Appunti di Teoria delle Distribuzioni Limite

Diego Lubian

13 marzo 1999

Le ipotesi del modello lineare classico con errori normali ci permettono di ottenere risultati sulla distribuzione di alcune statistiche di interesse: stimatori OLS e test di ipotesi statistiche. Lo stimatore OLS possiede una distribuzione (condizionata) normale multivariata, lo stimatore della varianza del termine di errore ha una distribuzione χ^2 con appropriati gradi di libertà, e queste due distribuzioni sono indipendenti. Da quest'ultimo risultato si ottiene che il test di Wald (ma anche LR e LM) per la verifica di un sistema di ipotesi sui parametri di interesse ha una distribuzione F. Se l'ipotesi di normalità degli errori è abbandonata (o si considerano modelli non-lineari - anche se questi ultimi esulano dal nostro corso) non è possibile ottenere risultati in forma chiusa per la distribuzione di stimatori e tests. Dobbiamo ricorrere a una qualche forma di approssimazione della distribuzione. In particolare, un metodo comunemente utilizzato consiste nell'approssimare con la distribuzione limite ottenuta quando la numerosità campionaria è infinita.

1 Convergenza in probabilità

Sia $\{X_t\}$ una successione infinita di v.c. $\{X_1, X_2, \ldots, X_t, \ldots\}$. Noi studiremo il comportamento limite di statistiche $h(X_1, X_2, \ldots, X_T)$ costruite a partire da un campione osservato di numerosità T, quando T tende a infinito, con la speranza che, se $h(\cdot)$ converge a qualcosa, questo sia collegato a qualche parametro di interesse. Ad es., nel modello lineare classico lo stimatore OLS $\hat{\beta}_T = \left(\sum_{t=1}^T x_t x_t'\right)^{-1} \left(\sum_{t=1}^T x_t y_t\right)$ è la statistica di interesse (indicizzato per la numerosità campionaria) e noi esamineremo se e sotto quali condizioni sia possibile ottenere un risultato generale a proposito di $\hat{\beta}_T$ quando più e più osserazioni sono aggiunte alla regressione.

Poiché le statistiche sono variabili casuali dobbiamo estendere il concetto standard di limite in senso probabilistico. Richiamo il concetto standard di limite:

Definizione 1.1 *La successione non aleatoria $\{x_t\}$ converge al limite x se per qualsiasi $\varepsilon > 0$ piccolo a piacere esiste un T^* tale che $T^* > T$ implica $|x_T - x| < \varepsilon$ e scriveremo $\lim_{T\to\infty} x_T = x$.*

Consideriamo ora una successione di v.c. $\{X_T\}$ e una v.c X. Poiché X_T è una v.c. l'evento $|X_T - X| < \varepsilon$ può sempre avere una probabilità non nulla di realizzarsi e non si può dire con certezza che X_T dista da X meno di ε, per ogni valore di X. Questo ci conduce alla seguente definizione di limite in probabilità.

Definizione 1.2 *La successione di v.c.* $\{X_T\}$ *converge in probabilità a* X *se per qualsiasi* $\varepsilon, \delta > 0$ *piccoli a piacere esiste un* T^* *tale che* $T^* > T$ *implica*

$$\Pr(|X_T - X| < \varepsilon) > 1 - \delta$$

che indichiamo con $\operatorname*{plim}_{T\to\infty} X_T = X$ *oppure con* $X_T \xrightarrow{p} X$.

In altri termini, il limite in probabilità è X se, per ogni intorno di X, la probabilità che X_T cada in quell'intorno è asintoticamente arbitrariamente alta. Quindi, quando T aumenta, X_T tende a essere uguale a X e di conseguenza prendere decisioni su X_T basandosi su X non produrrà risultati dissimili, con l'approssimazione che migliora quando la numerosità campionaria aumenta. In termini di definizioni standard per limiti, il limite in probabilità viene spesso scritto, per ε fissato, come

$$\lim_{T\to\infty} \Pr(|X_T - X| < \varepsilon) = 1.$$

Un altro modo di convergenza che può essere utile e di facile utilizzo in molte situazioni è la **convergenza in media quadratica**:

Definizione 1.3 *La successione di v.c.* $\{X_T\}$ *converge in media quadratica al limite* X *se*

$$\lim_{T\to\infty} E[(X_T - X)^2] = 0$$

che indichiamo con $X_T \overset{m.q.}{\rightarrow} X$.

Questa definizione di convergenza utilizza il concetto di *errore quadratico medio*. La successione $\{X_T\}$ converge al limite X se lo scarto al quadrato tra gli elementi della successione e il limite converge a zero. La convergenza in media quadratica differisce quindi sostanzialmente dalla convergenza in probabilità in quanto viene definita in termini di limite di un determinato momento della successione, viz.

$$\lim_{T\to\infty} \int Z_T^2 f(Z_T) \mathrm{d}Z_T = 0$$

dove $Z_T = X_T - X$, e non con riferimento a "probabilità di eventi", viz, $|X_T - X| < \varepsilon$. Una situazione in cui il limite in probabilità differisce dalla limite in media quadratica è data nell'esempio seguente.

Esempio 1.1. (McFadden) Sia $\Pr(X_T = 0) = 1 - 4^{-T}$ e sia $\Pr(X_T = \theta^T) = 4^{-T}$. Si ha che $X_T \xrightarrow{p} 0$ per qualsiasi θ. Infatti

$$\lim_{T\to\infty} \Pr(X_T = 0) = \lim_{T\to\infty} (1 - 4^{-T}) = 1$$
$$\lim_{T\to\infty} \Pr(X_T = \theta^T) = \lim_{T\to\infty} 4^{-T} = 0$$

per cui $X_T \xrightarrow{p} 0$. Tuttavia $E(X_T^2) = (\theta/2)^{2T}$ converge a zero solo per $|\theta| < 2$. Ne segue che $X_T \xrightarrow{m.q.} 0$ se e solo se $|\theta| < 2$. Questo risultato è una conseguenza del fatto che la probabilità limite non dipende da θ mentre il momento secondo dipende da θ.

L'esempio precedente mostra come la convergenza in probabilità non implichi la convergenza in media quadratica. La proposizione successiva indica invece che il viceversa è vero.

Proposizione 1.4 *Se $X_T \xrightarrow{m.q.} X$ allora $X_T \xrightarrow{p} X$.*

PROVA. La prova della proposizione si basa sulla disuguaglianza di Chebyshev di cui si può dare una formulazione generale: si consideri una v.c. Y, per ogni costante c, ε e r si ha che $\Pr(|Y-c| < \varepsilon) \geq 1 - \dfrac{E(|Y-c|^r)}{\varepsilon^r}$. Da una applicazione diretta ponendo $c = X$ e $r = 2$ si ottiene:

$$\Pr(|X_T - X| < \varepsilon) \geq 1 - \frac{E(|X_T - X|^2)}{\varepsilon^2}$$

da cui la proposizione segue immediatamente considerando che il lato destro tende a 1 se $X_T \xrightarrow{m.q.} X$.

Il teorema seguente è molto utile in diverse situazioni che incontreremo: esso ci dice che una trasformazione continua della successione di v.c. converge in probabilità alla medesima trasformazione della v.c. limite (ammesso che la successione originaria converga in probabilità).

Teorema 1.5 (Trasformazioni continue) *Sia $\{X_T\}$ una successione di v.c. (scalari) e sia $g(\cdot)$ una funzione continua da R in R e α una costante. Se $X_T \xrightarrow{p} \alpha$ allora $g(X_T) \xrightarrow{p} g(\alpha)$.*

Osservazioni:(i) questo risultato vale in termini più generali quando la convergenza in probabilità è a una v.c.; (ii) la proposizione vale anche in situazioni in cui $\{X_T\}$ è di dimensione k, $g(\cdot)$ è una funzione continua da R^k a R e α è un vettore di dimensione k; (iii) si ricordi che le trasformazioni inverse, logaritmiche, lineari, e potenze sono tutte continue.

Altri risultati che valgono per i limiti standard valgono pure per i limiti in probabilità. In particolare, se $X_T \xrightarrow{p} X$ e $Y_T \xrightarrow{p} Y$, allora

$$\begin{aligned} X_T + Y_T &\xrightarrow{p} X + Y \\ X_T \cdot Y_T &\xrightarrow{p} X \cdot Y. \end{aligned}$$

cioè, il plim di una somma di v.c. è uguale alla somma dei plim e il plim di un prodotto di v.c. è uguale al prodotto dei plim.

2 Consistenza

Gli stimatori sono successioni di variabili casuali indicizzate per la numerosità campionaria $h(X_1, X_2, \ldots, X_T) = h_T$. Ad es., nel modello lineare classico lo stimatore OLS

$\hat{\beta}_T = \left(\sum_{t=1}^T x_t x_t'\right)^{-1} \left(\sum_{t=1}^T x_t y_t\right)$ è una successione di v.c.. Una desiderabile caratteristica asintotica che uno stimatore deve possedere è la convergenza in probabilità al vero valore (nella popolazione) del parametro che viene stimato. Se uno stimatore possiede questa proprietà viene detto **consistente**.

Definizione 2.1 *Se uno stimatore h_T converge in probabilità a h, cioè $h_T \xrightarrow{p} h$, allora h_T è consistente.*

3 Leggi dei grandi numeri

Dimostrare la convergenza in probabilità di una successione di v.c. è un compito generalmente impossibile da svolgere a meno che non si vogliano fare delle ipotesi sulla distribuzione per T finito. Tuttavia, le nozioni di convergenza che abbiamo visto si basano su tale informazione. Ad es., per dimostrare la convergenza in media quadratica è necessario avere a disposizione il primo e secondo momento della distribuzione in campioni finiti; per dimostrare la convergenza in probabilità bisogna studiare il comportamento limite di X_T in un ε-intorno di X, che è pure una caratteristica della distribuzione in campioni finiti. Quindi, un approccio che richieda meno informazione per provare la convergenza in probabilità è necessario. Le leggi dei grandi numeri (Laws of large numbers), o brevemente LGN, svolgono esattamente questo ruolo, stabilendo la convergenza di determinate successioni costruite a partire dalla successione originaria. Due leggi dei grandi numeri sono riportate di seguito.

Teorema 3.1 (Khinchine) *Se $\{X_t\}$ è una successione di v.c. iid con $E(X_t) = \mu < \infty$ allora*

$$\bar{X}_T = \frac{1}{T}\sum_{t=1}^T X_t \xrightarrow{p} \mu.$$

Teorema 3.2 (Chebyshev) *Se $\{X_t\}$ è una successione di v.c. indipendenti con $E(X_t) = \mu_t$ e $E[(X_t - \mu_t)^2] = \sigma_t^2$ allora la condizione*

$$\lim_{T\to\infty} \frac{1}{T}\left(\frac{1}{T}\sum_{t=1}^T \sigma_t^2\right) = 0$$

implica

$$\bar{X}_T - \bar{\mu}_T = \frac{1}{T}\sum_{t=1}^T X_t - \frac{1}{T}\sum_{t=1}^T \mu_t \xrightarrow{p} 0.$$

PROVA DELLA LEGGE DEI GRANDI NUMERI DI CHEBYSHEV. Poiché $E(\bar{X}_T) = \bar{\mu}_T$ e $\mathrm{Var}(\bar{X}_T) = \frac{1}{T^2}\sum_{t=1}^T \sigma_t^2$, dalla disuguaglianza di Chebyshev si ha

$$\Pr[|\bar{X}_T - \bar{\mu}_T| > \varepsilon] \leq \frac{\sum_{t=1}^T \sigma_t^2}{T^2\varepsilon^2}.$$

Poiche il lato destro tende a zero per $T \to \infty$ per ipotesi, allora

$$\lim_{T\to\infty} \Pr[|\bar{X}_T - \bar{\mu}_T| > \varepsilon] = 0$$

e, quindi, $\bar{X}_T - \bar{\mu}_T \xrightarrow{p} 0$.
Osservazioni: (i) la LGN di Khinchine, per successioni iid, richiede solo l'esistenza di media finita; (ii) la LGN di Chebyshev, per successioni di v.c. indipendenti ma **non identicamente distribuite**, pone una ulteriore condizione sulla varianza: in particolare, si richiede che la media aritmetica delle varianze aumenti più lentamente della numerosità campionaria. Quindi si può rilasciare l'ipotesi di identica distribuzione a costo di aggiungere ulteriori condizioni sui momenti della successione di v.c.. Si può anche rimuovere l'ipotesi di indipendenza a costo di aggiungere condizioni ancora più stringenti sui momenti. In generale, si può dire che esiste un trade-off tra caratteristiche di indipendenza temporale e/o eterogeneità distributiva e condizioni sui momenti richieste per ottenere una LGN.

