ESTIMATEUR DU MAXIMUM DE VRAISEMBLANCE (G, H1, H3)
(09 / 04 / 2020, © Monfort, Dicostat2005, 2005-2020)

L'**estimateur du maximum de vraisemblance**, ou **estimateur de vraisemblance maximum**, est un estimateur à la fois très général, et dont le champ d'application est très étendu. Les méthodes fondées sur des moments légaux (par identification), ou sur des normes ou des distances (par minimisation), ou encore sur la vraisemblance (par maximisation) sont des méthodes de base dans la résolution d'un **problème d'estimation**.

(i) Soit ($\mathcal{X}$, $\mathcal{B}$, $(P_\theta^X)_{\theta \in \Theta}$) un **modèle statistique**. On suppose que c'est un **modèle dominé** par une **mesure positive** μ et l'on note $f = dP_\theta^X / d\mu$ sa **fonction de vraisemblance**.

La **méthode du maximum de vraisemblance** conduit à choisir un **estimateur** $t : \mathcal{X} \mapsto \Theta$ de θ, ie une **statistique** $T = t(X) : \Omega \mapsto \Theta$, tq :

(1) $f(x, t(x)) = \sup_{\theta \in \Theta} f(x, \theta)$ (μ-p.p.).

L'estimateur ainsi défini est appelé **estimateur du maximum de vraisemblance** (mv) de θ. On le note souvent $q^{\sim}$, θ^* ou parfois $\theta^{\wedge}$ (notamment lorsqu'il coïncide avec l'estimateur des moindres carrés) (cf **méthode des moindres carrés**).

Le rôle de la **fonction de vraisemblance** est donc important dans la détermination de cet estimateur. S'agissant d'une solution d'un **problème d'optimisation**, cet estimateur peut ne pas exister, ou n'est pas nécessairement unique. De plus, il implique la mise en oeuvre d'un **modèle dominé** par une mesure, puisqu'une vraisemblance n'est autre que la **dérivée de NIKODYM-RADON** des lois du modèle.

(ii) Diverses hypothèses peuvent être supposées, qui portent sur f. En pratique, on étudie souvent les cas où :

(a) $f > 0$ sur $\mathcal{X} \times \Theta$, sauf éventuellement sur une partie μ-négligeable $N \in \mathcal{B}$ ne dépendant pas de θ (cf **partie négligeable**) ;

(b) $D_2 f = \mathrm{Grad}_\theta f$ (**gradient** de f) existe μ-presque partout, $\forall \theta \in \Theta$, avec $\Theta = \mathbf{R}^Q$. Alors, t est solution de l'**équation de vraisemblance** :

(2) $\mathrm{Grad}_\theta f(X, \theta) = 0$,

ou de l'**équation de Log-vraisemblance** :

(3) $\mathrm{Grad}_\theta (\mathrm{Log} f(X, \theta)) = 0$;

(c) il existe une partie μ-négligeable $N \in \mathscr{B}$, indépendante de θ, tq la fonction $\theta \mapsto f(x, .)$ soit continue sur Θ, $\forall x \in N^c$. Cette hypothèse, assez forte, est souvent nécessaire pour étudier les **propriétés asymptotiques** de $T = t(X)$).

(iii) On établit les propriétés suivantes :

(a) si $(\mathscr{Y}, \mathscr{C})$ est un **espace mesurable** auxiliaire, $s : \mathscr{X} \mapsto \mathscr{Y}$ une **statistique exhaustive** définissant $S = s(X)$, $\Theta \subset \mathbf{R}^Q$ et si $T = t(X)$ est l'estimateur du mv de θ, il existe une application $(\mathscr{C}, \mathscr{B}_\Theta)$-mesurable $\varphi : \mathscr{Y} \mapsto \Theta$ tq :

(4) $\quad T = t(X) = \varphi(s(X)), \qquad P_\theta\text{-p.s.}, \forall \theta \in \Theta,$

où les P_θ sont les **probabilités** dont les images par X sont les P_θ^X (cf **mesure image**).

C'est la **propriété d'invariance de l'estimateur (ou de la méthode) du mv** (cf **invariance du maximum de vraisemblance**) : en effet, φ est alors estimateur du mv dans le modèle image $(\mathscr{Y}, \mathscr{C}, (P_\theta^S)_{\theta \in \Theta}\}$, où P_θ^S est la loi de S (ie l'image de P_θ par s).

Cette propriété existe encore pour un transformé de θ : en effet, si $\Theta = \mathbf{R}^Q$, si $g : \Theta \mapsto \mathbf{R}^K$ est une **application mesurable** donnée (avec $K \le Q$) et si $\theta_{mv}\tilde{}$ est un estimateur du mv de θ, alors $\tau_{mv}\tilde{} = g(\theta_{mv}\tilde{})$ est un estimateur du mv de $\tau = g(\theta)$.

