Econométrie non linéaire
Oraux

Tran Viet Chi

20 juillet 2005

1 Consistance d'un M-Estimateur

Enoncé 1. *Soit (Θ, d) un espace métrique compact. On considère une séquence $(M_n(\theta))_{n\in\mathbb{N}}$ de fonctions aléatoires réelles de la variable θ, convergeant uniformément en probabilité vers une fonction déterministe $M(\theta)$:*

$$\sup_{\theta\in\Theta} |M_n(\theta) - M(\theta)| \to^P_{n\to\infty} 0. \tag{1}$$

On suppose que M admet un maximum "bien séparé" en θ_0 :

$$\forall\varepsilon > 0, \quad \sup_{\theta \ | \ d(\theta,\theta_0)\geq\varepsilon} M(\theta) < M(\theta_0). \tag{2}$$

On s'intéresse à la convergence des estimateurs $\widehat{\theta}_n$ tels que :

$$M_n(\widehat{\theta}_n) \geq \sup_{\theta\in\Theta} M_n(\theta) - o_P(1). \tag{3}$$

1. *En quoi les estimateurs (3) sont-ils une généralisation des M-estimateurs classiques :*

$$\widehat{\theta}_n \in \arg\max_{\theta\in\Theta} M_n(\theta)\,?$$

2. *Citer quelques M-estimateurs.*
3. *Dessiner une fonction M qui ne satisfasse pas la condition (2).*
4. *Montrer que :*

$$M_n(\widehat{\theta}_n) + o_P(1) \geq M(\theta_0).$$

5. *En déduire une majoration de :*

$$M(\theta_0) - M(\widehat{\theta}_n).$$

6. *Conclure que :*

$$\widehat{\theta}_n \to^P_{n\to\infty} \theta_0.$$

■

Démonstration. 1.2. La définition (3) signifie que $\widehat{\theta}_n$ maximise presque M_n. On s'autorise ainsi des erreurs numériques ou des estimateurs imparfaits à distance finie. Un exemple de M-estimateur bien connu est l'estimateur de maximum de vraisemblance.
3. Si M admet plusieurs maxima, alors (2) n'est pas satisfaite.
4. Un cas particulier de (3) est :

$$M_n(\widehat{\theta}_n) \geq M_n(\theta_0) - o_P(1).$$

Comme $M_n(\theta_0) \to^P_{n\to\infty} M(\theta_0)$, alors :

$$M_n(\theta_0) = M(\theta_0) + o_P(1),$$

d'où :

$$M_n(\widehat{\theta}_n) \geq M(\theta_0) - o_P(1).$$

5. On en déduit :

$$\begin{aligned} 0 \leq M(\theta_0) - M(\widehat{\theta}_n) &\leq M_n(\widehat{\theta}_n) - M(\widehat{\theta}_n) + o_P(1) \\ &\leq \sup_{\theta\in\Theta} |M_n(\theta) - M(\theta)| + o_P(1). \end{aligned}$$

La positivité vient de la définition de θ_0 comme maximum de M. La première inégalité de droite vient de 4. tandis que la dernière est une conséquence de la convergence uniforme.
6. Soit $\varepsilon > 0$. Par (2), il existe $\eta_\varepsilon > 0$ tel que :

$$\sup_{\theta \mid d(\theta,\theta_0)\geq\varepsilon} M(\theta) < M(\theta_0) - \eta_\varepsilon.$$

Donc,

$$\{d(\theta,\theta_0) \geq \varepsilon\} \subset \{M(\theta) < M(\theta_0) - \eta_\varepsilon\}.$$

Comme on a montré en 5. que $M_n(\widehat{\theta}_n) - M(\theta_0) \to 0$, alors :

$$P\left\{M(\widehat{\theta}_n) < M(\theta_0) - \eta_\varepsilon\right\} \to_{n\to\infty} 0.$$

On en conclut que :

$$P\left\{d(\widehat{\theta}_n,\theta_0) \geq \varepsilon\right\} \to_{n\to\infty} 0.$$

