Econometrics II

(Thur., 8:50-10:20)

Room # 1 (法経講義棟)

- The prerequisites of this class are **Special Lectures in Economics (Statistical Analysis), 経済学特論（統計解析）** (last semester) and **Econometrics I (エコノメトリックス I)** (graduate level, last semester).

Teaching Assistant (TA)

- **Wang Xin (王 馨，オウ ケイ)**

  u459048k [at] ecs.osaka-u.ac.jp

- **Zheng Xuzhu (鄭 旭珠，テイ キョクシュ)**

  u088026j [at] ecs.osaka-u.ac.jp

TA Session

From　Oct. 16, 2023

AM10:30 – 12:00

Room　# 509 （法経研究棟）

If you have any questions, contact TAs.

- **Download the lecture notes from the following websites:**

http://www2.econ.osaka-u.ac.jp/˜tanizaki/class/2023/econome2/

http://stat.econ.osaka-u.ac.jp/˜tanizaki/class/2023/econome2/

Contents

1 Maximum Likelihood Estimation (MLE, 最尤法) — Review

1. We have random variables X_1, X_2, $\cdots$, X_n, which are assumed to be mutually independently and identically distributed.

2. The distribution function of $\{X_i\}_{i=1}^n$ is $f(x;\theta)$, where $x = (x_1, x_2, \cdots, x_n)$ and $\theta = (\mu, \Sigma)$.

   Note that X is a vector of random variables and x is a vector of their realizations (i.e., observed data).

   Likelihood function $L(\cdot)$ is defined as $L(\theta; x) = f(x;\theta)$.

   Note that $f(x;\theta) = \prod_{i=1}^n f(x_i;\theta)$ when X_1, X_2, $\cdots$, X_n are mutually independently and identically distributed.

The maximum likelihood estimator (MLE) of θ is θ such that:

$$\max_{\theta} \; L(\theta; X). \qquad \Longleftrightarrow \qquad \max_{\theta} \; \log L(\theta; X).$$

MLE satisfies the following two conditions:

(a) $\dfrac{\partial \log L(\theta; X)}{\partial \theta} = 0.$

(b) $\dfrac{\partial^2 \log L(\theta; X)}{\partial \theta \partial \theta'}$ is a negative definite matrix.

3. **Fisher's information matrix (フィッシャーの情報行列)** is defined as:

$$I(\theta) = -\mathrm{E}\Big(\frac{\partial^2 \log L(\theta; X)}{\partial \theta \partial \theta'}\Big),$$

where we have the following equality:

$$-\mathrm{E}\Big(\frac{\partial^2 \log L(\theta; X)}{\partial \theta \partial \theta'}\Big) = \mathrm{E}\Big(\frac{\partial \log L(\theta; X)}{\partial \theta}\frac{\partial \log L(\theta; X)}{\partial \theta'}\Big) = \mathrm{V}\Big(\frac{\partial \log L(\theta; X)}{\partial \theta}\Big)$$

Proof of the above equality:

$$\int L(\theta; x) \mathrm{d}x = 1$$

Take a derivative with respect to θ.

$$\int \frac{\partial L(\theta; x)}{\partial \theta} \mathrm{d}x = 0$$

(We assume that (i) the domain of x does not depend on θ and (ii) the derivative $\dfrac{\partial L(\theta; x)}{\partial \theta}$ exists.)

