Week 3: Nov. 17, 2022

Note Writer: Yu-Chieh Kuo[†]

[†]Department of Information Management, National Taiwan University

Recap

Asymptotic Theorems

We have a sequence of *i.i.d.* observations $z_1, z_2, \cdots, z_n$, which are vectors of $k \times 1$, and the expectation of z_i is $\mathbb{E}[z_i] \equiv \mu$. Thus, we have

$$\frac{1}{n}\sum_{i=1}^{n} z_i \xrightarrow{p} \mu \quad \text{and} \quad \sqrt{n}\frac{1}{n}\sum_{i=1}^{n}(z_i - \mu) \xrightarrow{d} \mathcal{N}\left(0, \mathbb{E}[(z_i-\mu)(z_i-\mu)']\right)$$

Least Squares

Given the criterion / objective function $Q_n(\beta)$ and parameters of interest β ($k \times 1$ vector), for any β, we have the convergence result

$$Q_n(\beta) \xrightarrow{p} Q_\infty(\beta) \quad \text{or} \quad \underset{n\to\infty}{\text{plim}}\, Q_n(\beta) = Q_\infty(\beta)$$

As usual, given the data

$$\begin{array}{lll} \underset{1\times 1}{y_1}, & y_2, \cdots, y_n & \text{dependent variables} \\ \underset{k\times 1}{x_1}, & x_2, \cdots, x_n & \text{independent variables,} \end{array}$$

we define

$$Y \equiv \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad \text{and} \quad X \equiv \begin{pmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{pmatrix},$$

where Y and X are $n \times 1$ and $n \times k$ matrices, respectively.

For example, in a linear least squares case, we have

$$Q_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - x_i'\beta\right)^2 \quad \xrightarrow{p} \quad Q_\infty(\beta) = \mathbb{E}\left[y_i - x_i'\beta\right]^2,$$

which leads to

$$\begin{aligned} \hat{\beta} &\equiv \arg\min_{\beta} Q_n(\beta) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} x_i y_i\right) \\ \xrightarrow{p} \quad \beta_\infty &\equiv \arg\min_{\beta} Q_\infty(\beta) = \left(\mathbb{E}\left[x_i x_i'\right]\right)^{-1}(\mathbb{E}[x_i y_i]). \end{aligned}$$

Here we define $e_i \equiv y_i - x_i\beta_\infty$, which leads to $\mathbb{E}[x_i e_i] = 0$. Next, after substituting $y_i = x_i'\beta_\infty + e_i$ into $\hat\beta$, we obtain

$$\hat\beta - \beta_\infty = \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} x_i e_i\right).$$

Some books call this sampling errors after multiply $\sqrt{n}$.

Statistical properties

Theorem.

$$\mathbb{E}\left[\hat\beta - \beta_\infty \mid X\right] = 0 \quad \text{if} \quad \mathbb{E}[e_i \mid X] = 0$$

We call this unbiasedess. □

Remark. Note that

$$\begin{array}{ccc} \mathbb{E}[e_i \mid x_i] = 0 & \Longrightarrow & \mathbb{E}[e_i x_i] = 0 \\ \text{(Stronger)} & \not\Leftarrow & \text{(Weaker).} \end{array}$$

□

Moreover, we have

$$\mathbb{E}\left[\left(\hat\beta - \beta_\infty\right)\left(\hat\beta - \beta_\infty\right)' \mid X\right] = \sigma^2\left(\sum_{i=1}^{n} x_i x_i'\right)^{-1} = \sigma^2 (X'X)^{-1} \quad \text{if} \quad \mathbb{E}\left[e_i^2 \mid X\right] = 0.$$

Another property is consistency. Clearly, $\hat\beta - \beta_\infty \xrightarrow{p} 0$.
Additionally, we derive the asymptotic covariance matrix as below.

