6 Convergence concepts. Asymptotics of OLS.

6.1 Convergence concepts

Definition 6.1.1: Convergence in probability

A sequence of random scalars $\{z_i\}_{i=1}^{\infty}$ converges in probability to z iff $\forall \varepsilon > 0$, $\lim_{n\to\infty} P(|z_n - z| \geq \varepsilon) = 0$, or equivalently $\lim_{n\to\infty} P(|z_n - z| < \varepsilon) = 1$. Written as $z_n \xrightarrow{p} z$ or $z_n - z = o_p(1)$ or $\text{plim}_{n\to\infty} z_n = z$.

This definition is extended to a sequence of random vectors or random matrices by requiring element-by-element convergence in probability. That is, a sequence of K-dimensional vectors $\mathbf{z_n}$ converges in probability to a K-dimensional vector $\mathbf{z}$ if, for any $\varepsilon > 0$

$$\lim_{n\to\infty} P(|z_{nk} - z_k| > \varepsilon) = 0 \quad \text{for all} k = 1, 2, ..., K$$

where z_{nk} is the k-th element of $\mathbf{z_n}$ and z_k the k-th element of $\mathbf{z}$.

Exercise 6.1.1. Let X_n be an IID sequence of continuous random variables having a uniform distribution over support

$$R_{X_n} = \left[-\frac{1}{n}, \frac{1}{n}\right]$$

with pdf

$$f_{X_n}(x) = \begin{cases} \frac{n}{2} & \text{if } x \in \left[-\frac{1}{n}; \frac{1}{n}\right] \\ 0 & \text{if } x \notin \left[-\frac{1}{n}; \frac{1}{n}\right] \end{cases}$$

Find the probability limit (if it exists) of the sequence X_n.

Solution:-

Intuitively as $n \to \infty$ the probability density becomes concentrated around $x = 0$; it seems reasonable to conjecture $X_n \xrightarrow{p} X = 0$. To show this formally, for any $\varepsilon > 0$:

$$\begin{aligned}
\lim_{n\to\infty} P(|X_n - X| > \varepsilon) &= \lim_{n\to\infty} P(|X_n - 0| > \varepsilon) \\
&= \lim_{n\to\infty} [1 - P(-\varepsilon \leq X_n \leq \varepsilon)] \\
&= 1 - \lim_{n\to\infty} \int_{-\varepsilon}^{\varepsilon} f_{X_n}(x)dx \\
&= 1 - \lim_{n\to\infty} \int_{\max(-\varepsilon,-1/n)}^{\min(\varepsilon,1/n)} \frac{n}{2}dx \quad (f(x) \text{ has no density outside } [-\tfrac{1}{n}, \tfrac{1}{n}]) \\
&= 1 - \lim_{n\to\infty} \int_{-1/n}^{1/n} \frac{n}{2}dx \quad (\text{when } n \text{ becomes large, } \frac{1}{n} < \varepsilon) \\
&= 1 - \lim_{n\to\infty} 1 \\
&= 0
\end{aligned}$$

Definition 6.1.2: Convergence in distribution

A sequence of random scalars $\{z_i\}_{i=1}^{\infty}$ converges in distribution to z iff, $\lim_{n\to\infty} F_{z_n}(z) = F_z(z)$ at all points where F_z is continuous. Written as $z_n \xrightarrow{d} z$ or $z_n - z = O_p(1)$ or as "z is the limiting distribution of z_n".

Convergence in distribution is also known as weak convergence or the convergence in law.

Theorem 6.1.1. $\mathbf{z_n} \xrightarrow{d} \mathbf{z}$ iff $\mathbb{E}f(\mathbf{z_n}) \to \mathbb{E}f(\mathbf{z})$ for all bounded, continuous functions f.

Claim 6.1.1. Convergence in probability implies convergence in distribution but not vice versa. The reverse only holds when the limit in distribution is a constant.

Example ($z_n \xrightarrow{d} z \not\Rightarrow z_n \xrightarrow{p} z$). Let $z \sim N(0,1)$. Let $z_n = -z$ for $n = 1, 2, 3, \ldots$; hence $z_n \sim N(0,1)$. z_n has the same distribution function as z for all n so, trivially, $\lim_{n\to\infty} F_n(x) = F(x)$ for all x. Therefore, $z_n \xrightarrow{d} z$. But $P(|z_n - z| > \varepsilon) = P(|2z| > \varepsilon) = P(|z| > \varepsilon/2) \neq 0$. So z_n does not tend to z in probability.

