ECON220C Discussion Section 6

M-Estimation

Lapo Bini

Roadmap

1. Refresh
2. General Consistency Theorem
3. Uniform Law of Large Numbers
4. Exercise on NLS and QMLE

Motivation

We develop a complex model that is a function of data and a parameter. The parameter is defined as the solution to the following **statistical problem**:

$$\theta = \arg\max_{c \in \Theta} M(c)$$

where $M(\cdot)$ is the **population criterion function**. What criterion functions do we use in econ?

1. **Loglikelihood**: $\{x_1 \ldots, x_n\} \sim \mathcal{L}_\theta$, $\theta \in \Theta$ then $M(c) = E[log(f(x_i; c))]$
2. **Distance/error**: we have our predictive model $g(x_i; \theta)$ and observed outcomes $\{y_i\}_{i=1}^n$, we maximize the negative expected prediction error: $M(c) = E[\rho(y_i - g(x_i; c))]$

Goal & Procedure

Goal: study the properties of an estimator that has no closed-form solution: consistency, asymptotic distribution, and asymptotic variance.

1. Replace the population criterion function with the **sample analogue**: $\hat{\theta} = \arg\max_{c \in \Theta} M_n(c)$.

2. Take **FOC** and **SOC** to derive the empirical score equation and empirical hessian.

3. **Taylor expansion** of first order around the true parameter θ.

4. Check and apply the **consistency** property of $\hat{\theta}$. Remember: $M_n(c) \xrightarrow{p} M(c)$ is not sufficient for $\hat{\theta} \xrightarrow{p} \theta$. Convergence involves two directions: $M_n(\hat{\theta}) \xrightarrow{p} M(\theta)$

General Consistency Theorem

Consider the setting:

$$\theta = \arg\max_{c \in \Theta} M(c) \quad \text{and} \quad \hat{\theta} = \arg\max_{c \in \Theta} M_n(c)$$

with Θ parameter space, **if the following conditions hold**:

1. **Identification**: $\forall \delta > 0, \exists \varepsilon > 0$ s.t. if $\|c - \theta\| > \delta$ then $M(c) < M(\theta) - \varepsilon$
2. **Uniform Convergence**: $sup_{c \in \Theta} \|M_n(c) - M(c)\| \xrightarrow{p} 0$

Then we have the following result:

$$\hat{\theta} \xrightarrow{p} \theta$$

General Consistency Theorem

Identification

Uniform Convergence

Uniform Law of Large Numbers

Let $m : \mathcal{X} \text{ x } \Theta \to \mathbb{R}$, with Θ parameter space, consider the setting:

$$\theta = \arg\max_{c \in \Theta} E[m(x_i; c)] \qquad \text{and} \qquad \hat{\theta} = \arg\max_{c \in \Theta} \frac{1}{n} \sum_{i=1}^{n} m(x_i; c)$$

if the following conditions hold:

1. Parameter space Θ is **compact**
2. Function $m(x_i; c)$ is **continuous** in the second argument
3. $E[sup_{c \in \Theta} \| m(x_i; c) \|] < \infty$ (**envelope condition**)

Then we have the following results

(i) $M_n(c) \xrightarrow{u} M(c)$

(ii) $M(c) = E[m(x_i; c)]$ is continuous.

Exercise (I/IV)

Suppose the observations $\{y_i, x_i\}_{i=1}^n$ are generated independently from the **binary probability model**:

$$y_i = \mathbb{1}\{x_i^T\beta + u_i > 0\}$$

where $u_i \perp x_i$ and $u_i \sim \mathcal{N}(0,1)$. **Questions**:

(1) Prove that $E[y_i|x_i] = \Phi(x_i^T\beta)$

(2) Let $m(x_i;\beta) = E[y_i|x_i]$, consider the following non linear least square estimator: $\hat{\beta}_{NLS} = \arg\max_{c\in\Theta} \; -(1/n)\sum_i [y_i - m(x_i;c)]^2$. Under what conditions is $\hat{\beta}_{NLS}$ consistent for β?

(3) Under the standard normal assumption on u_i, what is the functional form of $m(x_i;c)$? Verify the consistency requirements.

Exercise (II/IV)

Suppose the observations $\{y_i, x_i\}_{i=1}^{n}$ are generated independently from the **binary probability model**:

$$y_i = \mathbb{1}\{x_i^T\beta + u_i > 0\}$$

where $u_i \perp x_i$ and $u_i \sim \mathcal{N}(0,1)$. **Questions**:

(4) Now consider $\Lambda(x_i;\beta) = 1/[1 + exp\{-x_i\beta\}]$, and define the non linear least square estimator: $\tilde{\beta}_{NLS} = \arg\max_{c\in\Theta} \; -(1/n)\sum_i [y_i - \Lambda(x_i;c)]^2$. Where is $\tilde{\beta}_{NLS}$ converging to? In what sense is $\Lambda(x_i;\beta)$ close to $m(x_i;\beta)$?

(5) Next consider the QMLE estimator under the specification that $u_i|x_i \sim \mathcal{L}$ with CDF defined as $F_{U|X}(u|x) = 1/[1 + exp\{-u_i\}] \equiv \Lambda(u)$. Write down the maximization problem and the criterion function.

Exercise (III/IV)

Suppose the observations $\{y_i, x_i\}_{i=1}^n$ are generated independently from the **binary probability model**:

$$y_i = \mathbb{1}\{x_i^T \beta + u_i > 0\}$$

where $u_i \perp x_i$ and $u_i \sim \mathcal{N}(0, 1)$. **Questions**:

(6) Derive the score function, the variance of the score function, and the Hessian, for the value β_0 that maximizes the likelihood function. Note that $\Lambda'(u) = \Lambda(u)(1 - \Lambda(u))$

(7) Is $\hat{\beta}_{QMLE}$ consistent for the parameter that maximize the likelihood function β_0?

(8) Derive the Asymptotic linear representation $\sqrt{n}\left(\hat{\beta}_{QMLE} - \beta_0\right) = \dots$

Exercise (IV/IV)

Suppose the observations $\{y_i, x_i\}_{i=1}^n$ are generated independently from the **binary probability model**:

$$y_i = \mathbb{1}\{x_i^T \beta + u_i > 0\}$$

where $u_i \perp x_i$ and $u_i \sim \mathcal{N}(0, 1)$. **Questions**:

(9) Do you think that the probability limit of $\hat{\beta}_{QMLE}$ is the same as that of $\tilde{\beta}_{NLS}$?

Answer