Econometric Methods (Econometrics I)

Lecture 2:

Basic Asymptotic Theory

Prof. Dr. Kai Carstensen

Kiel University

Winter Term 2023/2024

Outline of this lecture

1. Convergence of deterministic sequences
2. Convergence in probability
3. Convergence in distribution
4. Limit theorems for random samples
5. Asymptotic properties of estimators
6. Asymptotic properties of test statistics

Reference: Wooldridge, Chapter 3.

1. Convergence of deterministic sequences

Convergence and boundedness

Definition

(1) A sequence of nonrandom numbers $\{a_N : N = 1, 2, \ldots\}$ *converges to a* (has limit a) if for all $\varepsilon > 0$, there exists N_ε such that if $N > N_\varepsilon$ then $|a_N - a| < \varepsilon$. We write $a_N \to a$ as $N \to \infty$.

(2) A sequence $\{a_N : N = 1, 2, \ldots\}$ is *bounded* if and only if there is some $b < \infty$ such that $|a_N| \leq b$ for all $N = 1, 2, \ldots$. Otherwise, we say that $\{a_N\}$ is *unbounded*.

These definitions apply to vectors and matrices element by element.

Examples:

- If $a_N = 2 + 1/N$, then $a_N \to 2$.
- If $a_N = (-1)^N$, then a_N does not have a limit, but it is bounded.
- If $a_N = N^{1/4}$, a_N is not bounded. Because a_N increases without bound, we write $a_N \to \infty$.

Convergence and boundedness

Definition

(1) A sequence $\{a_N\}$ is $O(N^\lambda)$ (*at most of order* N^λ) if $N^{-\lambda}a_N$ is bounded. When $\lambda = 0, \{a_N\}$ is bounded, and we also write $a_N = O(1)$ (*big oh one*).

(2) $\{a_N\}$ is $o(N^\lambda)$ if $N^{-\lambda}a_N \to 0$. When $\lambda = 0, a_N$, converges to zero, and we also write $a_N = o(1)$ (*little oh one*).

Notes:

- If $a_N = o(N^\lambda)$, then $a_N = O(N^\lambda)$; in particular, if $a_N = o(1)$, then $a_N = O(1)$.
- If each element of a sequence of vectors or matrices is $O(N^\lambda)$, we say the sequence of vectors or matrices is $O(N^\lambda)$, and similarly for $o(N^\lambda)$.

Examples:

- If $a_N = \log(N)$, then $a_N = o(N^\lambda)$ for any $\lambda > 0$.
- If $a_N = 10 + \sqrt{N}$, then $a_N = O(N^{1/2})$ and $a_N = o(N^{(1/2+\gamma)})$ for any $\gamma > 0$.

2. Convergence in probability

Probability limits

Definition

(1) A sequence of random variables $\{x_N : N = 1, 2, \ldots\}$ **converges in probability** to the constant a if for all $\varepsilon > 0$,

$$\mathbf{P}\left[|x_N - a| > \varepsilon\right] \to 0 \qquad \text{as } N \to \infty.$$

We write $x_N \xrightarrow{p} a$ and say that a is the **probability limit (plim)** of x_N: $\text{plim}\ x_N = a$.

(2) In the special case where $a = 0$, we also say that $\{x_N\}$ is $o_p(1)$ (*little oh p one*). We also write $x_N = o_p(1)$ or $x_N \xrightarrow{p} 0$.

(3) A sequence of random variables $\{x_N\}$ is **bounded in probability** if and only if for every $\varepsilon > 0$, there exists a $b_\varepsilon < \infty$ and an integer N_ε such that

$$\mathbf{P}\left[|x_N| \geq b_\varepsilon\right] < \varepsilon \qquad \text{for all } N \geq N_\varepsilon.$$

We write $x_N = \mathbf{O}_p(1)$ ($\{x_N\}$ is *big oh p one*).

This implies that if $x_N \xrightarrow{p} a$, then $x_N = \mathbf{O}_p(1)$. This also holds for vectors and matrices.

