INFORMATION DE FISHER (G06, H01, H05)
(21 / 04 / 2020, © Monfort, Dicostat2005, 2005-2020)

(i) Soit $(\Omega, \mathcal{T}, (P_\theta)_{\theta \in \Theta})$ un **modèle statistique**, $(\mathcal{X}, \mathcal{B})$ un **espace d'observation** et $X : \Omega \mapsto \mathcal{X}$ une **va** (eg un **échantillon**).

Le **modèle image** $(\mathcal{X}, \mathcal{B}, (P_\theta^X)_{\theta \in \Theta})$ ainsi défini est supposé tq :

(a) $\Theta \in \mathcal{O}(\mathbf{R}^Q)$ (ouvert de $\mathbf{R}^Q$) ;

(b) $(P_\theta^X)_{\theta \in \Theta}$ est une **famille dominée** par une **mesure positive** μ qui est aussi une **mesure σ-finie** sur $\mathcal{B}$;

(c) la **densité** (ou **vraisemblance**) (ie la **dérivée de NIKODYM-RADON**) est strictement positive sur $(\mathcal{X} \times \Theta)$ (**modèle homogène**), ie :

(1) $f > 0$, avec $f(x, \theta) = dP_\theta^X / d\mu\,(x), \quad \forall\,(x, \theta) \in \mathcal{X} \times \Theta$,

et que f est différentiable pr à θ, $\forall\,(x, \theta) \in \mathcal{X} \times \Theta$ (cf **dérivée**, **différentiabilité**). On note alors (cf **gradient**) :

(2) $D_2 f = \mathrm{Grad}_\theta f = (\partial f(x, \theta) / \partial\theta_1, \ldots, \partial f(x, \theta) / \partial\theta_Q)'$;

(d) la fonction $\varphi : \mathcal{B} \times \Theta \mapsto \mathbf{R}$ définie par :

(3) $\varphi(B, \theta) = \int \mathbf{1}_B(x)\, f(x, \theta)\, d\mu(x)$

est différentiable pr à θ, $\forall\,(B, \theta) \in \mathcal{B} \times \Theta$.

On appelle alors **information de R.A. FISHER** apportée par le modèle (ou par P_θ, ou encore par P_θ^X) la (Q,Q)-**matrice** de **dispersion** du **vecteur aléatoire** Grad_θ Log $(f(X, \theta))$, ie (comparer avec la notion d'**entropie**) :

(4) $I(\theta) = V_\theta\, \mathrm{Grad}_\theta (\mathrm{Log}\, f(X, \theta)) \in M_Q(\mathbf{R}), \qquad \forall\, \theta \in \Theta$.

Cette matrice est aussi notée $I_X(\theta)$ ou I_θ, ou encore $I_N(\theta)$ lorsque $\mathcal{X}$ est un espace produit de N espaces facteurs (ou un espace puissance N fois, eg $\mathcal{X} = \mathcal{X}_0^N$ ou $\mathcal{X} = \mathbf{R}^N$).

Comme $E_\theta\, \mathrm{Grad}_\theta\, \mathrm{Log}\, f(X, \theta) = 0$, la définition (4) se simplifie :

(4)' $I(\theta) = E_\theta \{\mathrm{Grad}_\theta\, \mathrm{Log}\, f(X, \theta) \,.\, (\mathrm{Grad}_\theta\, \mathrm{Log}\, f(X, \theta))'\}$.

Le modèle étant homogène, l'information $I(\theta)$ ne dépend pas de μ.

(ii) A titre d'exemple, si $X \sim \mathcal{N}_N(\theta, \Sigma)$ et si la **matrice de dispersion** Σ est une **matrice régulière**, on montre que l'information est égale à son inverse, ie :

(5) $\quad I(\theta) = \Sigma^{-1}$.

(iii) Dans le même cadre que précédemment, soit $S = s(X)$ une **statistique** fondée sur une fonction mesurable $s : \mathcal{X} \mapsto \mathcal{Y}$ (cf **application mesurable**) et à valeurs dans un **espace mesurable** $(\mathcal{Y}, \mathcal{C})$.

On appelle **information de R.A. FISHER** relative à (ou contenue dans, ou apportée par) S l'information de FISHER du modèle image de $(\mathcal{X}, \mathcal{B}, (P_\theta^X)_{\theta \in \Theta})$ par s, ie celle du modèle $(\mathcal{Y}, \mathcal{C}, (P_\theta^{s \circ X})_{\theta \in \Theta})$. On la note $I_S(\theta)$ ou I_θ^S.

(iv) On montre alors que :

(a) $I_S(\theta) \leq I(\theta), \forall \theta \in \Theta$ (**perte d'information**) ;

(b) si S est une **statistique libre**, alors $I_S(\theta) = 0$. Si S est une **statistique exhaustive**, alors $I_S(\theta) = I(\theta)$. Les réciproques de ces propriétés sont vraies ;

(c) si S et T sont deux statistiques indépendantes (cf **indépendance**), l'information de FISHER est additive, ie : $I_{(S,T)}(\theta) = I_S(\theta) + I_T(\theta), \forall \theta \in \Theta$, où (S, T) désigne le **couple aléatoire** $\Omega \mapsto \mathcal{Y} \times \mathcal{Z}$;

(d) si le modèle considéré est un **modèle d'échantillonnage** à distance finie $(\mathbf{R}, \mathcal{B}(\mathbf{R}), (P_\theta^\xi)_{\theta \in \Theta})^{\otimes N}$, alors $I(\theta) = N \cdot I_\xi(\theta)$, où $I_\xi(\theta)$ est l'information associée à la **variable parente** ξ dont la loi est P_θ^ξ.

(iii) La notion d'information de FISHER se définit aussi dans un cadre non paramétré.

En effet, soit $(\mathcal{X}_0, \mathcal{B}_0, P^\xi)$ un modèle non explicitement paramétré dans lequel $\xi : \Omega \mapsto \mathcal{X}_0$ est une **vars** dont l'une des lois possibles est P^ξ.

On appelle alors **information de R.A. FISHER** le scalaire :

(5) $\quad I(f_0) = \int (D f_0(x) / f_0(x))^2 \, dP^\xi(x) = \int (D f_0(x) / f_0(x))^2 \cdot f_0(x) \, d\mu_0(x)$,

où $f_0 = dP^\xi / d\mu_0$ est la **densité** de P^ξ pr à une **mesure** positive μ_0 définie sur $\mathcal{B}_0$.

De même, si $(\mathscr{X}, \mathscr{B}, \mathscr{P}^X)$ est un **modèle d'échantillonnage** dans lequel $X = (X_1, ..., X_N)$, $\mathscr{X} = \mathscr{X}_0^N$, $\mathscr{B} = \mathscr{B}_0^{\otimes N}$ et $P^X = (P^\xi)^{\otimes N} \in \mathscr{P}^X$ (cf **produit de convolution**), l'**information de FISHER** est définie par :

$$(6) \qquad I(f) = \begin{cases} \int \{D f(x) / f(x)\} \{D f(x) / f(x)\}' \, dP^X(x) \; f(x) / f(x)\}' \, f(x) \, d\mu(x), \\ \int \{D f(x) / f(x)\} \{D f(x) / f(x)\}' \, f(x) \, d\mu(x), \end{cases}$$

où l'on suppose que $(\mathscr{X}_0, \mathscr{B}_0) = (\mathbf{R}, \mathscr{B}(\mathbf{R}))$ et que μ domine P^X, avec $f = dP^X / d\mu$.