Efficacité semi-paramétrique pour la méthode des moments généralisée

P. Rochet

Université Toulouse 3
Institut de Mathématiques, Equipe Statistique et Probabilités

Mont Dore, mai 2010

Plan

Plan

Estimation d'un paramètre

Observations iid $X_1, ..., X_n$ de loi μ.
But : estimer un paramètre $\theta_0 = \theta(\mu)$.

Le problème se caractérise par

- **un modèle** $\mathcal{M}$: ensemble de mesures, valeurs possibles de μ (on suppose $\mu \in \mathcal{M}$),
- **un paramètre** θ : application $\theta : \mathcal{M} \to \Theta \subset \mathbb{R}^d$.

Estimation du paramètre

On note :

$$\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}.$$

Deux façons naturelles pour estimer θ_0.

- **1ère idée :** Construire un estimateur $\hat{\mu} \in \mathcal{M}$ de μ à partir des observations et définir $\hat{\theta} = \theta(\hat{\mu})$.
- **2ème idée :** Etendre θ à un modèle plus grand $\mathcal{P}$ contenant μ_n : $\overline{\theta} : \mathcal{P} \to \Theta$, et définir $\hat{\theta} = \overline{\theta}(\mu_n)$.

Agrandissement du modèle $\Longrightarrow$ perte d'information.

Efficacité semi-paramétrique

On distingue en statistique 3 types de modèles :

- **Paramétrique :** $\mathcal{M} = \{\mu_\theta, \theta \in \Theta \subset \mathbb{R}^d\}$.
- **Non-paramétrique :** $\mathcal{M} = \mathcal{P}$: l'ensemble des mesures de proba sur $\mathcal{X}$.
- **Semi-paramétrique :** entre les deux.
  ex : paramètre de dim. infinie, mesures à densité par rapport à λ, mesures de moyennes nulles, etc...

Information de Fisher

Calcul de l'Information de Fisher

- dans un modèle paramétrique $\{\mu_\theta, \theta \in \Theta\}$:

$$\mathcal{I}(\{\mu_\theta\}) = \int \left(\frac{\partial}{\partial \theta}_{|\theta_0} \log \frac{d\mu_\theta}{d\mu} \right)^2 d\mu.$$

- dans un modèle semi ou non-paramétrique :

$$\mathcal{I}(\mathcal{M}) = \inf_{\{\mu_\theta\} \subset \mathcal{M}} \mathcal{I}(\{\mu_\theta\}).$$

Remarques :

- \+ le modèle est grand, - on a d'information.
- \+ l'application $\theta(.)$ est "lisse", + on a d'information.

Borne d'efficacité

On note $\mathcal{B}(\mathcal{M}) = \mathcal{I}^{-1}(\mathcal{M})$.

Théorème (*Asymptotic statistics*, van der Vaart (1998))

Si $T = T(X_1, ..., X_n)$ est un estimateur asymptotiquement sans biais de θ dans un modèle $\mathcal{M}$, alors

$$\lim_{n\to\infty} n \operatorname{var}(T) \geq \mathcal{B}(\mathcal{M}).$$

Généralisation de l'inégalité de Cramer-Rao aux modèles semi et non-paramétriques.

$\mathcal{B}(\mathcal{M})$ est la *borne d'efficacité* du problème.

Plan

Le problème

- Echantillon $X_1, ..., X_n$ iid de loi μ inconnue, dans un espace probabilisé $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.
- On veut estimer $\theta_0 \in \Theta \subset \mathbb{R}^d$ tel que
$$F(\theta_0, \mu) = \int \Phi(x, \theta_0) d\mu = 0,$$
où $\Phi : \mathcal{X} \times \Theta \to \mathbb{R}^k$ connue. On suppose $k \geq d$.

Le problème

- Echantillon $X_1, ..., X_n$ iid de loi μ inconnue, dans un espace probabilisé $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.
- On veut estimer $\theta_0 \in \Theta \subset \mathbb{R}^d$ tel que
$$F(\theta_0, \mu) = \int \Phi(x, \theta_0) d\mu = 0,$$
où $\Phi : \mathcal{X} \times \Theta \to \mathbb{R}^k$ connue. On suppose $k \geq d$.

Ex : variance fonction de la moyenne,
$$\Phi(x, \theta) = \begin{pmatrix} x - \theta \\ x^2 - v(\theta) \end{pmatrix}.$$

Problème étudié en économétrie (Hansen (1982), Chamberlain (1987), Qin & Lawless (1994)).

