Teoria del modello lineare in grandi campioni

- Studieremo le proprietà di stimatori e test quando $n \to \infty$

- La teoria asintotica ci fornisce strumenti per approssimare le distribuzioni di stimatori e statistiche test in campioni finiti

Esempio $\quad X_1, X_2, \dots, X_n \sim \text{i.i.d.}(\mu, \sigma^2) \qquad \sigma^2 < \infty$

vogliamo testare $H_0 : \mu = 0$

Consideriamo la statistica test t_n:

$$t_n := \frac{\hat{\mu}}{s.e.(\hat{\mu})} \qquad s.e.(\hat{\mu}) = \text{standard error di } \hat{\mu}.$$

In genere $\hat{\mu} := \overline{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i$, il cui standard error è

$$s.e.(\hat{\mu}) = \sqrt{\frac{\hat{\sigma}_n^2}{n}}$$

dove

$$\hat{\sigma}_n^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

da cui

$$t_n = \sqrt{n}\,\frac{\overline{X}_n}{\hat{\sigma}_n}$$

<u>nota</u>: se $X_i \sim N(\mu, \sigma^2)$ allora, sotto H_0,

$$t_n \underset{H_0}{\sim} t(n-1)$$

dove $t(n-1)$ indica la v.c. t di Student con $n-1$ g.d.l.

Se X_i non è N, non conosco la distribuzione di t_n, né sotto H_0 né sotto l'alternativa.

Usiamo la teoria asintotica $(n \to \infty)$ per capire quale sia la distribuzione di $\bar{X}_n$ e t_n (in grandi campioni).

Consideriamo $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ e una legge dei grandi numeri [LGN].

Le LGN spiegano il comportamento asintotico di medie (o somme) di variabili casuali.

LGN per variabili i.i.d. Se $X_1 \dots X_n$ sono i.i.d. con $E(X_i) = \mu$, $|\mu| < \infty$, allora

$$\bar{X}_n \xrightarrow{q.c.} \mu \qquad P\left(\lim_{n \to \infty} \bar{X}_n \neq \mu\right) = 0$$

per $n \to \infty$.

È una LGN "forte" — la convergenza è quasi certa.
La legge forte implica la LGN debole:

$$\overline{X}_n \xrightarrow[n\to\infty]{P} \mu$$

Ricordiamo che questa espressione significa:

$$P\left(\left|\overline{X}_n - \mu\right| > \epsilon\right) \xrightarrow[n\to\infty]{} 0$$

per ogni $\epsilon > 0$.

Nota: Se $X_i \sim \text{iid}(\mu, \sigma^2)$, $\sigma^2 < \infty$, allora dimostrare che

$$\overline{X}_n \xrightarrow{P} \mu$$

è molto semplice.

Infatti:

$$V(\bar{X}_n) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} V\left(\sum_{i=1}^{n} X_i\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \longrightarrow 0$$

Quindi

$$E\left[(\bar{X}_n - \mu)^2\right] \xrightarrow[n\to\infty]{} 0$$

Nota: questa è la convergenza "in media quadratica", o, in L_2.

La convergenza in m.q. implica la convergenza in probabilità.

TCHEBYCHEV: $P\left(|\bar{X}_n - \mu| > \epsilon\right) \leq \frac{1}{\epsilon^2} V\left[(\bar{X}_n - \mu)\right]$ $\qquad E(\bar{X}_n - \mu) = 0$

Nel nostro caso,

$$P\left(|\overline{X}_n-\mu|>\varepsilon\right) \leq \frac{1}{\varepsilon^2}\frac{\sigma^2}{n} \longrightarrow 0 .$$

Riassumendo,

$$\overline{X}_n \xrightarrow{P} \mu.$$

Se H_0 è vero, allora $\mu=0$ e $\overline{X}_n \xrightarrow{P} 0$.

Consideriamo $t_n := \sqrt{n}\,\dfrac{\overline{X}_n}{\hat{\sigma}_n}$.