Esempio 3.1. Sia $\{Z_t\}$ una successione di v.c. iid con $E[Z_t] = \mu$ e $\text{Var}(Z_t) = \sigma^2 < \infty$. Consideriamo gli stimatori della media e della varianza

$$\begin{aligned} \bar{Z}_T &= \frac{1}{T}\sum_{t=1}^{T} Z_t \\ \hat{\sigma}_T^2 &= \frac{1}{T}\sum_{t=1}^{T}(Z_t - \bar{Z})^2. \end{aligned}$$

Sono stimatori consistenti? Sì. Infatti, per la LLN di Khinchine $\bar{Z}_T \xrightarrow{p} \mu$. Inoltre, la successione $\{Z_t^2\}$ è iid, essendo una trasformazione di una successione di v.c. iid, con $E(Z_t^2) = \sigma^2 + \mu^2$. Allora, per la LLN di Khinchine applicata alla successione $\{Z_t^2\}$ si ha che

$$\frac{1}{T}\sum_{t=1}^{T} Z_t^2 \xrightarrow{p} \sigma^2 + \mu^2.$$

Poiché lo stimatore della varianza può essere riscritto come

$$\hat{\sigma}_T^2 = \frac{1}{T}\sum_{t=1}^{T} Z_t^2 - \bar{Z}_T^2$$

possiamo applicare il teorema delle trasformazioni continue per dedurre $\hat{\sigma}_T^2 \xrightarrow{p} \sigma^2$. Lo stesso risultato si ottiene se si considera uno stimatore corretto della varianza del tipo s_T^2. Infatti

$$\begin{aligned} s_T^2 &= \frac{1}{T-1}\sum_{t=1}^{T}(Z_t - \bar{Z})^2 \\ &= \frac{T}{T-1}\hat{\sigma}_T^2 \\ &= \hat{\sigma}_T^2 + \frac{1}{T-1}\hat{\sigma}_T^2 \\ &\xrightarrow{p} \sigma^2. \end{aligned}$$

4 Convergenza in distribuzione

Le leggi dei grandi numeri e la loro applicazione allo studio della proprietà di consistenza di uno stimatore sembrano suggerire la convergenza a una costante. Tuttavia, questo può essere troppo restrittivo per i nostri obiettivi proprio perché noi vogliamo approssimare la distribuzione di variabili aleatorie. Quindi la nostra interpretazione di convergenza deve essere estesa alla convergenza di successioni di v.c. a qualcosa che è esso stesso una v.c.. A questo proposito è naturale soffermarsi sul comportamento della funzione di distribuzione degli elementi della successione di v.c.. Questo costituisce il punto di partenza per un'altra definizione di convergenza.

Definizione 4.1 *Sia $\{X_T\}$ una successione di v.c. con funzione di distribuzione $\{F_T(r)\}$. Se*

$$\lim_{T\to\infty} F_T(r) = F(r)$$

in ogni punto di continuità[1] *di $F(\cdot)$, allora si dice che X_T converge in distribuzione a X con funzione di distribuzione $F(r)$ e si indica con $X_T \xrightarrow{d} X$.*

La distribuzione F è detta distribuzione limite o asintotica di X_T. Poiché la nozione di convergenza in distribuzione è definita usando la funzione di distribuzione e non la v.c. stessa, questo modo di convergenza è fondamentalmente diverso dalla convergenza in probabilità. In quest'ultimo caso, una particolare v.c. è identificata come il limite della successione di v.c. mentre nella convergenza in distribuzione la v.c. limite è solo ipotetica: è "come se" la successione di v.c. stesse convergendo a una determinata v.c..

Di seguito sono proposti una serie di risultati (senza prova) molto importanti che verranno utilizzati nel seguito.

Teorema 4.2 *Se $X_T \xrightarrow{p} X$ allora $X_T \xrightarrow{d} X$. Se X è una costante vale anche il viceversa.*

Teorema 4.3 (Mann - Wald) *Sia $\{X_T\}$ una successione di v.c. (scalare) e sia $g(\cdot)$ una funzione continua da $\mathbb{R}$ a $\mathbb{R}$. Se $X_T \xrightarrow{d} X$ allora $g(X_T) \xrightarrow{d} g(X)$.*

Teorema 4.4 (Slutsky) *Siano $\{X_T\}$ e $\{Y_T\}$ due successioni di v.c. (scalari) e sia α una costante. Se $X_T \xrightarrow{d} X$ e $Y_T \xrightarrow{p} \alpha$, allora*

$$\begin{array}{lll} (i) & X_T + Y_T \xrightarrow{d} X + \alpha, & \\ (ii) & Y_T X_T \xrightarrow{d} \alpha X, & \\ (iii) & (X_T/Y_T) \xrightarrow{d} X/\alpha, & \text{se} \quad \alpha \neq 0 \end{array}$$

Osservazioni:

[1] Un punto di continuità di F è un punto r tale che $\lim_{\varepsilon\downarrow 0} F(r-\varepsilon) = F(r)$.

- il punto (ii) vale anche nel caso in cui $\alpha = 0$ per cui $X_T Y_T \xrightarrow{d} 0$;

- il teorema di Mann-Wald vale anche in situazioni in cui $g(\cdot)$ è una funzione continua da R^k a R;

- il teorema di Slutsky vale anche se Y_T è una successione di matrici aleatorie di dimensione $(k \times k)$ che converge in probabilità (elemento per elemento) a A, matrice di dimensione $(k \times k)$. Il punto (iii) del teorema si deve intendere come $Y_T^{-1} X_T \xrightarrow{d} A^{-1} X$, se A è invertibile.

Teorema 4.5 (Equivalenza asintotica) *Siano $\{X_T\}$ e $\{Y_T\}$ due successioni di v.c. scalari. Se $Y_T \xrightarrow{d} Y$ e $(Y_T - X_T) \xrightarrow{p} 0$ allora $X_T \xrightarrow{d} Y$.*

Come per i teoremi precedenti, quest'ultimo risultato vale anche per successioni di vettori aleatori, come nel seguente esempio.

Esempio 4.1. (Cappuccio - Orsi) Consideriamo la successione di v.c. $\{W_T\}$ di dimensione k e di matrici stocastiche $\{A_T\}$ di dimensione $(k \times k)$ con $W_T \xrightarrow{d} W \sim N(0, \Sigma)$ e $A_T \xrightarrow{p} \Sigma$. Consideriamo le due successioni di v.c. $X_T = A_T^{-1/2} W_T$ e $Y_T = \Sigma^{-1/2} W_T$.

Applicando il teorema delle trasformazioni continue si ottiene, poiché $W_T \xrightarrow{d} W$, la convergenza di $Y_T : Y_T = \Sigma^{-1/2} W_T \xrightarrow{d} \Sigma^{-1/2} W \sim N(0, I)$. Consideriamo ora il comportamento asintotico della differenza $X_T - Y_T$:

$$\begin{aligned}
X_T - Y_T &= A_T^{-1/2} W_T - \Sigma_T^{-1/2} W_T \\
&= (A_T^{-1/2} - \Sigma_T^{-1/2}) W_T \\
&= (A_T^{-1/2} \Sigma_T^{1/2} - I_k) \Sigma_T^{-1/2} W_T \\
&= (A_T^{-1/2} \Sigma_T^{1/2} - I_k) Y_T \\
&\xrightarrow{p} 0
\end{aligned}$$

dove l'ultimo riga segue applicando il teorema delle trasformazioni continue a $A_T \xrightarrow{p} \Sigma$, per cui $A_T^{-1/2} \xrightarrow{p} \Sigma^{-1/2}$ e $(A_T^{-1/2} \Sigma_T^{1/2} - I_k) \xrightarrow{p} 0$, e poi il teorema di Slutsky.

Per il teorema di equivalenza asintotica, si ottiene $X_T = A_T^{-1/2} W_T \xrightarrow{d} \Sigma^{-1/2} W$. Questo risultato può essere ottenuto direttamente applicando il teorema di Slutsky alla successione W_T.

Inoltre, consideriamo la trasformazione $g(Y) = Y'Y = W' \Sigma^{-1} W \sim \chi_k^2$. Per il teorema delle trasformazioni continue abbiamo che

$$g(Y_T) = Y_T' Y_T = W_T' \Sigma^{-1} W_T \xrightarrow{d} Y'Y \sim \chi_k^2$$

e per il teorema di Slutsky e il teorema delle trasformazioni continue

$$g(X_T) = X_T' X_T = W_T' A_T^{-1} W_T \xrightarrow{d} Y'Y \sim \chi_k^2.$$

Teorema 4.6 (Cramer-Wold device) *Sia $\{X_T\}$ una successione di vettori casuali di dimensione k. Se*

$$\lambda' X_T \xrightarrow{d} \lambda' X$$

per ogni $\lambda = (\lambda_1 \quad \lambda_2 \quad \cdots \quad \lambda_k)'$ e dove X è una v.c. di dimensione k con funzione di distribuzione congiunta $F(\cdot)$, allora la distribuzione limite di X_T esiste ed è uguale a $F(\cdot)$.

Questo risultato è molto utile perchè, come vedremo, i teoremi centrali del limite sono sempre dati per v.c. scalari, non per successioni di vettori. Determinare la distribuzione limite (marginale) di ciascuna componente della successione di vettori aleatori non è però sufficiente per determinare la distribuzione congiunta della successione di vettori aleatori. Poiché noi siamo quasi esclusivamente interessati a risultati di convergenza in distribuzione per succesioni di vettori casuali, in quanto stimiamo un insieme di parameteri simultaneamente e di questi vogliamo studiare la distribuzione congiunta, il teorema di cui sopra fornisce un utile "device" per passare da una distribuzione asintotica univariata a una distribuzione asintotica multivariata.

5 Teoremi centrali del limite

In generale, non è possibile determinare la distribuzione limite in modo diretto partendo dalle distribuzioni in campioni finiti e molto spesso la forma della distribuzione in campioni finiti non è nota. Abbiamo bisogno di un risultato generale, analogo alle leggi dei grandi numeri, anche per la convergena in distribuzione. Questi risultati vanno sotto il nome di Teoremi Centrali del Limite (Central limit theorems) o, brevemente, TCL. Due TCL rilevanti sono riportati di seguito (senza prova).