□

2 Condition du premier ordre et régression

Enoncé 2. *Soit $\Theta \subset \mathbb{R}$ et soit ψ une fonction réelle. On considère des couples i.i.d. d'observations réelles $(Y_i,X_i)_{1\leq i\leq n}$ et on s'intéresse aux M-estimateurs de la forme :*

$$\widehat{\theta}_n \in \arg\min_{\theta\in\Theta} \sum_{i=1}^{n} \psi(Y_i - m(X_i,\theta)). \tag{4}$$

1. *Interpréter le rôle de la fonction ψ.*
2. *Enoncer les hypothèses nécessaires à la convergence de l'estimateur $\widehat{\theta}_n$.*

Si on néglige les hypothèses de régularité (ce que l'on fera dans toute la suite de l'exercice), on peut s'intéresser plus particulièrement à la condition :

$$H\ :\ \theta_0\ \text{ vérifie la condition du premier ordre pour le problème limite.}$$

On supposera de plus que :

$$\mathbb{E}\left(\frac{\partial m}{\partial\theta}(X,\theta_0)\right) \neq 0. \tag{5}$$

3. *On suppose que le vrai modèle est :*
$$Y_i = m(X_i,\theta) + \varepsilon_i,$$
ou les variables aléatoires $(\varepsilon_i)_{1\leq i\leq n}$ sont indépendantes des variables aléatoires $(X_i)_{1\leq i\leq n}$. Montrer que $\widehat{\theta}_n$ vérifie la condition H si et seulement si :
$$E(\psi'(\varepsilon)) = 0.$$
4. *Supposons de plus que la loi de ε est absolument continue par rapport à la mesure de Lebesgue et symétrique autour de 0. Montrer que l'hypothèse H est satisfaite dès que ψ est paire.*
5. *Citer un exemple.*

■

Démonstration. 3. La condition H se réécrit :

$$\frac{\partial}{\partial\theta}\,\mathbb{E}\left(\psi(Y - m(X,\theta))\right)|_{\theta=\theta_0} = 0.$$

En supposant que l'on puisse intervertir dérivation et espérance, H est équivalente aux équations suivantes :

$$\begin{aligned}
& \mathbb{E}\left[\psi'(Y - m(X,\theta_0))\left(-\frac{\partial m}{\partial\theta}(X,\theta_0)\right)\right] = 0 \\
\Leftrightarrow\quad & \mathbb{E}\left[\psi'(\varepsilon)\left(\frac{\partial m}{\partial\theta}(X,\theta_0)\right)\right] = 0 \\
\Leftrightarrow\quad & \mathbb{E}\left[\frac{\partial m}{\partial\theta}(X,\theta_0)\right] = 0 \\
\Leftrightarrow\quad & \mathbb{E}\left[\psi'(\varepsilon)\right] = 0,
\end{aligned}$$

par (5).
4. En notant $f(\varepsilon)$ la densité de la loi de ε, $\mathbb{E}\left[\psi'(\varepsilon)\right] = 0$ est équivalente aux assertions suivantes :

$$\begin{aligned}
\int_{-\infty}^{\infty} \psi'(\varepsilon) f(\varepsilon) d\varepsilon &= 0 \\
\int_{0}^{\infty} \psi'(\varepsilon) f(\varepsilon) d\varepsilon + \int_{0}^{\infty} \psi'(-\varepsilon) f(-\varepsilon) d\varepsilon &= 0 \\
\int_{0}^{\infty} (\psi'(\varepsilon) + \psi'(-\varepsilon)) f(\varepsilon) d\varepsilon &= 0.
\end{aligned}$$

Il est clair que la parité de ψ (donc l'imparité de sa dérivée) est une condition suffisante pour que la dernière égalité soit replie.
5. Un exemple est la régression non linéaire (MCNL) gaussienne, avec des résidus normaux et $\psi(x) = x^2$. □

3 Stand de tir !!

Enoncé 3. *Dans une fête foraine, n joueurs indépendants ($1 \leq i \leq n$) sont passés au stand de tir. Chacune de ces personnes est caractérisée par une variable exogène Z_i scalaire (son acuité visuelle, sa résistance à l'alcool...). Le joueur i réalise un score X_i entier, distribué suivant une loi de Poisson de paramètre :*

$$\theta_i = e^{Z_i\beta}.$$

Le paramètre à estimer est $\beta \in \mathbb{R}_+^*$. On suppose qu'il est le même pour tous les joueurs.