Rewriting the above equation, we obtain:

$$\int \frac{\partial \log L(\theta; x)}{\partial \theta} L(\theta; x) \mathrm{d}x = 0,$$

i.e.,

$$\mathrm{E}\left(\frac{\partial \log L(\theta; X)}{\partial \theta} \right) = 0.$$

Again, differentiating the above with respect to θ, we obtain:

$$\begin{aligned}
&\int \frac{\partial^2 \log L(\theta; x)}{\partial\theta\partial\theta'} L(\theta; x)\mathrm{d}x + \int \frac{\partial \log L(\theta; x)}{\partial\theta} \frac{\partial L(\theta; x)}{\partial'\theta}\mathrm{d}x \\
&\qquad = \int \frac{\partial^2 \log L(\theta; x)}{\partial\theta\partial\theta'} L(\theta; x)\mathrm{d}x + \int \frac{\partial \log L(\theta; x)}{\partial\theta} \frac{\partial \log L(\theta; x)}{\partial\theta'} L(\theta; x)\mathrm{d}x \\
&\qquad = \mathrm{E}\Big(\frac{\partial^2 \log L(\theta; X)}{\partial\theta\partial\theta'}\Big) + \mathrm{E}\Big(\frac{\partial \log L(\theta; X)}{\partial\theta} \frac{\partial \log L(\theta; X)}{\partial\theta'}\Big) = 0.
\end{aligned}$$

Therefore, we can derive the following equality:

$$-\mathrm{E}\left(\frac{\partial^2 \log L(\theta; X)}{\partial\theta\partial\theta'}\right) = \mathrm{E}\left(\frac{\partial \log L(\theta; X)}{\partial\theta} \frac{\partial \log L(\theta; X)}{\partial\theta'}\right) = \mathrm{V}\left(\frac{\partial \log L(\theta; X)}{\partial\theta}\right),$$

where the second equality utilizes $\mathrm{E}\left(\frac{\partial \log L(\theta; X)}{\partial\theta}\right) = 0$.

4. **Cramer-Rao Lower Bound (クラメール・ラオの下限)**: $(I(\theta))^{-1}$

Suppose that an unbiased estimator of θ is given by $s(X)$.

Then, we have the following:

$$\mathrm{V}(s(X)) \geq (I(\theta))^{-1}$$

Proof:

The expectation of $s(X)$ is:

$$\mathrm{E}(s(X)) = \int s(x)L(\theta; x)\mathrm{d}x.$$

Differentiating the above with respect to θ,

$$\begin{aligned}
\frac{\partial \mathrm{E}(s(X))}{\partial \theta'} &= \int s(x)\frac{\partial L(\theta; x)}{\partial \theta'}\mathrm{d}x = \int s(x)\frac{\partial \log L(\theta; x)}{\partial \theta'}L(\theta; x)\mathrm{d}x \\
&= \mathrm{Cov}\left(s(X), \frac{\partial \log L(\theta; X)}{\partial \theta}\right)
\end{aligned}$$

For simplicity, let $s(X)$ and θ be scalars.

Then,

$$\begin{aligned}\left(\frac{\partial \mathrm{E}(s(X))}{\partial\theta}\right)^2 &= \left(\mathrm{Cov}\left(s(X), \frac{\partial \log L(\theta; X)}{\partial\theta}\right)\right)^2 = \rho^2 \mathrm{V}\left(s(X)\right) \mathrm{V}\left(\frac{\partial \log L(\theta; X)}{\partial\theta}\right) \\ &\leq \mathrm{V}\left(s(X)\right) \mathrm{V}\left(\frac{\partial \log L(\theta; X)}{\partial\theta}\right),\end{aligned}$$

where ρ denotes the correlation coefficient between $s(X)$ and $\dfrac{\partial \log L(\theta; X)}{\partial\theta}$, i.e.,

$$\rho = \frac{\mathrm{Cov}\left(s(X), \dfrac{\partial \log L(\theta; X)}{\partial\theta}\right)}{\sqrt{\mathrm{V}\left(s(X)\right)}\sqrt{\mathrm{V}\left(\dfrac{\partial \log L(\theta; X)}{\partial\theta}\right)}}.$$

Note that $|\rho| \leq 1$.