Probability limits

Definition

(1) A random sequence $\{x_N\colon N = 1, 2, \dots\}$ is $o_p(a_N)$, where $\{a_N\}$ is a nonrandom, positive sequence, if $x_N/a_N = o_p(1)$. We write $x_N = o_p(a_N)$.

(2) A random sequence $\{x_N\colon N = 1, 2, \dots\}$ is $\mathbf{O}_p(a_N)$, where $\{a_N\}$ is a nonrandom, positive sequence, if $x_N/a_N = \mathbf{O}_p(1)$. We write $x_N = \mathbf{O}_p(a_N)$.

Example: Let $x_N \equiv \sqrt{N}\, z$ where z is a random variable.

(a) Then $x_N/\sqrt{N} = x_N/N^{0.5} = z$. Hence, $x_N/N^{0.5}$ is bounded in probability, $x_N = \mathbf{O}_p(N^{1/2})$.

(b) Moreover, $x_N/N^{0.5+\gamma} \xrightarrow{p} 0$ for any $\gamma > 0$. Hence, $x_N = o_p(N^{0.5+\gamma}) = o_p(N^{\delta})$ for any $\delta > 0.5$.

Calculation rules

If $w_N = o_p(1), x_N = o_p(1), y_N = \mathbf{O}_p(1)$, and $z_N = \mathbf{O}_p(1)$, then

(1) $w_N + x_N = o_p(1) + o_p(1) = o_p(1)$

(2) $y_N + z_N = \mathbf{O}_p(1) + \mathbf{O}_p(1) = \mathbf{O}_p(1)$

(3) $x_N + z_N = o_p(1) + \mathbf{O}_p(1) = \mathbf{O}_p(1)$

(4) $w_N \cdot x_N = o_p(1) \cdot o_p(1) = o_p(1)$

(5) $y_N \cdot z_N = \mathbf{O}_p(1) \cdot \mathbf{O}_p(1) = \mathbf{O}_p(1)$

(6) $x_N \cdot z_N = o_p(1) \cdot \mathbf{O}_p(1) = o_p(1)$.

Application to vectors and matrices

All of the previous definitions apply element by element to sequences of random vectors or matrices.

(1) Let $\{\mathbf{x}_N\}$ be a sequence of $K \times 1$ random vectors.

- Then $\{\mathbf{x}_N\}$ converges in probability to the $K \times 1$ nonrandom vector $\mathbf{a}$, $\mathbf{x}_N \xrightarrow{p} \mathbf{a}$, if and only if $x_{Nj} \xrightarrow{p} a_j, j = 1, \ldots, K$.
- This is equivalent to $\| \mathbf{x}_N - \mathbf{a} \| \xrightarrow{p} 0$, where $\| \mathbf{b} \| \equiv (\mathbf{b}'\mathbf{b})^{1/2}$ denotes the Euclidean length of the $K \times 1$ vector $\mathbf{b}$.

(2) Let $\{\mathbf{Z}_N\}$ be a sequence of $M \times K$ random matrices.

- Then $\{\mathbf{Z}_N\}$ converges in probability to the $M \times K$ nonrandom matrix $\mathbf{B}$, $\mathbf{Z}_N \xrightarrow{p} \mathbf{B}$, if and only if $Z_{N,ij} \xrightarrow{p} B_{ij}, i = 1, \ldots, M; j = 1, \ldots, K$.
- This is equivalent to to $\| \mathbf{Z}_N - \mathbf{B} \| \xrightarrow{p} 0$, where $\| \mathbf{A} \| \equiv [\mathrm{tr}(\mathbf{A}'\mathbf{A})]^{1/2}$ and $\mathrm{tr}(\mathbf{C})$ denotes the trace of the square matrix $\mathbf{C}$.

Slutsky's theorem

Theorem

Let $\mathbf{g}\colon \mathbb{R}^K \to \mathbb{R}^J$ *be a function continuous at some point* $\mathbf{c} \in \mathbb{R}^K$.

Let $\{\mathbf{x}_N \colon N = 1, 2, \ldots\}$ *be a sequence of* $K \times 1$ *random vectors such that* $\mathbf{x}_N \xrightarrow{\text{p}} \mathbf{c}$.