The extension to a sequence of random vectors is immediate: $\mathbf{z_n} \xrightarrow{d} \mathbf{z}$ if the joint c.d.f. F_n of the random vector $\mathbf{z_n}$ converges to the joint c.d.f. F of $\mathbf{z}$ at every continuity point of F. However, element-by-element convergence does not necessarily imply convergence for the vector sequence (unlike with convergence in probability). Intuitively this is because different c.d.f.'s can have the same marginals.
A common way to establish the connection between scalar convergence in distribution and vector convergence in distribution is for every linear combination of z_{nk} to converge to the linear combination of z_n. Formally:

Definition 6.1.3: Cramer-Wold device

$\mathbf{z_n} \xrightarrow{d} \mathbf{z}$ if and only if $\lambda'\mathbf{z_n} \xrightarrow{d} \lambda'\mathbf{z}$ for every $\lambda \in \mathbb{R}^k$ with $\lambda'\lambda = 1$.

Note:-

Big O Little o notation

- Roughly speaking, a function is $o(z)$ iff it's of lower asymptotic order than z.
- $f(n) = o(g(n))$ iff $\lim_{n\to\infty} f(n)/g(n) = 0$.
- If $\{f(n)\}$ is a sequence of random variables, then $f(n) = o_p(g(n))$ iff $plim_{n\to\infty} f(n)/g(n) = 0$.
- We write $X_n - X = o_p(n^{-\gamma})$ iff $n^{\gamma}(X_n - X) \xrightarrow{\text{p}} 0$.

- Roughly speaking, a function is $O(z)$ iff it's of the same asymptotic order as z.
- $f(n) = O(g(n))$ iff $|f(n)/g(n)| < K$ for all $n > N$ and some positive integer N and some constant $K > 0$.
- If $\{f(n)\}$ is a sequence of random variables, then $f(n) = o_p(g(n))$ iff $plim_{n\to\infty} f(n)/g(n) = 0$.

Definition 6.1.4: Continuous mapping theorem (CMT)

Let f be continuous at every point $a \in C$ where $P(z \in C) = 1$. Then

1. If $\mathbf{z_n} \xrightarrow{p} \mathbf{z}$, then $f(\mathbf{z_n}) \xrightarrow{p} f(\mathbf{z})$
2. If $\mathbf{z_n} \xrightarrow{d} \mathbf{z}$, then $f(\mathbf{z_n}) \xrightarrow{d} f(\mathbf{z})$

Example. The CMT allows f to be discontinuous only if the probability of being at a discontinuity point is zero.
Consider $f(u) = u^{-1}$ is discontinuous at $u = 0$, but if $z_n \xrightarrow{d} z \sim N(0,1)$ then $P(z = 0) = 0$ so $z_n^{-1} \xrightarrow{d} z^{-1}$

Corollary 6.1.1 (Slutsky's theorem). If $z_n \xrightarrow{d} z$ and $c_n \xrightarrow{p} c$ as $n \to \infty$, then

1. $z_n + c_n \xrightarrow{d} z + c$
2. $z_n c_n \xrightarrow{d} zc$
3. $\frac{z_n}{c_n} \xrightarrow{d} \frac{z}{c}$ if $c \neq 0$.

The requirement that c_n converges to a constant is important. If it were to converge to a non-degenerate random variable, the theorem would be no longer valid. For example, let $z_n \sim \text{Uniform}(0,1)$ and $c_n = -z_n$. The sum $z_n + c_n = 0$ for all values of n. Moreover, $c_n \xrightarrow{d} c$ where $z \sim \text{Uniform}(0,1)$, $c \sim \text{Uniform}(-1,0)$, and z and c are independent.

Note:-
The theorem remains valid if we replace all convergences in distribution with convergences in probability.

