c.f) Question : the relationship between eigenvalues and testing power.

Answer) Testing $H_0 : C'\beta = a$.

$$\hat{\beta} | X \sim N(\beta, \sigma^2 (X'X)^{-1})$$

Then, $C'\hat{\beta} | X \sim N(C'\beta, \sigma^2 C'(X'X)^{-1} C)$

When we minimize the Variance $\sigma^2 C'(X'X)^{-1}C$, (i.e. increase the power)

$$\min_{C} \frac{C'(X'X)^{-1}C}{C'C} \Rightarrow \text{The solution} : \arg\min_{C} \frac{C'(X'X)^{-1}C}{C'C}$$

$\Rightarrow$ eigenvector of $(X'X)^{-1}$ corresponding to the smallest eigen value of $(X'X)^{-1}$

- Big $O_p(1)$, little $o_p(1)$

Note) $O_p(1)$ = stochastically bounded; $o_p(1)$ = $X_n \xrightarrow{P} 0$

① $o_p(1) + o_p(1) = o_p(1)$

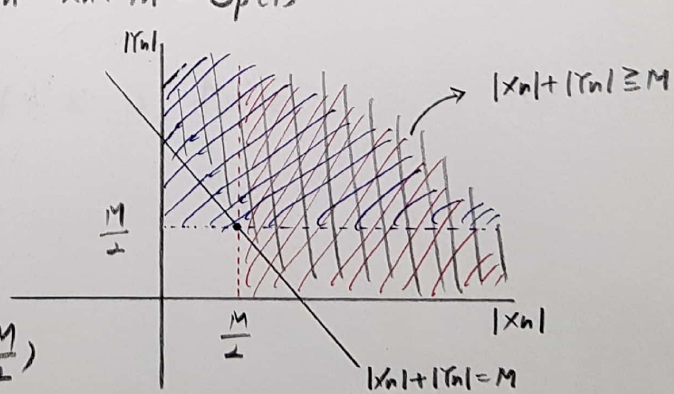
If $X_n \xrightarrow{P} 0$, and $Y_n \xrightarrow{P} 0$, then $X_n + Y_n \xrightarrow{P} 0$

② $O_p(1) + O_p(1) = O_p(1)$

If $X_n = O_p(1)$ and $Y_n = O_p(1)$, then $X_n + Y_n = O_p(1)$

We know $|X_n + Y_n| \leq |X_n| + |Y_n|$

Then,



$$Pr(|X_n + Y_n| \geq M) \leq Pr(|X_n| + |Y_n| \geq M)$$
$$\leq Pr(|X_n| \geq \tfrac{M}{2}) \cup Pr(|Y_n| \geq \tfrac{M}{2})$$
$$= Pr(|X_n| \geq \tfrac{M}{2}) + Pr(|Y_n| \geq \tfrac{M}{2})$$

Both can be made small by choosing large M.

③ $\left\{\begin{array}{l} \underline{o_p(1) + O_p(1) = O_p(1)} \\ \underline{o_p(1) \cdot O_p(1) = o_p(1)} \end{array}\right.$

If $X_n = o_p(1)$, then $X_n = O_p(1)$ holds.

(Because $X_n \xrightarrow{P} 0$ implies $\exists M_\varepsilon > 0 : \lim_{n\to\infty} Pr\{|X_n| > M_\varepsilon\} < \varepsilon$, $\forall \varepsilon > 0$.)

(converges to 0 in probability) (stochastically bounded)

④ $\left\{\begin{array}{l} \underline{o_p(1) \cdot o_p(1) = o_p(1)} \\ \underline{O_p(1) \cdot O_p(1) = O_p(1)} \end{array}\right.$ (By continuous mapping theorem)

ex) Let $\underset{k\times k}{A_n} = O_p(1)$ and $\underset{k\times 1}{B_n} = o_p(1)$

$$\underset{k\times k}{A_n}\underset{k\times 1}{B_n} = \begin{pmatrix} a_{n11} & \cdots & a_{n1k} \\ \vdots & \ddots & \vdots \\ a_{nk1} & \cdots & a_{nkk} \end{pmatrix} \begin{pmatrix} b_{n1} \\ \vdots \\ b_{nk} \end{pmatrix} = \begin{pmatrix} a_{n11} b_{n1} + \cdots + a_{n1k} b_{nk} \\ \vdots \\ \underbrace{a_{nk1} b_{n1}} + \cdots + a_{nkk} b_{nk} \end{pmatrix}$$

$O_p(1) \cdot o_p(1) = o_p(1)$

and $o_p(1) + \cdots + o_p(1) = o_p(1)$.

$$= \begin{pmatrix} o_p(1) \\ \vdots \\ o_p(1) \end{pmatrix} = \underset{k\times 1}{o_p(1)}$$ if k is finite.

exercise) Let $X_{ni} = o_p(1)$.