Therefore, we have the following inequality:

$$\left(\frac{\partial \mathrm{E}(s(X))}{\partial \theta}\right)^2 \leq \mathrm{V}(s(X))\ \mathrm{V}\left(\frac{\partial \log L(\theta; X)}{\partial \theta}\right),$$

i.e.,

$$\mathrm{V}(s(X)) \geq \frac{\left(\dfrac{\partial \mathrm{E}(s(X))}{\partial \theta}\right)^2}{\mathrm{V}\left(\dfrac{\partial \log L(\theta; X)}{\partial \theta}\right)}$$

Especially, when $\mathrm{E}(s(X)) = \theta$,

$$\mathrm{V}(s(X)) \geq \frac{1}{-\mathrm{E}\left(\dfrac{\partial^2 \log L(\theta; X)}{\partial \theta^2}\right)} = (I(\theta))^{-1}.$$

Even in the case where $s(X)$ is a vector, the following inequality holds.

$$\mathrm{V}(s(X)) \geq (I(\theta))^{-1},$$

where $I(\theta)$ is defined as:

$$
\begin{aligned}
I(\theta) &= -\mathrm{E}\left(\frac{\partial^2 \log L(\theta; X)}{\partial\theta\partial\theta'}\right) \\
&= \mathrm{E}\left(\frac{\partial \log L(\theta; X)}{\partial\theta}\frac{\partial \log L(\theta; X)}{\partial\theta'}\right) = \mathrm{V}\left(\frac{\partial \log L(\theta; X)}{\partial\theta}\right).
\end{aligned}
$$

The variance of any unbiased estimator of θ is larger than or equal to $(I(\theta))^{-1}$.

5. Asymptotic Normality of MLE:

Let $\tilde{\theta}$ be MLE of θ.

As n goes to infinity, we have the following result:

$$\sqrt{n}(\tilde{\theta} - \theta) \longrightarrow N\left(0, \lim_{n\to\infty}\left(\frac{I(\theta)}{n}\right)^{-1}\right),$$

where it is assumed that $\lim_{n\to\infty}\left(\frac{I(\theta)}{n}\right)$ converges.

That is, when n is large, $\tilde{\theta}$ is approximately distributed as follows:

$$\tilde{\theta} \sim N\left(\theta, (I(\theta))^{-1}\right).$$

Suppose that $s(X) = \tilde{\theta}$.

When n is large, $\mathrm{V}(s(X))$ is approximately equal to $(I(\theta))^{-1}$.

Practically, we utilize the following approximated distribution:

$$\tilde{\theta} \sim N\left(\theta, (I(\tilde{\theta}))^{-1}\right).$$

Then, we can obtain the significance test and the confidence interval for θ

6. **Central Limit Theorem:** Let X_1, X_2, $\cdots$, X_n be mutually independently distributed random variables with mean $\mathrm{E}(X_i) = \mu$ and variance $\mathrm{V}(X_i) = \sigma^2 < \infty$ for $i = 1, 2, \cdots, n$.

   Define $\overline{X} = (1/n) \sum_{i=1}^{n} X_i$.

   Then, the central limit theorem is given by:

$$\frac{\overline{X} - \mathrm{E}(\overline{X})}{\sqrt{\mathrm{V}(\overline{X})}} = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \longrightarrow N(0, 1).$$

   Note that $\mathrm{E}(\overline{X}) = \mu$ and $\mathrm{V}(\overline{X}) = \sigma^2/n$.

That is,

$$\sqrt{n}(\overline{X}-\mu) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-\mu) \longrightarrow N(0,\sigma^2).$$

Note that $\mathrm{E}(\overline{X}) = \mu$ and $n\mathrm{V}(\overline{X}) = \sigma^2$.

In the case where X_i is a vector of random variable with mean μ and variance $\Sigma < \infty$, the central limit theorem is given by:

$$\sqrt{n}(\overline{X}-\mu) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-\mu) \longrightarrow N(0,\Sigma).$$

Note that $\mathrm{E}(\overline{X}) = \mu$ and $n\mathrm{V}(\overline{X}) = \Sigma$.