Intervalli di Wald e metodo delta

Un esempio notevole

Luca La Rocca[1]

Dipartimento di Scienze Fisiche, Informatiche e Matematiche
Università degli Studi di Modena e Reggio Emilia

Insegnamento di Analisi Statistica dei Dati
Corsi di Laurea Magistrale in Informatica e Matematica
Anno Accademico 2018/2019

[1] http://personale.unimore.it/rubrica/dettaglio/llarocca

Teorema di Slutsky

Se T_n, $n \geq 1$, è una successione di stimatori asintoticamente normale (di ordine $1/\sqrt{n}$) per $\xi = h(\theta)$ e V_n, $n \geq 1$, una successione di stimatori consistenti (in probabilità) per la sua varianza limite $v_\infty(\theta)$, $\theta \in H$, la successione di combinanti $(T_n - \xi)/\sqrt{V_n/n}$, $n \geq 1$, converge in distribuzione alla distribuzione normale standard.

In pratica possiamo scrivere

$$\frac{T_n - \xi}{\sqrt{V_n/n}} \approx \mathrm{Norm}(0, 1),$$

quando n è "grande", di modo che il combinante $(T_n - \xi)/\sqrt{V_n/n}$ risulta essere un pivot asintotico.

Errore standard

Per uno stimatore T_n con distribuzione asintotica $\mathrm{Norm}(\xi, v_\infty(\theta)/n)$, se V_n stima in modo consistente la sua varianza limite $v_\infty(\theta)$, si dice errore standard asintotico la statistica

$$\mathrm{ASE}_{T_n} = \sqrt{\frac{V_n}{n}}$$

che stima asintoticamente $\mathrm{RMSE}_{T_n,\xi}(\theta) \simeq \sqrt{v_\infty(\theta)/n}$.

Per inciso, quando si voglia una stima non asintotica di $\mathrm{RMSE}_{T_n,\xi}(\theta)$, occorrerà non solo un errore standard $\mathrm{SE}_{T_n} = \sqrt{W_n}$ che stimi $\mathrm{Sd}_\theta(T_n)$, ma anche (in virtù della decomposizione dell'errore quadratico medio in distorsione quadratica e varianza) uno stimatore B_n per $B_\theta(T_n; \xi)$:

$$\widehat{\mathrm{RMSE}}_{T_n;\xi} = \sqrt{B_n^2 + W_n}.$$

Intervallo di Wald

Scelto $\gamma \in]1/2, 1[$, dalla valutazione asintotica (valida per ogni $\theta \in H$)

$$\mathbb{P}_\theta \left\{ -z_{\frac{1+\gamma}{2}} \leq \frac{T_n - \xi}{\mathrm{ASE}_{T_n}} \leq z_{\frac{1+\gamma}{2}} \right\} \simeq \gamma,$$

dove z_δ è il quantile di ordine δ della distribuzione normale standard, per ogni $\delta \in]0, 1[$, e abbiamo tenuto conto che $z_{(1-\gamma)/2} = -z_{(1+\gamma)/2}$, otteniamo il seguente intervallo di confidenza asintotico, al livello γ, per ξ:

$$\mathrm{WCI}_{\xi;\gamma} = \left[T_n - z_{\frac{1+\gamma}{2}} \mathrm{ASE}_{T_n}, \, T_n + z_{\frac{1+\gamma}{2}} \mathrm{ASE}_{T_n} \right];$$

il valore $z_{(1+\gamma)/2}$ prende il nome di coefficiente di confidenza.

Campione casuale da una popolazione Bernoulli

Se $X_1, X_2, \ldots, X_n$ sono i.i.d. come X tale che $\mathbb{P}\{X = 1\} = 1 - \mathbb{P}\{X = 0\} = \pi$, per un qualche $\pi \in]0, 1[$ ignoto, di modo che $S_n = X_1 + X_2 + \ldots X_n \sim \mathrm{Binom}(n, \pi)$ è sufficiente per π, lo stimatore di massima verosimiglianza $\hat{\Pi}_n = S_n/n$ per π è asintoticamente normale (di ordine $1/\sqrt{n}$) con varianza limite $\mathrm{Var}_\pi(X) = \pi(1 - \pi)$ per il Teorema Limite Centrale.