Abbiamo:

$$\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}_n\right)^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - n\,\overline{X}_n^2\right]$$

$$= \frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - \frac{n}{n-1}\,\overline{X}_n^2$$

Poiché $\bar{X}_n \xrightarrow{P} \mu$, $\frac{n}{n-1}\bar{X}_n^2 \xrightarrow{P} \mu^2$ per il seguente teorema:

<u>Teorema</u> (delle funzione continua): se $Y_n \xrightarrow{P} Y$ e se $g(\cdot)$ è una funzione continua tale che $E[g(Y)] < \infty$, allora

$$g(Y_n) \xrightarrow{P} g(Y)$$

$V(X) = EX^2 - (EX)^2$

$E(X^2) = V(X) + E(X)^2$

Inoltre, $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{P} E(X_i^2) = \sigma^2 + \mu^2 < \infty$ (per ipotesi) per la LGN.

Quindi:

$$\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_1^n X_i^2 - \frac{n}{n-1}\bar{X}_n^2$$

$$= \left(\frac{n}{n-1}\right)\frac{1}{n}\sum_1^n X_i^2 - \frac{n}{n-1}\bar{X}_n^2$$

$$\xrightarrow{P} (\sigma^2 + \mu^2) - \mu^2 = \sigma^2$$

Ho usato le seguente propietà:

$$\left.\begin{array}{l} X_n \xrightarrow{P} X \\ Y_n \xrightarrow{P} Y \end{array}\right\} \Longrightarrow \begin{array}{l} X_n + Y_n \xrightarrow{P} X + Y \\ X_n \cdot Y_n \xrightarrow{P} X \cdot Y \end{array}$$

Quindi, $\hat{\sigma}_n^2 \xrightarrow{P} \sigma^2$, da cui $\hat{\sigma}_n \xrightarrow{P} \sigma$ (teorema f. continua).

Torniamo a t_n e esprimiamola nel seguente modo:

$$t_n := \sqrt{n}\,\frac{\overline{X_n}}{\hat{\sigma}_n} = \sqrt{n}\,\frac{\overline{X_n}-\mu}{\hat{\sigma}_n} + \sqrt{n}\,\frac{\mu}{\hat{\sigma}_n}$$

Dobbiamo derivare le propietà (asintotiche) di

$$\sqrt{n}\,\frac{\overline{X_n}-\mu}{\hat{\sigma}_n}$$

e, in particolare, di $\sqrt{n}(\overline{X_n}-\mu)$

$V\left(\frac{1}{a}X\right) = \frac{1}{a^2}V(X)$

$V(aX) = a^2 V(X)$

Abbiamo già visto che $V(\bar{X}_n - \mu) = \dfrac{\sigma^2}{n}$

$$V\left(\sqrt{n}(\bar{X}_n - \mu)\right) = n\, V(\bar{X}_n - \mu) = n \cdot \frac{\sigma^2}{n} = \sigma^2$$

Quindi $\sqrt{n}(\bar{X}_n - \mu) \sim (\underbrace{0}_{\text{v.a.}}, \underbrace{\sigma^2}_{\text{varianza}})$

o, analogamente

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \sim (0, 1)$$

Questo risultato è indipendente dalla distribuzione di X_i.
Il teorema centrale del limite [TCL], dice che la distribuzione di

$$\sqrt{n}\,\frac{\bar{X}_n - \mu}{\sigma}$$

converge alla distribuzione normale standardizzata:

Teorema: $X_1, X_2, \dots, X_n$ iid, $E(X_i) = \mu$, $V(X_i) = \sigma^2 < +\infty$,
Allora:

$$\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \xrightarrow[n\to\infty]{d} N(0,1)$$

("$\xrightarrow{d}$" indica la convergenza in distribuzione)

Torniamo a t_n:

$$t_n := \sqrt{n}\,\frac{\bar{X}_n - \mu}{\hat{\sigma}_n} + \sqrt{n}\,\frac{\mu}{\hat{\sigma}_n}$$

$$= \underbrace{\left(\frac{\sigma}{\hat{\sigma}_n}\right)}_{\downarrow P \atop 1} \underbrace{\sqrt{n}\,\frac{\bar{X}_n - \mu}{\sigma}}_{\downarrow d \atop N(0,1)} + \underbrace{\frac{1}{\hat{\sigma}_n}}_{\downarrow P \atop \frac{1}{\sigma}} \sqrt{n}\,\mu.$$

| Proprietà: $X_n \xrightarrow{P} a$, $Y_n \xrightarrow{d} Y \Rightarrow X_n Y_n \xrightarrow{d} aY$

$$t_n = \underbrace{\left(\frac{\sigma}{\hat{\sigma}}\right) \sqrt{n}\, \frac{\bar{X}_n - \mu}{\sigma}}_{\xrightarrow{d} N(0,1)} + \frac{1}{\hat{\sigma}_n} \sqrt{n}\, \mu$$

