Econometrics I
TA Session 12

Jukina HATAKEYAMA *

July 2, 2024

Contents

*E-mail: u868710a@ecs.osaka-u.ac.jp

1 Review of the Asymptotic Theory

We review the central limit theorems and the asymptotic properties of the OLSE. At first, we will briefly review of the multivariate central limit theorem. The discussion of this topic is given in Chapter3 of the main textbook and the Appendix D of Greene(2011).

Theorem 1.1. Let $\{w_i : i = 1, 2, \dots\}$ be a sequence of i.i.d. and $G \times 1$ random vectors such that $\mathbb{E}(w_{ig}^2) < \infty$, $g = 1, 2, \dots, G$, and $\mathbb{E}(w_i) = \mu$. Then, $\{w_i : i = 1, 2, \dots\}$ satisfies the Lindeberg-Levy central limit theorem:

$$\sqrt{n}(\bar{w}_i - \mu) \xrightarrow[n\to\infty]{d} N(0, Q), \tag{1}$$

where $\bar{w}_i$ represents the sample level mean of w_i and $Q = \mathrm{Var}(w_i) = \mathbb{E}(w_i w_i')$ is positive definite.

We can relax the assumption of the sequence w_i.

Theorem 1.2. Suppose that each $\{w_i : i = 1, 2, \dots\}$ has original mean vector μ_i and variance-covariance matrix $\mathrm{Var}(w_i) = Q_i$. Also, all cross-product third moments of the multivariate distribution are finite. Let $\bar{\mu}_i = \frac{1}{n}\sum_{i=1}^n \mu_i$ and $\bar{Q}_i = \frac{1}{n}\sum_{i=1}^n Q_i$. We assume that $\lim_{n\to\infty} \bar{Q}_i = Q$, where Q is a finite, positive definite matrix, and that for every i,

$$\lim_{n\to\infty}(n\bar{Q}_i)^{-1}Q_i = \lim_{n\to\infty}(\sum_{i=1}^n Q_i)^{-1}Q_i = 0.$$

With these assumptions, the Lindeberg-Feller Central Limit Theorem is given as follows:

$$\sqrt{n}(\bar{w}_i - \bar{\mu}_i) \xrightarrow[n\to\infty]{d} N(0, Q),$$

where $\bar{w}_i$ is the mean of w_i.

2 Review of the Asymptotic Normality of the M-estimator

2.1 Asymptotic Normality of M-estimator

Recall that we can get the asymptotic normality of the M-estimator through the first order condition of the optimization problem such that:

$$\frac{1}{n}\sum_{i=1}^n s(X_i, \hat{\theta}) = 0,$$

where $s(X_i, \hat{\theta})$ is **the Jacobian of the objective function**. The expansion of the score around the true parameter becomes (multiplied by $\sqrt{n}$)

$$0 = \frac{1}{\sqrt{n}}\sum_{i=1}^n s(X_i, \hat{\theta}) = \frac{1}{\sqrt{n}}\sum_{i=1}^n s(X_i, \theta) + \left(\frac{1}{n}\sum_{i=1}^n \ddot{H}(X_i, \tilde{\theta}_i)\right)\sqrt{n}(\hat{\theta} - \theta)$$

The notation $\ddot{H}_i$ denotes the $p \times p$ **Hessian of the objective function** $m(X_i, \cdot)$ with respect to θ, but with each row of $H_i \equiv H(X_i, \theta) = \nabla^2_{\theta,\theta'} m(X_i, \theta)$. Each entrance of the Hessian is evaluated for a parameter $\tilde{\theta}$ between θ and $\hat{\theta}$ and we know that each must converge in probability to θ (since each is "trapped" between θ and $\hat{\theta}$). Now, we can apply Lemma 1.1 in the TA session #10(...?) to get

$$\frac{1}{n}\sum_{i=1}^{n} \ddot{H}(X_i, \tilde{\theta}_i) \xrightarrow[n\to\infty]{p} \mathbb{E}[H(X, \theta)] \tag{2}$$

(under some moment conditions).

Lemma 2.1. Suppose that $\hat{\theta} \xrightarrow[n\to\infty]{p} \theta$, and assume that a function $r \colon \mathbb{R}^k \times \Theta \to \mathbb{R}^q$ satisfies the same assumptions on $m(X, \theta)$ in TA session #10(...?). Then,

$$\frac{1}{n}\sum_{i=1}^{n} r(X_i, \hat{\theta}) \xrightarrow[n\to\infty]{p} \mathbb{E}[r(X, \theta)]. \tag{3}$$

That is, $\frac{1}{n}\sum_{i=1}^{n} r(X_i, \hat{\theta})$ is a consistent estimator of $\mathbb{E}[r(X, \theta)]$.