Proof. This theorem follows from the fact that if z_n converges in distribution to z and c_n converges in probability to a constant c, then the joint vector (z_n, c_n) converges in distribution to (z, c).
Next we apply the continuous mapping theorem, recognising the functions $g(z,c)$ such as $g(z,c) = z + c$, $g(z,c) = zc$, and $g(z,c) = zc^{-1}$ are continuous (for the last function to be continuous, c has to be invertible). $\square$

Definition 6.1.5: Khinchine's law of large numbers

If Y_i are i.i.d. with finite mean $\mathbb{E}Y_i = m < \infty$ then $\frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow{p} m$

Lemma 6.1.1 (Markov's inequality). Let ξ be a non-negative random variable and let $\varepsilon > 0$ be a positive number. Then for any real number $p > 0$, the following inequality holds:

$$P(|\xi| \geq \varepsilon) \leq \frac{E[|\xi|^p]}{\varepsilon^p}.$$

Proof. Let ξ be a non-negative random variable and $\varepsilon > 0$. For any positive integer p:

$$\begin{aligned}
E[|\xi|^p] &= \int_0^\infty x^p f_\xi(x)\,dx && \text{(expectation definition)}\\
&= \int_0^\varepsilon x^p f_\xi(x)\,dx + \int_\varepsilon^\infty x^p f_\xi(x)\,dx && \text{(splitting the integral)}\\
&\geq \int_\varepsilon^\infty \varepsilon^p f_\xi(x)\,dx && \text{(since } x^p \geq \varepsilon^p \text{ for } x \geq \varepsilon)\\
&= \varepsilon^p P(|\xi| \geq \varepsilon) && \text{(definition of probability)}\\
P(|\xi| \geq \varepsilon) &\leq \frac{E[|\xi|^p]}{\varepsilon^p} && \text{(Markov's inequality)}
\end{aligned}$$

□

Lemma 6.1.2 (Chebyshev's inequality). Let η be a random variable with $\mathbb{E}[\eta] = m$ and $\mathrm{Var}(\eta) < \infty$. Then for any $\varepsilon > 0$,

$$P(|\eta - \mathbb{E}[\eta]| \geq \varepsilon) \leq \frac{\mathrm{Var}(\eta)}{\varepsilon^2}.$$

Proof. Using Markov's inequality, for any random variable η with finite expectation $E[\eta]$ and finite non-zero variance $\mathrm{Var}(\eta)$, and for any $\varepsilon > 0$, we have:

$$\begin{aligned}
P(|\eta - E[\eta]| \geq \varepsilon) &= P((\eta - E[\eta])^2 \geq \varepsilon^2) && \text{(squaring both sides)}\\
&\leq \frac{E[(\eta - E[\eta])^2]}{\varepsilon^2} && \text{(applying Markov's inequality)}\\
&= \frac{\mathrm{Var}(\eta)}{\varepsilon^2}. && \text{(variance definition)}
\end{aligned}$$

□

Definition 6.1.6: Chebyshev's law of large numbers

If Y_i are uncorrelated, and $\mathbb{E}Y_i = m < \infty$, $Var(Y_i) = \sigma_i^2 < \infty$ and $\frac{1}{n^2}\sum_{i=1}^n \sigma_i^2 \to 0$, then $\frac{1}{n}\sum_{i=1}^n (Y_i - m) \xrightarrow{p} 0$

Proof. Let $Y_1, Y_2, \ldots, Y_n$ be uncorrelated random variables with $E[Y_i] = m$ and $\mathrm{Var}(Y_i) = \sigma_i^2 < \infty$. Assume that $\frac{1}{n^2}\sum_{i=1}^n \sigma_i^2 \to 0$ as $n \to \infty$. Define $S_n = \frac{1}{n}\sum_{i=1}^n (Y_i - m)$. We want to show that $S_n \to 0$ in probability. By Chebyshev's inequality, for any $\varepsilon > 0$,

$$P(|S_n - E[S_n]| \geq \varepsilon) \leq \frac{\mathrm{Var}(S_n)}{\varepsilon^2}.$$

Since $E[S_n] = 0$ and the Y_i's are uncorrelated, we have

$$\mathrm{Var}(S_n) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n (Y_i - m)\right) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(Y_i - m) = \frac{1}{n^2}\sum_{i=1}^n \sigma_i^2.$$

$$\Rightarrow P(|S_n| \geq \varepsilon) \leq \frac{1}{n^2}\frac{\sum_{i=1}^n \sigma_i^2}{\varepsilon^2}.$$

Since $\frac{1}{n^2}\sum_{i=1}^n \sigma_i^2 \to 0$, it follows that for any $\varepsilon > 0$,

$$P(|S_n| \geq \varepsilon) \to 0 \text{ as } n \to \infty.$$

Hence, $S_n \to 0$ in probability. $\square$

Definition 6.1.7: Univariate Lindeberg-Lévy Central Limit Theorem

If Y_i are i.i.d. random variables with finite mean m and variance σ^2, then

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n Y_i - m\right) \xrightarrow{d} N(0, \sigma^2)$$

