \inf \{x \in \mathbf{R} : F(x) \geq p\}$ désigne la **fonction quantile**, on montre que :

(12) $H(\xi) = \int \text{Log } \{\Psi(p)\}\, dp$, avec $\Psi(p) = d\,F^{-1}(p) / dp$.

D'autre part, un changement affine de va $\xi = \mu + \sigma\,\varepsilon$ (avec $\sigma > 0$) entraîne : $H(\xi) = H(\varepsilon) + \text{Log } \sigma$, ie $H(F) = H(G) + \text{Log } \sigma$, où F est la fr de ξ et G celle de ε.

La relation (12) permet de définir un **estimateur de l'entropie** ainsi qu'un **test de normalité** basé sur l'entropie (cf **test de VASICEK**).

(ix) Soit f la densité associée à la loi $\mathscr{L}(\xi)$ de la **vars** ξ pr à la **mesure de LEBESGUE** dx, et dont le **support** est le segment $[a, b] \subset \mathbf{R}$. Soit $(\mu, \nu) \in \mathbf{R} \times \mathbf{R}$ un paramètre donné et $h : [a, b] \mapsto \mathbf{R}$ une **fonction numérique** donnée. Alors, la maximisation de l'entropie :

(13) $H(f) = - \int \mathbf{1}_{([a,\, b])}\, x . f(x) . \text{Log } (x . f(x))\, dx$

sous les contraintes :

$$
(14) \quad \begin{cases} \int \mathbf{1}_{([a,\, b])}\, x . f(x)\, dx = \alpha & \text{(espérance de } \xi\text{)}, \\ \int \mathbf{1}_{([a,\, b])}\, h(x) . f(x)\, dx = \beta & \text{(espérance de } h(\xi)\text{)}, \end{cases}
$$

conduit à une solution de la forme :

(15) $f^* (x) = \{c (\alpha, \beta) / x\} . \exp \{- (\beta / x) (h (x) - \alpha)\}$, $\forall x \in \mathbf{R}$,

où $c (\alpha, \beta) > 0$ est une **constante** de **normalisation** qui dépend de (α, β).

(ix) Diverses extensions de l'entropie ont été définies, eg :

(a) on peut remplacer, dans les définitions, la mesure de LEBESGUE par une **mesure positive** σ-finie (cf **mesure σ-finie**) ;

(b) l'**entropie de A. RÉNYI**. Si X est une **variable discrète** et si Card $\mathscr{X}$ = I est fini, on appelle **entropie (au sens de RÉNYI) d'ordre α** (avec $\alpha \geq 0$) le nombre réel, $H_\alpha (\xi)$ défini, s'il existe, par (cf **information de C.E. SHANNON**) :

(16) $$H_\alpha (\xi) = \begin{cases} (1 - \alpha)^{-1} \{\mathrm{Log}\, \Sigma_{i=1}^{I} (f (x_i))^\alpha\} & \text{si } \alpha \neq 0, \\ H (\xi) & \text{si } \alpha = 0, \end{cases}$$

où l'on note $dP^\xi = \Pi_{i=1}^{I} f (x_i) dx_i$.

(x) L'entropie sert parfois de **fonction de perte** ou de pseudo « **distance** ». Ainsi, une (fonction de) **perte entropique** peut s'écrire (eg dans le cas gaussien) :

(17) $$L (\Sigma, \Sigma_0) = \mathrm{tr} (\Sigma_0 \Sigma^{-1}) - \mathrm{Log} |\Sigma_0 \Sigma^{-1}| - p,$$

où $\xi \sim \mathscr{N}_p (\mu, \Sigma)$ (**loi normale multidimensionnelle**) et où Σ est une **matrice définie positive**.

(xi) Enfin, on rencontre souvent, au lieu des **logarithmes népériens** utilisés dans les questions théoriques, des logarithmes exprimés dans une **base** différente de e (eg logarithmes de base 2, ou **logarithmes décimaux**, ie de base 10).

La relation circulaire classique $\log_a c = (\log_a b)(\log_b c)$ permet de passer d'une définition à une autre.

(xii) En **Statistique**, soit $(\Omega, \mathscr{T}, (P_\theta)_{\theta \in \Theta})$ un **modèle paramétrique** et $\xi : \Omega \mapsto \mathbf{R}^K$ un **vecteur aléatoire** réel de loi $P_\theta^{\ \xi} = \xi (P_\theta)$. On suppose que $P_\theta^{\ \xi}$ admet une **densité** f_θ ou $f (. , \theta)$ pr à la **mesure de LEBESGUE** λ_K définie sur $\mathscr{B}(\mathbf{R}^K)$ et l'on note F_θ sa **fr**.

On appelle **entropie de (la loi de)** ξ le nombre réel (s'il existe) :

(5) $$H_\theta (\xi) = E\, \mathrm{Log} (1 / f (\xi, \theta)) = \int f (x, \theta) \mathrm{Log} (1 / f (x, \theta))\, d\lambda_K (x),$$

qui est l'**espérance** de la va $\eta : \Omega \mapsto \mathbf{R}$ définie par $\eta = \mathrm{Log} (1 / f (\xi, \theta))$. On a donc $H_\theta (\xi) = - E_\theta \mathrm{Log}\, f (\xi, \theta)$.

On peut aussi écrire :

(6) $H_\theta(\xi) = \int \text{Log}\,(1 / f(x, \theta))\, dF(x) = \int (1 / f(x, \theta))^{-1} \text{Log}\,(1 / f(x, \theta))\, d\lambda_K(x)$

(entropie de la fr ou de la densité).

Comme précédemment, on note aussi, selon le cas, $H(P_\theta^{\xi})$, ou $H(f_\theta)$ ou encore $H(F_\theta)$.

(xiii) De nombreuses **procédures statistiques** (construction de **modèles**, **sélection de modèles**) font usage d'un **principe d'entropie maximum**, en particulier en **théorie bayésienne**.

Le concept d'entropie se retrouve aussi dans celles d'**entropie d'une partition**, d'**entropie d'un recouvrement**, d'**entropie relative** ou d'**information**.