Econometrics II TA Session #2

Royou Kiku

October 20th, 2021

Contents

1.Maximum Likelihood Estimator

We assume that random variables X_1, X_2,..., X_n are mutually independent and identically distributed. The distribution function of $\{X_i\}_{i=1}^{n}$ is $f(x;\theta)$, where $x=(x_1,x_2,\dots,x_n)$ and $\theta=(\mu;\Sigma)$.

The joint density function of $x_1,x_2,\dots,x_n$ is

$$f(x_1,x_2,\dots,x_n|\theta)=\prod_{i=1}^{n}f(x_i\,;\theta) \tag{1}$$

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

$$\prod_{i=1}^{n}f(x_i\,;\theta)=L(\theta;x_1,x_2,\dots,x_n) \tag{2}$$

The log likelihood function is

$$logL(\theta;x_1,x_2,\dots,x_n)=\sum_{i=1}^{n}logf(x_i;\theta) \tag{3}$$

1.1Definition of MLE

$MLE(\hat{\theta})$ maximize the likelihood function. MLE satisfies the following condition

$$\frac{\partial logL(\theta;x)}{\partial\theta}=0 \tag{4}$$

$$\frac{\partial^2 logL(\theta;x)}{\partial\theta\partial\theta'}\,is\ negative\ definite\ matrix. \tag{5}$$

1.2Fisher's information matrix

$$I(\theta)=-E\left(\frac{\partial^2 logL(\theta;x)}{\partial\theta\partial\theta'}\right)=V\left(\frac{\partial logL(\theta;x)}{\partial\theta}\right) \tag{6}$$

Before starting the proof:

$$\int \dots \int L(\theta; x_1, x_2, \dots, x_n) dx_1 .. dx_n = 1 \tag{7}$$

$$\int L(\theta; x) dx = 1 \tag{8}$$

We assume $\frac{\partial}{\partial\theta} \int L(\theta; x) dx = \int \frac{\partial}{\partial\theta} L(\theta; x) dx$

Then, we prove it from

$$\frac{\partial}{\partial\theta} \int L(\theta; x) dx = 0 \tag{9}$$

$$\int \frac{\partial}{\partial\theta} L(\theta; x) dx = 0 \tag{10}$$

This equation can be rewritten as

$$\int \frac{\partial \log L(\theta; x)}{\partial\theta} L(\theta; x) dx = 0 \tag{11}$$

Differentiation by θ

$$\int \frac{\partial^2 \log L(\theta; x)}{\partial\theta\partial\theta'} L(\theta; x) dx + \int \frac{\partial \log L(\theta; x)}{\partial\theta} \frac{\partial L(\theta; x)}{\partial\theta'} dx = 0 \tag{12}$$

$$\int \frac{\partial^2 \log L(\theta; x)}{\partial\theta\partial\theta'} L(\theta; x) dx + \int \frac{\partial \log L(\theta; x)}{\partial\theta} \frac{\partial log L(\theta; x)}{\partial\theta'} L(\theta; x) dx = 0 \tag{13}$$

$L(\theta; x)$ is a probability density function and $\int g(x) L(\theta; x) dx = E[g(x)]$

$$\left(\frac{\partial^2 \log L(\theta; x)}{\partial\theta\partial\theta'} \right) + E \left(\frac{\partial \log L(\theta; x)}{\partial\theta} \frac{\partial log(\theta; x)}{\partial\theta'} \right) = 0 \tag{14}$$

Therefore,

$$-E \left(\frac{\partial^2 \log L(\theta; x)}{\partial\theta\partial\theta'} \right) = E \left(\frac{\partial \log L(\theta; x)}{\partial\theta} \frac{\partial log(\theta; x)}{\partial\theta'} \right) = V \left(\frac{\partial \log L(\theta; x)}{\partial\theta} \right) \tag{15}$$

1.3The Cramér-Rao Lower Bound

Suppose that $s(x)$ is an unbiased estimator of θ, then we have the following:

$$Var[s(x)] \geq I(\theta)^{-1} \tag{16}$$

Proof:
Taking the expectation of s(x)

$$E[s(x)] = \theta \tag{17}$$

$$E[s(x)] = \int s(x)L(\theta; x)dx \tag{18}$$

Differentiation by θ

$$\begin{aligned} \frac{\partial E[s(x)]}{\partial \theta} &= \int s(x)\frac{\partial L(\theta; x)}{\partial \theta}dx = \int s(x)\frac{\partial \log L(\theta; x)}{\partial \theta}L(\theta; x)dx \\ &= E\left(s(x)\frac{\partial logL(\theta; x)}{\partial \theta}\right) = Cov\left(s(x), \frac{\partial logL(\theta; x)}{\partial \theta}\right) \end{aligned} \tag{19}$$