Definition 6.1.8: Multivariate Lindeberg-Lévy Central Limit Theorem

If Y_i are i.i.d. with mean m and variance-covariance Σ, then

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n Y_i - m\right) \xrightarrow{d} N(0, \Sigma).$$

Proof. Set $\mathbf{c} \in \mathbb{R}^k$ with $\mathbf{c}'\mathbf{c} = 1$ and define $u_i = \mathbf{c}'(\mathbf{y}_i - \mathbf{m})$. The u_i are i.i.d. with $E(u_i^2) = \mathbf{c}'\mathbf{\Sigma}\mathbf{c} < \infty$. By the univariate CLT,

$$\mathbf{c}'\sqrt{n}(\bar{\mathbf{y}} - \mathbf{m}) = \frac{1}{\sqrt{n}}\sum_{i=1}^n u_i \xrightarrow{d} N(0, \mathbf{c}'\mathbf{\Sigma}\mathbf{c})$$

Notice that if $\mathbf{z} \sim N(0, \mathbf{\Sigma})$ then $\mathbf{c}'\mathbf{z} \sim N(0, \mathbf{c}'\mathbf{\Sigma}\mathbf{c})$. Thus

$$\mathbf{c}'\sqrt{n}(\bar{\mathbf{y}} - \mathbf{m}) \xrightarrow{d} \mathbf{c}'\mathbf{z}.$$

Since this holds for all $\mathbf{c}$, we can use the Cramer-Wold device:

$$\sqrt{n}(\bar{\mathbf{y}} - \mathbf{m}) \xrightarrow{d} \mathbf{z} \sim N(0, \mathbf{\Sigma})$$

$\square$

6.2 OLS in large samples

(OLS0)	(y_i, x_i) is an i.i.d. sequence		
(OLS1)	$E(x_i x_i')$ is finite non-singular	(GM1)	rank $\mathbf{X} = k$
(OLS2)	$E(y_i \mid x_i) = x_i'\beta$	(GM2)	$E(\mathbf{Y} \mid \mathbf{X}) = \mathbf{X}'\beta$
(OLS3)	$\text{Var}(y_i \mid x_i) = \sigma^2$	(GM3)	$\text{Var}(\mathbf{Y} \mid \mathbf{X}) = \sigma^2\mathbf{I}$
(OLS4)	$E\varepsilon_i^4 < \infty, \quad E\|x_i\|^4 < \infty$		

Remarks
(OLS0): Equivalent to random sampling, tells us that the pairs (x_i, y_i) are independent across i.
(OLS1): Ensures $\mathbf{X}'\mathbf{X}$ is invertible, or comparatively in sample $\frac{1}{n}\sum_{i=1}^n x_i x_i'$ exists.
(OLS2): Since all other x's are independent, this is equivalent to conditioning on all x's
(OLS3): Homoskedasticity and no serial correlation
(OLS4): Implies the existence of $\mathbb{E}(\varepsilon_i^2 x_i x_i')$ via Cauchy-Schwartz. This is required to use the CLT.

Lemma 6.2.1 (Expectation inequality). For any random vector $Y \in \mathbb{R}^m$ with $\mathbb{E}\|Y\| < \infty$ then

$$\|\mathbb{E}[Y]\| \leq E\|Y\|$$

Lemma 6.2.2 (Holder's inequality). If $p > 1$ and $q > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$, then for any random $m \times n$ matrices X and Y,

$$(\mathbb{E}\|X'Y\|) \leq (\mathbb{E}\|X\|^p)^{1/p} (\mathbb{E}\|Y\|^q)^{1/q}$$

Corollary 6.2.1 (Cauchy-Schwartz inequality). For any random $m \times n$ matrices X and Y,

$$(\mathbb{E}\|X'Y\|) \leq \left(\mathbb{E}\|X\|^2\right)^{1/2} \left(\mathbb{E}\|Y\|^2\right)^{1/2}$$

To see that the elements of $\mathbb{E}(\varepsilon_i^2 x_i x_i')$ are finite:

$$\begin{aligned}
\|\mathbb{E}(\varepsilon_i^2 x_i x_i')\| &\leq \mathbb{E}\|\varepsilon_i^2 x_i x_i'\| && \text{(using Lemma 6.2.1)} \\
&= \mathbb{E}(\varepsilon_i^2 \|x_i\|^2) \\
&\leq \mathbb{E}\left(\varepsilon_i^4\right)^{1/2} \mathbb{E}\left(\|x_i\|^4\right)^{1/2} && \text{(using Corollary 6.2.1)} \\
&< \infty && \text{(using OLS4)}
\end{aligned}$$