Then $\mathbf{g}(\mathbf{x}_N) \xrightarrow{\text{p}} \mathbf{g}(\mathbf{c})$ *as* $N \to \infty$. *In other words,*

$$\text{plim } \mathbf{g}(\mathbf{x}_N) = \mathbf{g}(\text{plim } \mathbf{x}_N)$$

if $\mathbf{g}(\cdot)$ *is continuous at* $\text{plim}\, \mathbf{x}_N$.

3. Convergence in distribution

Convergence in distribution: scalar case

Definition

A sequence of random variables $\{x_N : N = 1, 2, \ldots\}$ **converges in distribution** to the continuous random variable x if and only if

$$F_N(\xi) \to F(\xi) \qquad \text{as } N \to \infty \text{ for all } \xi \in \mathbb{R},$$

where F_N is the cumulative distribution function (c.d.f.) of x_N and F is the (continuous) c.d.f. of x. We write

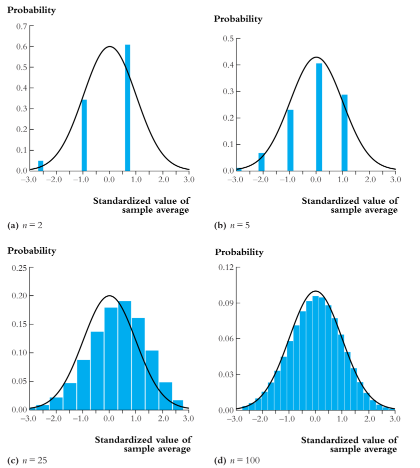
$$x_N \xrightarrow{\text{d}} x.$$

Comments:

- When $x \sim \text{Normal}(\mu, \sigma^2)$ we write $x_N \xrightarrow{\text{d}} \text{Normal}(\mu, \sigma^2)$ or $x_N \overset{\text{a}}{\sim} \text{Normal}(\mu, \sigma^2)$ meaning that x_N is **asymptotically normal**.
- Note that x_N is not required to be continuous for any N. Example: Demoivre-Laplace theorem which says that an appropriately centered and rescaled binomial random variable has a limiting standard normal distribution.

Convergence in distribution: Example

FIGURE 2.9 Distribution of the Standardized Sample Average of n Bernoulli Random Variables with $p = 0.78$



The sampling distribution of $\overline{Y}$ in Figure 2.8 is plotted here after standardizing $\overline{Y}$. This centers the distributions in Figure 2.8 and magnifies the scale on the horizontal axis by a factor of $\sqrt{n}$. When the sample size is large, the sampling distributions are increasingly well approximated by the normal distribution (the solid line), as predicted by the central limit theorem. The normal distribution is scaled so that the height of the distributions is approximately the same in all figures.

Source: Stock and Watson, Introduction to Econometrics, 3rd edition.

Definition

A sequence of $K \times 1$ random vectors $\{\mathbf{x}_N : N = 1, 2, \ldots\}$ converges in distribution to the continuous random vector $\mathbf{x}$ if and only if for any $K \times 1$ nonrandom vector $\mathbf{c}$ such that $\mathbf{c}'\mathbf{c} = 1, \mathbf{c}'\mathbf{x}_N \xrightarrow{\text{d}} \mathbf{c}'\mathbf{x}$, and we write

$$\mathbf{x}_N \xrightarrow{\text{d}} \mathbf{x}.$$

Comments:

- Rule: If $\mathbf{x}_N \xrightarrow{\text{d}} \mathbf{x}$, where $\mathbf{x}$ is any $K \times 1$ random vector, then $\mathbf{x}_N = \mathbf{O}_p(1)$.
- This is helpful because often it is straightforward to show that a sequence converges in distribution. By this rule, the sequence is also bounded in probability.