Se H_0 è vera, $\mu = 0$, $\frac{1}{\hat{\sigma}_n}\sqrt{n}\,\mu = 0$, e

$$t_n \xrightarrow[H_0]{d} N(0,1)$$

Se H_0 non è vera, $\mu \neq 0$, $\frac{1}{\hat{\sigma}_n}\sqrt{n}\,\mu \to \begin{cases} +\infty & \mu > 0 \\ -\infty & \mu < 0 \end{cases}$

$(\sigma \neq 0)$

da cui $|t_n| \longrightarrow +\infty$

Il risultato sotto H_0 implica che, ad esempio,

$$P\{-1{,}96 \le t_n \le 1{,}96 \mid H_0\} \longrightarrow 0.95$$

e che, se H_0 non è vera,

$$P\{-1{,}96 \le t_n \le 1{,}96 \mid H_0 \text{ falsa}\} \longrightarrow 0$$

(il test rifiuta l'H_0 falsa con probabilità che tende a 1 — il test è consistente)

<u>Nota</u>: Abbiamo visto che, se $X_1, \dots, X_n \sim N(\mu, \sigma^2)$, allora (sotto H_0),

$$t_n \underset{H_0}{\sim} t(n-1)$$

Abbiamo ora dimostrato che, sotto H_0, $t_n \xrightarrow{d} N(0,1)$
I due risultati non sono in disaccordo, in quanto la v.c. t di student converge (in distribuzione) alla $N(0,1)$ quando il numero di gradi di libertà diverge

$$t(n) \xrightarrow{d} N(0,1)$$

gradi di libertà $(n-1)$	valore critico $t(n-1)$	valore critico $N(0,1)$	ampiezza del test $N(0,1)$
5	± 2.571	± 1.96	0.107
10	± 2.228	± 1.96	0.078
30	± 2.042	± 1.96	0.059 ← $P(t_n \in [-1.96, 1.96])$
100	± 1.984	± 1.96	0.053
∞	± 1.96	± 1.96	0.05

Caso del modello lineare:

$$y_t = x_t' \beta + \varepsilon_t \qquad t = 1, 2, \dots, n \qquad \underset{n \times k}{X} = [x_1\ x_2 \dots, x_n]'$$

$$E[\varepsilon_t | X] = 0$$

$$V[\varepsilon_t | X] = \sigma^2 < \infty$$

$$C[\varepsilon_t, \varepsilon_{t'} | X] = 0 \ , \ t \neq t'$$

$$\hat{\beta} = (X'X)^{-1} X'y = \left(\sum_{t=1}^{n} x_t x_t' \right)^{-1} \left(\sum_{t=1}^{n} x_t y_t \right)$$

$$= \underbrace{\left(\frac{1}{n} \sum_{t=1}^{n} x_t x_t' \right)^{-1}}_{S_{xx}} \underbrace{\left(\frac{1}{n} \sum_{t=1}^{n} x_t y_t \right)}_{S_{xy}} =: S_{xx}^{-1} S_{xy}$$

Notiamo che

$$S_{xy} = \frac{1}{n} \sum_{t=1}^{n} x_t y_t = \frac{1}{n} \sum_{1}^{n} x_t (x_t' \beta + \varepsilon_t) = \underbrace{\frac{1}{n} \sum_{1}^{n} x_t x_t'}_{S_{xx}} \beta + \underbrace{\frac{1}{n} \sum_{1}^{n} x_t \varepsilon_t}_{S_{x\epsilon}}$$

$$= S_{xx} \beta + S_{x\epsilon}$$

da cui:

$$\hat{\beta} = S_{xx}^{-1} S_{xy} = S_{xx}^{-1} (S_{xx} \beta + S_{x\epsilon}) = \beta + S_{xx}^{-1} S_{x\epsilon}$$

$$\hat{\beta} - \beta = S_{xx}^{-1} S_{x\epsilon}, \qquad S_{x\epsilon} := \frac{1}{n} \sum_{t=1}^{n} x_t \varepsilon_t$$