Theorem 6.2.1. Under OLS0-4:

1. $\hat{\beta}_{OLS} \xrightarrow{p} \beta$
2. $\sqrt{n}(\hat{\beta}_{OLS} - \beta) \xrightarrow{d} \mathrm{N}\big(0, \sigma^2[\mathbb{E}(x_i x_i')]^{-1}\big)$

Proof. 1. We only require OLS0-2 for consistency[a]

$$\begin{aligned}
\hat{\beta}_{OLS} &= (X'X)^{-1}X'Y \\
&= \beta + (X'X)^{-1}X'\varepsilon \\
&= \beta + \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} x_i \varepsilon_i
\end{aligned}$$

Since $x_i \varepsilon_i$ is i.i.d. by OLS0[b] we can use Khinchine's LLN

$$\frac{1}{n}\sum_{i=1}^{n} x_i x_i' \xrightarrow{p} \mathbb{E}(x_i x_i') \quad \text{and} \quad \begin{aligned}[t] \frac{1}{n}\sum_{i=1}^{n} x_i \varepsilon_i &\xrightarrow{p} \mathbb{E}(x_i \varepsilon_i) \\ &= \mathbb{E}(\mathbb{E}(x_i \varepsilon_i | x_i)) \\ &= 0 \quad \text{(using OLS2)} \end{aligned}$$

By the Continuous Mapping Theorem,

$$\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \xrightarrow{p} [\mathbb{E}(x_i x_i')]^{-1} \quad \text{(exists due to OLS1)}$$

$$\Rightarrow \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} x_i \varepsilon_i \xrightarrow{p} 0$$

2.

$$\hat{\beta}_{OLS} - \beta = (X'X)^{-1}X'\varepsilon = \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} x_i\varepsilon_i$$

$$\Rightarrow \sqrt{n}\left(\hat{\beta}_{OLS} - \beta\right) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i\varepsilon_i$$

Using the CLT:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i\varepsilon_i \xrightarrow{d} N(0, Var(x_i\varepsilon_i)) = N\left(0, \sigma^2\mathbb{E}(x_i x_i')\right)$$

Where the second equality follows from:

$$\begin{aligned}
\text{Var}(x_i\varepsilon_i) &= E[x_i\varepsilon_i\varepsilon_i' x_i'] - E[x_i\varepsilon_i]E[x_i\varepsilon_i]' \\
&= E[\varepsilon_i^2 x_i x_i'] - E[x_i\varepsilon_i]E[x_i\varepsilon_i]' \quad \text{(since } \varepsilon_i \text{ scalar)} \\
&= E[E(\varepsilon_i^2 x_i x_i'|x_i)] - E[E(x_i\varepsilon_i|x_i)]E[x_i\varepsilon_i]' \quad \text{(first expectation exists by OLS4)} \\
&= E[E(\varepsilon_i^2|x_i)x_i x_i'] - E[x_i E(\varepsilon_i|x_i)]E[x_i\varepsilon_i]' \\
&= \sigma^2 E[x_i x_i']. \quad \text{(using OLS2)}
\end{aligned}$$

Using the CMT:

$$\begin{aligned}
\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i\varepsilon_i &\xrightarrow{d} [\mathbb{E}(x_i x_i')]^{-1} N\left(0, \sigma^2\mathbb{E}(x_i x_i')\right) \\
&\sim N\left(0, [\mathbb{E}(x_i x_i')]^{-1}\sigma^2\mathbb{E}(x_i x_i')[\mathbb{E}(x_i x_i')]^{-1}\right) \\
\sqrt{n}\left(\hat{\beta}_{OLS} - \beta\right) &\xrightarrow{d} N\left(0, \sigma^2[\mathbb{E}(x_i x_i')]^{-1}\right)
\end{aligned}$$

□

[a]Strictly we only need OLS0,1,2': $\mathbb{E}(x_i\varepsilon_i) = 0$

[b]$x_i\varepsilon_i = x_i(y_i - x_i'\beta)$ and we know (y_i, x_i) i.i.d.