If $H \equiv \mathbb{E}[H(X, \theta)]$ is nonsingular, then $n^{-1}\sum_{i=1}^{n} \ddot{H}_i$ is non–singular with probability approaching one and $[n^{-1}\sum_{i=1}^{n} \ddot{H}(X_i, \tilde{\theta}_i)]^{-1} \xrightarrow[n\to\infty]{p} H^{-1}$ (by **continuous mapping theorem**). Therefore, we can write

$$\sqrt{n}(\hat{\theta} - \theta) = H^{-1}\left[-\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s(X_i, \theta_0)\right] + o_p(1). \tag{4}$$

Theorem 2.1 (Asymptotic Normality of M–Estimators)**.** In addition to the assumptions to derive the Weak Law of Large Numbers and the consistency of M-Estimators in the lecture note of the TA session #10, assume

(a) θ is in the interior of Θ;

(b) $s(X, \cdot)$ is continuously differentiable on the interior of Θ for all $X \in \mathbb{R}^k$;

(c) Each element of $H(X, \theta)$ is bounded in absolute value by a function $b(X)$, where $\mathbb{E}[b(X)] < \infty$;

(d) $H \equiv \mathbb{E}[H(X, \theta)]$ is positive definite;

(e) $\mathbb{E}[s(X, \theta)] = 0$;

(f) Each element of $s(X, \theta)$ has finite second moment.

Then

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow[n\to\infty]{d} N\left(0, H^{-1}JH^{-1}\right) \tag{5}$$

where $H \equiv \mathbb{E}[H(X, \theta)]$ and $J \equiv \mathbb{E}[s(X, \theta)s(X, \theta)'] = \mathrm{Var}[s(X, \theta)]$.

3 M-estimator of the Linear Regression Model

In this section, we analyse the case of the M-estimator of the linear regression model. Also, we derive the estimator of the asymptotic variance for our convenience in the case of the statistical test.

3.1 Asymptotic Normality of the M-estimator of the Linear Regression Model

Suppose the linear regression model such that:

$$y_i = b_1 x_{i1} + b_2 x_{i2} + \cdots + b_k x_{ik} + u_i, \quad i = 1, 2, \cdots, n, \tag{6}$$

and we can rewrite this model as follows:

$$\underline{y} = \underline{x} b + \underline{u} \left(\iff \underbrace{\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}}_{\in \mathbb{R}^n} = \underbrace{\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}}_{\in \mathcal{M}_{n \times k}(\mathbb{R})} \underbrace{\begin{pmatrix} b_1 \\ \vdots \\ b_k \end{pmatrix}}_{\in \mathbb{R}^k} + \underbrace{\begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}}_{\in \mathbb{R}^n} \right),$$

where $\underline{x}_i = (x_{i,1}, \ldots, x_{i,k})$ is a $1 \times K$ vector for $i \in \{1, \ldots, n\}$ and $b = (b_1, \ldots, b_k)'$ is a $k \times 1$ vector. Then, the optimisation problem to derive the M-estimator is

$$\arg\max_{\theta} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{2} (y_i - x_{il} b_i)^2 = \frac{1}{n} \left[\frac{1}{2} (y - xb)'(y - xb) \right], \tag{7}$$

and be careful that the parameter vector is $\theta = (b, \sigma^2)'$. We may find that the objective function, which is represented as $m(X_i, \theta)$, is $\frac{1}{2n} \sum (y_i - x_{il} b_i)^2$. The first order condition and second order condition are [1]

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^{n} s(x_i, \theta) &= \frac{1}{n} \sum_{i=1}^{n} x_i u_i; \\ \frac{1}{n} \sum_{i=1}^{n} H(x_i, \theta) &= \frac{1}{n} \sum_{i=1}^{n} x_i x_i'. \end{aligned} \tag{8}$$

In addition, for a parameter $\tilde{\theta} \in (\hat{\theta}, \theta)$, the Hessian evaluated in $\tilde{\theta}$ is

$$\frac{1}{n} \sum_{i=1}^{n} \ddot{H}(x_i, \tilde{\theta}). \tag{9}$$

Therefore, in the linear regression case, we can say

$$\begin{aligned} \sqrt{n}(\hat{\theta} - \theta) &= \left(\frac{1}{n} \sum_{i=1}^{n} \ddot{H}(x_i, \tilde{\theta}) \right)^{-1} \left[-\frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i u_i \right] \\ &= H^{-1} \left[-\frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i u_i \right] + o_p(1) \end{aligned} \tag{10}$$

[1] $\sum_{i=1}^{n} s(x_i, \theta)$ and $\sum_{i=1}^{n} H(x_i, \theta)$ are the Jacobian and Hessian of $\left[\frac{1}{2}(y - x\beta)'(y - x\beta)\right]$.

