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 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 $f(\cdot\,;\theta_1,\dots,\theta_k)$ 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 c.c. di dim $= n$ $(X_1,\dots,X_n)$ i momenti campionari sono:

$$M'_J = \frac{1}{n}\sum_{i=1}^{n} X_i^J \qquad J = 1,\dots,k$$

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 c.c. $(X_1, \dots, X_n)$ di dim $= n$ estratto da una popolazione con densità esponenziale

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

determinare uno stimatore per θ col m. dei m.

Per una v.c. esponenziale X si sa $E[X] = \frac{1}{\theta}$.

$$\mu_1' = \mu = E[X] \implies \mu_1'(\theta) = \frac{1}{\theta}$$

mentre $M_1' = \bar{X}_n$.

Allora $M_1' = \mu_1' \implies \bar{X}_n = \frac{1}{\theta} \implies \bar{\theta} = \frac{1}{\bar{X}_n}$

2. Dato un c.c. $(X_1, \dots, X_n)$ di dim $= n$ estratto da una popolazione con densità normale $N(\mu, \sigma^2)$, determinare gli stimatori puntuali $\bar{\theta}_1, \bar{\theta}_2$ per i parametri $(\theta_1, \theta_2) = (\mu, \sigma)$ col m. dei m.

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:

$$\begin{cases} M_1' = \mu_1' = \mu_1'(\mu, \sigma) = \mu \\ M_2' = \mu_2' = \mu_2'(\mu, \sigma) = \mu^2 + \sigma^2 \end{cases}$$

ma

$$M_1' = \bar{X}_m \quad ; \quad M_2' = \frac{1}{m} \sum_{i=1}^{m} X_i^2$$

perciò

$\bar{\Theta}_1 = \bar{X}_m$ è lo stimatore di μ

$$\bar{\Theta}_2 = \sqrt{M_2' - \bar{X}_n^2}$$

$$= \sqrt{\frac{1}{m} \sum_{i=1}^{m} X_i^2 - \bar{X}_m^2}$$

$$= \sqrt{\frac{1}{m} \sum_{i=1}^{m} (X_i - \bar{X}_m)^2} \quad (*)$$

$= \sqrt{M_2}$ è lo stimatore di σ

$$(*) \quad \frac{1}{m} \sum_{i=1}^{m} (X_i - \bar{X}_m)^2 = \frac{1}{m} \sum_{i=1}^{m} (X_i^2 - 2X_i \bar{X}_m + \bar{X}_n^2)$$

$$= \frac{1}{m} \sum_{i=1}^{m} X_i^2 - \frac{2}{m} \bar{X}_m \cdot (m \bar{X}_m) + \frac{1}{m} \cdot \bar{X}_n^2 \cdot m$$

$$= \frac{1}{m} \sum_{i=1}^{m} X_i^2 - \bar{X}_n^2$$

PROPRIETÀ DEGLI STIMATORI PUNTUALI

Esistono stimatori che siano in qualche modo migliori di altri? Definiamo ora 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

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

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

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

DEFINIZIONE: Uno stimatore T si dice **CORRETTO** o **NON DISTORTO** di θ sse

$$E[T] = \theta$$

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

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

Se T è corretto $\Rightarrow D(T) = 0$.

PROPRIETÀ

Per ogni stimatore T del parametro θ vale la relazione:

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

Infatti:

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

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

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

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

$(E[T]-\theta)$ non dipende dalle variabili X_i del campione e quindi va considerato come una costante

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

$$E[(T-E[T])\cdot(E[T]-\theta)] = (E[T]-\theta)\cdot \underbrace{E[T-E[T]]}_{= E[T]-E[T] \equiv 0}$$

perciò

$$MSE[T] = E[(T-E[T])^2] + D^2[T]$$

$$= var[T] + D^2[T]$$

- se T è corretto allora:

$$MSE[T] = var[T]$$

Esempio Dato un c.c. di dim = m estratto da una popolazione con funzione di densità normale $N(\mu,\sigma^2)$ abbiamo ricavato col m. dei m.

