

11 Stima di parametri

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.

STIMA PUNTUALE

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 svariare. Noi discuteremo:

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

Per determinare uno stimatore puntuale ci sono vari metodi.

Noi discuteremo il **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

$$\begin{aligned} \bar{\theta}_2 &= \sqrt{M'_2 - \bar{X}_n^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2} = \\ &= \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2} = \sqrt{M_2} \text{ è lo stimatore di } \sigma \end{aligned}$$

La penultima uguaglianza deriva dalle seguenti espressioni:

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 &= \frac{1}{n} \sum_{i=1}^n (X_i^2 - 2X_i\bar{X}_n + \bar{X}_n^2) = \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 - \frac{2}{n} \bar{X}_n (n \cdot \bar{X}_n) + \frac{1}{n} \bar{X}_n^2 \cdot n = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2 \end{aligned}$$

PROPRIETÀ DEGLI STIMATORI PUNTUALI

Esistono stimatori che siano in qualche modo migliori di altri? Definiremo 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 mi-

nimo.

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

$$E\left[(E[T] - \theta)^2\right] = (E[T] - \theta)^2 = D[T]^2$$

$$E[(T - E[T])(E[T] - \theta)] = (E[T] - \theta)E[T - E[T]]$$

ma

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

quindi

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

- 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

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

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

DISUGUAGLIANZA DI RAO-CRAMER 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:

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

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

Esempio Dato un campione casuale (X_1, \dots, X_n) di dimensione n estratto da una popolazione con fun-

zione di densità esponenziale

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

si ha

$$\frac{\partial f}{\partial \theta} = e^{-\theta x} - \theta x e^{-\theta x} = (1 - \theta x) e^{-\theta x},$$

$$\left(\frac{1}{f} \frac{\partial f}{\partial \theta} \right) = \frac{1}{\theta} - x,$$

quindi

$$\begin{aligned} E \left[\left(\frac{1}{\theta} - X \right)^2 \right] &= E \left[\left(X - \frac{1}{\theta} \right)^2 \right] = E \left[\left(X - E[X] \right)^2 \right] = \\ &= \text{var}[X] = \frac{1}{\theta^2} \end{aligned}$$

dove abbiamo usato $E[X] = \mu = \frac{1}{\theta}$ e $\text{var}[X] = \frac{1}{\theta^2}$ per una variabile casuale esponenziale. Pertanto

$$\text{var}[T] \geq \frac{1}{n \frac{1}{\theta^2}} = \frac{\theta^2}{n} \quad \text{limite inferiore.}$$

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**.

ESEMPI

① Dato un campione casuale di dimensione n estratto da una popolazione con funzione di densità uniforme sull'intervallo $[0, \theta]$, trovare uno stimatore di θ col metodo dei momenti e stabilire se è corretto e consistente.

$$f(x, \theta) = \begin{cases} \frac{1}{\theta} & 0 \leq x \leq \theta \\ 0 & \text{altrove} \end{cases}$$

$$\mu = E[X] = \frac{a+b}{2} = \frac{\theta}{2}$$

$$\sigma^2 = \text{var}[X] = \frac{(b-a)^2}{12} = \frac{\theta^2}{12}$$

- Dall'equazione dei momenti $M'_1 = \mu'_1$ si ha:

$$\bar{X}_n = \frac{\theta}{2} \Rightarrow \bar{\theta} = 2\bar{X}_n$$

- $E[\bar{\theta}] \stackrel{???}{=} \theta$

$$E[2\bar{X}_n] = 2E[\bar{X}_n] = 2 \frac{\theta}{2} = \theta \Rightarrow \bar{\theta} \text{ è corretto}$$

$\bar{\theta}$ è consistente se $\text{MSE}[\bar{\theta}] \rightarrow 0$

$$\begin{aligned} \text{MSE}[\bar{\theta}] &= \text{var}[\bar{\theta}] + \underset{=0}{D[\bar{\theta}]^2} = \text{var}[2\bar{X}_n] = 4 \text{var}[\bar{X}_n] = \\ &= 4 \frac{\sigma^2}{n} = \frac{4}{n} \frac{\theta^2}{12} = \frac{\theta^2}{3n} \xrightarrow[n \rightarrow \infty]{=0} 0 \Rightarrow \bar{\theta} \text{ è consistente} \end{aligned}$$

② Dato un campione casuale di dimensione n estratto da una popolazione con densità normale $N(\mu, \sigma^2)$,

stabilire se lo stimatore per μ

$$T_n = \frac{X_1 + X_n}{2}$$

è corretto e consistente.

Per una variabile casuale normale X si ha $E[X] = \mu$
e $\text{var}[X] = \sigma^2$.

