Econometrics II TA Session #02*

Kenta KUDO †

October 15th, 2019

Contents

* All comments welcome!

† E-mail: vge011kk@student.econ.osaka-u.ac.jp

1 Preliminary

Today, we review the introductory topics of the maximum likelihood estimation and examples of the estimation.

2 Maximum Likelihood Estimator

Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. random variables with common probability density function $f(x;\theta)$. For now, assume that θ is an unknown vector parameter. The joint density of these i.i.d. observations obtained from this process is

$$f(x_1, x_2, \cdots, x_n|\theta) = \prod_{i=1}^{n} f(x_i;\theta) =: L(\theta; x). \tag{1}$$

We then have, by taking the logarithm, the following equation:

$$\log L(\theta; x) := \sum_{i=1}^{n} \log f(x_i;\theta). \tag{2}$$

This function is called **log likelihood function** of X.

2.1 Definition of Maximum Likelihood Estimator (MLE)

The definition of the maximum likelihood estimator (MLE) is given by as follows.

Definition 2.1 (Maximum Likelihood Estimator (MLE))**.** The maximum likelihood estimator (MLE), denoted by $\hat{\theta}$, maximizes the likelihood function. In other words, MLE satisfies the following conditions.

$$\frac{\partial \log L(\theta; x)}{\partial \theta}\Big|_{\theta=\hat{\theta}} = \mathbf{0};$$
$$\frac{\partial^2 \log L(\theta; x)}{\partial \theta \partial \theta'}\Big|_{\theta=\hat{\theta}} \prec \mathbf{0}.$$

In short, we can say

$$\log L(\hat{\theta}) \geq \log L(\theta)$$

is satisfied for any $\theta \in \Theta$ where Θ represents the set of all estimators obtained from the log likelihood function. Note that $\hat{\theta}$ also maximizes the likelihood function since the log function is an increasing function.

2.2 Fisher's information matrix

Assume that the log likelihood function is continuously twice differentiable and the integral of the log likelihood function is also continuously differentiated twice.

Definition 2.2. Fisher's information matrix is defined as

$$I(\theta) := -\mathbb{E}\left[\frac{\partial^2 \log L(\theta; X)}{\partial\theta\partial\theta'}\right] = \mathrm{Var}\left[\frac{\partial \log L(\theta; X)}{\partial\theta}\right].$$

Proof. We begin with the identity

$$\int L(\theta; x)dx = 1. \tag{3}$$

Take the derivative of both sides of Eq. (3) with respect to $\theta \in \mathbb{R}^{k\times 1}$, we have

$$\frac{\partial}{\partial\theta}\int L(\theta; x)dx = 0.$$

By changing the order of the integral, the above equation can be rewritten as

$$\int \frac{\partial}{\partial\theta} L(\theta; x)dx = 0.$$

This relationship can be rewritten as

$$\int \frac{\partial \log L(\theta; x)}{\partial\theta} L(\theta; x)dx = 0. \tag{4}$$

via the derivative of log function: $\frac{d}{dx}\log(x) = \frac{1}{x}$ for $x \in \mathbb{R}_{++} := (0, \infty)$. Writing the above equation as an expectation, we obtain

$$\mathbb{E}\left[\frac{\partial \log L(\theta; X)}{\partial\theta}\right] = 0. \tag{5}$$

Note that $L(\theta; x)$ is a probability density function and $\int g(x)L(\theta; x)dx = \mathbb{E}[g(X)]$. Again, defferentiating Eq. (4) with respect to $\theta' \in \mathbb{R}^{1\times k}$, we can derive

$$\int \frac{\partial^2 \log L(\theta; x)}{\partial\theta\partial\theta'} L(\theta; x)dx + \underbrace{\int \frac{\partial \log L(\theta; x)}{\partial\theta}\frac{\partial \log L(\theta; x)}{\partial\theta'} L(\theta; x)dx}_{I(\theta)} = 0.$$

Finally, we have

$$I(\theta) := -\mathbb{E}\left[\frac{\partial^2 \log L(\theta; X)}{\partial\theta\partial\theta'}\right] = \mathrm{Var}\left[\frac{\partial \log L(\theta; X)}{\partial\theta}\right],$$

because of Eq. (5). □

2.3 The Cramér–Rao Lower Bound

In this subsection, we establish a remarkable inequality called the **Cramér–Rao lower bound** which gives a lower bound on the variance of any unbiased estimator.