Theorem 6.2.2. Under OLS0-4:

1. $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$
2. $W \xrightarrow{d} \chi^2(p)$
3. $t \xrightarrow{d} N(0,1)$

Proof. 1.[a]

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-k}\varepsilon' M_X \varepsilon \\
&= \frac{1}{n-k}\varepsilon'(I - X(X'X)^{-1}X')\varepsilon \\
&= \frac{1}{n-k}\varepsilon'\varepsilon - \frac{1}{n-k}\varepsilon' X(X'X)^{-1}X'\varepsilon \\
&= \frac{n}{n-k}\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 - \frac{n}{n-k}\frac{1}{n}\sum_{i=1}^{n} x_i\varepsilon_i \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} x_i'\varepsilon_i
\end{aligned}$$

Using Khinchine's LLN:

$$\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 \xrightarrow{p} \mathbb{E}[\varepsilon_i^2], \quad \frac{1}{n}\sum_{i=1}^{n}x_i\varepsilon_i \xrightarrow{p} \mathbb{E}[x_i\varepsilon_i]=0, \quad \frac{1}{n}\sum_{i=1}^{n}x_ix_i' \xrightarrow{p} \mathbb{E}(x_ix_i'), \quad \frac{1}{n}\sum_{i=1}^{n}x_i'\varepsilon_i \xrightarrow{p} \mathbb{E}[x_i'\varepsilon_i]=0$$

Using CMT and Slutsky:

$$\begin{aligned}\hat{\sigma}^2 &\xrightarrow{p} \frac{n}{n-k}\mathbb{E}[\varepsilon_i^2] + \frac{n}{n-k}\times 0\\ &= \mathbb{E}[\varepsilon_i^2] \quad (\text{as } n\to\infty)\\ &= \sigma^2\end{aligned}$$

2.

$$W = \frac{\sqrt{n}\left(R\hat{\beta}-q\right)'\left(\sigma^2R\left(\frac{1}{n}X'X\right)^{-1}R'\right)^{-1}\sqrt{n}\left(R\hat{\beta}-q\right)}{\hat{\sigma}^2/\sigma^2}$$

We have seen that $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$, and

$$\begin{aligned}\sqrt{n}\left(R\hat{\beta}-q\right) &= \sqrt{n}(\hat{\beta}-\beta) \quad (\text{since } H_0: R\beta=q)\\ &\xrightarrow{d} RN\left(0,\sigma^2[\mathbb{E}(x_ix_i')]^{-1}\right)\\ &= N\left(0,\sigma^2R[\mathbb{E}(x_ix_i')]^{-1}R'\right)\\ &= \left(\sigma^2R[\mathbb{E}(x_ix_i')]^{-1}R'\right)^{1/2}N(0,I_p)\end{aligned}$$

Since $\frac{1}{n}X'X \xrightarrow{p} \mathbb{E}[x_ix_i']$, by the CMT,

$$\left(\sigma^2R\left(\frac{1}{n}X'X\right)^{-1}R'\right)^{-1} \xrightarrow{p} \left(\sigma^2R[\mathbb{E}(x_ix_i')]^{-1}R'\right)^{-1}$$

$$\begin{aligned}\Rightarrow W &\xrightarrow{d} \frac{\left(\left(\sigma^2R[\mathbb{E}(x_ix_i')]^{-1}R'\right)^{1/2}N(0,I_p)\right)'\left(\sigma^2R[\mathbb{E}(x_ix_i')]^{-1}R'\right)^{-1}\left(\sigma^2R[\mathbb{E}(x_ix_i')]^{-1}R'\right)^{-1/2}N(0,I_p)}{1}\\ &= (N(0,I_p))'\left(\sigma^2R[\mathbb{E}(x_ix_i')]^{-1}R'\right)^{1/2}\left(\sigma^2R[\mathbb{E}(x_ix_i')]^{-1}R'\right)^{-1}\left(\sigma^2R[\mathbb{E}(x_ix_i')]^{-1}R'\right)^{1/2}N(0,I_p)\\ &= (N(0,I_p))'I_pN(0,I_p)\\ &= \chi^2(p)\end{aligned}$$

3.

$$\begin{aligned}t &= \frac{\hat{\beta}_j-\beta}{\sqrt{\hat{\sigma}^2(X'X)^{-1}_{jj}}}\\ &= \frac{(\hat{\beta}_j-\beta)/\sqrt{\sigma^2(X'X)^{-1}_{jj}}}{\sqrt{\hat{\sigma}^2/\sigma^2}}\\ &= \frac{\xrightarrow{d} N(0,1)}{\sqrt{\xrightarrow{p} 1}} \quad (\hat{\sigma}^2 \xrightarrow{p} \sigma^2 \text{ and Theorem 6.2.1-2})\\ &\xrightarrow{d} N(0,1) \quad (\text{by Slutsky})\end{aligned}$$