Per capire il comportamento asintotico di $\hat{\beta}-\beta$ dobbiamo studiare S_{xx} e $S_{x\varepsilon}$

Ipotesi: $\{y_t, x_t\} \sim iid \quad \Rightarrow \quad \varepsilon_t = y_t - \beta' x_t \sim iid$

Consideriamo S_{xx}:

$$\underset{K\times K}{S_{xx}} := \frac{1}{n}\sum_{t=1}^{n} x_t x_t' \xrightarrow{p} \underset{K\times K}{E[x_t x_t']} =: \Sigma_{xx} \quad \text{se} \quad \| E[x_t x_t'] \| < \infty$$

per la LGN. Consideriamo $S_{x\varepsilon}$:

$$S_{x\varepsilon} := \frac{1}{n}\sum_{t=1}^{n} \underset{K\times 1}{x_t} \underset{1\times 1}{\varepsilon_t} \xrightarrow{p} E[\varepsilon_t x_t] = \underset{1\times 1}{E[\varepsilon_t]}\,\underset{K\times 1}{E[x_t]} = \underset{K\times 1}{0}$$

se $\| E[x_t] \| < \infty$ per la LGN.

Proprietà:

$$\left.\begin{array}{l} X_T \xrightarrow{P} X \\ U_T \xrightarrow{P} a \end{array}\right\} \Rightarrow U_T X_T \xrightarrow{P} aX$$

$$X_T \xrightarrow{P} X \} \Rightarrow X_T^{-1} \xrightarrow{P} X^{-1} \quad (X_T^{-1} \text{ e } X^{-1} \text{ devono esistere})$$

da cui:

$$\hat{\beta} = \beta + \underbrace{S_{XX}^{-1} S_{X\epsilon}}_{\substack{\downarrow P \quad \downarrow P \\ \Sigma_{XX}^{-1} \quad 0 \\ =0}} \xrightarrow{P} \beta$$

X ha rango max $= K$

Σ_{XX} ha rango max $= K$

Serve la condizione: Σ_{XX} ha rango massimo.

(oltre alla solita: S_{XX} ha rango massimo)

Sotto le ipotesi introdotte, $\hat{\beta}$ è consistente per β.

(esercizio: scrivere tutte le ipotesi usate)

Proprietà di $\hat\sigma^2$. Ricordiamo che $\hat\sigma^2 := \frac{1}{n-k}\sum_{t=1}^{n} \hat\varepsilon_t^2$, dove

$\hat\varepsilon_t := y_t - \hat\beta' x_t$ (residuo), $t = 1, 2, \dots, n$

$$\hat\varepsilon_t = y_t - \hat\beta' x_t = x_t'\beta + \varepsilon_t - \hat\beta' x_t = \varepsilon_t - (\hat\beta - \beta)' x_t$$

$$= \varepsilon_t - S_{\varepsilon x} S_{xx}^{-1} x_t \, , \quad \text{dove} \quad S_{\varepsilon x} = S_{x\varepsilon}'$$

$$\hat\sigma^2 = \frac{1}{n-k}\sum_1^n \hat\varepsilon_t^2 = \frac{1}{n-k}\sum_1^n \left(\varepsilon_t - (\hat\beta - \beta)' x_t\right)^2$$

$$= \frac{1}{n-k}\sum_1^n \left[\varepsilon_t^2 - 2(\hat\beta - \beta)' x_t \varepsilon_t + (\hat\beta - \beta)' x_t x_t' (\hat\beta - \beta)\right]$$

$$= \frac{1}{n-k}\sum_1^n \varepsilon_t^2 - 2(\hat\beta - \beta)' \frac{1}{n-k}\sum_1^n x_t \varepsilon_t + (\hat\beta - \beta)' \frac{1}{n-k}\sum x_t x_t' (\hat\beta - \beta)$$

$$= \frac{n}{n-k}\left\{ \frac{1}{n}\sum \varepsilon_t^2 - 2(\hat\beta - \beta)' \frac{1}{n}\sum x_t \varepsilon_t + (\hat\beta - \beta)' \frac{1}{n}\sum x_t x_t' (\hat\beta - \beta)\right\}$$

$$= \frac{n}{n-k}\left\{\frac{1}{n}\sum_{1}^{n}\varepsilon_t^2 - 2 S_{\varepsilon x} S_{xx}^{-1} S_{x\varepsilon} + S_{\varepsilon x} S_{xx}^{-1} \cancel{S_{xx}} \cancel{S_{xx}^{-1}} S_{x\varepsilon}\right\}$$

$$= \frac{n}{n-k}\left\{\frac{1}{n}\sum_{1}^{n}\varepsilon_t^2 - S_{\varepsilon x} S_{xx}^{-1} S_{x\varepsilon}\right\}$$

