VRAISEMBLANCE CONCENTRÉE (C5, H, J)

(11 / 09 / 2020, © Monfort, Dicostat2005, 2005-2020)

Une **vraisemblance concentrée** est le résultat d'une **modification** d'une vraisemblance comportant deux « paramètres ». La procédure consiste à optimiser cette vraisemblance pr au second paramètre et à réintroduire cette solution dans la vraisemblance. Celle-ci ne dépend alors plus que du premier paramètre.

(i) Soit $(\mathcal{X}, \mathcal{B}, (P_{\theta^{\xi}})_{\theta \in \Theta})$ un **modèle statistique** image. On suppose que c'est un **modèle dominé**, ie que $(P_{\theta^{\xi}})_{\theta \in \Theta}$ est une **famille de lois dominée** par une **mesure positive** μ et que Θ peut s'écrire sous la forme d'un produit cartésien $\Theta = \Theta' \times \Theta''$. On note alors $\theta = (\theta', \theta'')$ le paramètre considéré, où θ' est considéré comme un **paramètre principal** et θ'' comme un **paramètre secondaire**. On note alors :

$$(1) \quad \theta \in \Theta \mapsto f(x, \theta) \text{ ou } f(x, (\theta', \theta'')) = (dP_{\theta^{\xi}} / d\mu)(x) \in \mathbf{R}_+, \quad \forall x \in \mathcal{X},$$

la **vraisemblance** du modèle.

(ii) La maximisation partielle de la vraisemblance f pr à $\theta'' \in \Theta''$ conduit (en général) à une valeur $\theta''^{\sim}$ de θ'' tq, par définition :

$$(2) \quad \sup_{\theta'' \in \Theta''} f(x, (\theta', \theta'')) = f(x, (\theta', \theta''^{\sim})), \quad \forall (x, \theta') \in \mathcal{X} \times \Theta'.$$

Cette valeur dépend donc, en général, de x et de θ' , ie $\theta''^{\sim} = \phi(x, \theta')$.

On dit alors que la vraisemblance f est « concentrée » pr à θ'' ssi on remplace dans (1) le paramètre θ'' par la valeur $\theta''^{\sim}$ ainsi définie.

Par suite, on appelle **vraisemblance concentrée** pr à θ'' la fonction f_c suivante :

$$(3) \quad f_c(x, \theta') = f(x, (\theta', \theta''^{\sim})) = f(x, (\theta', \phi(x, \theta'))).$$

(iii) Souvent, $\Theta' \subset \mathbf{R}^{Q'}$ et $\Theta'' \subset \mathbf{R}^{Q''}$. Par suite, si f est différentiable (cf **différentiabilité**), la propriété (2) implique la condition du premier ordre suivante :

$$(4) \quad D_{\theta''} f(x, (\theta', \theta''^{\sim})) = 0,$$

où encore $D_3 f(x, \theta', \theta''^{\sim}) = 0$, en notant $f(x, \theta', \theta''^{\sim}) = f(x, \theta', \theta''^{\sim})$.

(iv) La **concentration d'une vraisemblance** est parfois utilisée dans la mise en oeuvre de la **méthode du maximum de vraisemblance**. On optimise alors pr à θ' la vraisemblance concentrée f définie en (3).

(v) Le procédé précédent est souvent utilisé dans le cadre du **modèle de régression** ou du **modèle d'interdépendance** (ou modèle à équations simultanées).

Ainsi, le **modèle d'interdépendance linéaire** suivant, écrit dans l'**espace de variables** (ξ, η) :

$$(5) \quad B \eta + C \xi = \varepsilon,$$

comporte un vecteur η constitué de G **variables endogènes**, un vecteur ξ constitué de K **variables exogènes**, un **vecteur aléatoire** ε à valeurs dans $\mathbf{R}^G$ (G équations), une **matrice** $B \in M_G(\mathbf{R})$ tq $b_{gg} = 1$ (pour tout $g = 1, \dots, G$) et une matrice $C \in M_{GK}(\mathbf{R})$.

On suppose que ce modèle vérifie :

$$E \varepsilon = 0 \in \mathbf{R}^G \quad (\text{espérance nulle}),$$

$$(6) \quad V \varepsilon = \Sigma = (\sigma_{gh})_{(g,h)} \in S_G(\mathbf{R}) \quad (\text{matrice symétrique}),$$

$$\varepsilon \sim \mathcal{N}_G(0, \Sigma) \quad (\text{loi gaussienne}),$$

et qu'il est « observé » dans un **espace d'observation** selon $B y_t + C x_t = u$, (où $t = 1, \dots, T$), ie (forme matricielle) :

$$(8) \quad Y' B + X' C = U' \quad (\text{notations évidentes}),$$

où (X, Y) est un **jeu d'observations** de (ξ, η) .

La vraisemblance (gaussienne) du modèle s'écrit :

$$(7) \quad f(X, Y, B, C, \Sigma) = |\det B|^T (2\pi)^{-GT/2} (\det \Sigma)^{-T/2} \cdot \exp \{-(1/2) \cdot Q(X, Y, B, C, \Sigma)\},$$

$$\text{avec } Q(X, Y, B, C, \Sigma) = \sum_{t=1}^T (B y_t + C x_t)' \Sigma^{-1} (B y_t + C x_t).$$

La maximisation de f (ou de $\log f$) partielle pr à Σ^{-1} conduit à une condition du premier ordre, d'où l'**estimateur** correspondant :

$$(9) \quad \Sigma \sim = (B C) M_{XY} (B C)',$$

où $M = T^{-1} [Y Y', Y X' / X Y', X X']$ est la $(2T, 2T)$ -matrice des **moments empiriques** « croisés ».

La **fonction de vraisemblance concentrée** s'écrit alors :

$$(10) \quad \tilde{f}(B, C) = f(B, C, \Sigma \sim),$$

et la **fonction de Log-vraisemblance concentrée** est de la forme :

$$(11) \quad \mathcal{L}^{\sim}(B, C) = \text{Log } f^{\sim}(B, C).$$

Ces deux fonctions $f^{\sim}$ et $\mathcal{L}^{\sim}$ sont obtenues en remplaçant Σ par $\Sigma^{\sim}$ dans (7). Elles ne dépendent que du couple de paramètres principaux (B, C) .

La **méthode du maximum de vraisemblance (ici concentrée)** revient alors à maximiser la (Log-)vraisemblance concentrée par à ce couple (B, C) , ie à maximiser :

$$(11) \quad \mathcal{L}^{\sim}(B, C) = -((G \cdot T)/2) \cdot \text{Log}(2\pi) + T \cdot \text{Log} |\det B| - (T/2) \cdot \text{Log } \Sigma^{\sim},$$

avec $\Sigma^{\sim} = (B \ C) M_{XY} (B \ C)'$ (d'après (9)).