$$\begin{aligned} \sqrt{n}\left(\hat\beta - \beta_\infty\right)\left(\hat\beta - \beta_\infty\right)' &= \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} x_i e_i \sum_{j=1}^{n} x_j' e_j\right)\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \\ &= \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i' e_i^2 + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} x_i x_j e_i e_j\right)\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \\ &\xrightarrow{p} \left(\mathbb{E}\left[x_i x_i'\right]\right)^{-1} \mathbb{E}\left[x_i x_i' e_i^2\right]\left(\mathbb{E}\left[x_i x_i'\right]\right)^{-1} \\ &= \sigma^2\left(\mathbb{E}\left[x_i x_i'\right]\right)^{-1} \quad \text{if } \mathbb{E}\left[x_i x_i' e_i^2 = \mathbb{E}\left[x_i x_i'\right]\sigma^2\right]. \end{aligned}$$

Therefore, we have the distribution

$$\sqrt{n}(\hat\beta - \beta_\infty) \xrightarrow{d} \mathcal{N}\left(0, \left(\mathbb{E}\left[x_i x_i'\right]\right)^{-1} \mathbb{E}\left[x_i x_i' e_i^2\right]\left(\mathbb{E}\left[x_i x_i'\right]\right)^{-1}\right).$$

- There is no assumption about what is the true model. We just show that the minimize of the finite sample criterion function $Q_n(\beta)$ converges to the minimizer of the asymptotic criterion function $Q_\infty(\beta)$ (as $\hat\beta \xrightarrow{p} \beta_\infty$).
- There is no endogeneity problem if we define consistency in this way.
- However, this is one of the key differences between statistics and econometrics. In econometrics, we start with a true model of a structural model. For example, given $y_i = x_i'\beta + e_i$ but $\mathbb{E}[x_i e_i] \neq 0$, we need to find instruments to solve the problem, so that the estimation is consistent in the sense of converging to the true parameters.

- Note that we may have a system of equations (or called structural form model)

  $$Y = AY + BX + E,$$

  where Y, X, and E are $m \times 1$, A and B are $m \times m$. Here we denote θ by the parameters of interest, and the matrix entries are functions of θ. To elaborate, we can denote A by $A(\theta)$ and B by $B(\beta)$. Therefore, the reduced form becomes

  $$Y = (I_m - A(\theta))^{-1}B(\theta)X + (I_m - A(\theta))^{-1}E.$$

Nonlinear case

Given $f(X;\beta)$ as a nonlinear function, we have

$$Q_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i;\beta))^2 \xrightarrow{p} Q_\infty(\beta) = \mathbb{E}[y_i - f(x_i;\beta)]^2$$

The FOC yeilds a $k \times 1$ matrix

$$\frac{\partial Q_n(\beta)}{\partial \beta} = \frac{-2}{n}\sum_{i=1}^{n}(y_i - f(x_i;\beta))\frac{\partial f_i}{\partial \beta} \xrightarrow{p} \frac{\partial Q_\infty(\beta)}{\partial \beta} = -2\,\mathbb{E}[y_i - f_i]\frac{\partial f_i}{\partial \beta},$$

(note that the FOC goes to 0 when $\beta = \beta_\infty$) and the second derivative ($k \times k$) is

$$\frac{\partial^2 Q_n(\beta)}{\partial \beta \partial \beta'} = \frac{2}{n}\sum_{i=1}^{n}\frac{\partial f_i}{\partial \beta}\frac{\partial f_i}{\partial \beta'} - \frac{2}{n}\sum_{i=1}^{n}(y_i - f_i)\frac{\partial^2 f_i}{\partial \beta \partial \beta'} \xrightarrow{p} \frac{\partial^2 Q_\infty \beta}{\partial \beta \partial \beta'} = 2\,\mathbb{E}\left[\frac{\partial f_i}{\partial \beta}\frac{\partial f_i}{\partial \beta'}\right] - 2\,\mathbb{E}[y_i - f_i]\frac{\partial^2 f_i}{\partial \beta \partial \beta'}.$$

Hence, the minimizer $\hat\beta$ of $Q_n(\beta)$ is defined by

$$\frac{\partial Q_n(\hat\beta)}{\partial \beta} = \frac{-2}{n}\sum_{i=1}^{n}\left(y_i - \hat{f}_i\right)\frac{\partial \hat{f}_i}{\partial \beta} = 0.$$

We here use computers to find the minimizer such as the Newton method.

