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 $t(X_1, ..., X_n)$ 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
$t_1(X_1, ..., X_n)$ e $t_2(X_1, ..., X_n)$ 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 svariate. 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 $f(\cdot; \theta_1, ..., \theta_k)$ con k parametri incogniti.
I momenti di ordine r della popolazione sono:

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

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

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

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

$$M'_j = \mu'_j(\theta_1, ..., \theta_k), \qquad j = 1, ..., k$$

nelle k incognite $\theta_1, ..., \theta_k$.
La soluzione unica di tale sistema $(\overline{\theta}_1, ..., \overline{\theta}_k)$ sarà lo stimatore puntuale cercato.

Esempi

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

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

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 $M_1' = \overline{X}_n$.

Allora $M_1' = \mu_1' \Rightarrow \overline{X}_n = \frac{1}{\theta} \Rightarrow \overline{\theta} = \frac{1}{\overline{X}_n}$

2 Dato un campione casuale $(X_1, ..., X_n)$ di dimensione n estratto da una popolazione con densità normale $N(\mu, \sigma^2)$, determinare gli stimatori puntuali $\overline{\theta}_1$, $\overline{\theta}_2$ 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

$$M_1' = \overline{X}_n \,; \qquad M_2' = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

perciò

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

$$\overline{\theta}_2 = \sqrt{M_2' - \overline{X}_n^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2} =$$

$$= \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2} = \sqrt{M_2} \text{ è lo stimatore di } \sigma$$

La penultima uguaglianza deriva dalle seguenti espressioni:

$$\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i^2 - 2X_i\overline{X}_n + \overline{X}_n^2\right) =$$

$$= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{2}{n}\overline{X}_n(n \cdot \overline{X}_n) + \frac{1}{n}\overline{X}_n^2 \cdot n = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2$$

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 $T = t(X_1, ..., X_n)$.

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

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

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

PROPRIETÀ

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

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

Infatti:

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

$= E[(T - E[T] + E[T] - \theta)^2] =$

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

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

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

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

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

$E\left[(T-E[T])(E[T]-\theta)\right] = (E[T]-\theta)E[T-E[T]]$
ma
$E[T-E[T]] = E[T]-E[T] = 0$
quindi

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

- Se T è corretto allora

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

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

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

Poichè $E[\overline{X}_n] = \mu \Rightarrow \overline{X}_n$ è uno stimatore corretto.
Inoltre
$\mathrm{MSE}[\overline{X}_n] = E[(\overline{X}_n-\mu)^2] = \mathrm{var}[\overline{X}_n] = \dfrac{\sigma^2}{n}$
Invece
$E[M_2] = \dfrac{n-1}{n}E[S^2]$,

ma $E[S^2] = \sigma^2$ (dal teorema 2 del campionamento), quindi
$E[M_2] = \dfrac{n-1}{n}\sigma^2 \neq \sigma^2 \Rightarrow M_2$ 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 $(X_1, ..., X_n)$ 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 $(X_1, ..., X_n)$ di dimensione n estratto da una popolazione con funzione di densità esponenziale

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

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

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

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:

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

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

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

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

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

$\Rightarrow \overline{X}_n$ 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.

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

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

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

- Dall'equazione dei momenti $M_1' = \mu_1'$ si ha:

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

- $E[\overline{\theta}] \overset{???}{=} \theta$

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

$\overline{\theta}$ è consistente se $\text{MSE}[\overline{\theta}] \to 0$

$\text{MSE}[\overline{\theta}] = \text{var}[\overline{\theta}] + \underbrace{D[\overline{\theta}]^2}_{=0} = \text{var}[2\overline{X}_n] = 4\,\text{var}[\overline{X}_n] =$

$= 4\frac{\sigma^2}{n} = \frac{4}{n}\frac{\theta^2}{12} = \frac{\theta^2}{3n} \xrightarrow[n\to\infty]{} 0 \Rightarrow \overline{\theta}$ è consistente

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 $\text{var}[X] = \sigma^2$.