Then $\frac{1}{n}\sum_{i=1}^{n} X_{ni} = o_p(1)$?

- What do we mean by "good estimator"?

So far conditional unbiasedness. ($E(\hat{\theta} \mid \mathbb{X}) = \theta_0, \ \forall \theta_0 \in \Theta$)

→ This is a kind of difficult one.

We can think of unbiasedness asymptotically.

① Asymtotic unbiasedness (weaker concept of unbiasedness)

$$E(\hat{\theta}_n \mid \underset{\Uparrow}{\mathbb{X}_n}) \longrightarrow \theta_0, \quad \forall \theta_0 \in \Theta$$

assume $\mathbb{X}_n$ is a deterministic sequence.

ex) $\hat{\sigma}^2_{MLE} = \frac{1}{N} \sum_{i=1}^{N} (Y_i - X_i' \hat{\beta}_{OLS})^2$

$$E(\hat{\sigma}^2_{MLE} \mid \mathbb{X}_N) = E\left\{ \frac{N-K}{N} \cdot \frac{1}{N-K} \sum_{i=1}^{N} (Y_i - X_i' \hat{\beta}_{OLS})^2 \mid \mathbb{X}_N \right\}$$

$$= \frac{N-K}{N} E(\hat{\sigma}^2 \mid \mathbb{X}_N) = \frac{N-K}{N} \sigma^2 \longrightarrow \sigma^2 \text{ as } n \to \infty$$

⇒ $\hat{\sigma}^2_{MLE}$: Not unbiased, but asymtotically unbiased.

$\hat{\sigma}^2_{OLS}$: unbiased (It is shown as a specific parameter)

② Consistency (Weak consistency $\xrightarrow{p}$, Strong consistency $\xrightarrow{a.s}$)

$\hat{\theta}_n$ is weakly consistent $\Leftrightarrow$ $\hat{\theta}_n \xrightarrow{p} \theta_0, \quad \forall \theta_0 \in \Theta$

$\hat{\theta}_n$ is strongly consistent $\Leftrightarrow$ $\hat{\theta}_n \xrightarrow{a.s} \theta_0, \quad \forall \theta_0 \in \Theta$

→ We need to check consistency in increasing sample size n.

→ Suppose two estimators are consistent.

In that case, their speeds to converge to θ_0 would be different,

so that the "convergence rate" is also an important property

as an estimator. ⇓

③ Convergence rate

If $r_n \cdot (\hat{\theta}_n - \theta_0) = O_p(1)$, then an estimator has convergence rate $\frac{1}{r_n}$ $(r_n \to \infty)$

Typically, $r_n = \sqrt{n}$ (not always)

↑

This is usually used for parametric / cross-section data cases.

→ Convergence rate only show how fast the convergence is.

If $O_p(1)$ has a distribution, then $\hat{\theta}_n - \theta_0$ also converges in the distribution.

Note: If $r_n(\hat{\theta}_n - \theta_0) = O_p(1)$ and $r_n'(\hat{\theta}_n - \theta_0) = O_p(1)$, and $\frac{r_n}{r_n'} \to 0$, then the rate is at least r_n'.

④ Convergence in Distribution

$r_n(\hat{\theta}_n - \theta_0) \xrightarrow{d} Z$ (a random variable)

(r_n is the best rate)

→ Let $\hat{\theta}_n$ be consistent and $r_n(\hat{\theta}_n - \theta_0) = O_p(1)$.

Then, $\hat{\theta}_n \xrightarrow{p} \theta_0$ and r_n is the convergence rate.

However, we cannot know to measure the inaccuracy of $\hat{\theta}_n$.

Thus, we can check its accuracy through the variance of Z.