1. *Donner ou calculer l'espérance et la variance de X_i conditionnellement aux Z_i.*
2. *Expliquer précisément comment estimer β par la méthode des MCNL pondérés. On donnera le programme de minimisation, et on précisera les différentes métriques utilisées.*

■

Démonstration. 1. On a :

$$\begin{aligned}
EX &= \sum_{k=0}^{\infty} k \frac{(e^{kZ_i\beta})}{k!} e^{-e^{Z_i\beta}} \\
&= e^{Z_i\beta} \sum_{k=1}^{\infty} \frac{(e^{(k-1)Z_i\beta})}{(k-1)!} e^{-e^{Z_i\beta}} \\
&= e^{Z_i\beta}.
\end{aligned}$$

et :

$$\begin{aligned}
Var(X) &= E(X^2) - (EX)^2, \\
E(X^2) &= \sum_{k=0}^{\infty} k^2 \frac{(e^{kZ_i\beta})}{k!} e^{-e^{Z_i\beta}} \\
&= \sum_{k=0}^{\infty} (k(k-1)+k) \frac{(e^{kZ_i\beta})}{k!} e^{-e^{Z_i\beta}} \\
&= e^{2Z_i\beta} \sum_{k=2}^{\infty} \frac{(e^{(k-2)Z_i\beta})}{(k-2)!} e^{-e^{Z_i\beta}} + e^{Z_i\beta} \\
&= e^{2Z_i\beta} + e^{Z_i\beta} \\
Var(X) &= e^{Z_i\beta}.
\end{aligned}$$

2. En écrivant X =espérance + bruit, on a :

$$X_i = e^{Z_i\beta} + \varepsilon_i.$$

On reconnaît une régression non-linéaire où la fonction de régression est :

$$m(z,\beta) \mapsto e^{z\beta}.$$

La régression non-linéaire, tout comme pour la régression linéaire, minimise la variance des résidus :

$$\widehat{\beta}_n = \arg\min_{\beta\in\mathbb{R}_+^*} \left\| \begin{pmatrix} X_1 - e^{Z_1\beta} \\ \vdots \\ X_n - e^{Z_n\beta} \end{pmatrix} \right\|_{S_n}^2 . \tag{6}$$

Lorsque $S_n = Id_{\mathbb{R}^n}$ la matrice identité de $\mathbb{R}^n$, on obtient les MCNL.

Il n'y a *a priori* pas de raison de choisir la norme euclidienne et on peut chercher à minimiser la norme calculée dans une autre métrique S_n. On obtient alors les MCNLQG (quasi-généralisés).

Lorsque la matrice S_n est diagonale par blocs de la forme :

$$S_n = \begin{pmatrix} a_n(Z_1) & & 0 \\ & \ddots & \\ 0 & & a_n(Z_n) \end{pmatrix}, \tag{7}$$

avec a_n tel que $a_n(Z) \to a^*(Z) = (Var(X \mid Z,\beta_0))^{-1}$ en probabilité, lorsque $n \to \infty$, on obtient un estimateur $\widehat{\beta}_n$ convergent et asymptotiquement efficace.

La mise en oeuvre des MCNLQG se donc fait en 2 étapes :

1. On commence par calculer l'estimateur MCNL non pondéré obtenu avec $S_n = Id_{\mathbb{R}^n}$. On obtient un estimateur convergent $\widetilde{\theta}_n$ de θ_0 (mais non optimal) :
$$\widetilde{\beta}_n = \arg\min_{\beta\in\mathbb{R}_+^*} \sum_{i=1}^n \left(X_i - e^{Z_i,\beta}\right)^2 .$$
Ce paramètre $\widetilde{\beta}_n$ nous permet d'obtenir une métrique permettant d'avoir un esimateur asymptotiquement efficace :
$$S_n = e^{-Z_i\widetilde{\beta}_n}.$$
2. Dans un second temps, on minimise ainsi :
$$\widehat{\beta}_n = \arg\min_{\beta\in\mathbb{R}_+^*} \sum_{i=1}^n e^{-Z_i\widetilde{\beta}_n} \left(y_i - m(x_i,\beta)\right)^2 .$$