(scomposizione varianza)

Poiché $S_{x\varepsilon} \xrightarrow{P} 0$ e $S_{xx}^{-1} \xrightarrow{P} \Sigma^{-1}$, $S_{\varepsilon x} S_{xx}^{-1} S_{x\varepsilon} \xrightarrow{P} 0$

$$\hat{\sigma}^2 = \frac{n}{n-k}\,\frac{1}{n}\sum_{1}^{n}\varepsilon_t^2 + o_p(1)$$

notazione:

$$\boxed{\begin{array}{c} X_T \xrightarrow[n\to\infty]{P} 0 \\ X_T = o_p(1) \end{array}}$$

Per la LGN:

$$\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^2 \xrightarrow{P} E[\varepsilon_t^2] = \sigma^2$$

se $\sigma^2 < \infty$.

Quindi: $\hat{\sigma}^2 \xrightarrow{P} \sigma^2$ (nota: $\frac{n}{n-k} \longrightarrow 1$ per k finito)

Lo stimatore $\hat{\sigma}^2$ è consistente per σ^2.

Distribuzione asintotica di $\hat{\beta}$

$$\hat{\beta}-\beta = S_{XX}^{-1} S_{X\epsilon} = \left(\frac{1}{n}\sum_{t=1}^{n} X_t X_t'\right)^{-1}\left(\frac{1}{n}\sum_{t=1}^{n} X_t \varepsilon_t\right)$$

$$\sqrt{n}\left(\hat{\beta}-\beta\right) = S_{XX}^{-1}\sqrt{n}\, S_{X\epsilon} = \underbrace{\left(\frac{1}{n}\sum_{t=1}^{n} X_t X_t'\right)^{-1}}_{\xrightarrow{p} \Sigma_{XX}^{-1}}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_t \varepsilon_t\right)$$

Consideriamo $\sqrt{n}\, S_{X\epsilon}$:

$$\sqrt{n}\, S_{X\epsilon} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_t \varepsilon_t \quad \text{dove } \{X_t, \varepsilon_t\} \text{ sono iid}$$

Inoltre, $\underset{K\times 1}{E\left[X_t \varepsilon_t\right]} = \underset{K\times 1}{0}$:

$$V(X) = E(X^2) - E(X)^2$$
$$V(X) = E(XX') - E(X)E(X)'$$

$$V\left[X_t \varepsilon_t\right] = E\left[(X_t\varepsilon_t)(X_t\varepsilon_t)'\right] - \underbrace{E(X_t\varepsilon_t)}_{0}\underbrace{E(X_t\varepsilon_t)'}_{0}$$

$$= E\left[X_t X_t' \cdot \varepsilon_t^2\right]$$

$$= E\left\{ E\left[X_t X_t' \varepsilon_t^2 \mid X_t \right] \right\}$$

legge dei valori attesi iterati

$$= E\left\{ X_t X_t' \underbrace{E\left[\varepsilon_t^2 \mid X_t \right]}_{\sigma^2} \right\}$$

$$V[X_t \varepsilon_t] = \sigma^2 E[X_t X_t'] =: \sigma^2 \Sigma_{XX}$$

Possiamo applicare il TCL (se $\|\Sigma_{xx}\| < \infty$):

$$\underset{K\times 1}{\sqrt{n}\, S_{xt}} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t \varepsilon_t \xrightarrow{d} N(\underset{K\times 1}{0}, \underset{K\times 1}{\sigma^2 \Sigma_{XX}})$$

[nota: abbiamo applicato una versione multivariata del TCL]

<u>Proprietà</u>: $\left.\begin{array}{l} X_T \xrightarrow{p} a \\ Y_T \xrightarrow{d} Y \end{array}\right\} \Rightarrow X_T Y_T \xrightarrow{d} aY$