Theorem 2.3 (Cramér–Rao Lower Bound)**.** Suppose that $s(X)$ is a unbiased estimator of θ (i.e. $\mathbb{E}[s(X)] = \theta$), then we have the following inequality:

$$\mathrm{Var}[s(X)] \geq I(\theta)^{-1}. \tag{6}$$

Proof. For simplicity, let θ and $s(X)$ be scalar. First, taking the expectation of $s(X)$, we have

$$\mathbb{E}[s(X)] = \int s(x)L(\theta; x)dx.$$

By taking the derivative of $\mathbb{E}[s(X)]$ with respect to $\theta \in \mathbb{R}$, the following equalities hold:

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

thanks for the following relations: since $\mathbb{E}\left[\frac{d\log L(\theta;x)}{d\theta}\right] = 0$,

$$\begin{aligned}
\mathrm{Cov}\left(s(X), \frac{d\log L(\theta; x)}{d\theta}\right) &= \mathbb{E}\left[s(X)\frac{d\log L(\theta; X)}{d\theta}\right] - \mathbb{E}[s(X)]\mathbb{E}\left[\frac{d\log L(\theta; X)}{d\theta}\right] \\
&= \mathbb{E}\left[s(X)\frac{d\log L(\theta; X)}{d\theta}\right].
\end{aligned}$$

Recall that $s(X)$ is a unbiased estimator of θ, so that $\mathbb{E}[s(X)] = \theta$, and thereby

$$1 = \mathrm{Cov}\left(s(X), \frac{d\log L(\theta; X)}{d\theta}\right)$$

Remind that we have

$$\begin{aligned}
-1 \leq \frac{\mathrm{Cov}\left(s(X), \frac{d\log L(\theta;X)}{d\theta}\right)}{\sqrt{\mathrm{Var}[s(X)]}\sqrt{\mathrm{Var}\left[\frac{d\log L(\theta;X)}{d\theta}\right]}} &\leq 1 \\
\iff -1 \leq \frac{1}{\sqrt{\mathrm{Var}[s(X)]}\sqrt{\mathrm{Var}\left[\frac{d\log L(\theta;X)}{d\theta}\right]}} &\leq 1,
\end{aligned}$$

Therefore, we can derive the following inequality:

$$\mathrm{Var}[s(X)] \geq V\left[\frac{d\log L(\theta; X)}{d\theta}\right]^{-1} = I(\theta)^{-1}.$$

The similar derivation yields the same inequality for the multivariate case. □

2.4 Asymptotic Distribution of MLE

The MLE has asymptotic normality as stated in the following theorem.

Theorem 2.4 (Asymptotic Distribution of MLE)**.** Suppose that $\hat{\theta}$ is the MLE and θ is the true value of the parameter. Then, the asymptotic distribution of the MLE is represented as follows:

$$\sqrt{n}(\hat{\theta} - \theta) \to N\left(0, \Sigma^{-1}\right), \tag{7}$$

where $\frac{1}{n} I(\theta) \to \Sigma$ as $n \to \infty$.

Proof. By the first–order approximation of $\frac{\partial \log L(\hat{\theta};x)}{\partial \theta} = 0$ around $\hat{\theta} = \theta$ by the Taylor expansion, we have

$$\frac{\partial \log L(\theta; x)}{\partial \theta} + \frac{\partial^2 \log L(\theta; x)}{\partial \theta \partial \theta'}(\hat{\theta} - \theta) = 0.$$

Rewriting the above equation, we establish the following equation

$$\sqrt{n}(\hat{\theta} - \theta) = \left(-\frac{1}{n}\frac{\partial^2 \log L(\theta; x)}{\partial \theta \partial \theta'}\right)^{-1} \frac{1}{\sqrt{n}} \frac{\partial \log L(\theta; x)}{\partial \theta}. \tag{8}$$

Here, by applying the following **Lindeberg–Feller Central Limit Theorem (Lindeberg–Feller CLT)**, we can derive the asymptotic distribution of MLE.

Theorem 2.5 (Lindeberg–Feller Central Limit Theorem for a Multivariate Random Variable)**.** In the case where $X_i \in \mathbb{R}^k$ is a vector of random variable with mean $\mu \in \mathbb{R}^k$ and variance $\Sigma_i \in \mathbb{R}^k$, the Lindeberg–Feller CLT is given by

$$\sqrt{n}(\bar{X} - \mu) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \xrightarrow{d} N(0, \Sigma), \tag{9}$$

where

$$\frac{1}{n}\sum_{i=1}^{n} X_i =: \overline{X}; \quad \lim_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Sigma_i = \Sigma < \infty. \tag{10}$$

Note that $\mathbb{E}(\bar{X}) = \mu$ and $n\mathrm{Var}(\bar{X}) \to \Sigma$ as n goes to infinity.