On peut remarquer que $e^{Z_i\widetilde{\beta}_n}$ est la variance de X_i conditionnellement à Z_i, et pas la variance tout court. Ainsi, il n'est pas possible de court-circuiter la première étape de l'estimation en deux étapes en remplaçant $a_n(Z)$ par un estimateur de l'inverse de la variance empirique des X_i. □

4 Au laboratoire...

Enoncé 4. *Un savant réalise n expériences indépendantes dans son laboratoire. Dans chaque boîte de Pétri, une cellule mère est posée sur un substrat nourrissant. Au bout de trois heures, il ouvre chacune des boîtes et relève le nombre de cellules filles qui y vivent.*

Avec probabilité λ la boîte contient des cellules filles (vivantes ou mortes). On notera alors $Y_i = 1$ pour marquer l'événement "il y a des cellules filles dans la boîte i" et le nombre de cellules filles vivantes X_i est distribué suivant une loi de Poisson de paramètre θ (ce nombre est éventuellement 0, mais on sait que la cellule s'est divisée en retrouvant les restes des filles...) Avec probabilité $1-\lambda$, la cellule mère est morte sans donner de fille ($X_i = 0, Y_i = 0$).

1. *Calculer la loi de la variable aléatoire X.*
2. *Calculer EX.*
3. *Dans ce modèle, les équations de vraisemblance ne se résolvent pas explicitement. On se propose donc d'estimer θ par MCA.*
   (a) *Calculer pour $x \geq 1$:*
   $$\frac{P(X = x+1)}{P(X = x)}.$$
   (b) *Soient $p_j = P(X = j)$ pour $j = 1,2,3$. Ecrire deux relations estimantes entre p_1, p_2 et p_3. On résumera ces équations estimantes sous l'écriture :*
   $$h(p_1,p_2,p_3,\theta) = 0_{\mathbb{R}^2}.$$

(c) *Soient n_j pour $j = 1{,}2{,}3$ le nombre de boîtes où le savant trouve j filles vivantes :*

$$n_j = card\{i \mid X_i = j\}.$$

Expliquer comment mettre en oeuvre les MCA pour obtenir un estimateur $\widehat{\theta}$ de θ en fonction de n_1, n_2 et n_3 (on ne fera pas le calcul de la métrique optimale).

(d) *Expliquer sans calcul comment tester $h(p_1,p_2,p_3,\theta) = 0_{\mathbb{R}^2}$. (On dira un mot de la statistique de test et on donnera la loi limite).*

(e) *Calculer l'estimateur MCA de θ pour la métrique particulière $S_n = Id_{\mathbb{R}^2}$.*

■

Démonstration. 1. On a :

$$\begin{aligned} P(X=k) &= P(X=k\,|\,Y=0)P(Y=0) + P(X=k\,|\,Y=1)P(Y=1) \\ &= (1-\lambda)\delta_0(k) + \lambda\frac{\theta^k e^{-\theta}}{k!}. \end{aligned}$$

2. On en déduit :

$$EX = 0 \times (1-\lambda+\lambda e^{-\theta}) + \sum_{k=1}^{\infty} k\lambda\frac{\theta^k e^{-\theta}}{k!} = \lambda\theta.$$

3. On déduit de la question 1 :

$$\frac{P(X=x+1)}{P(X=x)} = \frac{\lambda\theta^{x+1}e^{-\theta}}{(x+1)!}\frac{x!}{\lambda\theta^x e^{-\theta}} = \frac{\theta}{x+1}.$$

