

8 Stima puntuale di parametri

8.1 Stima puntuale

Abbiamo già osservato che quando si fa della probabilità si suppone che le distribuzioni siano completamente note, mentre in statistica si fa dell'inferenza su parametri sconosciuti utilizzando i dati osservati. L'inferenza statistica può essere divisa in due aree principali: la *stima* e la *verifica di ipotesi*.

Un tipo di stima è la *stima puntuale*, che consiste nel trovare una statistica $t(X_1, \dots, X_n)$ detta *stimatore puntuale*, che permette di stimare il parametro incognito " θ " della popolazione. Un secondo tipo di stima è la *stima intervallare*, che consiste nel definire due statistiche $t_1(X_1, \dots, X_n)$ e $t_2(X_1, \dots, X_n)$ con $t_1 < t_2$ in modo che (t_1, t_2) costituisca un intervallo di valori plausibili per θ per il quale si può calcolare la probabilità che θ vi appartenga.

Gli stimatori sono delle variabili casuali.

Il valore deterministico assunto da uno stimatore si chiama stima.

Problema: individuare la forma opportuna dello stimatore e calcolare la sua distribuzione;

- trovare una statistica da usare come stimatore puntuale;
- scegliere criteri per definire e ottenere uno stimatore “*ottimale*”, fra i molti possibili.

Le proprietà che uno stimatore può possedere sono svariate. Noi discuteremo:

- la **correttezza** o **non distorsione**
- la **consistenza**
- l'**efficienza**

8.2 Metodi di stima

Le proprietà di uno stimatore sono importanti perchè specificano quali sono le caratteristiche desiderabili: uno stimatore con proprietà migliori di un altro è chiaramente preferibile. D'altra parte le proprietà indicate non dicono molto circa il problema come fare a trovare uno stimatore (magari anche buono). Occorre allora trattare i metodi di stima puntuale proposti in letteratura verificando se, e sotto quali

condizioni operative, producono stimatori che soddisfano le proprietà indicate. Sono stati proposti molti metodi di stima, ciascuno adatto in specifici contesti. Ne presentiamo due:

- 1) *il metodo dei momenti*
- 2) *il metodo della massima verosimiglianza.*

8.2.1 Metodo dei momenti

Supponiamo che una popolazione sia caratterizzata da una funzione di densità $f(\cdot; \theta_1, \dots, \theta_k)$ con k parametri incogniti.

I momenti di ordine r della popolazione sono:

$$\mu'_r = E[X^r] = \mu'_r(\theta_1, \dots, \theta_k)$$

Dato un campione casuale (X_1, \dots, X_n) di dimensione n , i momenti campionari sono:

$$M'_j = \frac{1}{n} \sum_{i=1}^n X_i^j, \quad j = 1, \dots, k$$

Il metodo consiste nell'uguagliare i momenti della popolazione con i momenti campionari corrispondenti, cioè nel costruire il sistema di k equazioni

$$M'_j = \mu'_j(\theta_1, \dots, \theta_k), \quad j = 1, \dots, k$$

nelle k incognite $\theta_1, \dots, \theta_k$.

La soluzione unica di tale sistema $(\bar{\theta}_1, \dots, \bar{\theta}_k)$ sarà lo stimatore puntuale cercato.

Esempi

1) Dato un campione casuale (X_1, \dots, X_n) di dimensione n estratto da una popolazione con densità esponenziale

$$f(x, \theta) = \theta e^{-\theta x}, \quad x \geq 0,$$

determinare uno stimatore per θ con il metodo dei momenti.

Per una variabile casuale esponenziale X si sa che $E[X] = \frac{1}{\theta}$.

$$\mu'_1 = \mu = E[X] \Rightarrow \mu'_1(\theta) = \frac{1}{\theta}$$

mentre $M'_1 = \bar{X}_n$.

$$\text{Allora } M'_1 = \mu'_1 \Rightarrow \bar{X}_n = \frac{1}{\theta} \Rightarrow \bar{\theta} = \frac{1}{\bar{X}_n}$$

2) Dato un campione casuale (X_1, \dots, X_n) di dimensione n estratto da una popolazione con densità normale $N(\mu, \sigma^2)$, determinare gli stimatori puntuali $\bar{\theta}_1$,

$\bar{\theta}_2$ per i parametri $(\theta_1, \theta_2) = (\mu, \sigma)$ con il metodo dei momenti.