Continuous mapping theorem

Theorem

Let $\{\mathbf{x}_N\}$ *be a sequence of* $K \times 1$ *random vectors such that*

$$\mathbf{x}_N \xrightarrow{\text{d}} \mathbf{x}.$$

If $\mathbf{g} : \mathbb{R}^K \to \mathbb{R}^J$ *is a continuous function, then*

$$\mathbf{g}(\mathbf{x}_N) \xrightarrow{\text{d}} \mathbf{g}(\mathbf{x}).$$

Continuous mapping theorem: Application to the normal distribution

Let $\{\mathbf{z}_N\}$ is a sequence of $K \times 1$ random vectors such that $\mathbf{z}_N \xrightarrow{\text{d}} \mathbf{z}$, where

$$\mathbf{z} \sim \text{Normal}(\mathbf{0}, \mathbf{V}).$$

This is equivalent to

$$\mathbf{z}_N \xrightarrow{\text{d}} \text{Normal}(\mathbf{0}, \mathbf{V}).$$

Then the following results hold:

(1) For any $K \times M$ nonrandom matrix $\mathbf{A}$, $\mathbf{A}'\mathbf{z} \sim \text{Normal}(\mathbf{0}, \mathbf{A}'\mathbf{V}\mathbf{A})$.

(2) For any $K \times M$ nonrandom matrix $\mathbf{A}$, $\mathbf{A}'\mathbf{z}_N \xrightarrow{\text{d}} \text{Normal}(\mathbf{0}, \mathbf{A}'\mathbf{V}\mathbf{A})$.

(3) $\mathbf{z}'\mathbf{V}^{-1}\mathbf{z} \sim \chi^2_K$.

(4) $\mathbf{z}_N'\mathbf{V}^{-1}\mathbf{z}_N \xrightarrow{\text{d}} \chi^2_K$ (or $\mathbf{z}_N\mathbf{V}^{-1}\mathbf{z}_N \overset{\text{a}}{\sim} \chi^2_K$).

Asymptotic equivalence lemma

Lemma

Let $\{\mathbf{x}_N\}$ *and* $\{\mathbf{z}_N\}$ *be sequences of* $K \times 1$ *random vectors. If*

$$\mathbf{z}_N \xrightarrow{\text{d}} \mathbf{z} \qquad \textit{and} \qquad \mathbf{x}_N - \mathbf{z}_N \xrightarrow{\text{p}} \mathbf{0},$$

then

$$\mathbf{x}_N \xrightarrow{\text{d}} \mathbf{z}.$$

Lemma

Let $\{x_N\}$ and $\{b_N\}$ be sequences of scalar random variables. If $x_N \xrightarrow{\text{d}} x$ and $b_N \xrightarrow{\text{p}} b$, where x is a random variable and b is a constant, then

(1) $x_N + b_N \xrightarrow{\text{d}} x + b$,

(2) $x_N \cdot b_N \xrightarrow{\text{d}} x \cdot b$

(3) $x_N / b_N \xrightarrow{\text{d}} x/b$, *provided* $b \neq 0$.

- Unfortunately, theorems are labeled differently in different textbooks.
- This one can also be found as Slutzky's theorem.
- In Wooldridge's textbook, Cramer's theorem is not explicitly mentioned because it is a consequence of the asymptotic equivalence lemma.
- For example, to prove (1), define $y_N = x_N + b_N$ and $z_N = x_N + b$. By the continuous mapping theorem, $z_N = x_N + b \xrightarrow{\text{d}} x + b$. In addition, $y_N - z_N = b_N - b \xrightarrow{\text{p}} 0$. Hence, by the asymptotic equivalence lemma, $y_N = x_N + b_N \xrightarrow{\text{d}} x + b$.

4. Limit theorems for random samples

Weak law of large numbers (WLLN)

Theorem

Let $\{\mathbf{w}_i : i = 1, 2, \ldots\}$ be a sequence of independent, identically distributed (i.i.d.) $G \times 1$ random vectors such that $\mathrm{E}(|w_{ig}|) < \infty, g = 1, \ldots, G$. Then the sequence satisfies the **weak law of large numbers (WLLN)***:*

$$N^{-1} \sum_{i=1}^{N} \mathbf{w}_i \xrightarrow{\text{p}} \boldsymbol{\mu}_w,$$

where $\boldsymbol{\mu}_w \equiv \mathrm{E}(\mathbf{w}_i)$.