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

$$E[T_n] = E\left[\frac{X_1+X_n}{2}\right] = \frac{1}{2}\{E[X_1] + E[X_n]\} =$$
$$= E[X] = \mu \Rightarrow T_n \text{ è } \text{corretto}.$$

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

$$\text{MSE}[T_n] = \text{var}[T_n] + \underbrace{D[T_n]^2}_{=0} = \text{var}[T_n] =$$
$$= \text{var}\left[\frac{X_1+X_n}{2}\right] = \frac{1}{4}\text{var}[X_1 + X_n]$$
$$= \left[\text{var}[X_1] + \text{var}[X_n] + 2\underbrace{\text{cov}(X_1, X_n)}_{=0,\text{ indipendenti}}\right] =$$
$$= \underbrace{\frac{1}{4}(2\,\text{var}[X])}_{\text{ident. distr.}} = \frac{\text{var}[X]}{2} = \frac{\sigma^2}{2} \neq 0 \Rightarrow$$

$\Rightarrow T_n$ 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 $E[X]$, var$[X]$
- Determinare uno stimatore di "a" col metodo dei momenti.

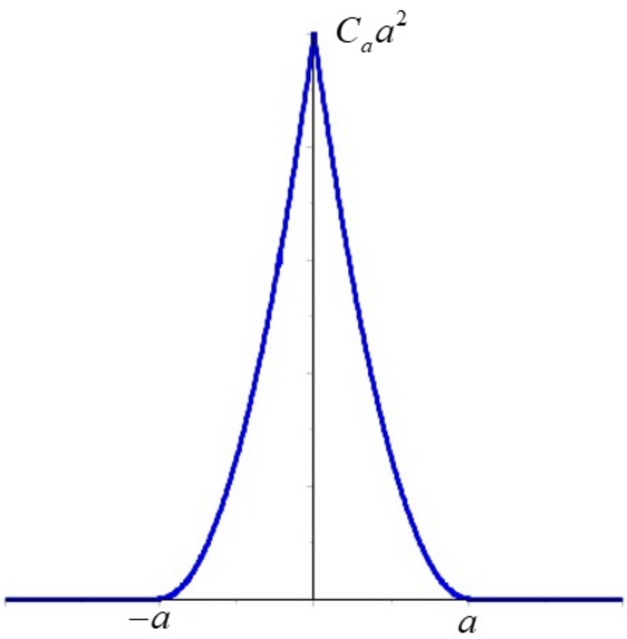

Figura 38

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

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

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

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

- per simmetria $E[X] = 0$
$\Rightarrow \mathrm{var}[X] = E[X^2]$

$$E[X^2] = 2C_a \int_0^a x^2(x-a)^2 dx =$$
$$\frac{3}{a^3}\int_0^a (x^4 - 2ax^3 + a^2x^2)dx =$$
$$= \frac{3}{a^3}\left(\frac{1}{5}a^5 - \frac{1}{2}a^5 + \frac{1}{3}a^5\right) = \frac{a^2}{10}$$

- L'equazione dei momenti è $M_1' = \mu_1'$.

Ma in questo caso $\mu_1' = \mu = E[X] = 0 \Rightarrow \overline{X}_n = 0!!!$ pertanto devo passare all'ordine 2: $M_2' = \mu_2'$.

$$\mu_2' = E[X^2] = \frac{a^2}{10} \Rightarrow M_2' = \frac{a^2}{10} \Rightarrow$$

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

$E[\overline{a}^2] = E[10M'_2] = 10E[M'_2] = 10E[\frac{1}{n}\sum\limits_{i=1}^{n} X_i^2] =$

$= \frac{10}{n}\sum\limits_{i=1}^{n} E[X_i^2] = \dfrac{10}{n}\, nE[X^2] = 10\,\dfrac{a^2}{10} = a^2$

è il quadrato dello stimatore ad essere corretto.