Teorema 5.1 (Lindeberg-Levy) *Se $\{X_t\}$ è una successione di v.c. IID con $E(X_t) = \mu < \infty$ e $E[(X_t - \mu)^2] = \sigma^2 < \infty$, allora*

$$\frac{\sqrt{T}(\bar{X}_T - \mu)}{\sigma} \xrightarrow{d} X \sim N(0,1)$$

Osservazioni: (i) rispetto alla convergenza in probabilità per processi iid (LGN di Khinchine), la convergenza in distribuzione richiede l'esistenza di media e varianza finita; (ii) occorre trasformare la sequenza originaria $\bar{X}_T$ per una appropriata funzione di T per poter ottenere la convergenza in distribuzione a una v.c. non-degenere (infatti, per la LGN $\bar{X}_T$ converge in probabilità a una costante e quindi anche in distribuzione). Queste trasformazioni prendono il nome di *trasformazioni stabilizzatrici* in quanto impediscono che la varianza di $\bar{X}_T$ vada a zero quando $T \to \infty$ e quindi che la distribuzione limite degeneri in un punto; (iii) se $E(X_t) = 0$ allora si ottiene il risultato standard

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} X_t \xrightarrow{d} N(0, \sigma^2).$$

L'ipotesi di identica distribuzione della successione di v.c. deve essere rimossa se si vuole analizzare il modello di regressione lineare classico con variabili non-stocastiche. Come per le leggi dei grandi numeri questo può essere fatto a condizione di imporre vincoli più restrittivi sui momenti della successione di v.c. $\{X_t\}$. Il TCL seguente fornisce questa estensione.

Teorema 5.2 (Lyapunov) *Se $\{X_t\}$ è una successione di v.c. indipendenti con $E(X_t) = \mu_t < \infty$ e $E[(X_t-\mu_t)^2] = \sigma_t^2 < \infty$ e $\sigma_t^2 \neq 0$ per ogni t. Se, per ogni t, $E[|X_t-\mu_t|^{2+\delta}] < \Delta < \infty$ per un qualunque $\delta > 0$ e*

$$\lim_{T\to\infty} \bar{\sigma}_T^2 = \bar{\sigma}^2$$

dove $\bar{\sigma}_T^2 = \frac{1}{g(T)}\sum_{t=1}^{T}\sigma_t^2$, con $\bar{\sigma}^2$ positivo e finito, allora

$$\frac{\sqrt{T}(\bar{X}_T-\bar{\mu})}{\bar{\sigma}_T} \xrightarrow{d} X \sim N(0,1)$$

Osservazioni: (i) rispetto al TCL di Lindeberg-Levy, viene aggiunta un'ipotesi sui momenti della successione di v.c. e viene rilasciata l'ipotesi di identica distribuzione; (ii) la condizione su $\bar{\sigma}_T^2$ implica che all'aumentare della numerosità campionaria (che fa tendere all'infinito il denominatore), le varianze che si aggiungono al numeratore sono non nulle e, di conseguenza, le v.c. della successione sono nondegeneri cosicché il loro contenuto informativo diventa rilevante; (iii) la funzione $g(T)$ è tipicamente una potenza di T (nel caso standard si ha $g(T) = T$); (iv) se $E(X_t) = 0$ attraverso semplici manipolazioni si ottiene l'utile risultato

$$\frac{1}{\sqrt{g(T)}}\sum_{t=1}^{T} X_t \xrightarrow{d} N(0,\bar{\sigma}^2);$$

(v) se noi fissiamo $\delta = 1$ e poniamo $E[|X_t-\mu_t|^3] = m_t$, la condizione su $\bar{\sigma}_T^2$ può essere riscritta come

$$\lim_{T\to\infty}\frac{\left(\sum_{t=1}^{T} m_t\right)^{1/3}}{\left(\sum_{t=1}^{T}\sigma_t^2\right)^{1/2}} = 0$$

che può essere così interpretata: l'eterogeneità nelle distribuzioni della successione di v.c. (misurata da momenti di X_t che variano con l'indice t), standardizzata per la deviazione standard che standardizza la media campionaria, deve essere insignificante per $T \to \infty$. Non è possibile che le distribuzioni siano così eterogenee che osservazioni estratte da una determinata distribuzione dominino il comportamento della varianza campionaria.

Esempio 5.1. Riprendiamo l'esempio 3.1 e consideriamo la statistica

$$Y_T = \frac{\sqrt{T}(\bar{Z}_T-\mu)}{s_T}$$

che possiamo riscrivere come

$$Y_T = \frac{\sqrt{T}(\bar{Z}_T - \mu)/\sigma}{s_T/\sigma}.$$

Per il teorema delle trasformazioni continue per il denominatore si ha $\frac{s_T}{\sigma} \xrightarrow{p} 1$. Per il numeratore, per il TCL di Lindeberg-Levy si ha che

$$\frac{\sqrt{T}(\bar{Z}_T - \mu)}{\sigma} \xrightarrow{d} N(0,1).$$

Infine per il teorema di Slutsky, si ha $Y_T \xrightarrow{d} Y \sim N(0,1)$ (e per il teorema delle trasformazioni continue $Y_T^2 \xrightarrow{d} \chi_1^2$). Si deve notare che per ottenere la distribuzione limite di Y_T non è necessario disporre della distribuzione limite di s_T^2 (a differenza di quanto accade in campioni finiti, cioè per T fissato) ma basta semplicemente mostrare la sua convergenza in probabilità a σ^2.

La statistica Y_T può essere modificata molto semplicemente per essere interpretata come una statistica test dell'ipotesi nulla $H_0 : \mu = \mu_0$ contro l'ipotesi alternativa $H_1 : \mu \neq \mu_0$. Infatti, se sostituiamo μ con μ_0 nell'espressione per Y_T otteniamo proprio la statistica test sotto l'ipotesi nulla. Niente di quanto derivato sopra muta con questa sostituzione e, di conseguenza, si contiua ad avere $\frac{\sqrt{T}(\bar{Z}_T - \mu_0)}{s_T} \xrightarrow{d} N(0,1)$. A questo punto possiamo fissare un livello di significatività α e individuare, dalle tavole della normale standardizzata, il valore critico c_α tale che, per $T \to \infty$, $\Pr(|Y| < c_\alpha) = 1 - \alpha$. La decisione a proposito della validità dell'ipotesi nulla si baserà quindi sul valore della statistica test Y_T, e si rifiuterà l'ipotesi nulla quando essa assume valori elevati. Poiché le ipotesi di partenza non ci permettono di determinare la distribuzione esatta in campioni finiti della statistica test, dobbiamo usare la v.c. normale standardizzata come approssimazione asintotica. Si noti la differenza con quanto accade in campioni finiti con popolazione normale dove, per T è fissato, Y_T è distribuito come una t di Student con $T-1$ gradi di libertà. Ma questo è, in realtà, coerente con il risultato asintotico in quanto è noto che la distribuzione t di Student coincide con la distribuzione normale standardizzata quando i gradi di libertà tendono a infinito.

Esempio 5.2. Consideriamo la successione di v.c. iid $\{Z_t\}$ con $E(Z_t) = \mu$, $\mathrm{Var}(Z_t) = \sigma$ e $E[(Z_t - \mu)^4] = \mu_4 < \infty$. Vogliamo determinare la distribuzione asintotica dello stimatore della varianza $\hat{\sigma}_T^2 = (1/T)\sum_{t=1}^T (Z_t - \bar{Z})^2$.

Inizio considerando il comportamento asintotico di $\tilde{\sigma}_T^2 = (1/T)\sum_{t=1}^T (Z_t - \mu)^2$. A questo proposito, definisco la successione di v.c. $\{w_t\}$ con $w_t = (Z_t - \mu)^2$. Poiché

$$\begin{aligned} E(w_t) &= E[(Z_t - \mu)^2] = \mathrm{Var}(Z_t) = \sigma^2 \\ \mathrm{Var}(w_t) &= E(w_t^2) - [E(w_t)]^2 = E[(Z_t - \mu)^4] - \sigma^4 \\ &= \mu_4 - \sigma^4 \end{aligned}$$

posso applicare il TCL di Lindeberg-Levy alla successione di v.c. iid $\{w_t\}$:

$$\frac{\sqrt{T}(\tilde{\sigma}_T^2 - \sigma^2)}{\sqrt{\mu_4 - \sigma^4}} \xrightarrow{d} N(0,1)$$

che, applicando il teorema delle trasformazioni continue, possiamo scrivere come

$$\sqrt{T}(\tilde{\sigma}_T^2 - \sigma^2) \xrightarrow{d} N(0, \mu_4 - \sigma^4).$$

A partire da quest'ultimo risultato, se si riesce a dimostrare che

$$\sqrt{T}(\hat{\sigma}_T^2 - \sigma^2) - \sqrt{T}(\tilde{\sigma}_T^2 - \sigma^2) \xrightarrow{p} 0$$

allora si ottiene anche la distribuzione limite desiderata per $\hat{\sigma}_T^2$. Riscrivendo $\hat{\sigma}_T^2$ come

$$\hat{\sigma}_T^2 = (1/T)\sum_{t=1}^{T}(Z_t - \bar{Z})^2 = \tilde{\sigma}_T^2 - (\bar{Z}_T - \mu)^2$$

si ha

$$\sqrt{T}(\hat{\sigma}_T^2 - \sigma_T^2) - \sqrt{T}(\tilde{\sigma}_T^2 - \sigma_T^2) = \sqrt{T}(\hat{\sigma}_T^2 - \tilde{\sigma}_T^2) = -\sqrt{T}(\bar{Z}_T - \mu)^2$$

per cui, in ultima analisi, si deve mostrare che il lato destro dell'espressione di cui sopra converge in probabilità a zero. Già sappiamo che, per il TCL di Lindeberg-Levy,

$$\frac{\sqrt{T}(\bar{Z}_T - \mu)}{\sigma} \xrightarrow{d} N(0,1)$$

da cui

$$\frac{T(\bar{Z}_T - \mu)^2}{\sigma^2} \xrightarrow{d} \chi_1^2.$$

Quindi si ha che

$$\sqrt{T}(\bar{Z}_T - \mu)^2 = \sqrt{T}(\bar{Z}_T - \mu)^2 \frac{\sigma^2}{\sigma^2}\frac{\sqrt{T}}{\sqrt{T}} = \frac{T(\bar{Z}_T - \mu)^2}{\sigma^2}\frac{\sigma^2}{\sqrt{T}} \xrightarrow{p} 0.$$

A questo punto, per il teorema di equivalenza asintotica, si ha che

$$\sqrt{T}(\hat{\sigma}_T^2 - \sigma^2) \xrightarrow{d} N(0, \mu_4 - \sigma^4).$$

Un altro risultato utile è dato dal seguente teorema.

Teorema 5.3 (Delta Method generale) *Si supponga che siano soddisfatte le condizioni per cui si ha che $S_T \xrightarrow{p} S$ e $\sqrt{T}(S_T - S) \xrightarrow{d} X$. Sia $g(\cdot)$ è una funzione continua in un intorno di S e sia il vettore di derivate di $g(\cdot)$ valutato in S dato da $g'(\cdot)$, allora $\sqrt{T}(g(S_T) - g(S)) \xrightarrow{d} g'(S)X$.*

PROOF. Il teorema si prova a partire da un'approssimazione di Taylor del primo ordine attorno a S o anche utilizzando il teorema del valor medio: infatti, da quest'ultimo

$$g'(S_T^*) = \frac{g(S_T) - g(S)}{S_T - S}$$

si ha che

$$g(S_T) = g(S) + g'(S_T^*)(S_T - S)$$

dove S_T^* è un punto tra S_T e S, $S_T^* = \alpha S_T + (1-\alpha)S$, con $\alpha \in [0,1]$. Naturalmente, $S_T^* \xrightarrow{p} S$ poiché $S_T \xrightarrow{p} S$ e, per il teorema delle trasformazioni continue $g'(S_T^*) \xrightarrow{p} g'(S)$. Di conseguenza,

$$g'(S_T^*)\sqrt{T}(S_T - S) \xrightarrow{d} g'(S)X$$

per il teorema di Slutsky e si ha che

$$g'(S_T^*)\sqrt{T}(S_T - S) = \sqrt{T}(g(S_T) - g(S)) \xrightarrow{d} g'(S)X.$$

Una versione specializzata del teorema precedente è utile nel seguito.