In this case, remind that we need the following expectation and variance:

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \frac{\partial \log f(X_i; \theta)}{\partial \theta}\right]; \tag{11}$$

$$\mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} \frac{\partial \log f(X_i; \theta)}{\partial \theta}\right], \tag{12}$$

where

$$\sum_{i=1}^{n} \frac{\partial \log f(X_i;\theta)}{\partial \theta} = \frac{\partial \log L(\theta;x)}{\partial \theta}.$$

In addition, define the variance of $\frac{\partial \log f(X_i;\theta)}{\partial \theta}$ as Σ_i, then we can say $I(\theta) = \sum_{i=1}^{n} \Sigma_i$ in the case that all X_is are mutually independent. Note also that

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

Moreover, $n\mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} \frac{\partial \log L(\theta;X_i)}{\partial \theta}\right] = \frac{1}{n} I(\theta) \to \Sigma$ as $n \to \infty$.
In Eq. (8), we can calculate

$$\frac{1}{n}\frac{\partial^2 \log L(\theta;x)}{\partial \theta \partial \theta'} \xrightarrow{p} \frac{1}{n}\mathbb{E}\left[\frac{\partial^2 \log L(\theta;X)}{\partial \theta \partial \theta'}\right]; \tag{13}$$
$$\frac{1}{\sqrt{n}}\frac{\partial \log L(\theta;x)}{\partial \theta} \xrightarrow{d} N(0,\Sigma).$$

Recall that we use the **Weak Law of Large Numbers** in Eq. (13) and $\frac{1}{n} I(\theta) \to \Sigma$ as $n \to \infty$. Therefore, we can derive the asymptotic distribution by the **Slutsky's theorem** as follows:

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \Sigma^{-1}).$$

□

2.5 Example of the ML Method

The following discussion is explained in Chapter14, Example 14.2 & 14.3 of Greene (2012). Suppose the case that $X_i \sim N(\mu, \sigma^2)$ for $i \in \{1, \ldots, n\}$. The likelihood of the each observed variable x_i $(i = 1, 2, \cdots, n)$ is given by

$$L(\theta; x_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x_i - \mu)^2}{2\sigma^2}\right\},$$

Here, we assume that the parameter vector is $\theta = (\mu, \sigma^2)$. By taking the logarithm, the above equation is rewritten as follows:

$$\log L(\theta; x_i) = -\frac{1}{2}\log 2\pi - \log\sigma - \frac{(x_i - \mu)^2}{2\sigma^2}.$$

Recall that we must optimize $\sum_{i=1}^{n} \log L(\theta; x_i)$ such that:

$$\sum_{i=1}^{n} \log L(x_i;\theta) = (constant) - n\log\sigma - \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2}.$$

Therefore, when we estimate μ, the first order condition is given as follows:

$$\frac{d\sum_{i=1}^{n}(x_i-\mu)^2}{d\mu} = -2\sum_{i=1}^{n}(x_i-\mu) = 0,$$

and $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$, which coincides with the OLS estimator. In the same manner, we have an estimator of the variance as

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2.$$

Note that the MLE of the variance is not the same as the OLS estimator and therefore this is not an unbiased estimator (or this estimator is a biased one). The second order conditions are:

$$\begin{aligned}
\frac{d^2 \log L(x;\theta)}{d\mu d\mu} &= -\frac{n}{\sigma^2};\\
\frac{d^2 \log L(x;\theta)}{d\sigma^2 d\sigma^2} &= \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}(x_i-\mu)^2;\\
\frac{d^2 \log L(x;\theta)}{d\mu d\sigma^2} &= -\frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i-\mu).
\end{aligned}$$

By deriving the second order conditions, we have the informaton matrix as follows:

$$\left(\mathbb{E}\left[\frac{\partial^2 \log L(x;\theta)}{\partial\theta\partial\theta'}\right]\right)^{-1} = \begin{pmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/n \end{pmatrix}.$$

References

[1] Greene, W. H. (2012) *"Econometric analysis Seventh Edition",* Pearson.