Comments:

- The condition $\mathrm{E}(|w_{ig}|) < \infty$ implies that the distribution of w_{ig} has a well-defined mean.
- The WLLN states that for an i.i.d. sample, the sample mean is a consistent estimator of the population mean.
- Under certain conditions, a WLLN also holds for samples of independent, not identically distributed random variables.

Lindeberg-Levy central limit theorem

Theorem

Let $\{\mathbf{w}_i : i = 1, 2, \ldots\}$ *be a sequence of independent, identically distributed* $G \times 1$ *random vectors such that* $\mathrm{E}(w_{ig}^2) < \infty, g = 1, \ldots, G$*, and* $E(\mathbf{w}_i) = \mathbf{0}$*. Then* $\{\mathbf{w}_i \colon i = 1, 2, \ldots\}$ *satisfies the* **central limit theorem (CLT)***; that is,*

$$N^{-1/2} \sum_{i=1}^{N} \mathbf{w}_i \xrightarrow{\mathrm{d}} \mathrm{Normal}(\mathbf{0}, \mathbf{B})$$

where $\mathbf{B} = Var(\mathbf{w}_i) = E(\mathbf{w}_i \mathbf{w}_i')$ *is necessarily positive semidefinite. For our purposes,* $\mathbf{B}$ *is almost always positive definite.*

Discussion

Consider an iid sequence of random variables $\mathbf{w}_i$ with possibly nonzero mean $\boldsymbol{\mu}_w$.

Based on the WLLN and the CLT, the sample mean is consistent

$$\bar{\mathbf{w}} = N^{-1} \sum_{i=1}^{N} \mathbf{w}_i \xrightarrow{\text{p}} \boldsymbol{\mu}_w,$$

while $N^{1/2}$ times the sample mean converges in distribution,

$$N^{1/2}(\bar{\mathbf{w}} - \boldsymbol{\mu}_w) = N^{-1/2} \sum_{i=1}^{N} (\mathbf{w}_i - \boldsymbol{\mu}_w) \xrightarrow{\text{d}} \text{Normal}(\mathbf{0}, \mathbf{B}),$$

where $\text{E}(\mathbf{w}_i - \boldsymbol{\mu}_w) = 0$ and $\mathbf{B} = \text{Var}(\mathbf{w}_i - \boldsymbol{\mu}_w) = \text{Var}(\mathbf{w}_i) = \text{E}[(\mathbf{w}_i - \boldsymbol{\mu}_w)(\mathbf{w}_i - \boldsymbol{\mu}_w)']$.

5. Asymptotic properties of estimators

Consistent estimators

Definition

Let $\{\hat{\boldsymbol{\theta}}_N : N = 1, 2, \ldots, \}$ be a sequence of estimators of the $P \times 1$ vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where N indexes the sample size. If

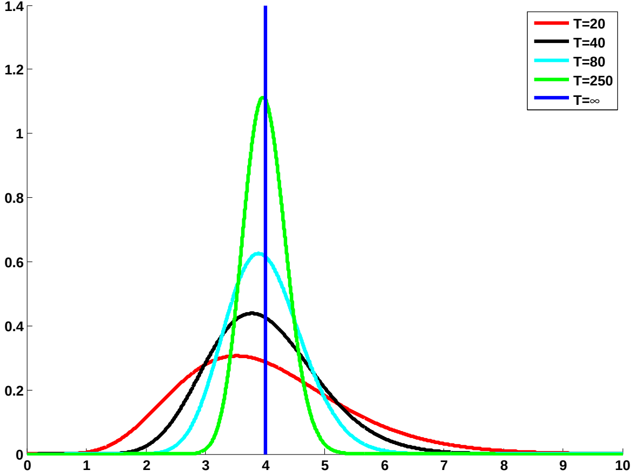
$$\hat{\boldsymbol{\theta}}_N \xrightarrow{p} \boldsymbol{\theta}$$

for any value of $\boldsymbol{\theta}$, then we say $\hat{\boldsymbol{\theta}}_N$ is a **consistent estimator** of $\boldsymbol{\theta}$.

Comment:

- This consistency concept is also called *weak* consistency.