$X \sim N(0,V)$

$AX \sim N(0, AVA')$

$A'X \sim N(0, A'VA)$

$$\sqrt{n}(\hat{\beta}-\beta) = S_{xx}^{-1} \underbrace{\sqrt{n}\, S_{x\epsilon}}_{} \xrightarrow{d} \Sigma_{xx}^{-1} \cdot N(0, \sigma^2 \Sigma_{xx}) \equiv N(0, \sigma^2 \Sigma_{xx}^{-1} \cancel{\Sigma_{xx}} \cancel{\Sigma_{xx}^{-1}})$$

$$S_{xx}^{-1} \xrightarrow{p} \Sigma_{xx}^{-1} \qquad \sqrt{n}\, S_{x\epsilon} \xrightarrow{d} N(0, \sigma^2 \Sigma_{xx})$$

e quindi:

$$\sqrt{n}(\hat{\beta}-\beta) \xrightarrow{d} N(0, \sigma^2 \Sigma_{xx}^{-1})$$

Lo stimatore $\hat{\beta}$ è asintoticamente normale.

$V_{\hat{\beta}} := \sigma^2 \Sigma_{xx}^{-1}$ si chiama matrice di var.-cov. asintotica.

Sotto le ipotesi forte, $V_{\hat{\beta}}$ è la più piccola possibile (Rao-Cramer), quindi $\hat{\beta}$ è asintoticamente efficiente.

Nota: $V_{\hat{\beta}}$ non è nota; la possiamo stimare:

$$\hat{V}_{\hat{\beta}} = \hat{\sigma}^2 \hat{\Sigma}_{xx}^{-1} = \hat{\sigma}^2 S_{xx}^{-1}$$

che risulta consistente:

$$\hat{V}_{\hat{\beta}} = \underset{\substack{\downarrow p \\ \sigma^2}}{\hat{\sigma}^2} \ \underset{\substack{\downarrow p \\ \Sigma_{XX}^{-1}}}{S_{XX}^{-1}} \xrightarrow{p} \sigma^2 \Sigma_{XX}^{-1}$$

Nota sull'interpretazione della normalità asintotica:

$$\sqrt{n}(\hat{\beta}-\beta) \xrightarrow{d} N(0, \sigma^2 \Sigma_{XX}^{-1}) \qquad n \to \infty$$

$$\sqrt{n}(\hat{\beta}-\beta) \approx N(0, \sigma^2 \Sigma_{XX}^{-1}) \qquad \text{approssimazione per } n \text{ fisso}$$

$$\sqrt{n}(\hat{\beta}-\beta) \approx N(0, \hat{\sigma}^2 S_{XX}^{-1}) \qquad \text{approssimazione "calcolabile"}$$

$$\hat{\beta}-\beta \approx N\Big(0, \underbrace{\hat{\sigma}^2 \frac{1}{n} S_{XX}^{-1}}_{\hat{\sigma}^2 \frac{1}{n}\left(\frac{X'X}{n}\right)^{-1} = \hat{\sigma}^2 (X'X)^{-1}}\Big)$$

$$\hat\beta \approx N\left(\beta, \hat\sigma^2 (X'X)^{-1}\right)$$

Applicazione ai test di ipotesi.

esempio: test t per $H_0: \beta_i = 0$

$$t_n := \frac{\hat\beta_i}{s.e.(\hat\beta_i)}, \qquad s.e.(\hat\beta_i)^2 = \hat\sigma^2 \hat a_{ii},$$ dove $\hat a_{ii}$ è l'elemento (i,i) di $\hat A := (X'X)^{-1}$

Se $\varepsilon | X \sim N$, allora sotto H_0, $t_n \sim t(n-K)$

Proprietà asintotiche:

$$t_n = \frac{\hat\beta_i}{\hat\sigma \sqrt{\hat a_{ii}}} = \frac{1}{\hat\sigma} \frac{\sqrt{n}\,\hat\beta_i}{\sqrt{n\, \hat a_{ii}}}$$

dove $\hat\sigma \xrightarrow{P} \sigma$, $n\hat a_{ii}$ è l'elemento (i,i) di $n(X'X)^{-1} = S_{XX}^{-1} \xrightarrow{P} \Sigma_{XX}^{-1}$, converge all'elemento (i,i) di $\Sigma_{XX}^{-1} =: A$:

$$n\hat{a}_{ii} \to a_{ii}, \text{ elemento } (i,i) \text{ di } A := \Sigma_{xx}^{-1}$$