□

[a]See Lecture 5 for derivation of the first step

The distribution of the Wald statistic is as expected, recall $W/p|x \sim F(p, n-k)$ under normal regression, and thus we see $W|x \sim pF(p, n-k) \xrightarrow{d} \chi^2(p)$. Why?

$$\begin{aligned} p \times F &= p \frac{\chi^2(p)/p}{\chi^2(n-k)/(n-k)} \\ &= \frac{\chi^2(p)}{\chi^2(n-k)/(n-k)} \end{aligned}$$

$$\frac{\chi^2(n-k)}{n-k} = \frac{1}{n-k} \sum_{i=1}^{n-k} Z_i^2 \xrightarrow{p} \mathbb{E}[Z_i^2] = 1$$

$$\Rightarrow pF \xrightarrow{d} \chi^2(p)$$

Asymptotic confidence intervals and sets

Since $t \xrightarrow{d} N(0,1)$ we can build asymptotic confidence intervals for β_j. From the critical values of $N(0,1)$:

$$Pr\left(\left|\frac{\sqrt{n}(\hat{\beta}_j - \beta)}{\sqrt{\hat{\sigma}^2(\frac{1}{n}X'X)^{-1}_{jj}}}\right| \leq 1.96\right) \approx 0.95$$

$$\Rightarrow Pr\left(\left|\hat{\beta}_j - \beta\right| \leq 1.96\sqrt{\hat{\sigma}^2(X'X)^{-1}_{jj}}\right) \approx 0.95 \quad \text{(cancel n's and rearrange)}$$

$$\Rightarrow \left[\hat{\beta}_j - 1.96\sqrt{\hat{\sigma}^2(X'X)^{-1}_{jj}}, \hat{\beta}_j + 1.96\sqrt{\hat{\sigma}^2(X'X)^{-1}_{jj}}\right] \quad \text{Asymptotic confidence interval}$$

This gives us the set all all values of β_j that are not rejected by he t-test with asymptotic size 5%. We say that the confidence interval is obtained by inversion of the test. We can similarly invert the Wald test, consider a test of the entire vector $\beta = b$ (i.e. $R = I_p$):

$$\begin{aligned} W &= (\hat{\beta} - b)'(\hat{\sigma}^2(X'X^{-1}))^{-1}(\hat{\beta} - b) \\ &= \frac{(\hat{\beta} - b)'X'X(\hat{\beta} - b)}{\hat{\sigma}^2} \end{aligned}$$

The asymptotic 95% confidence set for β is the ellipsoid with centre $\hat{\beta}$:

$$\Rightarrow \left\{\beta : \frac{(\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta)}{\hat{\sigma}^2} \leq \chi^2_{0.95}(k)\right\}$$

6.3 Delta method

Sometimes we need to know confidence intervals or sets for some (possibly nonlinear) function of regression parameters. We can do this with the delta method.

Definition 6.3.1: Delta method

Suppose $\hat{\theta}$ is a k-dimensional vector where $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \xi$, and suppose $g : \mathbb{R}^k \to \mathbb{R}$ has continuous first dervatives. Denote by $G(\theta)$ the $r \times k$ matrix of first derivatives evaluated at θ: $G(\theta) \equiv \frac{\partial g(\theta)}{\partial \theta'}$ then as $n \to \infty$

$$\sqrt{n}(g(\hat{\theta}) - g(\theta)) \xrightarrow{d} G(\theta)\xi$$

. In particular, if $\xi \sim N(0, V)$ then as $n \to \infty$

$$\sqrt{n}(g(\hat{\theta}) - g(\theta)) \xrightarrow{d} N(0, GVG')$$

Proof. By the mean value theorem, there exists a k-dimensional vector $\bar{\theta}$ between $\hat{\theta}$ and θ such that

$$g(\hat{\theta}) - g(\theta) = \underset{r\times k}{G(\bar{\theta})}\underset{k\times 1}{(\hat{\theta} - \theta)}$$

$$\Rightarrow \sqrt{n}(g(\hat{\theta}) - g(\theta)) = G(\bar{\theta})\sqrt{n}(\hat{\theta} - \theta)$$