Consistent estimators



Asymptotic normality

Definition

Let $\{\hat{\boldsymbol{\theta}}_N \colon N = 1, 2, \dots\}$ be a sequence of estimators of the $P \times 1$ vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

Suppose that

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}\right) \xrightarrow{\mathrm{d}} \mathrm{Normal}(\mathbf{0}, \mathbf{V}) \tag{1}$$

where $\mathbf{V}$ is a $P \times P$ positive semidefinite matrix.

Then we say that

- $\hat{\boldsymbol{\theta}}_N$ is **$\sqrt{\mathbf{N}}$-asymptotically normally distributed** and
- $\mathbf{V}$ is the **asymptotic variance** of $\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta})$, denoted

$$\mathrm{Avar}\left[\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta})\right] = \mathbf{V}.$$

Approximate normality

In applied work we need to know the finite-sample distribution of the estimator, $\hat{\boldsymbol{\theta}}_N$, rather than the asymptotic distribution of $\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta})$.

Unfortunately, the *exact* finite-sample distribution of $\hat{\boldsymbol{\theta}}_N$ is generally unknown.

Therefore, we use the asymptotic distribution to approximate the finite-sample distribution of $\hat{\boldsymbol{\theta}}_N$:

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}) \approx \mathbf{z} \sim \text{Normal}(\mathbf{0}, \mathbf{V}).$$

This implies that

$$\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta} \approx \mathbf{z}/\sqrt{N} \sim \text{Normal}(\mathbf{0}, \mathbf{V}/N)$$

and thus

$$\hat{\boldsymbol{\theta}}_N \approx \mathbf{z}/\sqrt{N} + \boldsymbol{\theta} \sim \text{Normal}(\boldsymbol{\theta}, \mathbf{V}/N).$$

Hence, we treat $\hat{\boldsymbol{\theta}}_N$ *as if*

$$\hat{\boldsymbol{\theta}}_N \sim \text{Normal}(\boldsymbol{\theta}, \mathbf{V}/N).$$

Asymptotic variance

Using the approximation

$$\hat{\boldsymbol{\theta}}_N \sim \text{Normal}(\boldsymbol{\theta}, \mathbf{V}/N),$$

we call $\mathbf{V}/N$ the **asymptotic variance** of $\hat{\boldsymbol{\theta}}_N$,

$$\mathbf{V}_{\hat{\boldsymbol{\theta}}} \equiv \text{Avar}(\hat{\boldsymbol{\theta}}_N) = \mathbf{V}/N.$$

Comments:

- In a strict sense, this statement is not meaningful because $\mathbf{V}/N$ converges to zero.
- It is just a short form to saying that $\mathbf{V}$ is the asymptotic variance of $\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta})$ which is used to derive the approximate normal distribution of $\hat{\boldsymbol{\theta}}_N$.
- How good is the approximation? Generally, it should be fairly good for samples of $N > 30$ and really good for samples of $N > 100$.

Consistent estimation of the asymptotic variance

Somewhat sloppily, we briefly call

$$\hat{\mathbf{V}}_{\hat{\boldsymbol{\theta}}} \equiv \widehat{\text{Avar}}(\hat{\boldsymbol{\theta}}_N) = \hat{\mathbf{V}}_N/N$$

a consistent estimator of $\text{Avar}(\hat{\boldsymbol{\theta}}_N) = \mathbf{V}/N$ whenever we really mean that

$$\hat{\mathbf{V}}_N \xrightarrow{\text{p}} \mathbf{V} = \text{Avar}\left[\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta})\right].$$

This is again an imprecise shorthand because, in a strict sense, the statement "$\hat{\mathbf{V}}_N/N$ is a consistent estimator of $\text{Avar}(\hat{\boldsymbol{\theta}}_N)$" is not very meaningful. To see this note that

$$\hat{\mathbf{V}}_N \xrightarrow{\text{p}} \mathbf{V}$$

is not necessary for

$$\hat{\mathbf{V}}_N/N \xrightarrow{\text{p}} \mathbf{V}/N$$

because $\mathbf{V}/N$ converges to zero. There are obviously many estimators $\hat{\mathbf{V}}_N$ which do not converge to $\mathbf{V}$ but for which $\hat{\mathbf{V}}_N/N$ converge to zero.