Possiamo allora prendere $V_n = \hat{\Pi}_n(1 - \hat{\Pi}_n)$, consistente (in probabilità) per il teorema della mappa continua, e ottenere così

$$\mathrm{WCI}_{\pi;\gamma} = \left[\hat{\Pi}_n - z_{\frac{1+\gamma}{2}}\sqrt{\hat{\Pi}_n(1 - \hat{\Pi}_n)/n}, \hat{\Pi}_n + z_{\frac{1+\gamma}{2}}\sqrt{\hat{\Pi}_n(1 - \hat{\Pi}_n)/n}\right],$$

una volta esplicitato che $\mathrm{ASE}_{T_n} = \sqrt{V_n/n} = \sqrt{\hat{\Pi}_n(1 - \hat{\Pi}_n)/n}$.

Intervallo di confidenza osservato

Se $n = 100$ e $s_n^\bullet = 2$, di modo che $\hat{\pi}_n^\bullet = 0.02$
e $\mathrm{ase}^\bullet_{\hat{\Pi}_n} = \sqrt{\hat{\pi}_n^\bullet(1 - \hat{\pi}_n^\bullet)/n} = 0.014$, preso per esempio $\gamma = 0.95$,
troviamo $z_{(1+\gamma)/2} = z_{0.975} = 1.96$ e quindi

$$\begin{aligned} \mathrm{wci}^\bullet_{\pi;0.95} &= [0.02 - 1.96 \times 0.014, 0.02 + 1.96 \times 0.014] \\ &= [-0.007, 0.047] \end{aligned}$$

che diventerà senz'altro l'intervallo aperto a sinistra
$\mathrm{wci}^\bullet_{\pi;0.95} =]0, 0.047]$, perché sappiamo a priori che $\pi > 0$.

<u>Attenzione</u>: non possiamo affermare che $0 < \pi \leq 0.047$ con probabilità 95%, perché π non ha una distribuzione, ma confidiamo al 95% che i due confini valgano, perché li abbiamo calcolati con un metodo che copre il parametro nel 95% dei casi.

Trasformazioni monotone

Se $\left[\underline{T}, \overline{T}\right]$ è un intervallo di confidenza al livello γ per ξ
e $\varphi = k(\xi)$ con $\xi \mapsto k(\xi)$ strettamente crescente (per fissare le idee)
possiamo ricavare un intervallo di confidenza al livello γ per φ
come $\left[k\left(\underline{T}\right), k\left(\overline{T}\right)\right]$: $\mathbb{P}_\theta\left\{k\left(\underline{T}\right) \leq \varphi \leq k\left(\overline{T}\right)\right\} = \mathbb{P}_\theta\left\{\underline{T} \leq \xi \leq \overline{T}\right\}$.

Per esempio, nel caso del campione casuale da una popolazione
Bernoulli, con $n = 100$ e $s_n^\bullet = 2$, se $\varphi = \mathrm{logit}(\pi) = \log\{\pi/(1-\pi)\}$,
$\pi \in]0, 1[$, prendendo $\gamma = 0.95$, troveremo

$$\mathrm{wci}^\bullet_{\mathrm{logit}(\pi);0.95} =]-\infty, -3.01]$$

che però non appare di grande soddisfazione; in effetti, se definiamo
il margine di errore come la semiampiezza dell'intervallo di confidenza,
abbiamo un intervallo osservato con margine di errore infinito.

Metodo delta

Se T_n è uno stimatore con distribuzione asintotica $\mathrm{Norm}(\xi, v_\infty(\theta)/n)$ e $\xi \mapsto k(\xi)$ è un diffeomorfismo (trasformazione liscia con inversa liscia) allora abbiamo

$$\frac{k(T_n) - k(\xi)}{\sqrt{\dot{k}^2(\xi) v_\infty(\theta)/n}} \approx \mathrm{N}(0, 1),$$

dove $\dot{k}(\xi) = \frac{dk}{d\xi}(\xi) = 1/\frac{dk^{-1}}{d\xi}(\xi) \neq 0$.

Possiamo quindi prendere come errore standard asintotico[2]

$$\mathrm{ASE}_{k(T_n)} = \sqrt{\dot{k}^2(T_n) V_n/n} = |\dot{k}(T_n)| \mathrm{ASE}_{T_n},$$

visto che T_n con distribuzione asintotica $\mathrm{Norm}(\xi, v_\infty(\theta)/n)$ è necessariamente consistente (in probabilità).