Recall that $\mathbb{E}[s(X_i,\theta)] = \mathbb{E}[x_i u_i] = 0$ (exogenous assumption). Therefore, $-n^{-1/2}\sum_{i=1}^n x_i u_i$ satisfies the **Lindeberg-Lévy central limit theorem** under suitable conditions, because it is the average of i.i.d. random vectors with zero mean, multiplied by the usual $\sqrt{n}$. By applying this theorem into the last term of the Eq. (10), we have

$$\sqrt{n}\left(-\frac{1}{n}\sum_{i=1}^n x_i u_i - 0\right) \xrightarrow[n\to\infty]{d} N(0,\sigma^2 H);$$
$$-\frac{1}{\sqrt{n}}\sum_{i=1}^n x_i u_i \xrightarrow[n\to\infty]{d} N(0,\sigma^2 H). \tag{11}$$

Recall that $\sigma^2 H = \mathrm{Var}(x'\underline{u})$. Thus, we can get the asymptotic distribution of the MLE. The Slutsky's Theorem is used to derive it.

Lemma 3.1 (Slutsky's Theorem)**.** Suppose a sequence of random vector $\mathbf{x}_n \xrightarrow[n\to\infty]{\mathbb{P}} \mathbf{x}$ and $\mathbf{y}_n \xrightarrow[n\to\infty]{d} c$, respectively. Then, we have:

$$(\mathbf{x}_n,\ \mathbf{y}_n) \xrightarrow[n\to\infty]{d} (\mathbf{x},\ c). \tag{12}$$

Example.

$$\mathbf{x_n} + \mathbf{y_n} \xrightarrow[n\to\infty]{d} \mathbf{x} + \mathbf{c},$$
$$\frac{\mathbf{x_n}}{\mathbf{y_n}} \xrightarrow[n\to\infty]{d} \frac{\mathbf{x}}{c}$$

Therefore, we can derive the following result, $\sqrt{n}(\hat{\theta}-\theta) \xrightarrow[n\to\infty]{d} N(0,\sigma^2 H^{-1})$.
Remind that we use the formula of the conditional variance like in the case of the OLS, because covariate x are stochastic.

3.2 The Estimator of the Asymptotic Variance of the M-estimator

Now we estimate the asymptotic variance of the M-estimaor in the general case. Recall that we have Asy. $\mathrm{Var}[\sqrt{n}(\hat{\theta}-\theta)] = H^{-1}JH^{-1}$, so we can rewrite Asy. $\mathrm{Var}(\hat{\theta}) = \frac{1}{n}H^{-1}JH^{-1}$. By the Eq. (2), the estimator of H and J can be derived by the following relationship.

$$\frac{1}{n}\sum_{i=1}^n H(x_i,\hat{\theta}) \equiv \frac{1}{n}\sum_{i=1}^n \hat{H}_i \xrightarrow[n\to\infty]{p} H$$
$$\frac{1}{n}\sum_{i=1}^n s(X_i,\hat{\theta})s(X_i,\hat{\theta})' \equiv \frac{1}{n}\sum_{i=1}^n \hat{s}_i\hat{s}_i' \xrightarrow[n\to\infty]{p} J \tag{13}$$

Thus, the estimator of the asymptotic variance is

$$\mathrm{Var}(\hat{\theta}) = \frac{1}{n}(\sum_{i=1}^n \hat{H}_i)^{-1}(\sum_{i=1}^n \hat{s}_i\hat{s}_i')(\sum_{i=1}^n \hat{H}_i)^{-1}. \tag{14}$$

In the case of the linear regression model,

$$\mathrm{Var}(\hat{\theta}) = \frac{1}{n}(\sum_{i=1}^n x_i x_i')^{-1}(\sum_{i=1}^n x_i x_i' \hat{u}_i^2)(\sum_{i=1}^n x_i x_i')^{-1}. \tag{15}$$