Definition

If

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}) \overset{a}{\sim} \text{Normal}(0, \mathbf{V}),$$

where $\mathbf{V}$ is positive definite with jth diagonal element v_{jj}, and if there is a consistent estimator

$$\hat{\mathbf{V}}_N \xrightarrow{p} \mathbf{V},$$

then the **asymptotic standard error** of $\hat{\boldsymbol{\theta}}_{Nj}$, is

$$\text{se}(\hat{\boldsymbol{\theta}}_{Nj}) = (\hat{\mathbf{v}}_{Njj}/N)^{1/2}.$$

Root-N consistency

Note that if $\mathbf{x}_N \xrightarrow{\text{d}} \mathbf{x}$, where $\mathbf{x}$ is any $K \times 1$ random vector, then $\mathbf{x}_N = \mathbf{O}_p(1)$.

Hence, from

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}\right) \xrightarrow{\text{d}} \text{Normal}(\mathbf{0}, \mathbf{V})$$

follows that

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}\right) = \mathbf{O}_p(1)$$

and thus

$$\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta} = \mathbf{O}_p(N^{-1/2}).$$

This implies that $\hat{\boldsymbol{\theta}}_N$ converges to $\boldsymbol{\theta}$ at a rate of $\sqrt{N}$. This is called $\sqrt{N}$-consistency.

In this course, we will only study $\sqrt{N}$-consistent estimators.

Definition

Let $\hat{\boldsymbol{\theta}}_N$ and $\tilde{\boldsymbol{\theta}}_N$ be estimators of $\boldsymbol{\theta}$ satisfying

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}\right) \xrightarrow{\text{d}} \text{Normal}(\mathbf{0}, \mathbf{V})$$

and

$$\sqrt{N}\left(\tilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}\right) \xrightarrow{\text{d}} \text{Normal}(\mathbf{0}, \mathbf{D}),$$

respectively. Then

(1) $\hat{\boldsymbol{\theta}}_N$ is **asymptotically efficient relative to** $\tilde{\boldsymbol{\theta}}_N$ if $\mathbf{D} - \mathbf{V}$ is positive semidefinite for all $\boldsymbol{\theta}$;

(2) $\hat{\boldsymbol{\theta}}_N$ and $\tilde{\boldsymbol{\theta}}_N$ are **$\sqrt{\mathbf{N}}$-equivalent** if $\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \tilde{\boldsymbol{\theta}}_N) = o_p(1)$.

6. Asymptotic properties of test statistics

Asymptotic size of a test

Definition

The **asymptotic size** of a testing procedure is defined as the limiting probability of rejecting H_0 when it is true. Mathematically, we can write this as

$$\text{asymptotic size} = \lim_{N \to \infty} \mathrm{P}_N(\text{reject } H_0 | H_0),$$

where the N subscript indexes the sample size.

Consistency of a test

Definition

A test is said to be **consistent** against the alternative H_1 if the null hypothesis is rejected with probability approaching one when H_1 is true:

$$\lim_{N\to\infty} \mathrm{P}_N(\text{reject } H_0|H_1) = 1.$$

t tests

Assume

$$\sqrt{N}(\hat{\theta}_N - \theta) \overset{a}{\sim} \text{Normal}(0, v^2).$$

We want to test the null hypothesis $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$.

Under H_0, the distribution of the t-statistic for known variance v^2 is

$$\sqrt{N}(\hat{\theta}_N - \theta_0)/v \overset{a}{\sim} \text{Normal}(0, 1).$$

The drawback of this statistic is that the variance v^2 is typically unknown. Therefore, replace it by a consistent estimator $\hat{v}^2$. Then, by Cramer's theorem, the asymptotic distribution of the resulting t-statistic remains unchanged:

$$\sqrt{N}(\hat{\theta}_N - \theta_0)/\hat{v} \overset{a}{\sim} \text{Normal}(0, 1).$$

To decide the test, the t-statistic has to be compared with the appropriate critical value of the standard normal distribution.