[2] in pratica adottando la formula per la propagazione dell'errore $d\varphi = |\dot{k}(\xi)| d\xi$

Altro intervallo di Wald

Nell'esempio del campione casuale da una popolazione Bernoulli, se $\varphi = \operatorname{logit}(\pi) = \log\{\pi/(1-\pi)\}$, $\pi \in \,]0,1[$, troviamo

$$\frac{d \operatorname{logit}}{d\pi}(\pi) = \frac{1}{\pi(1-\pi)}, \qquad \pi \in \,]0,1[,$$

quindi

$$\mathrm{ASE}_{\operatorname{logit}(\hat{\Pi}_n)} = \sqrt{\frac{1}{n\hat{\Pi}_n(1-\hat{\Pi}_n)}} = \sqrt{\frac{1}{S_n} + \frac{1}{n-S_n}}$$

e in definitiva

$$\mathrm{WCI}_{\varphi;\gamma} = \left[\operatorname{logit}(\hat{\Pi}_n) - z_{\frac{1+\gamma}{2}}\mathrm{ASE}_{\operatorname{logit}(\hat{\Pi}_n)}, \operatorname{logit}(\hat{\Pi}_n) + z_{\frac{1+\gamma}{2}}\mathrm{ASE}_{\operatorname{logit}(\hat{\Pi}_n)}\right].$$

Altro intervallo osservato

Se $n = 100$ e $s_n^\bullet = 2$, di modo che $\hat\pi_n^\bullet = 0.02$ e $\frac{d\,\mathrm{logit}}{d\pi}(\hat\pi_n^\bullet) = 51.02$, otterremo

$$\mathrm{ase}^\bullet_{\hat\Pi_n} = \sqrt{\frac{1}{2} + \frac{1}{98}} = 0.714 = 51.02 \times 0.014 = \left| \frac{d\,\mathrm{logit}}{d\pi}(\hat\pi_n^\bullet) \right| \mathrm{ase}^\bullet_{\hat\Pi_n}$$

e, prendendo $\gamma = 0.95$, ricaveremo

$$\mathrm{wci}^\bullet_{\varphi;0.95} =]\mathrm{logit}(0.02) - 1.96 \times 0.714, \mathrm{logit}(0.02) + 1.96 \times 0.714],$$

trovandoci così nella posizione di avere…

Mancata equivarianza

…due intervalli di confidenza (stimati) per $\varphi = \mathrm{logit}(\pi)$ allo stesso livello $\gamma = 0.95$:

$$
\begin{aligned}
\mathrm{wci}^{\bullet}_{\mathrm{logit}(\pi);0.95} &= \,]-\infty, -3.01];\\
\mathrm{wci}^{\bullet}_{\varphi;0.95} &= [-5.29, -2.49].
\end{aligned}
$$

Il secondo intervallo appare di maggiore soddisfazione, avendo margine di errore finito (pari a 1.4); trasformandolo tramite

$$
\mathrm{logit}^{-1}(\varphi) = \frac{e^{\varphi}}{1 + e^{\varphi}}, \qquad \varphi \in \mathbb{R},
$$

troviamo $\mathrm{wci}^{\bullet}_{\mathrm{logit}^{-1}(\varphi);0.95} = [0.01, 0.076]$ che, se non altro, ha il pregio di avere il confine inferiore strettamente maggiore di zero.

Morale

L'intervallo di Wald ha il punto debole di non essere equivariante per trasformazioni monotone del parametro di interesse, ma possiamo fare di questo punto debole un punto di forza scegliendo una parametrizzazione che ci torni "comoda".

Nell'esempio del campione casuale da una popolazione Bernoulli, la scelta $\varphi = \operatorname{logit}(\pi)$ ci fornisce un intervallo di confidenza staccato da zero (usando l'approssimazione normale su un parametro libero di variare in $\mathbb{R}$ invece che su π vincolato tra 0 e 1).

Un'altra possibilità è scegliere $\varphi = \operatorname{arsin}(\sqrt{\pi})$ per ottenere una varianza asintotica che non dipenda da π e quindi non debba essere stimata; in questo caso si parla di trasformazione che stabilizza la varianza.