Assume we can patition x into x_1 and x_2, and that θ indexes some feature of the distribution of x_2 given x_1. Define

$$A(x_1, \theta) \equiv \mathbb{E}[H(x, \theta)|x_1] \tag{16}$$

While $H(x, \theta)$ is generally a function only of x_1 and x_2, $A(x_1, \theta)$ is a function only of x_1. By the law of iterated expectations, we have $\mathbb{E}[A(x_1, \theta)] = \mathbb{E}[A(x, \theta)] = J$. Therefore, we can derive another estimator of J as

$$\frac{1}{n}\sum_{i=1}^{n} A(x_{1i}, \hat{\theta}) \equiv \frac{1}{n}\sum_{i=1}^{n} \hat{A}_i \xrightarrow[n\to\infty]{p} H. \tag{17}$$

The above estimator is useful in the case where $\mathbb{E}[H(x, \theta)|x]$ can be obtained in closed form or is easily approximated. In the above case, the estimator of the asymptotic variance is

$$\mathrm{Var}(\hat{\theta}) = \frac{1}{n}(\sum_{i=1}^{n} \hat{A}_i)^{-1}(\sum_{i=1}^{n} \hat{s}_i\hat{s}_i')(\sum_{i=1}^{n} \hat{A}_i)^{-1}. \tag{18}$$

This estimator is used in the case of the NLS (non-linear least square).

4 R Excercise

In this section, we will explain how to use R.[2] Here, we use data of the real estate price and the location of each building in HongKong.[3]From this data, we analyze the regression model by using the ML method. We derive the MLE of the following regression model.

$$\mathrm{(price)_i = \alpha + \beta_1(a\ number\ of\ convinience\ stores)_i + \beta_2(housage)_i + u_i}, \tag{19}$$

where $u_i \sim N(0, \sigma^2)$.

```
rm(list=ls(all=TRUE))

variableset<-read.csv("Real estate valuation data set.csv",header=T)
variableset<-data.frame(variableset)
houseage<-variableset[,3]
convstore<-variableset[,5]
price<-variableset[,8]
y<-as.vector(price)
x_1<-as.vector(convstore)
x_2<-as.vector(houseage)
NROW(price)
const<-rep(1,NROW(price))
#By using rep function, we can make "i" vector, whose all elements are 1.
#Capital "NROW" can be used in the vec. structure.
x<-cbind(const,x_1,x_2)#Make a matrix "x" of the stacked model.


#E(u_i)=0,Var(u_i)=sigma^2 for all i.
#Parameter vector=(beta',sigma^2)'
```

[2]The data set used in this class is uploaded in the UCI Machine Learning Repository.

[3]Yeh, I. C., & Hsu, T. K. (2018). Building real estate valuation models with comparative approach through case-based reasoning. Applied Soft Computing, 65, 260-271.

```
#This is the original code to make the log-likelihood function.
log_likelihood<-function(para){
k<-ncol(x)
n<-nrow(x)
beta<-para[1:k]
sigma_2<-para[k+1]
-(n/2)*log(2*pi*sigma_2)-(1/(2*sigma_2))*t((y-x%*%beta))%*%(y-x%*%beta)
}
k<-ncol(x)
n<-nrow(x)
opt<-optim(par=rep(1,k+1),fn=log_likelihood,
control=list(fnscale=-1),hessian=TRUE)

#optim function:searching parameters
opt

#Derive the standard error.
se<- diag(sqrt(abs(solve(opt$hessian[1:(k+1), 1:(k+1)]))))
se

lm_house<-lm(y~x_1+x_2)
summary(lm_house)

hat_u<-residuals(lm_house)

hat_var<-solve(n-k)*t(hat_u)%*%hat_u
hat_var
#Results of the mle and olse are similar.
```

The result of this model is given as bellow. The test statistics of the MLE and how to test by using R will be explained in the TA session of Large Sample Tests.

Table 1: The Result of the ML Method

	Estimate	Standard Error
(Intercept)	32.0262912	1.1893170
x_1	2.6885753	0.1782664
x_2	-0.2853355	0.0460913
(σ^2)	113.5939028	7.8736729

Table 2: The Result of the OLS

	Estimate	Std. Error	t value	Pr($>|t|$)
(Intercept)	32.02251	1.19529	26.791	2e-16 ***
x_1	2.69251	0.17916	15.028	2e-16***
x_2	-0.28601	0.04632	-6.174	1.6e-09 ***
(σ^2)	114.738			

Signif. codes: 0 ‘ ***’0.001 ‘ ** ’ 0.01 ‘ * ’ 0.05 ‘ . ’ 0.1 ‘ ’ 1