Since $\bar{\theta}$ is between $\hat{\theta}$ and θ and since $\hat{\theta} \xrightarrow{p} \theta$ we know $\bar{\theta} \xrightarrow{p} \theta$. $G()$ is assumed continuous, so by CMT:

$$G(\bar{\theta}) \xrightarrow{p} G(\theta)$$

$$\Rightarrow \sqrt{n}(g(\hat{\theta}) - g(\theta)) = G(\bar{\theta})\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{p} G(\theta)\xi$$

□

Exercise 6.3.1. Let $\{\hat{\theta}_n\}$ be a sequence of 2x1 random vectors satisfying $\sqrt{n}(\hat{\theta}_0 - \theta_0) \xrightarrow{d} N(0, V)$ where the asymptotic mean is $\theta_0 = [0, 1]'$ and the asymptotic covariance matrix is I_2. Denote the two entries of $\hat{\theta}_n$ by $\hat{\theta}_{n,1}$ and $\hat{\theta}_{n,2}$. Derive the asymptotic distribution of the sequence of products $\{\hat{\theta}_{n,1}\hat{\theta}_{n,2}\}$

Solution:-

We can apply the delta method because the function

$$g(\theta) = g(\theta_1, \theta_2) = \theta_1\theta_2$$

is continuously differentiable. The asymptotic mean of the transformed sequence is

$$g(\theta_0) = \theta_{0,1}\theta_{0,2} = 0 \times 1 = 0$$

The Jacobian of the function is

$$G(\theta) = \begin{bmatrix} \frac{\partial g(\theta_1,\theta_2)}{\partial \theta_1} & \frac{\partial g(\theta_1,\theta_2)}{\partial \theta_2} \end{bmatrix} = [\theta_2, \theta_1]$$

By evaluating at θ_0 we obtain $G(\theta_0) = [1, 0]$.
Therefore the asymptotic covariance matrix is

$$G(\theta_0)VG(\theta_0)' = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 1$$

And we can write $\sqrt{n}\hat{\theta}_{n,1}\hat{\theta}_{n,2} \xrightarrow{d} N(0, 1)$

Example (Nerlove's returns to scale).

$$\log TC_i = \beta_1 + \beta_2 \log Q_i + \beta_3 \log p_{C_i} + \beta_4 \log p_{L_i} + \beta_5 \log p_{F_i} + \varepsilon_i$$

Suppose we want to study the asymptotic confidence region of the normalised regression with coefficients $\alpha = (\beta_3/\beta_2, \beta_4/\beta_2, \beta_5/\beta_2)'$ (i.e. the powers of the Cobb-Douglas production function). Define

$$g(\beta) = \begin{bmatrix} \beta_3/\beta_2 \\ \beta_4/\beta_2 \\ \beta_5/\beta_2 \end{bmatrix}$$

$$G(\beta) = \frac{\partial g(\beta)}{\partial \theta'} = \begin{bmatrix} 0 & -\beta_3/\beta_2^2 & 1/\beta_2 & 0 & 0 \\ 0 & -\beta_4/\beta_2^2 & 0 & 1/\beta_2 & 0 \\ 0 & -\beta_5/\beta_2^2 & 0 & 0 & 1/\beta_2 \end{bmatrix}$$

Thus considering the Wald statistic with $H_0 : \hat{\alpha} = \alpha$, i.e.: $R = I_3, q = \alpha$:

$$\begin{aligned} W &= \frac{\left(R\hat{\beta} - q\right)' \left(\sigma^2 R \left(X'X\right)^{-1} R'\right)^{-1} \left(R\hat{\beta} - q\right)}{\hat{\sigma}^2/\sigma^2} \\ &= \frac{\sqrt{n}\left(\hat{\alpha} - \alpha\right)' \left(\sigma^2 R \left(\frac{1}{n} X'X\right)^{-1} R'\right)^{-1} \sqrt{n}\left(\hat{\alpha} - \alpha\right)}{\hat{\sigma}^2/\sigma^2} \\ &\xrightarrow{d} [N(0, I_3)]' I_3 N(0, I_3) \quad \text{(using theorem 6.2.2)} \\ &= \chi^2(3) \end{aligned}$$

Hence the asymptotic 95% confidence set for α is the ellipsoid

$$\left\{\alpha : (\hat{\alpha} - \alpha)' \left(G(\hat{\beta})\hat{\sigma}^2 (X'X)^{-1} G(\hat{\beta})'\right) (\hat{\alpha} - \alpha) \leq \chi^2_{0.95}(3)\right\}$$