A natural question we need to care is when we will have $\hat\beta \xrightarrow{p} \beta_\infty$ to allow the asymptotic derivation. (Reference: Ch7, Hayashi)

- If $Q_n(\beta) \xrightarrow{p} Q_\infty(\beta)$ uniformly, the set of candidate parameters θ is compact. Therefore, $\hat\beta \xrightarrow{p} \beta_\infty$.
- If $Q_n(\beta) \xrightarrow{p} Q_\infty(\beta)$, and $Q_n(\beta)$ is a convex and continuous function, then $\hat\beta \xrightarrow{p} \beta_\infty$.

Asymptotic Distribution

We apply the mean value theorem to derive the asymptotic distribution.

$$\frac{\partial Q_n\left(\hat\beta\right)}{\partial \beta} = \frac{\partial Q_n(\beta_\infty)}{\partial \beta} + \frac{\partial^2 Q_n(\beta_m)}{\partial \beta \partial \beta'}\left(\hat\beta - \beta\right),$$

where $\beta_m \in [\hat\beta, \beta_\infty]$. Since $\hat\beta \xrightarrow{p} \beta_\infty$, therefore, it gives $\beta_m \xrightarrow{p} \beta_\infty$. Moreover, re-writting and re-scaling the NLS problem as an asymptotic form gives

$$\sqrt{n}\left(\hat\beta - \beta_\infty\right) = \left(\frac{\partial^2 Q_n(\beta_m)}{\partial \beta \partial \beta'}\right)^{-1}\left(-\sqrt{n}\frac{\partial Q_n(\beta_\infty)}{\partial \beta}\right).$$

Now, we assume that

$$\sqrt{n}\frac{\partial Q_n(\beta_\infty)}{\partial \beta} \xrightarrow{d} \mathcal{N}\left(0, \operatorname*{plim}_{n\to\infty}\left(n\frac{\partial Q_n(\beta_\infty)}{\partial \beta}\frac{\partial Q_n(\beta_\infty)}{\partial \beta'}\right)\right),$$

if such the assumption is correct, then the distribution of $\sqrt{n}\left(\hat{\beta}-\beta_\infty\right)$ asymptotes to

$$\sqrt{n}\left(\hat{\beta}-\beta_\infty\right) \xrightarrow{d} \mathcal{N}\left(0, \left(\operatorname*{plim}_{n\to\infty}\frac{\partial^2 Q_n(\beta_\infty)}{\partial\beta\partial\beta'}\right)^{-1}\left(\operatorname*{plim}_{n\to\infty} n\frac{\partial Q_n(\beta_\infty)}{\partial \beta}\frac{\partial Q_n(\beta_\infty)}{\partial \beta'}\right)\left(\operatorname*{plim}_{n\to\infty}\frac{\partial^2 Q_n(\beta_\infty)}{\partial\beta\partial\beta'}\right)^{-1}\right).$$

Nonlinear case

For the nonlinear case, we have

$$\frac{\partial Q_n(\beta_\infty)}{\partial \beta} = \frac{-2}{n}\sum_{i=1}^{n} e_i \frac{\partial f_i(\beta_\infty)}{\partial \beta},$$

where $e_i \equiv y_i - f(x_i;\beta)$.

Cancelling -2 for convenience and assuming the assumption is satisfied, we obtain

$$\sqrt{n}\frac{1}{n}\sum_{i=1}^{n}\frac{\partial f_i(\beta_\infty)}{\partial \beta}e_i \xrightarrow{d} \mathcal{N}\left(0, \mathbb{E}\left[\frac{\partial f_i(\beta_\infty)}{\partial \beta}\frac{\partial f_i(\beta_\infty)}{\partial \beta'}e_i^2\right]\right).$$

Moreover,

$$\frac{\partial^2 Q_n(\beta_\infty)}{\partial\beta\partial\beta'} \xrightarrow{p} 2\,\mathbb{E}\left[\frac{\partial f_i(\beta_\infty)}{\partial \beta}\frac{\partial f_i(\beta_\infty)}{\partial \beta'}\right] + 0,$$

and

$$\sqrt{n}(\hat{\beta}-\beta_\infty) \xrightarrow{d} \mathcal{N}\left(0, Cov\right),$$