Cette propriété vaut même si g n'est pas une **application bijective** (eg si $K < Q$) ;

(b) si le modèle considéré est un **modèle exponentiel** réduit, ie tq :

(5) $\quad f(x, \theta) = c(\theta) \,.\, e^{<\theta, t(x)>}, \qquad$ avec $\Theta \in \mathscr{O}(\mathbf{R}^Q)$ et $\Theta \neq \varnothing$,

et si :

$(b)_1$ $V_\theta\, t(X) = V_\theta\, T = I_\theta \geq 0$ (**matrice définie positive**), $\forall \theta \in \Theta$;

$(b)_2$ $t(\mathscr{X}) \subset (D_\theta\, \psi)(\Theta)$, avec $\psi(\theta) = -\operatorname{Log} c(\theta)$, $\forall \theta \in \Theta$,

alors il existe un estimateur du mv unique de θ, qui est aussi une **statistique exhaustive minimale** pour θ. De plus, cette statistique est une **statistique totale** ;

(c) sous certaines hypothèses, s'il existe un estimateur $T = t(X)$ sans biais de θ qui est efficace (cf **estimateur efficace**) et tq $t(X) \in \Theta$, alors T est l'unique estimateur du mv de θ ;

(d) on suppose que le modèle considéré est un **modèle d'échantillonnage** de la forme $(\mathcal{X}_0 , \mathcal{B}_0 , P_\theta^\xi)^{\otimes N}$, ie l'image, $\forall$ N $\in$ **N***, du modèle d'échantillonnage asymptotique $(\mathcal{X}_0 , \mathcal{B}_0 , P_\theta^\xi)^{\otimes \mathbf{N}^*}$ par la projection $pr_{1\ldots N} : \mathcal{X}_0^{\mathbf{N}^*} \mapsto \mathcal{X}_0^N$ (cf **modèle asymptotique**), et l'on pose :

(6) $\quad f_{\xi,\theta}$ ou $f_{\xi,\theta} = dP_\theta^\xi / d\mu_0$

(où μ_0 est une mesure donnée sur $\mathcal{B}_0$). On peut alors étudier les propriétés asymptotiques de l'estimateur du mv. Sous des hypothèses générales, on montre que, quelle que soit la **vraie valeur** $\theta \in \Theta$ du paramètre :

(d_1) si Q = 1, la propriété d'**efficacité asymptotique** suivante est valide :

(7) $\quad N^{1/2} (T_N - \theta) \to^{\mathcal{L}}{}_{N \to +\infty} \mathcal{N}(0, N / I_N (\theta))$,

où l'on note $T_N = t_N (X_1 ,..., X_N) = t_N (X)$ au lieu de $T = t (X)$ l'estimateur du mv associé au modèle d'échantillonnage à distance finie N, et $I_N (\theta)$ l'**information de FISHER** du modèle ;

(d_2) si Q > 1, alors :

(8) $\quad N^{1/2} (T_N - \theta) \to^{\mathcal{L}}{}_{N \to +\infty} \mathcal{N}_Q (0, N . (I_N (\theta))^{-1})$

(**loi normale multidimensionnelle** centrée). Cette propriété est analogue à (7) pour l'estimateur vectoriel T_N , où $I_N (\theta)$ est la **matrice d'information** de FISHER ;

(d_3) si $g : \Theta \mapsto \mathbf{R}^K$ est continue, alors :

(9) $\quad g (T_N) \to_{N \to +\infty} g (\theta), \qquad (P_\theta^\xi)^{\otimes \mathbf{N}^*}$-p.s.,

où les P_θ sont les probabilités dont les images par ξ sont les P_θ^ξ ;

(d_4) si $g : \Theta \mapsto \mathbf{R}^K$ est différentiable sur Θ (cf **différentiabilité**), alors :

(1) $\quad N^{1/2} (g (T_N) - g (\theta)) \to^{\mathcal{L}}{}_{N \to +\infty} \mathcal{N}_K (0, \Sigma)$,

où $\Sigma = N J_\theta (I_N (\theta))^{-1} J_\theta'$ et J_θ est la **matrice jacobienne** associée à la **dérivée** D g (θ).

(iv) L'estimateur du mv est en général un estimateur biaisé à distance finie (cf **biais**, **estimateur sans biais**).

L'existence de paramètres secondaires peut mettre en défaut certaines propriétés de l'estimateur du mv (cf **paramètre importun**, **paramètre principal**) : en effet, l'optimisation d'une fonction tq $(\theta, \tau) \mapsto L(\theta, \tau)$ pr à θ (à τ donné, inconnu) n'est pas équivalente à l'optimisation de L pr à (θ, τ).

Il en est de même lorsque la **vraisemblance** comporte des **paramètres incidents**, ie des paramètres qui dépendent des observations (et, en particulier, de leur nombre), auquel cas l'estimateur du mv est généralement inconvergent (cf eg **estimateur convergent**, **modèle à erreurs sur les variables**, **régression pondérée**).

Lorsque les hypothèses portant sur Θ ne sont pas vérifiées, cet estimateur peut aussi avoir un comportement non « régulier ».