4. On en déduit les relations estimantes :

$$\begin{aligned} 2p_2 - \theta p_1 &= 0 \\ 3p_3 - \theta p_2 &= 0, \end{aligned}$$

qui est de la forme annoncée dans l'énoncé avec :

$$h \;:\; (p_1,p_2,p_3,\theta) \mapsto \begin{pmatrix} 2p_2 - \theta p_1 \\ 3p_3 - \theta p_2 \end{pmatrix}.$$

5. On a une relation estimante entre le paramètre d'intérêt θ et les paramètres auxiliaires p_1, p_2, p_3 que l'on sait estimer (par n_1/n, n_2/n et n_3/n respectivement). Nous disposons plus précisément du résultat suivant lié au Théorème Central Limite :

$$\sqrt{n}\left[\begin{pmatrix} \frac{n_1}{n} - p_1 \\ \frac{n_2}{n} - p_2 \\ \frac{n_3}{n} - p_3 \end{pmatrix}\right] \rightsquigarrow_{n\to\infty} \mathcal{N}\left(0_{\mathbb{R}^3}, \Omega_0\right),$$

où :

$$\Omega_0 = \begin{pmatrix} p_1(1-p_1) & -p_1p_2 & -p_1p_3 \\ -p_1p_2 & p_2(1-p_2) & -p_2p_3 \\ -p_1p_3 & -p_2p_3 & p_3(1-p_3) \end{pmatrix}.$$

On peut donc mettre en oeuvre les MCA en estimant θ par :

$$\widehat{\theta}_n = \arg\min_{\theta\in\Theta} \left\| h\left(\frac{n_1}{n},\frac{n_2}{n},\frac{n_3}{n},\theta\right)\right\|^2_{S_n}.$$

Pour S_n convergeant en probabilité vers :

$$S_\infty := \left(\frac{\partial h}{\partial(p_1,p_2,p_3)}(p_1^0,p_2^0,p_3^0)\; \Omega_0\; \frac{\partial h'}{\partial\begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}}(p_1^0,p_2^0,p_3^0) \right)^{-1},$$

où p_1^0, p_2^0 et p_3^0 sont les vrais paramètres, on a un estimateur asymptotiquement optimale au sens de la variance. Comme S_∞ dépend des vrais paramètres p_1^0, p_2^0 et p_3^0 inconnus, on peut mettre en oeuvre une méthode d'estimation en deux étapes.
6. La statistique de test est :

$$\xi_n = n\ h\left(\frac{n_1}{n},\frac{n_2}{n},\frac{n_3}{n},\widehat{\theta}_n\right)' S_n h\left(\frac{n_1}{n},\frac{n_2}{n},\frac{n_3}{n},\widehat{\theta}_n\right).$$

Elle converge en loi sous $H_0\ :\ h(p_1,p_2,p_3,\theta) = 0_{\mathbb{R}^2}$ vers un $\chi^2(1)$ et diverge vers ∞ sous l'hypothèse alternative. La région de rejet est :

$$W = \left\{\xi_n \geq q_{95\%}(\chi^2(1))\right\}.$$

7. Le calcul donne :

$$\widehat{\theta}_n = \frac{2n_1n_2 + 3n_2n_3}{n_1^2 + n_2^2}.$$

□

5 Une loi normale et une régression

Enoncé 5. *Soit n couples indépendants d'observations scalaires $(Y_i,X_i)_{1\leq i\leq n}$ tels que :*

$$Y_i \rightsquigarrow \mathcal{N}\left(m(X_i,\theta), \sigma^2(\theta)\right).$$

Les fonctions m et σ^2 sont connues. Le paramètre à estimer θ est supposé être dans un compact Θ.