Ricordiamo che:

$$\begin{cases} \mu = \mu'_1 \\ \sigma^2 = \mu'_2 - (\mu'_1)^2 \end{cases}$$

quindi:

$$\begin{cases} \mu'_1 = \mu \\ \mu'_2 = \sigma^2 + (\mu'_1)^2 = \sigma^2 + \mu^2 \end{cases}$$

Le equazioni dei momenti sono

$$\begin{cases} M'_1 = \mu'_1(\mu, \sigma) = \mu \\ M'_2 = \mu'_2(\mu, \sigma) = \mu^2 + \sigma^2 \end{cases}$$

ma

$$M'_1 = \bar{X}_n; \quad M'_2 = \frac{1}{n} \sum_{i=1}^n X_i^2$$

perciò

$\bar{\theta}_1 = \bar{X}_n$ è lo stimatore di μ

$\bar{\theta}_2 = \sqrt{M'_2 - \bar{X}_n^2} = \sqrt{M_2}$ è lo stimatore di σ

8.2.2 Metodo della massima verosimiglianza

È uno dei metodi più rilevanti, la cui importanza è dovuta:

1. alle particolari proprietà di cui godono gli stimatori ottenuti con tale metodo;
2. alla generalità e alla relativa semplicità del metodo dal punto di vista del calcolo (per poterlo applicare basta riuscire a calcolare la verosimiglianza);
3. alla sua notevole diffusione, legata ai due punti precedenti: circa l'85 – 90% dei modelli sono stimati con tale metodo.

Data una popolazione, con funzione di densità, $f(x; \theta)$, e un c.c. (X_1, \dots, X_n) estratto da essa, sappiamo che la distribuzione del campione è data dalla funzione di densità

$$f(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta).$$

La funzione di densità $f(x_1, \dots, x_n; \theta)$ è funzione delle osservazioni $x = (x_1, \dots, x_n)$ del campione per un certo θ fissato.

La *verosimiglianza* $\mathcal{L}(\theta)$ è la funzione di densità del campione (X_1, \dots, X_n) vista però come funzione del

parametro per un fissato campione, cioè

$$\mathcal{L}(\theta) = f(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta).$$

Pertanto la verosimiglianza coincide in termini formali con la funzione di densità del campione; si tratta infatti della stessa espressione interpretata:

1. come funzione delle osservazioni $x = (x_1, \dots, x_n)$ che variano nell'universo dei campioni per θ fissato;
2. come funzione del parametro θ per un campione $x = (x_1, \dots, x_n)$ prefissato.

Nella prima interpretazione (a priori), si fa riferimento all'universo dei campioni e le variabili che interessano sono, appunto, le variabili casuali campionarie (X_1, \dots, X_n) .

Nella seconda interpretazione (a posteriori), le variabili campionarie hanno assunto particolari determinazioni (x_1, \dots, x_n) e sono, pertanto, costanti note.

Risulta, allora, ragionevole interpretare l'espressione come funzione del parametro θ (o più in generale dei parametri) che, pur essendo una costante, assume la veste di variabile essendo incognito il suo valore.

Il metodo di stima della massima verosimiglianza consiste nello scegliere il valore $\tilde{\theta}$ che massimizza

la funzione $\mathcal{L}(\theta)$. Se $\mathcal{L}(\theta)$ è una funzione differenziabile, condizione necessaria affinché essa abbia un massimo rispetto a θ è che

$$\frac{d}{d\theta} \mathcal{L}(\theta) = 0.$$

Poichè derivare $\mathcal{L}(\theta)$ significa fare la derivata di un prodotto, procedimento in generale abbastanza laborioso, in genere si preferisce massimizzare non la verosimiglianza $\mathcal{L}(\theta)$, ma il logaritmo naturale della verosimiglianza

$$\ln \mathcal{L}(\theta) = \sum_{i=1}^n \ln f(x_i; \theta),$$

detto *log-verosimiglianza*. Essendo il logaritmo naturale una funzione monotona, è garantito che il valore $\tilde{\theta}$ che massimizza $\ln \mathcal{L}(\theta)$ sia identico a quello che massimizza $\mathcal{L}(\theta)$.