Teorema 5.4 (Il nostro Delta Method) *Se sono soddisfatte le condizioni per cui si ha che $\bar{X}_T \xrightarrow{p} \mu$ e $\sqrt{T}(\bar{X}_T - \mu) \xrightarrow{d} N(0,\Sigma)$ e la funzione $g(\cdot)$ è differenziabile nel punto μ e il vettore di derivate valutato in μ è dato da $G(\mu)$, allora*

$$\sqrt{T}(g(\bar{X}_T) - g(\mu)) \xrightarrow{d} N(0, G\Sigma G').$$

6 Proprietà asintotiche dello stimatore OLS

Consideriamo il modello lineare classico $y_t = x_t\beta + u_t$ con $t = 1,\ldots,T$. Il vettore di coefficienti β ha dimensione k, $E(u_t) = 0$ e $E(u_t^2) = \sigma^2$ per ogni t mentre $E(u_t u_s) = 0$ per $t \neq s$. Lo stimatore OLS di β può essere scritto come

$$\begin{aligned}
\hat{\beta}_T &= \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1}\left(\sum_{t=1}^{T} x_t y_t\right) \\
&= \beta + \left(\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} x_t u_t\right)
\end{aligned}$$

La nostra attenzione si soffermerà su quest'ultima formulazione che ci indica come le proprietà asintotiche di $\hat{\beta}$ dipendano dal comportamento in grandi campioni delle matrici di momenti campionari $T^{-1}\sum_{t=1}^{T} x_t x_t'$ e $T^{-1}\sum_{t=1}^{T} x_t u_t$.

6.1 Consistenza di $\hat{\beta}_T$

6.1.1 Regressori non-stocastici

Consideriamo il caso i cui le variabili esplicative sono non-stocastiche e introduciamo la seguente ipotesi

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} x_t x_t' = \lim_{T\to\infty} \frac{X'X}{T} = Q$$

dove Q è una matrice finita e nonsingolare. Questa ipotesi implica che (i) quando $T \to \infty$, gli elementi di $\sum_{t=1}^{T} x_t x_t'$ non crescono più velocemente di T e (ii) le variabili esplicative sono linearmente indipendenti per $T \to \infty$, ma non necessariamente per T finito. Dall'ipotesi si ottiene, per il teorema delle trasformazioni continue,

$$\lim_{T\to\infty} \left(\frac{1}{T} \sum_{t=1}^{T} x_t x_t' \right)^{-1} = Q^{-1}.$$

Consideriamo ora il termine $T^{-1} \sum_{t=1}^{T} x_t u_t$ e definiamo la successione di vettori aleatori $\{w_t\}$, dove $w_t = x_t u_t$, con $E(w_t) = 0$ e $E(w_t w_t') = E(x_t u_t^2 x_t') = \sigma^2 x_t x_t'$, cioè la matrice di varianza-covarianza cambia con l'indice temporale. La LGN di Chebyshev può essere applicata se la condizione

$$\lim_{T\to\infty} \frac{1}{T^2} \sum_{t=1}^{T} \sigma^2 x_t x_t' = 0$$

è soddisfatta. Sotto l'ipotesi precedente $\lim_{T\to\infty}(X'X/T) = Q$, abbiamo che

$$\lim_{T\to\infty} \frac{\sigma^2}{T} \sum_{t=1}^{T} \frac{x_t x_t'}{T} = \lim_{T\to\infty} \frac{\sigma^2}{T} \frac{X'X}{T} = 0$$

Quindi, si può concludere che

$$\frac{1}{T} \sum_{t=1}^{T} w_t = \frac{1}{T} \sum_{t=1}^{T} x_t u_t \xrightarrow{p} 0$$

e, per il teorema delle trasformazioni continue,

$$\hat{\beta} \xrightarrow{p} \beta + Q^{-1} \cdot 0 = \beta$$

ottenendo la consistenza di $\hat{\beta}$ nel modello lineare classico con regressori non-stocastici.

6.1.2 Regressori stocastici

Consideriamo il caso in cui termine di errore u_t e regressori x_s sono indipendenti per ogni t e s e assumiamo inoltre che le successioni $\{u_t\}$ e $\{x_t\}$ siano iid con media nulla e varianza pari rispettivamente a σ^2 (scalare) e pari a Ω (matrice di dimensione $(k \times k)$) con elemento generico ω_{ij}. Consideriamo la matrice di momenti campionari $T^{-1} \sum_{t=1}^{T} x_t x_t'$ il cui elemento generico è dato da $T^{-1} \sum_{t=1}^{T} x_{ti} x_{tj}$.

Poiché l'elemento generico, $x_{ti} x_{tj}$, della successione di matrici aleatorie $\{x_t x_t'\}$ ha media costante pari a ω_{ij}, elemento generico di Ω, per la LGN di Khinchine si ottiene

$$\frac{1}{T} \sum_{t=1}^{T} x_{ti} x_{tj} \xrightarrow{p} \omega_{ij}$$

che implica $T^{-1} \sum_{t=1}^{T} x_t x_t' \xrightarrow{p} \Omega$ e, per il teorema delle trasformazioni continue,

$$\left(\frac{1}{T} \sum_{t=1}^{T} x_t x_t' \right)^{-1} = \left(\frac{X'X}{T} \right)^{-1} \xrightarrow{p} \Omega^{-1}.$$

Consideriamo ora il vettore di momenti campionari $T^{-1}\sum_{t=1}^{T} x_t u_t$ il cui elemento generico è dato da $T^{-1}\sum_{t=1}^{T} x_{ti} u_t$. Poiché $\{x_{ti}u_t\}$ è una successione iid, in quanto trasformazione di v.c. iid, con $E(x_{ti}u_t) = 0$, per la LGN di Khinchine, applicata elemento per elemento alla successione di vettori aleatori $\{x_t u_t\}$, si ottiene

$$\frac{1}{T}\sum_{t=1}^{T} x_t u_t \xrightarrow{p} 0.$$

Combinando questi i risultati di convergenza in probabilità ottenuti si deduece che $\hat{\beta} \xrightarrow{p} \beta + \Omega^{-1} \cdot 0 = \beta$ e, quindi, lo stimatore OLS è consistente.

6.2 Distribuzione asintotica di $\hat{\beta}_T$

Poiché $\hat{\beta} \xrightarrow{p} \beta$ si ha anche che $\hat{\beta} \xrightarrow{d} \beta$, una v.c. degenere che ha tutta la densità concentrata in un punto e varianza nulla. Per ottenere una distribuzione asintotica nondegenere abbiamo già visto come sia necessario introdurre una trasformazione stabilizzatrice che impedisca alla varianza di andare a zero per T che tende a infinito. Come per la consistenza analizziamo separatamente i due casi di regressori non-stocastici e stocastici.

6.2.1 Regressori non-stocastici

Nel caso di regressori non-stocastici, abbiamo visto sopra come la successione di vettori aleatori $\{x_t u_t\}$ sia indipendente ma non identicamente distribuita in quanto $E(x_t u_t^2 x_t') = \sigma^2 x_t x_t'$ che varia con t. Per ottenere la distribuzione limite dello stimatore OLS di β, appropriatamente standardizzato, dobbiamo quindi far riferimento al TCL di Lyapunov e verificare la validità delle sue condizioni nel nostro contesto. Per prima cosa dobbiamo considerare la successione di v.c. indipendenti ma non identicamente distribuite $\{v_t\}$, dove $v_t = \lambda' x_t u_t$ per un qualche vettore λ di dimensione k, che ha media nulla e varianza $E(v_t^2) = \sigma^2 \lambda' x_t x_t' \lambda$. L'ipotesi

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} x_t x_t' = Q$$

con Q finito e nonsingolare implica

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \lambda' x_t x_t' \lambda = \lambda' Q \lambda$$

dove il limite è finito e diverso da zero. Questo a sua volta implica che la condizione sulla somma delle varianze che compare tra le ipotesi del TCL di Lyapunov, che nel nostro caso si riduce a,

$$\lim_{T\to\infty} \bar{\sigma}_T^2 = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \sigma^2 \lambda' x_t x_t' = \sigma^2 \lambda' Q \lambda$$

è soddisfatta. Dobbiamo quindi imporre l'ulteriore condizione che v_t abbia momenti di ordine leggermente superiore a due finiti, cioè $E(|v_t|^{2+\delta}) < \infty$ per poter concludere che

$$\frac{\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T}\lambda' v_t\right)}{\sqrt{\sigma^2\lambda' Q\lambda}} \xrightarrow{d} N(0,1)$$

che possiamo anche riscrivere come

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\lambda' v_t = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\lambda' x_t u_t \xrightarrow{d} N(0,\sigma^2\lambda' Q\lambda).$$

Per il teorema "Cramer-Wold device" possiamo concludere che

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} v_t = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_t u_t \xrightarrow{d} N(0,\sigma^2 Q).$$

Congiuntamente all'ipotesi $\lim_{T\to\infty}(X'X/T) = Q$, abbiamo

$$\begin{aligned}
\sqrt{T}(\hat{\beta}-\beta) &= \left(\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_t u_t\right) \\
&\xrightarrow{d} Q^{-1}N(0,\sigma^2 Q) \\
&\xrightarrow{d} N(0,\sigma^2 Q^{-1})
\end{aligned}$$

che è la distribuzione asintotica dello stimatore OLS di β nel modello lineare classico con regressori non-stocastici.

6.2.2 Regressori stocastici

Supponiamo ancora che $\{x_t\}$ sia una successione di vettori casuali iid con media nulla e matrice di varianza-covarianza Ω indipendenti dalla successione di v.c. $\{u_t\}$ che possiede media nulla e varianza σ^2. Abbiamo già visto come $(X'X/T)^{-1} \xrightarrow{p} \Omega^{-1}$. Vogliamo ora studiare esaminare la distribuzione limite di $T^{-1}\sum_{t=1}^{T} x_t u_t$ appropriatamente standardizzato.

Definiamo la successione di v.c. iid $\{v_t\}$ (scalare), dove $v_t = \lambda' x_t u_t$ per un qualsiasi vettore λ di dimensione k, con media nulla e varianza $\lambda'\sigma^2\Omega\lambda$ (per l'indipendenza tra x_t e u_t). Per il TCL di Lindeberg-Levy abbiamo che

$$\frac{\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T}\lambda' x_t u_t\right)}{\sqrt{\lambda'\sigma^2\Omega\lambda}} \xrightarrow{d} N(0,1)$$

o, anche,

$$\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T}\lambda' x_t u_t\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\lambda' x_t u_t \xrightarrow{d} N(0,\lambda'\sigma^2\Omega\lambda)$$

e, quindi, per il teorema "Cramer-Wold device" si ha

$$\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T}x_t u_t\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}x_t u_t \xrightarrow{d} N(0,\sigma^2\Omega).$$

Congiuntamente al risultato di convergenza in probabilità richiamato sopra, per il teorema di Slutsky, abbiamo

$$\begin{aligned}\sqrt{T}(\hat{\beta}-\beta) &= \left(\frac{1}{T}\sum_{t=1}^{T}x_t x_t'\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}x_t u_t\right)\\ &\xrightarrow{d} \Omega^{-1}N(0,\sigma^2\Omega)\\ &\xrightarrow{d} N(0,\sigma^2\Omega^{-1})\end{aligned}$$

che è la distribuzione asintotica dello stimatore OLS di β.

