Econometrics II

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

Room # 4 (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).

TA Session (by Yonekura and Miura):

From	**Oct. 7(?), 2015**
Wed.,	**14:40 - 16:10**
Room	**# 4 (**法経講義棟**)**

Web site

`https://sites.google.com/site/ougseeconometricsi/`

Statistics Test (統計検定) on Nov. 29 (Sun.)

- **Exams：** Level 1 (1 級) – Level 4 (4 級)

  Note that Level 4 is Junior high school level,

  Level 3 is High school level, and

  Level 2 is the 1st or 2nd year statistics in undergraduate school.

  Level 1 is the 3rd or 4th year statistics in undergraduate school (or the 1st year in graduate school).

  See **http://www.toukei-kentei.jp/** in more detail.

- **Qualification for Exam (受験資格)：**

  Undergraduate and Graduate Students in Osaka University

- **Application Period (受験申込期間)：** September 9 (Wed.), AM10:00 — October 14 (Wed.), PM15:00

  Go to **http://qajss.org/jinse/kentei201511.html** for application.

- **Application Fee (受験料)：** Free

受験料は，平成 24 年度に採択された文部科学省の大学間連携共同推進事業「データに基づく課題解決型人材育成に資する統計教育質保証」から支払われる。

連携校： 東京大学，大阪大学，総合研究大学院大学，青山学院大学（代表校），多摩大学， 立教大学，早稲田大学，同志社大学

ちなみに、連携大学以外の人の受験料は，

1級「統計数理」	10:30～12:00	6,000円
1級「統計応用」	13:30～15:00	6,000円
2級	10:30～12:00	5,000円
3級	13:30～14:30	4,000円
4級	10:30～11:30	3,000円
統計調査士	13:00～14:30	5,000円
専門統計調査士	10:30～12:00	10,000円

となる。ただし，1級「統計数理」と「統計応用」の両方受験の場合，受験料は10,000円となる。

- **Exam Date (試験日)：** Nov. 29 (Sun.)
- **Exam Place (場所)：** 人間科学研究科
  本館（44講義室）・東館（303，404講義室）

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:

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

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

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:

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

Proof:

The expectation of $s(X)$ is:

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

Differentiating the above with respect to θ,

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

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

Then,

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

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.,

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

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

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

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 $n\mathrm{V}(\overline{X}) = \Sigma$.

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

   Assume:

$$\sigma^2 = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2 < \infty.$$

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

   The central limit theorem is given by:

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

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

In the case where X_i is a vector of random variable with mean μ and variance Σ_i, the central limit theorem is given by:

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

where $\Sigma = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \Sigma_i < \infty$.

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