→ Suppose $r_n(\hat{\theta}_n - \theta_0) \xrightarrow{d} Z$ hold. and $\exists \{r_n'\}$.

$$r_n'(\hat{\theta}_n - \theta_0) = \underbrace{\left(\frac{r_n'}{r_n}\right)}_{\Downarrow} \cdot \underbrace{r_n(\hat{\theta}_n - \theta_0)}_{\xrightarrow{d} Z}$$

0 or ∞ as $n \to \infty$

(Because they are deterministic)

Therefore, this result is used to evaluate the errors we make by using $\hat{\theta}_n$.

- If we have two estimators that converge with different rate, then the one that converges faster is the better estimator.
- If two estimators converge at the same rate, then the one which has the smaller inaccuracy measure (typically variance) is the better estimator.

ex) Typically, $\sqrt{n}(\hat{\theta}_{1n} - \theta_0) \xrightarrow{d} N(0, \sigma_1^2)$

$\sqrt{n}(\hat{\theta}_{2n} - \theta_0) \xrightarrow{d} N(0, \sigma_2^2)$

then if $\sigma_1^2 < \sigma_2^2$, then we say $\hat{\theta}_{1n}$ is more efficient then $\hat{\theta}_{2n}$.

and $\frac{\sigma_2^2}{\sigma_1^2}$ is the relative efficiency of the first estimator.

ex2)

Consider estimation of $E(X)$ by $\frac{1}{N}\sum_{i=1}^{N} X_i$

$Var\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) = \frac{1}{N}\sigma_X^2$

$Var(\hat{\theta}) = \frac{1}{N}\sigma_1^2$

Suppose $\frac{\sigma_1^2}{\sigma_X^2} = 2$. and assume that $\exists$ sample size $\frac{N}{2}$

Then $Var\left(\frac{1}{\left(\frac{N}{2}\right)}\sum_{i=1}^{\frac{N}{2}} X_i\right) = \frac{1}{2N}\cdot\sigma_X^2$

At the viewpoint of sample size problem,

"2" means sample size of $\hat{\theta}$ is double more than $\frac{1}{N}\sum_{i=1}^{N} X_i$.

- Asymtotic Properties of the OLS estimator.

$$\hat{\beta} = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{Y} = \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1}\sum_{i=1}^{N} x_i y_i = \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1}\sum_{i=1}^{N} x_i (x_i'\beta + u_i)$$

$$= \cancel{\left(\sum_{i=1}^{N} x_i x_i'\right)^{-1}\sum_{i=1}^{N} x_i x_i'}\beta + \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1}\sum_{i=1}^{N} x_i u_i$$

$$= \beta + \underbrace{\left(\sum_{i=1}^{N} x_i x_i'\right)^{-1}}_{:=A_N}\underbrace{\sum_{i=1}^{N} x_i u_i}_{:=B_N}$$

By Kolmogrov's SLLN.

① k-l element of A_N is $\frac{1}{N}\sum_{i=1}^{N} x_{ki}x_{li} \xrightarrow{a.s.} E(x_{ki}x_{li})$ Under iid

Therefore, $\underline{A_N \xrightarrow{a.s} E\{x_i x_i'\}}$ (Because each element converges to $E(x_{ki}x_{li})$ for each k, l)

$(\Rightarrow) \; A_N^{-1} \xrightarrow{a.s} E\{x_i x_i'\}^{-1}) \; \cdots$ ⊛

② kth element of B_N is $\frac{1}{N}\sum_{i=1}^{N} x_{ki}u_i \xrightarrow{a.s.} E(x_{ki}u_i)$

$$= E(E(x_{ki}u_i \mid x_i))$$

$$= E(x_{ki} \cdot \underbrace{E(u_i \mid x_i)}_{=0}) = 0$$

(Since x_{ki} is constant)

Therefore, $\frac{1}{N}\sum_{i=1}^{N} x_{ki}u_i \xrightarrow{a.s.} 0$

From ⊛, $A_N^{-1} = O_p(1)$, and $B_N = o_p(1)$. Therefore, $O_p(1)\cdot o_p(1) = o_p(1)$

$\hat{\beta} = \beta + o_p(1)$, therefore, this shows strong consistency of OLS estimator.

※ Important assumptions

① iid sampling

② $E(x_i x_i')$ is invertible

③ $E(u_i \mid x_i) = 0$

(We need $E(x_i u_i) = 0$)

- To show asym. property of OLS estimator, We don't need $\mathbb{X}'\mathbb{X}$ is invertible. but need $E(x_i x_i')$ is invertible
- For consistency, Homoskedasticity is not important.