Since $E\left[\frac{\partial logL(\theta;x)}{\partial \theta}\right] = 0$,we can obtain the following equation

$$\begin{aligned} Cov\left(s(x), \frac{\partial logL(\theta; x)}{\partial \theta}\right) &= E\left(s(x)\frac{\partial logL(\theta; x)}{\partial \theta}\right) - E[s(x)]E\left(\frac{\partial logL(\theta; x)}{\partial \theta}\right) \\ &= E\left(s(x)\frac{\partial logL(\theta; x)}{\partial \theta}\right) \end{aligned} \tag{20}$$

For simplicity, let θ and s(x) be scalars

$$\left(\frac{\partial E(s(x))}{\partial \theta}\right)^2 = \left(Cov\left(s(x), \frac{\partial logL(\theta; x)}{\partial \theta}\right)\right)^2 \tag{21}$$

$$= \frac{\left(Cov\left(s(x), \frac{\partial logL(\theta; x)}{\partial \theta}\right)\right)^2}{\left(\sqrt{V(s(x))}\sqrt{V\left(\frac{\partial logL(\theta; x)}{\partial \theta}\right)}\right)^2} V(s(x)V\left(\frac{\partial logL(\theta; x)}{\partial \theta}\right)$$

$$= \rho^2 V(s(x)V\left(\frac{\partial logL(\theta; x)}{\partial \theta}\right)$$

Since ρ denotes the correlation coefficient, $|\rho| \leq 1$

$$\rho^2 V(s(x)V\left(\frac{\partial logL(\theta;x)}{\partial\theta}\right) \leq V(s(x)V\left(\frac{\partial logL(\theta;x)}{\partial\theta}\right) \tag{22}$$

$$\left(\frac{\partial E(s(x))}{\partial\theta}\right)^2 \leq V(s(x)V\left(\frac{\partial logL(\theta;x)}{\partial\theta}\right) \tag{23}$$

Then, when $E(s(x)) = \theta$,we get

$$V(s(x)) \geq \frac{\left(\frac{\partial E(s(x))}{\partial\theta}\right)^2}{V\left(\frac{\partial logL(\theta;x)}{\partial\theta}\right)} = \frac{1}{V\left(\frac{\partial logL(\theta;x)}{\partial\theta}\right)} = (I(\theta))^{-1} \tag{24}$$

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

1.4Asymptotic Normality of MLE

Review1: Central Limit Theorem

We assume that $X_1, X_2, \dots, X_n$ are iid with $E(X_i) = \mu\ and\ V(X_i) = \sigma^2 < \infty\ for\ i = 1,2,\dots,n$. Define $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$,

Then,

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

Which implies

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

Review2: Central Limit Theorem II

We assume that X_i is iid and a vector of random variable with $E(X_i) = \mu\ and\ V(X_i) = \Sigma_i$ $for\ i = 1,2,\dots,n.$

Then,

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

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

Review3: Asymptotic Theories

Convergence in Probability: $X_n \longrightarrow a$
X converges in probability to a.
Convergence in Distribution: $X_n \longrightarrow X$
The distribution of X_n converges to the distribution of X as n goes to infinity.

Review4: Weak Law of Large Numbers

. We assume that $X_1, X_2, \dots, X_n$ are iid with $E(X_i) = \mu \; and \; V(X_i) = \sigma^2 < \infty \; for \; i = 1,2, \dots, n.$
Then, $\bar{X} \longrightarrow \mu$ as n $\longrightarrow \infty$ which is called the week law of large numbers.

Proof: Asymptotic Normality of MLE

Suppose that $\hat{\theta}$ is the MLE and θ is the true value of the parameter,

$$\max \quad logL(\theta; x) \tag{25}$$

$$\frac{\partial logL(\theta; x)}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial logf(x_i; \theta)}{\partial \theta} = 0 \tag{26}$$

When we are using the Central Limit Theorem II, we use X_i instead of x_i.

$$\sqrt{n}(\bar{X} - \mu) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \quad \longrightarrow N(0, \Sigma) \tag{27}$$

Where $\frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}, \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \Sigma_i = \sum < \infty.$

$$E\left(\frac{1}{n}\frac{\partial logL(\theta; X)}{\partial \theta}\right) = 0 \tag{28}$$

$$V\left(\frac{1}{n}\frac{\partial logL(\theta; X)}{\partial \theta}\right) = \frac{1}{n^2} I(\theta) \tag{29}$$

Thus, the asymptotic distribution of $\frac{1}{n}\frac{\partial logL(\theta;X)}{\partial\theta}$ is given by:

$$\sqrt{n}\left(\frac{1}{n}\frac{\partial logL(\theta;X)}{\partial\theta} - E\left(\frac{1}{n}\frac{\partial logL(\theta;X)}{\partial\theta}\right)\right)$$
$$= \frac{1}{\sqrt{n}}\frac{\partial logL(\theta;X)}{\partial\theta} \longrightarrow N(0,\Sigma) \qquad (30)$$