Il valore $\tilde{\theta}$ che massimizza la verosimiglianza o la log-verosimiglianza è detto *stima di massima verosimiglianza* del parametro incognito θ . Esso deve verificare la condizione:

$$\frac{d^2}{d\theta^2} \ln \mathcal{L}(\tilde{\theta}) = \frac{1}{\mathcal{L}(\tilde{\theta})} \frac{d^2}{d\theta^2} \mathcal{L}(\tilde{\theta}) < 0.$$

Se nella soluzione si sostituiscono alle determinazioni (x_1, \dots, x_n) le corrispondenti variabili casuali (X_1, \dots, X_n) si ottengono gli stimatori di massima verosimiglianza. Ovviamente se la distribuzione della popolazione, è caratterizzata da più parametri $\theta_1, \dots, \theta_k$, per trovare il massimo occorrerà uguagliare a 0 ciascuna delle derivate parziali rispetto ad ogni singolo parametro e poi ricavare la stima dei parametri facendo il sistema fra le equazioni ottenute.

Esempio

Dato un c.c. di ampiezza n estratto da una popolazione con funzione di densità:

$$f(x; p) = p^x (1 - p)^{1-x}$$

con $0 \leq p \leq 1$, la funzione di verosimiglianza è data da:

$$\mathcal{L}(p) = \prod_{i=1}^n p^{x_i} (1 - p)^{1-x_i} = p^{\sum_{i=1}^n x_i} (1 - p)^{(n - \sum_{i=1}^n x_i)}$$

Posto $y = \sum_{i=1}^n x_i$ allora

$$\mathcal{L}(y; p) = p^y (1 - p)^{n-y}$$

e

$$\ln \mathcal{L}(y; p) = y \ln p + (n - y) \ln (1 - p)$$

Pertanto

$$\frac{d}{dp} \ln \mathcal{L}(y; p) = \frac{y - np}{p(1 - p)}$$

e tale derivata si annulla per

$$\tilde{p} = \frac{y}{n} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}_n.$$

Tale soluzione è un massimo poichè la condizione

$$\frac{d^2}{dp^2} \ln \mathcal{L}(p)|_{p=\tilde{p}} = -\frac{n^3}{y(n - y)} < 0$$

è sempre vera. Quindi $\tilde{p} = \bar{x}_n$ è la stima di massima verosimiglianza per p .

8.3 Proprietà degli stimatori puntuali

Esistono stimatori che siano in qualche modo migliori di altri? Definiamo adesso alcune proprietà che uno stimatore può possedere o meno, utili per decidere se uno stimatore è da preferirsi ad un altro.

Definizione: Si definisce **errore quadratico medio** (= **MSE**) di uno stimatore T del parametro θ la quantità

$$\text{MSE}[T](\theta) = E[(T - \theta)^2]$$

dove $T = t(X_1, \dots, X_n)$.

Esso misura la dispersione dei valori di T rispetto a θ (come la varianza di una variabile casuale X misura la sua dispersione attorno alla media)

Definizione: Uno stimatore T del parametro θ si dice **corretto** o **non distorto** se e solo se

$$E[T] = \theta$$

Poichè trovare uno stimatore con MSE minimo è difficile, restringendoci alla classe degli stimatori non distorti c'è la speranza di trovare quello con MSE minimo.

Definizione: Si definisce **distorsione** di uno stimatore T la quantità

$$D[T](\theta) = \theta - E[T] \quad (\geq 0)$$

Se T è corretto $\Rightarrow D[T] = 0$.

