

Inferenza statistica

Un'introduzione minimale

Luca La Rocca¹

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>

Idea di base

Supponiamo che il dato osservato z^\bullet sia la realizzazione di una variabile aleatoria Z la cui **distribuzione incognita** sia l'oggetto di interesse; per esempio

$$Z \sim \text{Binom}(n, \pi)$$

con π incognita nella **stima di una proporzione** e

$$Z \sim \text{Hyper}(n, N, M)$$

con N incognita nel **metodo di cattura-ricattura**.

In concreto, potremmo avere $z^\bullet = 2$ su $n = 10$ neonati con la Sindrome di Klinefelter² o $z^\bullet = 5$ pesci marcati su $n = 63$ pesci estratti da un lago con $M = 26$ pesci marcati in tutto.

²https://en.wikipedia.org/wiki/Klinefelter_syndrome

Modelli statistici

Si dice **modello statistico** l'insieme delle distribuzioni ipotizzate per Z ; per esempio

$$\{\text{Binom}(n, \pi) \mid \pi \in]0, 1[\}$$

nel caso della **stima di una proporzione** e

$$\{\text{Hyper}(n, N, M) \mid N \in \mathbb{N}, N \geq \max\{n, M\}\}$$

nel caso del **metodo di cattura-ricattura**.

Supporremo senz'altro di avere **ben specificato** il modello statistico, cioè che il dato osservato sia stato effettivamente generato da una distribuzione appartenente al modello; questo non vuol dire che il caso contrario non ci interessi (essendo ogni modello una nostra ipotesi).

Parametri statistici

Si dice **parametro statistico** il valore assunto da un certo funzionale, definito sul modello statistico, in corrispondenza della distribuzione incognita che ha generato il dato; per esempio

$$\pi \quad \& \quad \mu = \mathbb{E}_{\pi}[Z] = n\pi$$

nel caso della **stima di una proporzione** e

$$N \quad \& \quad \mu = \mathbb{E}_N[Z] = n \frac{M}{N}$$

nel caso del **metodo di cattura-ricattura**.

Si noti il ruolo speciale del **parametro indice** che definisce il modello (il primo nei due esempi) e di cui ogni altro parametro è funzione.

Modelli dominati

Supporremo in genere che le distribuzioni nel modello statistico scelto ammettano tutte una **funzione di densità** (per il momento discreta):

$$\mathbb{P}_\theta\{Z \in B\} = \sum_{z \in B} f(z|\theta), \quad B \subseteq \mathcal{Z}, \quad \theta \in H,$$

se Z prende valori in \mathcal{Z} e il parametro indice θ varia in H ; per esempio

$$f(z|\pi) = \binom{n}{z} \pi^z (1 - \pi)^{n-z}, \quad z = 0, 1, \dots, n,$$

nel caso della **stima di una proporzione** e

$$f(z|N) = \frac{\binom{M}{z} \binom{N-M}{n-z}}{\binom{N}{n}}, \quad z = 0, 1, \dots, n,$$

nel caso del **metodo di cattura-ricattura** (con $\binom{m}{k} = 0$ se $k > m$).

Funzione di verosimiglianza

La **funzione di verosimiglianza**, alla luce del dato z^\bullet , è definita come la probabilità di osservare z^\bullet in funzione del parametro indice:

$$L^\bullet(\theta) = \mathbb{P}_\theta\{Z = z^\bullet\} = f(z^\bullet|\theta), \quad \theta \in H;$$

per esempio

$$L^\bullet(\pi) = f(z^\bullet|\pi) = \binom{n}{z^\bullet} \pi^{z^\bullet} (1 - \pi)^{n - z^\bullet}, \quad 0 < \pi < 1,$$

nel caso della **stima di una proporzione** e

$$L^\bullet(N) = f(z^\bullet|N) = \frac{\binom{M}{z^\bullet} \binom{N-M}{n-z^\bullet}}{\binom{N}{n}}, \quad N \geq \max\{n, M + (n - z^\bullet)\},$$

nel caso del **metodo di cattura-ricattura**.

Stima di massima verosimiglianza

Il valore che rende massimamente probabile il dato osservato,

$$\hat{\theta}^\bullet = \operatorname{argmax}_{\theta \in H} L^\bullet(\theta),$$

se esiste (unico) è detto **stima di massima verosimiglianza** del parametro indice alla luce del dato osservato; per esempio

$$\hat{\pi}^\bullet = \frac{z^\bullet}{n}$$

nel caso della **stima di una proporzione** (sul bordo per $z^\bullet \in \{0, 1\}$) e

$$\hat{N}^\bullet = \left\lfloor \frac{Mn}{z^\bullet} \right\rfloor$$

nel caso del **metodo di cattura-ricattura** (infinita per $z^\bullet = 0$).