Abbiamo visto come lo stimatore OLS sia consistente e asintoticamente normale sotto diverse ipotesi sul processo generatore delle osservazioni (risultati analoghi si possono ottenere con regressori stocastici dipendenti ed eterogeneamente distribuiti) senza dover far ricorso alle ipotesi del modello lineare classico. Si deve ora notare che $T^{1/2}(\hat{\beta}_T-\beta) = T^{1/2}(X'X)^{-1}X'u$ non dipende in alcun modo dal valore preso da β. Questo implica che la convergenza in distribuzione di questo prodotto tra momenti campionari non è funzione di β e quindi la convergenza è *uniforme* in β. In questo caso $\hat{\beta}$ è uno stimatore consistente e uniformemente asintoticamente normale (CUAN, da Consistent and Uniformly Asymptotically Normal).

6.3 Un caso particolare: regressione con costante e trend

Consideriamo il modello $y_t = \alpha + \beta t + u_t$ con $t = 1,\ldots,T$ e $\{u_t\}$ una successione di v.c. iid con media nulla e varianza σ^2. Lo stimatore dei minimi quadrati ordinari è:

$$\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\end{pmatrix} = \begin{pmatrix} T & \sum_{t=1}^{T} t\\ \sum_{t=1}^{T} t & \sum_{t=1}^{T} t^2\end{pmatrix}^{-1}\begin{pmatrix}\sum_{t=1}^{T} y_t\\ \sum_{t=1}^{T} t y_t\end{pmatrix}$$

da cui otteniamo immediatamente

$$\begin{pmatrix}\hat{\alpha}-\alpha\\ \hat{\beta}-\beta\end{pmatrix} = \begin{pmatrix} T & \sum_{t=1}^{T} t\\ \sum_{t=1}^{T} t & \sum_{t=1}^{T} t^2\end{pmatrix}^{-1}\begin{pmatrix}\sum_{t=1}^{T} u_t\\ \sum_{t=1}^{T} t u_t\end{pmatrix}. \tag{1}$$

Se procediamo in modo standard, dividendo il primo e il secondo termine a destra per T, abbiamo, per il primo termine

$$\frac{1}{T}\begin{pmatrix} T & \sum_{t=1}^{T} t\\ \sum_{t=1}^{T} t & \sum_{t=1}^{T} t^2\end{pmatrix}$$

Consideriamo il comportamento asintotico di ciascuna componente di questa matrice separatamente. Si ha che

$$\frac{1}{T}\sum_{t=1}^{T} t = \frac{1}{T}\frac{T(T+1)}{2}$$

diverge per $T \to \infty$, così come

$$\frac{1}{T}\sum_{t=1}^{T} t^2 = \frac{1}{T}\frac{T(T+1)(2T+1)}{6}$$

diverge per $T \to \infty$. Conclusione: dividere per T il primo termine di (1) non è sufficiente per ottenere la convergenza a un limite finito. Poiché se dividiamo per T^2 si ripresenta il problema della divergenza di $\sum_{t=1}^{T} t^2$, proviamo a dividere per T^3, in modo da assicuraci un limite finito per $\sum_{t=1}^{T} t^2$. In questo caso, si ottiene

$$\begin{aligned}
\lim_{T\to\infty} \quad & \frac{1}{T^3}\sum_{t=1}^{T} t = \frac{1}{T^3}\frac{T(T+1)}{2} = 0 \\
\lim_{T\to\infty} \quad & \frac{1}{T^3}\sum_{t=1}^{T} t^2 = \frac{1}{T}\frac{T(T+1)(2T+1)}{6} = \frac{1}{3}
\end{aligned}$$

e quindi

$$\lim_{T\to\infty} \frac{1}{T^3}\begin{pmatrix} T & \sum_{t=1}^{T} t \\ \sum_{t=1}^{T} t & \sum_{t=1}^{T} t^2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1/3 \end{pmatrix}$$

con una matrice limite finita ma non invertibile. Il problema risiede nelle differenti velocità di convergenza dei vari elementi della matrice. Questo problema si risolve definendo una matrice diagonale che tenga conto di queste diverse velocità di convergenza. Sia

$$D_T = \begin{pmatrix} T^{1/2} & 0 \\ 0 & T^{3/2} \end{pmatrix}$$

questa matrice diagonale. La relazione (1) può essere riscritta, premoltiplicando ambo i lati per D_T, come

$$\begin{aligned}
D_T \begin{pmatrix} \hat{\alpha}-\alpha \\ \hat{\beta}-\beta \end{pmatrix} &= D_T \begin{pmatrix} T & \sum_{t=1}^{T} t \\ \sum_{t=1}^{T} t & \sum_{t=1}^{T} t^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^{T} u_t \\ \sum_{t=1}^{T} t u_t \end{pmatrix} \\
&= \left[D_T^{-1} \begin{pmatrix} T & \sum_{t=1}^{T} t \\ \sum_{t=1}^{T} t & \sum_{t=1}^{T} t^2 \end{pmatrix} D_T^{-1} \right]^{-1} D_T^{-1} \begin{pmatrix} \sum_{t=1}^{T} u_t \\ \sum_{t=1}^{T} t u_t \end{pmatrix} \qquad (2)
\end{aligned}$$

Considerando il primo termine a destra, abbiamo che

$$\begin{aligned}
\lim_{T\to\infty} D_T^{-1} \begin{pmatrix} T & \sum_{t=1}^{T} t \\ \sum_{t=1}^{T} t & \sum_{t=1}^{T} t^2 \end{pmatrix} D_T^{-1} &= \lim_{T\to\infty} \begin{pmatrix} 1 & (1/T^2)\sum_{t=1}^{T} t \\ (1/T^2)\sum_{t=1}^{T} t & (1/T^3)\sum_{t=1}^{T} t^2 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{pmatrix} = Q
\end{aligned}$$

dove la matrice Q è finita e invertibile. Consideriamo ora il secondo termine a destra della (2),

$$D_T^{-1} \begin{pmatrix} \sum_{t=1}^{T} u_t \\ \sum_{t=1}^{T} t u_t \end{pmatrix} = \begin{pmatrix} \dfrac{1}{\sqrt{T}} \displaystyle\sum_{t=1}^{T} u_t \\ \dfrac{1}{T^{3/2}} \displaystyle\sum_{t=1}^{T} t u_t \end{pmatrix} = \frac{1}{\sqrt{T}} \begin{pmatrix} \sum_{t=1}^{T} u_t \\ \sum_{t=1}^{T} (t/T) u_t \end{pmatrix},$$

dove il primo elemento di questo vettore aleatorio converge in distribuzione alla v.c. $N(0,1)$ per il TCL di Lindeberg-Levy per prcessi iid. Per il secondo elemento si osservi che la successione di v.c.$\{(t/T)u_t\}$ possiede media nulla e varianza pari a $(t/T)^2\sigma^2$. Poiché si tratta di una successione di v.c. indipendenti ma non identicamente distribuite, dobbiamo verificare le condizioni del TCL di Lyapunov. Assumiamo che tu_t soddisfi la condizione sui momenti e concentriamoci invece sulla condizione sul limite della somma delle varianze,appropriatamente standardizzata. Poiché

$$\sum_{t=1}^{T}(t/T)^2\sigma^2 = \frac{\sigma^2}{T^2}\frac{T(T+1)(2T+1)}{6}$$

ponendo $g(T)=T$ si ottiene

$$\lim_{T\to\infty}\bar{\sigma}_T^2 = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}(t/T)^2\sigma^2 = \lim_{T\to\infty}\frac{\sigma^2}{T^3}\frac{T(T+1)(2T+1)}{6} = \frac{\sigma^2}{3} = \bar{\sigma}^2$$

e quindi, possiamo applicare il TCL di Lyapunov (alla luce dell'osservazione (iv)), ottenendo

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(t/T)u_t = \frac{1}{T^{3/2}}\sum_{t=1}^{T}tu_t \xrightarrow{d} N(0,\sigma^2/3).$$

La conoscenza delle distribuzioni marginali asintotiche appena derivate non è però sufficiente per stabilire la convergenza in distribuzione della successione di vettori casuali $\{(u_t \quad (t/T)u_t)'\}$ e a questo scopo dobbiamo utilizzare il teorema "Cramer-Wold device". Definiamo la v.c. $w_{tT}=\lambda_1 u_t+\lambda_2(t/T)u_t$. È facile verificare che w_{tT} possiede media nulla e varianza pari a $\lambda_1^2\sigma^2+\lambda_2^2(t/T)^2\sigma^2+2\lambda_1\lambda_2(t/T)\sigma^2$. Inoltre,

$$\begin{aligned}
\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\left[\lambda_1^2\sigma^2+\lambda_2^2\left(\frac{t}{T}\right)^2\sigma^2+2\lambda_1\lambda_2\frac{t}{T}\sigma^2\right] &= \lambda_1^2\sigma^2+\frac{\lambda_2^2\sigma^2}{3}+\frac{2\lambda_1\lambda_2\sigma^2}{2}\\
&= \sigma^2\,(\lambda_1 \quad \lambda_2)\begin{pmatrix}1 & 1/2\\ 1/2 & 1/3\end{pmatrix}\begin{pmatrix}\lambda_1\\ \lambda_2\end{pmatrix}\\
&= \sigma^2\lambda' Q\lambda
\end{aligned}$$

per cui si può applicare il TCL di Lyapunov (valido anche se w_{tT} dipende da T) e dedurre che

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\lambda_1 u_t+\lambda_2(t/T)u_t) \xrightarrow{d} N(0,\sigma^2\lambda' Q\lambda).$$

Infine, per il teorema "Cramer-Wold device", si ottiene

$$\frac{1}{\sqrt{T}}\begin{pmatrix}\sum_{t=1}^{T}u_t\\ \sum_{t=1}^{T}(t/T)u_t\end{pmatrix} = D_T^{-1}\sum_{t=1}^{T}\begin{pmatrix}u_t\\ tu_t\end{pmatrix} \xrightarrow{d} N(0,\sigma^2 Q).$$

e, per il teorema di Slutsky, si ricava la distribuzione limite

$$\begin{pmatrix}\sqrt{T}(\hat{\alpha}-\alpha)\\ T^{3/2}(\hat{\beta}-\beta)\end{pmatrix} \xrightarrow{d} N(0,\sigma^2 Q^{-1}).$$

6.4 Consistenza di s_T^2

Consideriamo lo stimatore corretto di σ^2 nel modello lineare classico $s_T^2 = (T-k)^{-1}\sum_{t=1}^{T}\hat{u}_t^2$ che possiamo riscrivere come $s_T^2 = (T-k)^{-1}\hat{u}'\hat{u} = (T-k)^{-1}[u'u - u'X(X'X)^{-1}X'u]$. Chiaramente, per k fissato, non fa alcuna differenza asintoticamente se $(T-k)$ viene sostituito con T come faremo nel seguito. Poiché il processo $\{u_t\}$ è iid, anche il processo $\{u_t^2\}$ sarà iid con $E(u_t^2) = \sigma^2$ costante. Possiamo quindi appllicare la LGN di Khinchine per dedurre che

$$\frac{1}{T}\sum_{t=1}^{T} u_t^2 \xrightarrow{p} \sigma^2.$$

Rimane da analizzare il secondo termine e, in particolare, per mostrare la consistenza di s_T^2 bisogna mostrare che $T^{-1}u'X(X'X)^{-1}X'u \xrightarrow{p} 0$. Se assumiamo, come in precedenza, che $(X'X)^{-1} \xrightarrow{p} Q^{-1}$ e se teniamo conto del risultato già ottenuto

$$\frac{1}{\sqrt{T}}X'u = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_t u_t \xrightarrow{d} N(0,\sigma^2 Q)$$

che implica $T^{-1}X'u \xrightarrow{p} 0$, allora si ottiene

$$\frac{1}{T}u'X(X'X)^{-1}X'u = \left(\frac{u'X}{T}\right)\left(\frac{X'X}{T}\right)^{-1}\left(\frac{X'u}{T}\right) \xrightarrow{p} 0\cdot Q^{-1}\cdot 0 = 0$$

che è il risultato desiderato.