Proprietà

Per ogni stimatore T del parametro θ vale la seguente relazione:

$$\text{MSE}[T](\theta) = \text{var}[T] + D[T]^2$$

Infatti:

$$\begin{aligned}
 \text{MSE}[T] &= E[(T - \theta)^2] = \\
 &= E[(T - E[T] + E[T] - \theta)^2] = \\
 &= E\left[\left[(T - E[T]) + (E[T] - \theta)\right]^2\right] = \\
 &= E\left[(T - E[T])^2\right] + E\left[2(T - E[T])(E[T] - \theta)\right] + \\
 &+ E\left[(E[T] - \theta)^2\right]
 \end{aligned}$$

Si noti che $E[T] - \theta$ non dipende dalle variabili X_i del campione e quindi va considerato una costante, quindi

$$\begin{aligned}
 E\left[(E[T] - \theta)^2\right] &= (E[T] - \theta)^2 = D[T]^2 \\
 E\left[(T - E[T])(E[T] - \theta)\right] &= (E[T] - \theta)E[T - E[T]]
 \end{aligned}$$

ma

$$E[T - E[T]] = E[T] - E[T] = 0$$

quindi

$$\begin{aligned}
 \text{MSE}[T] &= E\left[(T - E[T])^2\right] + D[T]^2 = \\
 &= \text{var}[T] + D[T]^2
 \end{aligned}$$

Se T è **corretto** allora

$$\text{MSE}[T] = \text{var}[T]$$

Esempio

Dato un campione casuale di dimensione n estratto da una popolazione con funzione di densità normale $N(\mu, \sigma^2)$, abbiamo ricavato col metodo dei momenti

$$\begin{cases} \bar{\theta}_1 = \mu = \bar{X}_n \\ \bar{\theta}_2 = \sigma = \sqrt{M_2} \sim \sigma^2 = M_2 \end{cases}$$

Poichè $E[\bar{X}_n] = \mu \Rightarrow \bar{X}_n$ è uno stimatore **corretto**.

Inoltre

$$\text{MSE}[\bar{X}_n] = E[(\bar{X}_n - \mu)^2] = \text{var}[\bar{X}_n] = \frac{\sigma^2}{n}$$

Invece

$$E[M_2] = \frac{n-1}{n} E[S^2],$$

ma $E[S^2] = \sigma^2$ (dal teorema 2 del campionamento), quindi

$$E[M_2] = \frac{n-1}{n} \sigma^2 \neq \sigma^2 \Rightarrow M_2 \text{ è uno stimatore } \textbf{distorto}.$$

Cercare uno stimatore con MSE minimo tra quelli non distorti equivale a cercare uno stimatore a varianza minima nella stessa classe (**più efficiente**). Un limite inferiore della varianza di stimatori non distorti è dato dalla seguente disuguaglianza.

Disuguaglianza di Rao-Cramér

Dato un campione casuale (X_1, \dots, X_n) estratto da una popolazione con funzione di densità $f(\cdot, \theta)$ e T uno stimatore non distorto di θ , si ha:

$$\text{var}[T] \geq \frac{1}{nE \left[\left(\frac{1}{f} \frac{\partial f}{\partial \theta} \right)^2 \right]}$$

$$\frac{1}{f} \frac{\partial f}{\partial \theta} = \frac{\partial}{\partial \theta} [\ln(f)]$$

Poichè uno stimatore di θ dipende dal numero di campionamenti, enunciamo ora una proprietà definita in termini di ampiezza crescente del campione.

Definizione: Uno stimatore T_n del parametro θ è detto **consistente** in media quadratica se e solo se:

$$\lim_{n \rightarrow \infty} E [(T_n - \theta)^2] = 0$$

Poichè $E[(T_n - \theta)^2] = \text{MSE}[T_n] = \text{var}[T_n] + D[T_n]^2$ si ha

$$\text{var}[T_n] \xrightarrow[n \rightarrow \infty]{} 0; \quad D[T_n] \xrightarrow[n \rightarrow \infty]{} 0.$$

Esempio

Abbiamo visto che \bar{X}_n è uno stimatore non distorto per μ per un campione casuale di dimensione n estratto da una popolazione con densità normale $N(\mu, \sigma^2)$.

$$\lim_{n \rightarrow \infty} E [(\bar{X}_n - \mu)^2] = \lim_{n \rightarrow \infty} \text{var}[\bar{X}_n] = \lim_{n \rightarrow \infty} \frac{\sigma^2}{n} = 0$$

$\Rightarrow \bar{X}_n$ è uno stimatore **consistente**.