Note that this test has correct size only asymptotically. In finite samples, it holds approximately.

Also note that the t_∞-distribution is equivalent to the standard normal distribution.

To see that the t test is consistent, consider the case that H_1 is true, i.e., the true θ is $\theta_1 \neq \theta_0$. Then

$$\sqrt{N}(\hat{\theta}_N - \theta_1)/\hat{v} \overset{a}{\sim} \text{Normal}(0, 1).$$

The t statistic is

$$\sqrt{N}(\hat{\theta}_N - \theta_0)/\hat{v} = \sqrt{N}(\hat{\theta}_N - \theta_1 + \theta_1 - \theta_0)/\hat{v} = \sqrt{N}(\hat{\theta}_N - \theta_1)/\hat{v} + \sqrt{N}(\theta_1 - \theta_0)/\hat{v}.$$

Note that the first part converges to a standard normal random variable while the second part is $\mathbf{O}_p(\sqrt{N})$ because $\theta_1 - \theta_0 \neq 0$ and $\hat{v} \xrightarrow{p} v$.

Thus, under H_1 the t statistic (absolutely) grows over all limits as $T \to \infty$. This leads to a rejection with probability one if the test statistic is compared to a critical value from the standard normal distribution.

Wald tests

Assume

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}) \overset{a}{\sim} \text{Normal}(0, \mathbf{V}),$$

where $\mathbf{V}$ is positive definite. Now define a nonrandom $Q \times P$ matrix $\mathbf{R}$ with (a) $Q \leq P$ and (b) $\text{rank}(\mathbf{R}) = Q$. Then the asymptotic distribution of the linear function $\mathbf{R}\boldsymbol{\theta}$ is given by

$$\sqrt{N}\mathbf{R}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}) \overset{a}{\sim} \text{Normal}(0, \mathbf{RVR}').$$

Consequently, the distribution of the quadratic form is

$$\left[\sqrt{N}\mathbf{R}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta})\right]' \left[\mathbf{RVR}'\right]^{-1} \left[\sqrt{N}\mathbf{R}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta})\right] \overset{a}{\sim} \chi^2_Q.$$

The drawback of this quadratic form is that the variance $\mathbf{V}$ is typically unknown. Therefore, replace it by a consistent estimator $\hat{\mathbf{V}}_N$. Then, by Cramer's theorem, the asymptotic distribution of the resulting quadratic form remains unchanged:

$$\left[\mathbf{R}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta})\right]' \left[\mathbf{R}(\hat{\mathbf{V}}_N/N)\mathbf{R}'\right]^{-1} \left[\mathbf{R}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta})\right] \overset{a}{\sim} \chi^2_Q.$$

Wald tests

We can use this result to test the general null hypothesis $H_0 : \mathbf{R}\boldsymbol{\theta} = \mathbf{r}$ against $H_1 : \mathbf{R}\boldsymbol{\theta} \neq \mathbf{r}$.

Under H_0, $\mathbf{R}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}) = \mathbf{R}\hat{\boldsymbol{\theta}}_N - \mathbf{R}\boldsymbol{\theta} = \mathbf{R}\hat{\boldsymbol{\theta}}_N - \mathbf{r}$. Thus, under H_0 the Wald statistic

$$W_N \equiv \left[\mathbf{R}\hat{\boldsymbol{\theta}}_N - \mathbf{r}\right]' \left[\mathbf{R}(\hat{\mathbf{V}}_N/N)\mathbf{R}'\right]^{-1} \left[\mathbf{R}\hat{\boldsymbol{\theta}}_N - \mathbf{r}\right]$$

is again asymptotically χ^2_Q distributed:

$$W_N \overset{a}{\sim} \chi^2_Q.$$

To decide the test, the Wald statistic has to be compared with the appropriate critical value of the χ^2_Q distribution.

Relationship between the Wald statistic and the F-statistic:

$$W_N = Q \times F.$$

Hence, the asymptotic distribution of the F statistic is $\chi^2_Q/Q = F_{Q,\infty}$.