- $E[T_n] \stackrel{???}{=} \mu$

$$E[T_n] = E\left[\frac{X_1 + X_n}{2}\right] = \frac{1}{2} \{E[X_1] + E[X_n]\} =$$

$$= E[X] = \mu \Rightarrow T_n \text{ è corretto.}$$

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

$$\text{MSE}[T_n] = \text{var}[T_n] + \underbrace{D[T_n]^2}_{=0} = \text{var}[T_n] =$$

$$= \text{var}\left[\frac{X_1 + X_n}{2}\right] = \frac{1}{4} \text{var}[X_1 + X_n]$$

$$= \left[\text{var}[X_1] + \text{var}[X_n] + 2 \text{cov}(X_1, X_n) \right] =$$

$=0, \text{ indipendenti}$

$$= \frac{1}{4} (2 \text{var}[X]) \stackrel{\text{ident. distr.}}{=} \frac{\text{var}[X]}{2} = \frac{\sigma^2}{2} \neq 0 \Rightarrow$$

$\Rightarrow T_n$ non è consistente

③ Sia X una variabile casuale distribuita con la legge

$$f(x) = C_a \begin{cases} (x+a)^2 & -a \leq x < 0 \\ (x-a)^2 & 0 \leq x \leq a \\ 0 & \text{altrove} \end{cases}$$

- Calcolare C_a
- Calcolare $E[X]$, $\text{var}[X]$
- Determinare uno stimatore di “ a ” col metodo dei momenti.

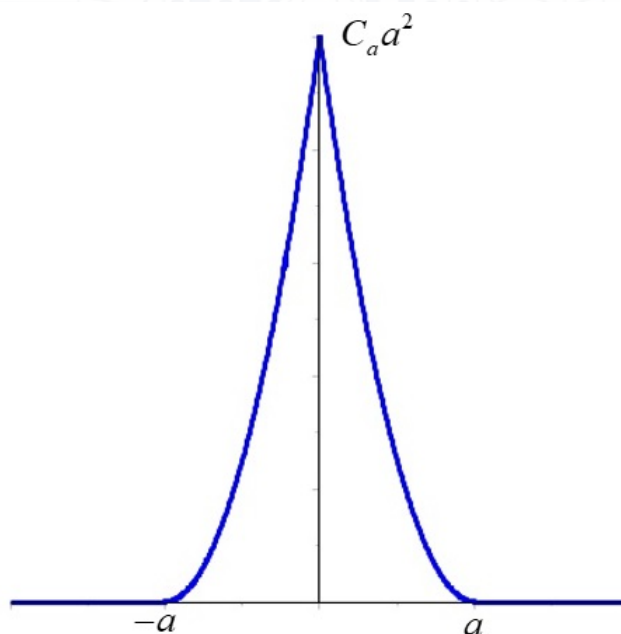


Figura 1

- Bisogna richiedere che la funzione $f(x)$ sia normalizzata a 1 \Rightarrow

$$\int_{-\infty}^{\infty} f(x) dx = 2 \int_0^a C_a (x-a)^2 dx = 2C_a \int_0^a (x^2 - 2ax + a^2) dx$$

↓
per simmetria

$$= 2C_a \left(\frac{1}{3}x^3 - ax^2 + a^2x \right)_0^a = \frac{2}{3} C_a a^3 = 1$$

$$\Rightarrow C_a = \frac{3}{2a^3}$$

- per simmetria $E[X] = 0$

$$\Rightarrow \text{var}[X] = E[X^2]$$

$$E[X^2] = 2C_a \int_0^a x^2 (x-a)^2 dx =$$

$$\frac{3}{a^3} \int_0^a (x^4 - 2ax^3 + a^2x^2) dx =$$

$$= \frac{3}{a^3} \left(\frac{1}{5}a^5 - \frac{1}{2}a^5 + \frac{1}{3}a^5 \right) = \frac{a^2}{10}$$

- L'equazione dei momenti è $M'_1 = \mu'_1$.

Ma in questo caso $\mu'_1 = \mu = E[X] = 0 \Rightarrow \bar{X}_n = 0!!!$
 pertanto devo passare all'ordine 2: $M'_2 = \mu'_2$.

$$\mu'_2 = E[X^2] = \frac{a^2}{10} \Rightarrow M'_2 = \frac{a^2}{10} \Rightarrow$$

$\Rightarrow \bar{a} = \sqrt{10M'_2}$ è lo stimatore di a , con $M'_2 = \frac{1}{n} \sum_{i=1}^n X_i^2$

$$\begin{aligned} E[\bar{a}^2] &= E[10M'_2] = 10E[M'_2] = 10E\left[\frac{1}{n} \sum_{i=1}^n X_i^2\right] = \\ &= \frac{10}{n} \sum_{i=1}^n E[X_i^2] = \frac{10}{n} nE[X^2] = 10 \frac{a^2}{10} = a^2 \end{aligned}$$

è il quadrato dello stimatore ad essere corretto.