1. *Vous paraît-il naturel de calibrer θ par la régression non linéaire :*

$$\widehat{\theta}_n \in \arg\min_{\theta\in\Theta} \|Y_i - m(X_i,\theta)\|^2_{S_n} \tag{8}$$

2. *Ecrire les conditions du premier ordre pour une matrice S_n diagonale :*

$$S_n = \begin{pmatrix} a_1^n & & 0 \\ & \ddots & \\ 0 & & a_n^n \end{pmatrix}. \tag{9}$$

3. *Quel choix optimal (dans quel sens ?) peut-on faire pour les a_i^n ?*
4. *Expliciter la mise en oeuvre par MCNL pondérés.*
5. *Ecrire la log-vraisemblance des observations.*
6. *Ecrire les conditions du premier ordre pour l'estimateur du maximum de convergence. Les comparer avec celles obtenues pour les moindres carrés.*

■

Démonstration. 1.2. Pour S_n choisie comme dans l'énoncé :

$$\|Y_i - m(X_i,\theta)\|^2_{S_n} \quad = \quad \sum_{i=1}^n a_i^n (Y_i - m(X_i,\theta))^2.$$

Les conditions du premier ordre s'écrivent :

$$0 \quad = \quad \sum_{i=1}^n 2\, a_i^n\, (Y_i - m(X_i,\theta))\frac{\partial m}{\partial\theta}(X_i,\theta).$$

Cette relation peut s'interpréter comme une relation d'orthogonalité entre les résidus de la régression et le vecteur de composantes $\frac{\partial m}{\partial\theta}(X_i,\theta)$.
3. Un choix optimal (au sens de la variance asymptotique de l'estimateur des MCNL) de a_i^n est obtenu lorsque a_i^n converge en probabilité vers $Var(Y|X,\theta)^{-1} = (\sigma^2(\theta))^{-1}$ lorsque n tend vers l'infini. Cette dernière quantité dépend de θ inconnu.
4. Pour contourner ce problème, on peut classiquement mettre en oeuvre un méthode d'estimation en

deux étapes. On commence par mettre en oeuvre la maximisation (8) avec $S_n = Id$. On obtient alors un estimateur $\widetilde{\theta}_n$ convergent mais non optimal de θ. Ensuite, on réitère la maximisation (8) avec :

$$S_n = (\sigma^2(\widetilde{\theta}_n))^{-1}\, Id_{\mathbb{R}^n}.$$

On peut remarquer que $\sigma^2(\theta)$ n'étant que la variance de Y conditionnellement aux X_i, on ne peut pas court-circuiter la première étape en choisissant par exemple la variance empirique des observations.
5. La log-vraisemblance des observations est :

$$\log L_n(\theta) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log(\sigma^2(\theta)) - \sum_{i=1}^{n} \frac{(Y_i - m(X_i,\theta))^2}{2\sigma^2(\theta)}.$$

6. Les équations de vraisemblance sont :

$$\begin{aligned}
0 &= \frac{\partial \log L_n}{\partial \theta}(\theta) \\
&= -\frac{n}{2}\frac{1}{\sigma^2(\theta)}\frac{\partial \sigma^2}{\partial \theta}(\theta) - \sum_{i=1}^{n} \frac{-2(Y_i - m(X_i,\theta))\frac{\partial m}{\partial \theta}(X_i,\theta) \times 2\sigma^2(\theta) - 2\frac{\partial \sigma^2}{\partial \theta}(\theta)(Y_i - m(X_i,\theta))^2}{4(\sigma^2(\theta))^2} \\
&= \frac{n}{2\sigma^2(\theta)}\frac{\partial \sigma^2}{\partial \theta}(\theta) + \sum_{i=1}^{n} \frac{(Y_i - m(X_i,\theta))}{\sigma^2(\theta)}\frac{\partial m}{\partial \theta}(X_i,\theta) + \sum_{i=1}^{n} \frac{1}{2\sigma^2(\theta)}\frac{\partial \sigma^2}{\partial \theta}(\theta)\frac{(Y_i - m(X_i,\theta))^2}{\sigma^2(\theta)} \\
&= \frac{1}{2\sigma^2(\theta)}\frac{\partial \sigma^2}{\partial \theta}(\theta)\left(\sum_{i=1}^{n}\left(\frac{(Y_i - m(X_i,\theta))^2}{\sigma^2(\theta)} - 1\right)\right) + \sum_{i=1}^{n} \frac{(Y_i - m(X_i,\theta))}{\sigma^2(\theta)}\frac{\partial m}{\partial \theta}(X_i,\theta).
\end{aligned}$$

□