Inoltre

$$\sqrt{n}\,\hat{\beta}_i \xrightarrow[H_0]{d} N(0, \sigma^2 a_{ii})$$

Quindi:

$$t_n = \frac{\sqrt{n}\,\hat{\beta}_i}{\hat{\sigma}\sqrt{\hat{a}_{ii} n}} \xrightarrow{d} \frac{N(0, \sigma^2 a_{ii})}{\sigma\sqrt{a_{ii}}} \equiv N(0,1)$$

Sotto H_0, $t_n \xrightarrow{d} N(0,1)$

Esempio: test F / χ^2 per $H_0: \underset{k\times 1}{R'\beta} - r = \underset{q\times 1}{0}$

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 \Sigma_{xx}^{-1})$$

$$\sqrt{n}(R'\hat{\beta} - R'\beta) \underset{H_0}{=} \sqrt{n}(R'\hat{\beta} - r) \xrightarrow[H_0]{d} N(\underset{q\times 1}{0}, \sigma^2 R'\Sigma_{xx}^{-1} R)$$

$$n(R'\hat{\beta}-r)'\left[\sigma^2 R'\Sigma_{XX}^{-1}R\right]^{-1}(R'\hat{\beta}-r) \xrightarrow[H_0]{d} \chi^2(q), \qquad \underset{k\times q}{R}$$

$$(\text{se rango}(R)=q)$$

$$W_n := n(R'\hat{\beta}-r)'\left[\hat{\sigma}^2 R' S_{XX}^{-1}R\right]^{-1}(R'\hat{\beta}-r)$$

$$= (R'\hat{\beta}-r)'\left[\hat{\sigma}^2 R'(X'X)^{-1}R\right]^{-1}(R'\hat{\beta}-r)$$

Poiché $\hat{\sigma}^2 \xrightarrow{P} \sigma^2$, $S_{XX} = \frac{X'X}{n} \xrightarrow{P} \Sigma_{XX}$, abbiamo che

$$W_n \xrightarrow[H_0]{d} \chi^2(q)$$

W_n prende il nome di statistica test di Wald.

Nota: $W_n = \frac{1}{\hat\sigma^2}(R'\hat\beta - r)'[R'(X'X)^{-1}R]^{-1}(R'\hat\beta - r)$

somiglia alla statistica Test F:

$$F_n = \frac{1}{\hat\sigma^2}(R'\hat\beta - r)'[R'(X'X)^{-1}R]^{-1}(R'\hat\beta - r)/q$$

ovvero $F_n = W_n/q$ oppure $W_n = q F_n$

Di conseguenza otteniamo che, per $n \to \infty$,

$$q \cdot F_n = W_n \xrightarrow[H_0]{d} \chi^2(q)$$

(sotto l'ipotesi di normalità, $F_n \underset{H_0}{\overset{d}{\sim}} F(q, n-k)$; non c'è contraddizione in quanto

$$q\,F(q, n-k) \longrightarrow \chi^2(q) \text{ per } n \to \infty$$
)

Teoria asintotica — caso non iid

$\{y_t, x_t\}$ iid $\qquad E[y_t | x_t] = x_t'\beta$

$\qquad V[y_t | x_t] = \sigma^2 < \infty$

LGN: $\frac{1}{n}\sum_{t=1}^{n} x_t x_t' \xrightarrow{P} \Sigma_{xx} := E[x_t x_t'] < \infty$, rang $\Sigma_{xx} = \max$

$\frac{1}{n}\sum_{t=1}^{n} x_t \varepsilon_t \xrightarrow{P} \Sigma_{x\varepsilon} := E[x_t \varepsilon_t] = 0 \qquad (\varepsilon_t := y_t - x_t'\beta)$

TCL: $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} x_t \varepsilon_t \xrightarrow{d} N(0, \sigma^2 \Sigma_{xx})$