Application au problème

Contrainte de moment

$$F(\theta_0, \mu) = \int \Phi(x, \theta_0) d\mu = 0.$$

- **Le modèle :** $\mathcal{M} = \left\{\nu : \exists\theta = \theta(\nu) \in \Theta, \int \Phi_\theta d\nu = 0\right\}$
- **Le paramètre :** $\theta(\nu)$: zéro de $\theta \mapsto \int \Phi_\theta d\nu$.

Théorème (Qin et Lawless (1994))

Soit $D = \int \frac{\partial}{\partial\theta}\Phi(., \theta_0) d\mu$ et $V = \int \Phi(., \theta_0)\Phi(., \theta_0)^t d\mu$,

$$\mathcal{B}(\mathcal{M}) = \left[DV^{-1}D^t\right]^{-1}.$$

Vraisemblance empirique généralisée (GEL)

1ère idée : Construire $\hat{\mu} \in \mathcal{M}$ et définir $\hat{\theta} = \theta(\hat{\mu})$.

Estimateur GEL

Soit f convexe avec $f(1) = f'(1) = 0$, on définit

$$\hat{\mu} = \operatorname*{argmin}_{\nu \in \mathcal{M}} \int f\left(\frac{d\nu}{d\mu_n}\right) d\mu_n$$

- Atteint la borne semi-paramétrique asymptotiquement.
- $\hat{\mu}$ est parfois difficile à déterminer explicitement.

Méthode des moments généralisée (GMM)

2ème idée : Etendre θ, $\overline{\theta} : \mathcal{P} \to \Theta$ et définir $\hat{\theta} = \overline{\theta}(\mu_n)$.

Estimateur GMM

Soit M une matrice symétrique définie positive ,

$$\overline{\theta}(\nu) = \underset{\theta \in \Theta}{\text{argmin}} \, \|F(\theta, \nu)\|_M = \underset{\theta \in \Theta}{\text{argmin}} \, F(\theta, \nu)^t M F(\theta, \nu).$$

- Estimateur plus facile à implémenter.
- Modèle plus grand $\Longrightarrow$ moins d'efficacité.

Résultat

Méthode des moments généralisée :

- **Le modèle :** $\mathcal{P}$ (modèle non-paramétrique).
- **Le paramètre :** $\overline{\theta}_M(\nu) = \underset{\theta \in \Theta}{\text{argmin}} \, \|F(\theta, \nu)\|_M$.

Théorème

Dans le modèle $\mathcal{P}$, la borne semi-paramétrique pour estimer $\overline{\theta}_M$ est

$$\mathcal{B}(\mathcal{P}) = \left[DMD^t\right]^{-1} \left[DMVMD^t\right] \left[DMD^t\right]^{-1}.$$

- $\mathcal{B}(\mathcal{P})$ dépend du choix de la matrice M.
- Choisir M telle que l'extension $\overline{\theta}_M$ soit efficace.

Résultats

On rappelle : $\mathcal{B}(\mathcal{M}) = \left[DV^{-1}D^t\right]^{-1}$

Lemme

$$\left[DMD^t\right]^{-1}\left[DMVMD^t\right]\left[DMD^t\right]^{-1} \geq \left[DV^{-1}D^t\right]^{-1},$$

avec égalité :
- pour tout M si $d = k$,
- pour $M = V^{-1}$ si $k > d$.

- Comme on pouvait s'y attendre $\mathcal{M} \subseteq \mathcal{P} \Longrightarrow \mathcal{B}(\mathcal{P}) \geq \mathcal{B}(\mathcal{M})$.
- GMM optimal : $M = V^{-1}$ (cf Chamberlain (1987)).

Résultats

GMM avec contrainte approchée :
On observe une approximation de la contrainte $\Phi_m \to \Phi$.

Théorème

Sous les hypothèses :

- $$\int \Big(\sqrt{n} \sup_{\theta \in \theta} \|\Phi_m(.,\theta) - \Phi(.,\theta)\|\Big)^{12} d\mu \to 0$$
- $\exists\, \mathcal{V}$ voisinage de θ_0,
$$\int \Big(\sqrt{n} \sup_{\theta \in \mathcal{V}} \|\nabla\Phi_m(.,\theta) - \nabla\Phi(.,\theta)\|\Big)^{12} d\mu \to 0,$$

le GMM optimal reste efficace avec la contrainte approchée.

MERCI