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

Poiché $E[\bar{X}_m] = \mu \Rightarrow \bar{X}_m$ è uno stimatore CORRETTO
Inoltre

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

Invece

$$E[M_2] = \frac{(m-1)}{m} E[S^2] = \frac{(m-1)}{m}\sigma^2 \neq \sigma^2 \Rightarrow M_2 \text{ è uno stimatore DISTORTO}$$

$E[S^2] = \sigma^2$ (TH.2 del campionamento)

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-CRAMÈR

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

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

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

Esempio Dato un c.c. $X_1, \dots, X_n$ estratto da una popolazione con funzione di densità esponenziale

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

si ha:

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

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

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

$E[X] = \mu = \frac{1}{\theta}$ per una v.c. exp.

quindi

$$var[T] \geq \frac{1}{n \cdot \frac{1}{\theta^2}} = \frac{\theta^2}{n} \qquad \text{LIMITE INFERIORE}$$

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

DEFINIZIONE: Uno stimatore è detto CONSISTENTE in media quadratica sse:

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

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

$$\Rightarrow var[T_n] \xrightarrow[n \to \infty]{} 0 \quad ; \quad D[T_n] \xrightarrow[n \to \infty]{} 0$$

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

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

$\Rightarrow \bar{X}_n$ è uno stimatore CONSISTENTE

Esempi

1. Dato un c.c. di dim=n estratto da una popolazione con funzione di densità uniforme sull'intervallo $[0, \theta]$.
Trovare uno stimatore di θ col m. dei m e stabilire se è corretto e consistente.

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

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

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

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

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

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

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

- $\bar{\theta}$ è consistente se $MSE[\bar{\theta}] \to 0$

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

$$= 4\frac{\sigma^2}{n} = \frac{4}{n} \cdot \frac{\theta^2}{12} = \frac{\theta^2}{3n} \xrightarrow[n\to\infty]{} 0 \implies \bar{\theta} \text{ è consistente}$$

2. Dato un c.c. di dim=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 v.c. normale X si sa $E[X]=\mu$ e $var[X]=\sigma^2$.

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

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

$$= E[X] = \mu \implies T_n \text{ è CORRETTO}$$

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

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

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

$$= \frac{1}{4}\left[var[X_1]+var[X_n]+2\underbrace{cov(X_1,X_n)}_{0}\right]$$

perchè indipendenti

$$= \frac{1}{4}\cdot(2\, var[X])$$

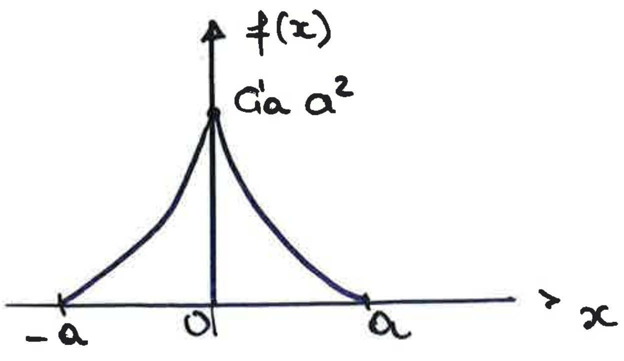
perchè identicamente distribuite

$$= \frac{var[X]}{2} = \frac{\sigma^2}{2} \neq 0 \implies T_n \text{ non è CONSISTENTE}$$

③ Sia X una v.c. distribuita con la legge:

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

- calcolare C_a
- calcolare $E[X]$, $var[X]$.
- determinare uno stimatore di "a" col m. dei m.



$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

$$1 = \int_{-\infty}^{+\infty} = 2\int_0^{\infty} = 2\int_0^a C_a (x-a)^2 dx = 2C_a \int_0^a (x^2 - 2ax + a^2)dx$$

per simmetria

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

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

- per simmetria $E[X] \equiv 0$

  $\Rightarrow var[X] \equiv E[X^2]$

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

$$= \frac{3}{a^3} \left(\frac{1}{5}a^5 - 2a \cdot \frac{1}{4}a^4 + a^2 \cdot \frac{1}{3}a^3 \right) = \frac{3}{a^3} \cdot \frac{a^5}{30} = \frac{a^2}{10}$$

- l'eq. dei momenti è: $M_1' = \mu_1'$

  ma in questo caso $\mu_1' \equiv \mu = E[X] \equiv 0 \Rightarrow \bar{X}_n = 0$ ↯

  allora passo all'ordine 2: $M_2' = \mu_2'$

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

  quindi $M_2' = \frac{a^2}{10} \Rightarrow \bar{a} = \sqrt{10 M_2'}$ stimatore di a.

  dove $M_2' = \frac{1}{n}\sum_{i=1}^{n} X_i^2$

$$E[\bar{a}^2] = E[10\, M_2'] = 10\, E[M_2'] = 10\, E\left[\frac{1}{n}\sum_i X_i^2\right]$$

$$= \frac{10}{n}\sum_1 E[X_i^2] = \frac{10}{n} \cdot n \cdot E[X^2] = 10 \cdot \frac{a^2}{10} = a^2$$

è il quadrato dello stimatore ad essere corretto.