where

$$\begin{aligned} Cov &= \left(\mathbb{E}\left[\frac{\partial f_i(\beta_\infty)}{\partial \beta}\frac{\partial f_i(\beta_\infty)}{\partial \beta'}\right]\right)^{-1}\left(\mathbb{E}\left[\frac{\partial f_i(\beta_\infty)}{\partial \beta}\frac{\partial f_i(\beta_\infty)}{\partial \beta'}\right]e_i^2\right)\left(\mathbb{E}\left[\frac{\partial f_i(\beta_\infty)}{\partial \beta}\frac{\partial f_i(\beta_\infty)}{\partial \beta'}\right]\right)^{-1} \\ &= \sigma^2\left(\mathbb{E}\left[\frac{\partial f_i(\beta_\infty)}{\partial \beta}\frac{\partial f_i(\beta_\infty)}{\partial \beta'}\right]\right)^{-1} \end{aligned}$$

if we assume that $\mathbb{E}\left[\frac{\partial f_i(\beta_\infty)}{\partial \beta}\frac{\partial f_i(\beta_\infty)}{\partial \beta'}\right] = \sigma^2$.

Honho: Trust me, all 2 and -2 will be canceled since there are inverse in the distribution, just algebra operation.

Prediction Errors and Model Selection

Assume the true model is $y_i = f(x_i;\beta_0) + e_i$, where we denote β_0 by true parameters for convenience. Applying 2nd-order Taylor approximations, we have

$$Q_n(\hat{\beta}) = Q_n(\beta_0) + \frac{\partial Q_n(\beta_0)}{\partial \beta'}\left(\hat{\beta}-\beta_0\right) + \frac{1}{2}\left(\hat{\beta}-\beta_0\right)\frac{\partial^2 Q_n(\beta_0)}{\partial\beta\partial\beta'}\left(\hat{\beta}-\beta_0\right) + \text{ higher order terms.}$$

Be careful that $Q_n(\cdot)$ is 1×1, $\frac{\partial Q_n(\beta_0)}{\partial \beta'}$ is $1\times k$, and $\left(\hat{\beta}-\beta_0\right)$ is $k\times k$.

Moreover, the out-sample case is in the similar form:

$$Q_n^{out}(\hat{\beta}) = Q_n^{out}(\beta_0) + \frac{\partial Q_n^{out}(\beta_0)}{\partial \beta'}\left(\hat{\beta} - \beta_0\right) + \frac{1}{2}\left(\hat{\beta} - \beta_0\right)\frac{\partial^2 Q_n^{out}(\beta_0)}{\partial \beta \partial \beta'}\left(\hat{\beta} - \beta_0\right) + \text{ higher order terms.}$$

By FOC (**why??????**), therefore,

$$\begin{aligned} 0 = \frac{\partial Q_n(\hat{\beta})}{\partial \beta} &= \frac{\partial Q_n(\beta_0)}{\partial \beta} + \frac{\partial^2 Q_n(\beta_0)}{\partial \beta \partial \beta'}\left(\hat{\beta} - \beta_0\right) + \cdots \\ \iff \left(\hat{\beta} - \beta_0\right) &= \left(\frac{\partial^2 Q_n(\beta_0)}{\partial \beta \partial \beta'}\right)^{-1}\left(-\frac{\partial Q_n(\beta_0)}{\partial \beta}\right). \end{aligned}$$

Substituting $\left(\hat{\beta} - \beta_0\right)$ into the original 2nd-order Taylor approximation, we have

$$\begin{aligned} Q_n(\hat{\beta}) &= Q_n(\beta_0) - \frac{1}{2}\left(\frac{\partial Q_n(\beta_0)}{\partial \beta}\right)' \frac{\partial^2 Q_n(\beta_0)}{\partial \beta \partial \beta'}^{-1} \frac{\partial Q_n(\beta_0)}{\partial \beta} \\ Q_n^{out}(\hat{\beta}) &= Q_n^{out}(\beta_0) - \left(\frac{\partial Q_n^{out}(\beta_0)}{\partial \beta}\right)' \frac{\partial^2 Q_n^{out}(\beta_0)}{\partial \beta \partial \beta'}^{-1} \frac{\partial Q_n^{out}(\beta_0)}{\partial \beta} + \frac{1}{2}(pQ_0)'(ppQ_0)^{-1}(ppQ_0)(ppQ)^{-1}(pQ). \end{aligned}$$

Notice a trick that taking expectation w.r.t. X to all terms above is available, therefore, it alters