Questi risultati sono sufficienti per dimostrare

$$\hat\beta - \beta = \left(\frac{1}{n}\sum x_t x_t'\right)^{-1} \frac{1}{n}\sum x_t \varepsilon_t \xrightarrow{P} 0$$

$$\sqrt{n}(\hat\beta - \beta) = \left(\qquad\qquad\right)^{-1} \frac{1}{\sqrt{n}}\sum x_t \varepsilon_t \xrightarrow{d} N(0, \sigma^2 \Sigma_{xx}^{-1})$$

Esistono casi in cui l'ipotesi di dati iid è violata

$$y_t = \beta x_t + \varepsilon_t , \qquad x_t := y_{t-1} \qquad AR(1)$$

ovvero

$$y_t = \beta y_{t-1} + \varepsilon_t .$$

In questo caso $\{y_t, x_t\} \equiv \{y_t, y_{t-1}\}$ non è in generale iid

Cerchiamo di valutare cosa serve per dimostrare $\hat\beta \xrightarrow{P} \beta$

e $\sqrt{n}(\hat\beta - \beta) \xrightarrow{d} N$.

$$\hat\beta - \beta = \left(\frac{1}{n}\sum x_t x_t'\right)^{-1}\left(\frac{1}{n}\sum x_t \varepsilon_t\right)$$

(i) Dobbiamo avere $\frac{1}{n}\sum_t x_t x_t' \xrightarrow{P} \Sigma_{XX}$, def. pos. e di rango max.

Una condizione sufficiente è che:

(i) x_t è stazionario ed ergodico per $E[x_t x_t'] =: \Sigma_{XX}$

(ii) rango Σ_{XX} = massimo. (serve a garantire l'esistenza di Σ_{XX}^{-1})

② Dobbiamo avere $\frac{1}{n}\sum_{t=1}^{n} x_t \varepsilon_t \xrightarrow{P} E[\varepsilon_t x_t] = 0$.

Una condizione sufficiente è:

(iii) $\{x_t, \varepsilon_t\}$ è stazionario ed ergodico per $E[\varepsilon_t x_t] = \Sigma_{x\varepsilon}$

(iv) $\Sigma_{x\varepsilon} = E[x_t \varepsilon_t] = 0$

La (iv) è una condizione essenziale per l'uso di OLS.

Le condizioni (i) - (iv) garantiscono che $\hat{\beta} - \beta \xrightarrow{P} 0$ (consistenza)

Nota: la condizione (iii) si può anche scrivere come:

$\{y_t, x_t\}$ è stazionario ed ergodico per $E[y_t x_t] = \Sigma_{yx}$.

Per la normalità asintotica possiamo usare il seguente teorema (Mann - Wald, 1943).

Teorema: Se X_t stazionario ed ergodico e se $\varepsilon_t \sim iid(0,\sigma^2)$, con $E[\varepsilon_t X_t]=0$, allora

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_t \varepsilon_t \xrightarrow{d} N(0, \sigma^2 \Sigma_{xx}), \text{ dove}$$

$$\Sigma_{xx} := E[X_t X_t']$$

Per poterlo applicare possiamo fare la seguente ipotesi:

(v) $\{\varepsilon_t\}$ è i.i.d.$(0,\sigma^2)$

che, unitamente alle precedenti, implica che

$$\sqrt{n}(\hat{\beta}-\beta) \xrightarrow{d} N(0, \sigma^2 \Sigma_{xx}^{-1})$$

Riassumendo:

Sotto le ipotesi (i) – (iv), $\hat{\beta}-\beta \xrightarrow{p} 0$ (consistenza)

" " " (i) – (v), $\sqrt{n}(\hat{\beta}-\beta) \xrightarrow{d} N(0, \sigma^2 \Sigma_{xx}^{-1})$ (normalità asintotica)

Esercizio: dato il modello $y_t = \beta y_{t-1} + \varepsilon_t$ $(AR(1))$

dove $|\beta| < 1$ e $\varepsilon_t \sim iid(0, \sigma^2)$,

dimostrare che lo stimatore OLS di β,

$$\hat{\beta} = \left(\frac{1}{n}\sum_{t=1}^{n} y_{t-1}^2\right)^{-1}\left(\frac{1}{n}\sum_{t=1}^{n} y_t y_{t-1}\right),$$

è consistente e asintoticamente normale.

In particolare, mostrare che la varianza asintotica di $\sqrt{n}(\hat{\beta} - \beta)$ è $(1-\beta^2)$.