6.5 Distribuzione asintotica di s_T^2

Esaminiamo ora la distribuzione asintotica di s_T^2, sempre dividendo per T e non per $T-k$, data l'equivalenza asintotica. Poiché il processo $\{u_t^2\}$ è iid con media pari a σ^2, per poter applicare il TCL di Lindeberg.Levy dobbiamo assumere che la varianza di u_t^2 esista e sia finita che, a sua volta, implica l'esistenza di momenti quarti finiti per u_t. Assumiamo quindi che $E(u_t^4) = \mu_4$ da cui otteniamo $\mathrm{Var}(u_t^2) = \mu_4 - \sigma^4$. Il risultato cercato segue immediatamente da

$$\begin{aligned}\sqrt{T}(s_T^2 - \sigma^2) &= \frac{u'u}{\sqrt{T}} - \frac{u'X(X'X)^{-1}X'u}{\sqrt{T}} - \sqrt{T}\sigma^2 \\ &= \sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T} u_t^2 - \sigma^2\right) - \frac{u'X(X'X)^{-1}X'u}{\sqrt{T}} \xrightarrow{d} N(0,\mu_4 - \sigma^4)\end{aligned}$$

dove l'ultima riga segue dal fatto che

$$\frac{u'X(X'X)^{-1}X'u}{\sqrt{T}} = \frac{u'X}{T}\left(\frac{X'X}{T}\right)^{-1}\frac{X'u}{\sqrt{T}} \xrightarrow{d} 0\cdot Q^{-1}\cdot N(0,\sigma^2 Q) = 0.$$

7 Verifica di ipotesi e intervalli di confidenza

Data l'approssimazione asintotica per la distribuzone di $\hat{\beta}$ è immediato costruire sia intervalli di confidenza approssimati che determinare la distribuzione asintotica di statistiche tests costruite a partire dagli stimatori OLS del modello di regressione. Supponiamo che il modello di regressione sia

$$y_t = x_t'\beta + u_t$$

con $t = 1, \ldots, T$ e $E(u_t) = 0$ e $E(u_t^2) = \sigma^2$. Inoltre, (i) la successione di vettori casuali $\{x_t, u_t\}$ è indipendente ma non necessariamente identicamente distribuita (accomodando quindi il caso di regressori non-stocastici), con (ii) momenti di ordine quarto finiti, (iii) x_t e u_t sono indipendenti per ogni t. Sotto queste condizioni sappiamo che $\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1})$, dove $T^{-1}\sum_{t=1}^{T} x_t x_t' \xrightarrow{p} Q$.

Questa approssimazione richiede due quantità non note: σ^2 e Q. La procedura naturale consiste nel sostituire queste due quantità con degli stimatori consistenti. Abbiamo già visto come σ^2 sia stimato consistentemente da s_T^2 (ma anche da $\hat{\sigma}_T^2 = (T/(T-k))s_T^2$) e come $T^{-1}\sum_{t=1}^{T} x_t x_t' \xrightarrow{p} Q$. A questo punto non ci rimane che applicare il teorema di Slutsky per dedurre

$$\sigma_T^{-1}\left[\left(\frac{X'X}{T}\right)\right]^{1/2}\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, I_k).$$

Consideriamo ora l'ipotesi nulla $H_0 : R\beta = r$ dove la matrice R di dimensione $(q \times k)$ è di rango pieno pari a q. Procedendo come sopra si ottiene

$$\hat{\sigma}_T^{-1}\left[R\left(\frac{X'X}{T}\right)R'\right]^{1/2}\sqrt{T}R(\hat{\beta} - \beta) \xrightarrow{d} N(0, I_q)$$

da cui si ricava

$$(R\hat{\beta}_T - R\beta)[\hat{\sigma}_T^2 R(X'X)R']^{-1}(R\hat{\beta} - R\beta) \xrightarrow{d} \chi_q^2$$

che cositutisce il punto di partenza per la costruzione di intervalli di confidenza multivariati asintotici nel modello lineare classico (con la differenza rispetto all'analisi in campioni finiti che $\hat{\sigma}_T^2$ viene trattato come una costante e non una v.c.).

In modo del tutto analogo, sotto l'ipotesi nulla H_0 la statistica test di Wald si distribuisce asinoticamente come un χ^2 con gradi di libertà pari a q (cioè al numero di restrizioni sotto l'ipotesi nulla)

$$W = (R\hat{\beta}_T - r)[\hat{\sigma}_T^2 R(X'X)R']^{-1}(R\hat{\beta} - r) \xrightarrow{d} \chi_q^2.$$

Si può mostrare, vedi seguito, che anche le statistiche test dei moltiplicatori di Lagrange e del rapporto di verosimiglianza si distribuiscono asintoticamente come un χ^2 con gradi di libertà pari a q.

8 Proprietà asintotiche dello stimatore ML

Supponiamo di estrarre un campione casuale Z di dimensione T da una popolazione con fdp $f(z;\theta)$ completamente caratterizzata dal vettore θ di dimensione k di parametri ignoti[2]. La funzione di verosimiglianza è definita da

$$L(\theta;Z_1,\ldots,Z_T)=\prod_{t=1}^{T}f(z_t;\theta)$$

anche se noi lavoreremo con la log-verosimiglianza

$$l(\theta;Z)=\ln L(\theta;Z)=\sum_{t=1}^{T}\ln f(z_t;\theta).$$

Lo stimatore di massima verosimiglianza (ML) è definito come il vettore di parametri che massimizza la funzione di (log)verosimiglianza e lo indichiamo con $\hat{\theta}_T$. Inoltre, definiamo lo score per la singola osservazione

$$s(\theta;z_t)=\frac{\partial \ln f(\theta;z_t)}{\partial\theta}$$

per cui valgono i seguenti risultati

$$\begin{aligned} E[s(\theta;z_t)] &= 0,\\ E[s(\theta;z_t)^2] &= -E\left[\frac{\partial^2 \ln f(\theta;z_t)}{\partial\theta\partial\theta'}\right]=I_t(\theta) \end{aligned}$$

dove $I_t(\theta)$ è la matrice di informazione per la singola osservazione t.

Esempio 8.1. (ML nel modello lineare) Per il modello lineare con errori distribuiti normalmente, la funzione di log-verosimiglianza per β e σ^2 è

$$\begin{aligned} l(\beta,\sigma^2;X,y) &= -\frac{T}{2}\ln(2\pi\sigma^2)-\frac{T}{2\sigma^2}(y-X\beta)'(y-X\beta)\\ &= -\frac{T}{2}\ln(2\pi\sigma^2)-\frac{T}{2\sigma^2}\sum_{t=1}^{T}(y_t-x_t'\beta)^2 \end{aligned}$$

il vettore dello *score* è dato da

$$s(\beta,\sigma^2;X,y)=\begin{pmatrix} \frac{1}{\sigma^2}\sum_{t=1}^{T}x_t(y_t-x_t'\beta)\\ -\frac{T}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{t=1}^{T}(y_t-x_t'\beta)^2 \end{pmatrix},$$

la matrice hessiana è data da

$$\begin{pmatrix} -\frac{1}{\sigma^2}\sum_{t=1}^{T}x_tx_t' & -\frac{1}{\sigma^4}\sum_{t=1}^{T}x_t(y_t-x_t'\beta)\\ -\frac{1}{\sigma^4}\sum_{t=1}^{T}x_t(y_t-x_t'\beta) & \frac{T}{2\sigma^4}-\frac{1}{\sigma^6}\sum_{t=1}^{T}(y_t-x_t'\beta)^2 \end{pmatrix}$$

[2]Formalmente, devono valere le cosidette “condizioni di regolarità”: (i) $\theta\in A\subset\mathbb{R}^k$ dove A è un insieme chiuso e limitato; (ii) la funzione di distribuzione $F(z;\theta)$ è continua e derivabile tre volte rispetto a θ; (iii) il dominio di z non dipende da θ.

da cui si ottiene la matrice di informazione

$$I(\beta,\sigma^2) = \begin{pmatrix} \frac{1}{\sigma^2}\sum_{t=1}^{T} x_t x_t' & 0 \\ 0 & \frac{T}{2\sigma^4} \end{pmatrix}.$$

A volte lo stimatore di ML, se corretto, possiede varianza campionaria pari all'estremo inferiore di Cramer-Rao, che è pari all'inversa della matrice di informazione attesa. In questo caso, si dice che nella classe degli stimatori corretti, lo stimatore di ML è efficiente.

Esempio 8.2. (ML nel modello lineare - continua) L'inversa della matrice di informazione nel modello lineare è abbastanza semplice in quanto essa è diagonale a blocchi tra β e σ^2. Per quanto riguarda lo stimatore di β, la disuguaglianza di Cramer-Rao afferma che non esiste stimatore con varianza più piccola di $\sigma^2(X'X)^{-1}$, che è la varianza sia dello stimatore ML che OLS, che sono quindi efficienti. Lo stimatore di ML per σ^2 dato da $\hat{u}'\hat{u}/T$ non è corretto mentre lo stimatore OLS dato da $\hat{u}'\hat{u}/(T-k)$ è corretto con varianza pari a $2\sigma^4/(T-k)$ che eccede l'estremo inferiore di Cramer-Rao. In realtà, non esiste stimatore corretto di σ^2 che raggiunge questo estremo inferiore.

8.1 Consistenza

Consideriamo lo stimatore di ML di β_0, vero valore nella popolizione, nel modello lineare con errori normali nel caso in cui la varianza σ^2 sia nota. Questo semplice caso ci permette di evidenziare in modo semplice ma efficace i problemi connessi, in generale, alla dimostrazione della consistenza dello stimatore ML. In questo contesto la log-verosimiglianza media è proporzionale a

$$\begin{aligned} \frac{1}{T} l(\beta;X,y) \quad = \quad & \text{costante} - \frac{1}{2T}\sum_{t=1}^{T}\left(\frac{y_t - x_t'\beta_0}{\sigma^2}\right)^2 + \frac{1}{T\sigma^2}(\beta-\beta_0)'\sum_{t=1}^{T} x_t(y_t - x_t\beta_0) - \\ & \frac{1}{T2\sigma^2}\sum_{t=1}^{T}(\beta-\beta_0)'x_t x_t'(\beta-\beta_0). \end{aligned}$$

Considerando i termini a destra separatamente, a parte il termine costante che tralasciamo, abbiamo:

- $T^{-1}\sum_{t=1}^{T}[(y_t - x_t'\beta_0)/\sigma^2]^2 \xrightarrow{p} 1$, in quanto $\{(y_t - x_t'\beta_0)/\sigma^2\}$ è un successione di v.c. iid $\sim \chi_1^2$ con media pari a 1 e possiamo applicarea la LGN di Khinchine;
- $T^{-1}\sum_{t=1}^{T} x_t(y_t - x_t\beta_0) \xrightarrow{p} 0$ per la LGN di Khinchine applicata alla successione iid $\{x_t(y_t - x_t\beta_0)\} = \{x_t u_t\}$ con media nulla;

Di conseguenza, asintoticamente, la log-verosimiglianza è una funzione di β solo attraverso l'ultimo termine, una forma quadratica in β centrata sul suo vero valore β_0. Questa forma quadratica è massimizzata per $\beta = \beta_0$. Poiché sappiamo che $\hat{\beta}_T =$

$(X'X)^{-1}X'y$ massimizza la log-verosimiglianza per ogni numerosità T ed è uno stimatore ML e poiché la log-verosimiglianza media converge a una funzione con un massimo in β_0, è plausibile attendersi che $\hat{\beta}_T$ converga a β_0. Questo, in realtà, è vero come abbiamo visto da un esame diretto di $\hat{\beta}_T$.