Where $nV\left(\frac{1}{n}\frac{\partial logL(\theta;X)}{\partial\theta}\right) = \frac{1}{n}V\left(\frac{\partial logL(\theta;X)}{\partial\theta}\right) = \frac{1}{n}I(\theta) \rightarrow \Sigma$

Consider the asymptotic distribution of

$$\frac{1}{\sqrt{n}}\frac{\partial logL(\hat{\theta};X)}{\partial\theta}, \qquad (31)$$

Which is expanded around $\hat{\theta} = \theta$ as follows:

$$0 = \frac{1}{\sqrt{n}}\frac{\partial logL(\hat{\theta};X)}{\partial\theta} \approx \frac{1}{\sqrt{n}}\frac{\partial logL(\theta;X)}{\partial\theta} + \frac{1}{\sqrt{n}}\frac{\partial^2 logL(\theta;X)}{\partial\theta\partial\theta'}(\hat{\theta} - \theta) \qquad (32)$$

$$-\frac{1}{\sqrt{n}}\frac{\partial^2 logL(\theta;X)}{\partial\theta\partial\theta'}(\hat{\theta} - \theta) \approx \frac{1}{\sqrt{n}}\frac{\partial logL(\theta;X)}{\partial\theta} \longrightarrow N(0,\Sigma) \qquad (33)$$

$$-\frac{1}{\sqrt{n}}\frac{\partial^2 logL(\theta;X)}{\partial\theta\partial\theta'}(\hat{\theta} - \theta) = \sqrt{n}\left(-\frac{1}{n}\frac{\partial^2 logL(\theta;X)}{\partial\theta\partial\theta'}\right)(\hat{\theta} - \theta) \qquad (34)$$

$$\sqrt{n}\left(-\frac{1}{n}\frac{\partial^2 logL(\theta;X)}{\partial\theta\partial\theta'}\right)(\hat{\theta} - \theta) = \frac{1}{\sqrt{n}}\frac{\partial logL(\theta;X)}{\partial\theta} \qquad (35)$$

$$\sqrt{n}(\hat{\theta} - \theta) \approx \left(-\frac{1}{n}\frac{\partial^2 logL(\theta;X)}{\partial\theta\partial\theta'}\right)^{-1}\left(\frac{1}{\sqrt{n}}\frac{\partial logL(\theta;X)}{\partial\theta}\right) \qquad (36)$$

$$\longrightarrow N(0,\Sigma^{-1}\Sigma\Sigma^{-1}) = N(0,\Sigma^{-1}) \qquad (37)$$

Using the law of large number

$$-\frac{1}{n}\frac{\partial^2 logL(\theta;X)}{\partial\theta\partial\theta'} \longrightarrow \lim_{n\to\infty}\frac{1}{n}\left(-E\left(\frac{\partial^2 logL(\theta;X)}{\partial\theta\partial\theta'}\right)\right) \qquad (38)$$
$$= \lim_{n\to\infty}\frac{1}{n}\left(V\left(\frac{\partial logL(\theta;X)}{\partial\theta}\right)\right) = \lim_{n\to\infty}\frac{1}{n}I(\theta) = \Sigma$$

1.5Optimization

Optimization procedure

$$0 = \frac{\partial logL(\theta; x)}{\partial \theta} \approx \frac{\partial logL(\theta^*; x)}{\partial \theta} + \frac{\partial^2 logL(\theta^*; x)}{\partial \theta \partial \theta'}(\theta - \theta^*) \tag{39}$$

$$\theta = \theta^* - \left(\frac{\partial^2 logL(\theta^*; x)}{\partial \theta \partial \theta'}\right)^{-1} \frac{\partial logL(\theta^*; x)}{\partial \theta} \tag{40}$$

Newton-Raphson method

$$\theta^{(i+1)} = \theta^{(i)} - \left(\frac{\partial^2 logL\left(\theta^{(i)}; x\right)}{\partial \theta \partial \theta'}\right)^{-1} \frac{\partial logL(\theta^{(i)}; x)}{\partial \theta} \tag{41}$$

Method of Scoring

$$\theta^{(i+1)} = \theta^{(i)} - \left(E\left[\frac{\partial^2 logL\left(\theta^{(i)}; x\right)}{\partial \theta \partial \theta'}\right]\right)^{-1} \frac{\partial logL(\theta^{(i)}; x)}{\partial \theta} \tag{42}$$

Note that

$$I(\theta) = -E\left[\frac{\partial^2 logL(\theta; x)}{\partial \theta \partial \theta'}\right] \tag{43}$$

is the Fisher information matrix.