Sfortunatamente, metodi diretti di prova come nel caso del modello lineare con errori normali non sono generalmente disponibili, in quanto spesso non esiste una espressione in forma chiusa per lo stimatore di ML e l'unica informazione disponibile consiste nel fatto che $\hat{\theta}_T$ massimizza una qualche funzione obiettivo. La consistenza dello stimatore di ML viene di solito stabilita mostrando la convergenza di una successione di funzioni obiettivo a una funzione limite che possiede un massimo globale unico. Questo implica che la successione dei massimi globali associata alla successione di funzioni obiettivo converge all'unico massimo globale della funzione limite. Un'analisi formalizzata di questo approccio esula dagli obiettivi del nostro corso (si veda Amemiya (1985) per una discussione).

La dimostrazione della consistenza dello stimatore di ML ha bisogno di due elementi:

1. la convergenza della successione di funzioni di log-verosimiglianza standardizzate per la numerosità campionaria,cioè

$$l_T(\theta;z) = \frac{1}{T}l(\theta;z) = \frac{1}{T}\sum_{t=1}^{T}\ln f(\theta;z_t)$$

   a una funzione limite $l(\theta,z)$. Ad es., se $\{z_t\}$ è una successione di vettori aleatori iid con media finita, per la LGN di Khinchine la successione $\{l_T(\theta,z)\}$ converge in probabilità a questa media finita.

2. l'unicità del massimo θ_0 della funzione limite. Questo risultato deriva dalla disuguaglianza[3]

$$E[l(\theta_0,z)] \geq E[l(\theta,z)] \qquad \forall\theta \neq \theta_0.$$

8.2 Distribuzione asintotica

Per determinare il comportamento asintotico dello stimatore ML dobbiamo, al solito, fare ipotesi addizionali che ci permettano di individuare statistiche che sono somme di v.c. appropriatemente standardizzate attraverso trasformazioni stabilizzatrici in modo da poter applicare un qualche TCL.

Consideriamo il caso semplice in cui la successione $\{z_t\}$ è iid con media e varianza finita. Partendo dalle equazioni normali $s(\hat{\theta}_T;z) = 0$ e considerando uno sviluppo in

[3] Dalla disuguaglianza di Jensen per funzioni convesse, come la funzione di verosimiglianza, $E[g(x)] \geq g(E(x))$, si ha che

$$E[-\ln[L(\theta;z)/L(\theta_0;z)] \geq -\ln(E[L(\theta;z)/L(\theta_0;z)]).$$

Quindi, poiché $E[L(\theta;z)/L(\theta_0;z)] = 1$, la disuguaglianza segue immediatamente.

serie di Taylor del primo ordine[4]

$$0 = s(\theta_0;z) + \frac{\partial^2 l(\bar{\theta}_T;z)}{\partial\theta\partial\theta'}(\hat{\theta}_T - \theta_0)$$

dove $\bar{\theta}_T = \alpha\hat{\theta}_T + (1-\alpha)\theta_0$, con $\alpha \in [0,1]$, è un punto tra $\hat{\theta}_T$ e θ_0. Chiaramente la consistenza di $\hat{\theta}_T$ implica la consistenza di $\bar{\theta}_T$. Standardizzando per $T^{-1/2}$ possiamo applicare un TCL al primo termine a destra ottenendo

$$\frac{1}{\sqrt{T}} s(\theta_0;z) \xrightarrow{d} N(0, I(\theta_0)).$$

Infatti, partendo dalla successione di vettori casuali $\left\{\frac{\partial \ln f(z_t;\theta)}{\partial\theta}\right\}$, che possiede valore atteso nullo e matrice di varianza-covarianza pari a $I_t(\theta_0)$, possiamo costruire la successione di v.c. $\{\lambda' \frac{\partial \ln f(z_t;\theta)}{\partial\theta}\}$ con media nulla e varianza pari a $\lambda' I_t(\theta_0)\lambda$. Per il TCL di Lindeberg-Levy si ha che

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \lambda' \left(\frac{\partial \ln f(z_t;\theta)}{\partial\theta}\right) \xrightarrow{d} N(0, \lambda' I(\theta_0)\lambda)$$

e, infine, per il teorema "Cramer-Wold device" si ottiene il risultato desiderato che corrisponde anche alla distribuzione limite di

$$-\frac{1}{T} \frac{\partial^2 l(\bar{\theta}_T;z)}{\partial\theta\partial\theta'} \sqrt{T}(\hat{\theta}_T - \theta_0).$$

Dobbiamo ora esaminare il termine

$$\frac{1}{T} \frac{\partial^2 l(\bar{\theta}_T;z)}{\partial\theta\partial\theta'} = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial^2 \ln f(\bar{\theta}_T;z_t)}{\partial\theta\partial\theta'}.$$

e studiare se la successione $\left\{\frac{\partial^2 \ln f(\bar{\theta}_T;z_t)}{\partial\theta\partial\theta'}\right\}$ soddisfa la LGN di Khinchine. Poiché

$$E\left[\frac{\partial^2 \ln f(\bar{\theta}_T;z_t)}{\partial\theta\partial\theta'}\right] = -E\left[\left(\frac{\partial \ln f(\bar{\theta}_T;z_t)}{\partial\theta}\right)^2\right] = -I(\bar{\theta})$$

costante per tutte le osservazioni (ricordate le osservazioni sono iid) e $\bar{\theta} \xrightarrow{p} \theta_0$ (poiché $ol\theta \xrightarrow{p} \alpha\theta_0 + (1-\alpha)\theta_0 = \theta_0$), per la LGN di Khinchine abbiamo che

$$\frac{1}{T} \frac{\partial^2 l(\bar{\theta}_T;z)}{\partial\theta\partial\theta'} \xrightarrow{p} -I(\theta_0).$$

[4] In generale, lo sviluppo in serie di Taylor di ordine n di una funzione $f(x)$ in un intorno del punto x_0 è dato da

$$f(x) = f(x_0) + (x-x_0) f^{(1)}(x_0) + \frac{1}{2!}(x-x_0)^2 f^{(2)}(x_0) + \frac{1}{3!}(x-x_0)^3 f^{(3)}(x_0) + \cdots \frac{1}{n!}(x-x_0)^n f^{(n)}(x^*)$$

dove $f^{(i)}(x_0)$ è la derivata di ordine i-esimo della funzione $f(x)$ valutata in x_0 e $f^{(n)}(x^*)$ è la derivata di ordine i-esimo della funzione $f(x)$ valutata in x^*, un punto compreso tra x e x_0.

A questo punto è sufficiente assemblare il rilsultato sulla convergenza in distribuzione del vettore dello score e sulla convergenza in probabilità della marice Hessiana. Infatti, abbiamo che

$$\left[\frac{1}{T}\frac{\partial^2 l(\overline{\theta}_T;z)}{\partial\theta\partial\theta'}\right]^{-1}\frac{1}{\sqrt{T}}s(\theta_0;z)\xrightarrow{d}N(0,I(\theta_0)^{-1})$$

che, riscrivendo l'approssimazione di Taylor come

$$\sqrt{T}(\hat{\theta}-\theta_0)=-\left[\frac{1}{T}\frac{\partial^2 l(\overline{\theta}_T;z)}{\partial\theta\partial\theta'}\right]^{-1}\frac{1}{\sqrt{T}}s(\theta_0;z)$$

ci permette di ottenere il risultato desiderato

$$\sqrt{T}(\hat{\theta}_T-\theta_0)\xrightarrow{d}N(0,I(\theta_0)^{-1}).$$

Osservazioni: (i) il risultato di normalità asintotica implica che quando T è grande, la matrice di varianza-covarianza dello stimatore ML può essere approssimata da meno il valore atteso della matrice Hessiana valutata nel vero vettore dei parametri. Ma questo è esattamente l'estremo inferiore di Cramer-Rao per uno stimatore non distorto. Si dice che uno stimatore corretto la cui varianza raggiunge asintoticamente l'estremo inferiore di Cramer-Rao è *asintoticamente efficiente*. Lo stimatore ML è per definizione asintoticamente efficiente.

Esempio 8.3. (ML nel modello lineare classico - continua) Introduciamo l'ulteriore ipotesi che $x_t \sim \text{iid}(0,\Omega)$ sia indipendente da $u_t \sim N(0,\sigma_0^2)$ con $E(u_t u_s)=0$ per ogni t e s. Iniziamo considerando il vettore dello score standardizzato per $\sqrt{T}$ e valutato in $\theta_0=(\beta_0 \quad \sigma_0^2)'$, vero valore dei parametri:

$$\frac{1}{\sqrt{T}}s(\beta_0,\sigma_0^2;X,y)=\begin{pmatrix}\dfrac{1}{\sqrt{T}}\displaystyle\sum_{t=1}^{T}\frac{x_t u_t}{\sigma_0^2}\\ \dfrac{1}{\sqrt{T}}\displaystyle\sum_{t=1}^{T}\left(\frac{u_t^2-\sigma_0^2}{2\sigma_0^4}\right)\end{pmatrix}.$$

Consideriamo la successione di vettori casuali

$$\left\{\begin{matrix}x_t u_t/\sigma_0^2\\(u_t^2-\sigma_0^2)/2\sigma_0^4\end{matrix}\right\}$$

che per ogni t possiede media nulla e varianza pari alla matrice di informazione

$$I(\beta_0,\sigma_0^2)=\begin{pmatrix}\dfrac{\Omega}{\sigma_0^2} & 0\\ 0 & \dfrac{1}{2\sigma_0^4}\end{pmatrix}.$$

Possiamo quindi applicare il teorema di Lindeberg-Levy alla successione

$$\{v_t\}=\left\{\lambda'\begin{pmatrix}x_t u_t/\sigma_0^2\\(u_t^2-\sigma_0^2)/2\sigma_0^4\end{pmatrix}\right\}$$

per cui

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}v_t\xrightarrow{d}N(0,\lambda' I(\beta_0,\sigma_0^2)\lambda)$$

e, infine, per il teorema Cramer-Wold device, si ottiene

$$\frac{1}{\sqrt{T}}s(\beta_0,\sigma_0^2;X,y) = \begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{x_t u_t}{\sigma_0^2} \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{u_t^2-\sigma_0^2}{2\sigma_0^4}\right) \end{pmatrix} \xrightarrow{d} N(0,I(\beta_0,\sigma_0^2)).$$

Dobbiamo ora volgere la nostra attenzione alla studio della convergenza in probabilità della matrice Hessiana. Considerando elemento per elemento della matrice Hessiana, valutata nel vero vettore di parametri $(\beta_0\ \sigma_0^2)'$ e cambiato di segno, abbiamo che

$$\begin{aligned} \frac{1}{T}\frac{1}{\sigma_0^2}\sum_{t=1}^{T}x_t x_t' &\xrightarrow{p} \Omega/\sigma_0^2 \\ \frac{1}{T}\frac{1}{\sigma_0^4}\sum_{t=1}^{T}x_t(y_t - x_t'\beta_0) &\xrightarrow{p} 0 \\ -\frac{1}{2\sigma^4}+\frac{1}{\sigma_0^6}\frac{1}{T}\sum_{t=1}^{T}u_t^2 &\xrightarrow{p} \frac{1}{2\sigma_0^4} \end{aligned}$$

che è pari alla matrice di informazione $I(\beta_0,\sigma_0^2)$. Quindi possiamo concludere che

$$\sqrt{T}\begin{pmatrix} (\hat{\beta}_T-\beta_0) \\ (\hat{\sigma}_T^2-\sigma_0^2) \end{pmatrix} \xrightarrow{d} N\left(0,\begin{pmatrix} \sigma_0^2\Omega^{-1} & 0 \\ 0 & 2\sigma_0^4 \end{pmatrix}\right).$$

Osservazioni: (i) $\hat{\beta}_T$ e $\hat{\sigma^2}_T$ sono asintoticamente efficienti; (ii) $\hat{\beta}_T$ e $\hat{\sigma^2}_T$ sono asintoticamente indipendenti, come è giusto attendersi visto che sono indipendenti per ogni T data la diagonalità a blocchi della matrice di informazione.

9 Proprietà asintotiche dei tests di Wald, LR e LM

Lo studio delle proprietà asintotiche dei tre tests connessi alla verosimiglianza avviene in due fasi. Primo, mostreremo che se la funzione di log-verosimiglianza per il vettore di parametri θ è quadratica allora le tre statistiche test di Wald, LR e LM per restrizioni lineari del tipo $R\theta = r$ sono identiche (e quindi condividono la stessa distribuzione limite). Secondo, mostreremo che la funzione di log-verosimiglianza è "approssimativamente" quadratica per T che tende a infinito in un intorno di $\hat{\theta}_T$ lo stimatore di ML, dove "approssimativamente" deve intendersi nel senso che ciò che della funzione di log-verosimiglianza viene tralasciato converge a zero per T che tende a infinito. (Naturalmente, i risultati di consistenza e normalita asintotica ottenuti sono ancora validi sotto le ipotesi fatte in precedenza.)

Iniziamo analizzando quest'ultimo punto: la funzione di log-verosimiglianza è "approssimativamente" quadratica per T che tende a infinito. Sia $l(\theta_0;z) = \sum_{t=1}^{T}\ln f(z_t,\theta_0)$ la funzione di log-verosimiglianza valutata in θ_0, il vero valore del parametro e sia $\hat{\theta}_T$ lo stimatore di ML per un campione di T osservazioni. Consideriamo lo sviluppo in serie di Taylor, del terzo ordine, di $l(\theta_0; z)$ in un intorno di $\hat{\theta}_T$:

$$\begin{aligned} l(\theta_0; z) &= l(\hat{\theta}_T, z) + \left(\frac{\mathrm{d}l(\hat{\theta}_T, z)}{\mathrm{d}\theta}\right)(\theta_0 - \hat{\theta}_T) + \\ &+ \frac{1}{2}(\theta_0 - \hat{\theta}_T)^2 \left(\frac{\mathrm{d}^2 l(\hat{\theta}_T, z)}{\mathrm{d}\theta^2}\right) + \frac{1}{6}\frac{\mathrm{d}^3 l(\bar{\theta}, z)}{\mathrm{d}\theta^3}(\theta_0 - \hat{\theta}_T)^3 \end{aligned}$$

dove $\bar{\theta} = \alpha\hat{\theta}_T + (1-\alpha)\theta_0$ è un punto tra $\hat{\theta}_T$ e θ_0. Chiaramente il termine $\frac{\mathrm{d}l(\hat{\theta}_T, z)}{\mathrm{d}\theta}$ vale zero per la definizione stessa di stimatore di ML. L'ultimo termine di questa espansione può essere riscritto come (trascurando la costante (1/6)):

$$\frac{1}{\sqrt{T}}\frac{1}{T}\frac{\mathrm{d}^3 l(\bar{\theta}, z)}{\mathrm{d}\theta^3}\sqrt{T}(\theta_0 - \hat{\theta}_T)\sqrt{T}(\theta_0 - \hat{\theta}_T)\sqrt{T}(\theta_0 - \hat{\theta}_T)$$

dove $\sqrt{T}(\theta_0 - \hat{\theta}_T) \xrightarrow{d} N(0, I(\theta_0)^{-1})$ per la normalità asintotica dello stimatore di ML. A questo punto, poiché $\bar{\theta} \xrightarrow{p} \theta_0$, è sufficiente assumere che

$$E\left[\frac{\mathrm{d}^3 \ln f(z, \bar{\theta})}{\mathrm{d}\theta^3}\right]$$

esista e sia finito per poter applicare una legge dei grandi dei numeri e stabilire che

$$\frac{1}{T}\frac{\mathrm{d}^3 l(\bar{\theta}, z)}{\mathrm{d}\theta^3}$$

converge in probabilità a questo valore atteso e concludere che tutto l'ultimo termine dell'espansione converge a zero a causa del fattore $T^{-1/2}$ che tende a zero per T che tende a infinito.

Di conseguenza, per T che tende a infinito la log-verosimiglianza è una funzione quadratica

$$l(\theta_0; z) = l(\hat{\theta}_T, z) + \frac{1}{2}(\theta_0 - \hat{\theta}_T)^2 \left(\frac{\mathrm{d}^2 l(\hat{\theta}_T, z)}{\mathrm{d}\theta^2}\right) \qquad \text{per} \quad T \to \infty$$

che è relativamente semplice da massimizzare. Infatti, poiché il fattore $\frac{\mathrm{d}^2 l(\hat{\theta}_T, z)}{\mathrm{d}\theta^2}$ è non-negativo (una matrice positiva semidefinita nel caso in cui θ sia un vettore di parametri in quanto è l'inversa della matrice di varianza-covarianza), è immediato concludere che $\hat{\theta}_T$ massimizza l'approssimazione quadratica.

Visto che la log-verosimiglianza è quadratica in θ_0 per T che tende a infinito, consideriamo ora la forma delle statistiche test di Wald, LR e LM quando la log-verosimiglianza è quadratica e si vuole sottoporre a verifica di ipotesi vincoli lineari del tipo $R\theta = r$ dove R ha rango pieno pari a q. Supponiamo quindi che la log-verosimiglianza per θ, un vettore di parametri di dimensione k, sia data da

$$l(\theta; z) = b - \frac{1}{2}(\theta - \hat{\theta}_T)' A_T (\theta - \hat{\theta}_T)$$

dove A_T è una matrice simmetrica definita positiva che, in generale, può essere funzione delle osservazioni o di parametri noti, b è uno scalare, e $\hat{\theta}_T$ è uno stimatore (funzione delle osservazioni) di θ. Sotto queste condizioni abbiamo che

$$\begin{aligned} s(\theta) &= -A_T'(\theta - \hat{\theta}_T) \\ I_T(\theta) &= -E\left[\frac{\partial^2 l\theta; z)}{\partial\theta\partial\theta'}\right] = E(A_T) \end{aligned}$$

dove si deve notare che

- la matrice di informazione è indipendente da θ, e quindi rimane invariata nel problema di stima non vincolata e vincolata, cioè sotto l'ipotesi nulla. Inoltre, se A_T non è casuale allora $I(\theta) = A_T$;
- $\hat{\theta}_T$ è lo stimatore ML di θ in quanto soddisfa le condizioni del primo ordine $s(\hat{\theta}_T) = 0$;
- di conseguenza:
$$\sqrt{T}(\hat{\theta}_T - \theta) \xrightarrow{d} N(0, I(\theta)^{-1})$$
dove $I(\theta)$ è pari al limite di $T^{-1}A_T$ per T che tende a infinito .

Poiché sotto l'ipotesi nulla abbiamo che

$$\sqrt{T}(R\hat{\theta}_T - r) \xrightarrow{d} N(0, RI(\theta)^{-1}R')$$

possiamo derivare immediatamente la distribuzione limite del test di Wald

$$\begin{aligned} W &= (R\hat{\theta}_T - r)'\left[RA_T^{-1}R'\right]^{-1}(R\hat{\theta}_T - r) \\ &= \sqrt{T}(R\hat{\theta}_T - r)'\left[R\left(\frac{A_T}{T}\right)^{-1}R'\right]^{-1}\sqrt{T}(R\hat{\theta}_T - r) \xrightarrow{d} \chi^2_q. \end{aligned}$$

Consideriamo ora la statistica test LM. Sotto il vincolo $R\theta = r$ bisogna risolvere il problema di massimizzazione

$$\max_{\theta,\lambda} b - \frac{1}{2}(\theta - \hat{\theta}_T)'A_T(\theta - \hat{\theta}_T) - \lambda'(R\theta - r)$$

da cui si ottengono le condizioni del primo ordine

$$\begin{aligned} -A_T(\theta - \hat{\theta}_T) + R'\lambda &= 0 \\ R\theta - r &= 0 \end{aligned}$$

che possono essere risolte per $\tilde{\theta}_T$ lo stimatore di θ vincolato e per $\tilde{\lambda}_T$

$$\begin{aligned} \tilde{\theta}_T &= \hat{\theta}_T + A_T^{-1}R'(RA_T^{-1}R')^{-1}(R\hat{\theta}_T - r) \\ \tilde{\lambda}_T &= -(RA_T^{-1}R')^{-1}(R\hat{\theta}_T - r) \end{aligned}$$

da cui si ottiene

$$\sqrt{T}\tilde{\lambda}_T \xrightarrow{d} N(0, (RI(\theta)^{-1}R')^{-1})$$

A questo punto è immediato costruire il test LM

$$\begin{aligned}
LM &= \tilde{\lambda}_T' R A_T^{-1} R' \tilde{\lambda}_T \\
&= (R\hat{\theta}_T - r)' \left[RA_T^{-1}R'\right]^{-1} (R\hat{\theta}_T - r) \\
&= \sqrt{T}(R\hat{\theta}_T - r)' \left[R\left(\frac{A_T}{T}\right)^{-1} R'\right]^{-1} \sqrt{T}(R\hat{\theta}_T - r) \xrightarrow{d} \chi^2_q
\end{aligned}$$

che coincide con il test di Wald.

Infine, consideriamo il test del rapporto di verosimiglianza: $LR = 2[\hat{l} - \tilde{l}]$ dove $\hat{l}$ e $\tilde{l}$ rappresentano il massimo raggiunto dalla funzione di verosimiglianza non vincolata e vincolata, rispettivamente. Attraverso semplici sostituzioni otteniamo che

$$LR = 2[\hat{l} - \tilde{l}] = (\tilde{\theta}_T - \hat{\theta}_T)' A_T (\tilde{\theta}_T - \hat{\theta}_T).$$

Sostituendo per $\tilde{\theta}_T$ la sua espressione in termini di $\hat{\theta}_T$, e tenendo conto dell'espressione per $\tilde{\lambda}_T$, si ricava che

$$\begin{aligned}
LR &= (R\hat{\theta}_T - r)' \left[RA_T^{-1}R'\right]^{-1} RA_T^{-1}A_T A_T^{-1}R' \left[RA_T^{-1}R'\right]^{-1} (R\hat{\theta}_T - r) \\
&= (R\hat{\theta}_T - r)' \left[RA_T^{-1}R'\right]^{-1} (R\hat{\theta}_T - r) \xrightarrow{d} \chi^2_q
\end{aligned}$$

come per il test di Wald e per il test LM.

In conclusione, le tre procedure di test sono asintoticamente equivalenti quando si considerano vincoli lineari del tipo $R\theta = r$.
Osservazioni:

1. In conclusione, le tre procedure di test sono asintoticamente equivalenti quando si considerano vincoli lineari del tipo $R\theta = r$.

2. Bisogna notare come un risultato analogo, ma in campioni finiti, si ottiene anche nel modello lineare con errori normali e varianza nota. Questo si verifica proprio perché in entrambe le situazioni la funzione di log-verosimiglianza è quadratica.

3. Se l'ipotesi nulla è formulata in termini di vincoli non lineari del tipo $G(\theta) = r$, abbiamo che, per $T \to \infty$:

   - $LR = 2[\hat{l} - \tilde{l}] \sim \chi^2_q$;
   - $W = TR(\hat{\theta})'[R(\hat{\theta})\left(\frac{A_T}{T}\right)^{-1} R(\hat{\theta})']^{-1} R(\hat{\theta}) \sim \chi^2_q$;
   - $LM = \sqrt{T}s(\tilde{\theta})' \left(\frac{A_T}{T}\right)^{-1} \sqrt{T}s(\tilde{\theta}) \sim \chi^2_q$

   dove $R(\theta) = (\partial G/\partial\theta)$.