Corso di Statistica

francesco mola

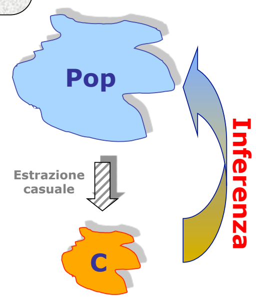
Lezione n° 14: Campionamento e distribuzioni campionarie

sommario

- Cosa si intende per *campionamento*
- Stima puntuale o intervallare per un parametro incognito della popolazione
- Proprietà degli stimatori
- Teorema del Limite Centrale

Cosa si intende per *campionamento*

Si definisce campionamento un procedimento attraverso il quale da un insieme di unità costituenti l'oggetto dello studio, si estrae un numero ridotto di casi scelti con criteri tali da consentire la generalizzazione all'intera popolazione dei risultati ottenuti.



Sitma puntuale o intervallare per un parametro incognito della popolazione

- Ricorre spesso l'esigenza di stimare un parametro incognito della popolazione. Non sempre si dispone di tutti i dati e/o spesso i vincoli di tempo e costi sono tali da indurci al campionamento
- L'obiettivo è trovare uno stimatore che produca stime quanto più vicine possibile al parametro incognito.

Sitma puntuale o intervallare per un parametro incognito della popolazione

- La stima può essere di tipo puntuale o intervallare.
  - Es1. la variabile casuale media campionaria, stimatore della media, quando lavora con i dati campionari genera la stima della media della popolazione
  - Es2. Considerando la tecnica degli intervalli di confidenza, ed utilizzando come funzione *pivot* la media campionaria, ottengo una stima intervallare della media della popolazione

- Es3. la variabile casuale proporzione campionaria, stimatore della proporzione, quando lavora con i dati campionari genera la stima della proporzione della popolazione
- Es4. Considerando la tecnica degli intervalli di confidenza, ed utilizzando come funzione *pivot* la proporzione campionaria, ottengo una stima intervallare della media della popolazione

Inferenza statistica

Logica induttiva

"Dal particolare al generale"

1) Popolazione $[v.c.X]$

Se si conosce F(X) si può fare inferenza

$$X = f(X,\theta)$$

$$\theta = \theta_1, \theta_2, \ldots \theta_p =$$ p parametri incogniti

$\Omega(\theta) =$ Spazio parametrico

2) Campione casuale = realizzazione di n variabili casuali indipendenti

$x_i =$ i-esima osservazione

$X_i =$ v.c. generatrice

3) Stima dei parametri

Intervalli di confidenza

Test delle ipotesi

Cosa produce un processo inferenziale?
•Un risultato numerico
•Un giudizio di validità dello stesso

$(x_1, x_2, \ldots x_n)$ Campione osservato

$(X_1, X_2, \ldots X_n)$ Campione casuale

n.b.: le X sono supposte indipendenti ed identicamente distribuite (iid) come

$$X \sim f(x, \theta)$$

Due problematiche:
1. Bontà dello stimatore (proprietà dello stimatore)
2. Metodi di costruzione degli stimatori

Proprietà degli stimatori

Caratterizzazione in media (non distorsione)

•Uno stimatore T_n si dice non distorto (**unbiased**) per θ se:

$$E(T_n) = \theta$$

$$b(T_n) = E(T_n) - \theta =$$ Distorsione (**bias**)

$b(T_n) > 0 \Rightarrow$ Distorsione positiva (**sovrastima**)
$b(T_n) < 0 \Rightarrow$ Distorsione negativa (**sottostima**)

In generale se $b(T_n) \neq 0 \Rightarrow$ Errore sistematico

Se T_n è non distorto allora θ è baricentro (media) della distribuzione di T_n

Per vedere come si accentra uno stimatore rispetto a θ si può considerare la varianza.

ma se T_n è distorto ($b(T_n) \neq 0$) la varianza non può essere usata!

Si usa l'errore quadratico medio o Mean Square Error (MSE)

$$MSE(T_n) = E(T_n - \theta)^2 = \underbrace{VAR(T_n)}_{\text{Varianza dello stimatore}} + \underbrace{b^2(T_n)}_{\text{Bias al quadrato}}$$

n.b. se $b(T_n) = 0 \Rightarrow MSE = VAR(T_n)$

Efficienza di uno stimatore

Siano T_{1n} e T_{2n} due stimatori per il parametro incognito

Se $MSE(T_{1n}) < MSE(T_{2n})$ allora

T_{1n} è più efficiente di T_{2n}

Efficienza relativa

$$eff(T_{1n}|T_{2n}) = \frac{MSE(T_{2n})}{MSE(T_{1n})}$$

se $eff(T_{1n}|T_{2n}) = \frac{MSE(T_{2n})}{MSE(T_{1n})} > 1 \Rightarrow T_{1n}$ È preferibile a T_{2n}

se $eff(T_{1n}|T_{2n}) = \frac{MSE(T_{2n})}{MSE(T_{1n})} < 1 \Rightarrow T_{2n}$ È preferibile a T_{1n}

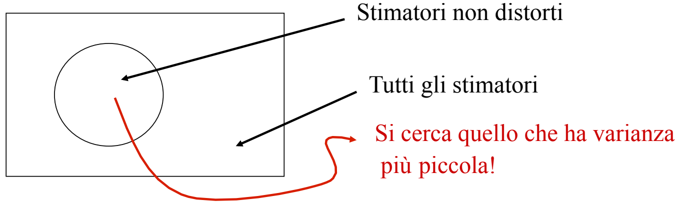
se $eff(T_{1n}|T_{2n}) = \frac{MSE(T_{2n})}{MSE(T_{1n})} = 1 \Rightarrow$ **Indifferenza nella scelta (rispetto a MSE)**

$$b(T_{1n}) = b(T_{2n}) = 0 \Rightarrow eff(T_{1n}|T_{2n}) = \frac{VAR(T_{2n})}{VAR(T_{1n})}$$

• *Perché si considera la varianza più piccola?*

Perché la varianza è funzione diretta di n e quindi significa Che quando si lavora con gli stimatori più efficienti si possono avere campioni più piccoli!

Esiste un limite inferiore (non dipendente da θ) per la Variabilità di uno stimatore?



Consistenza

T_n È consistente in media quadratica per θ se:

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

$$T_n \xrightarrow{m} \theta$$

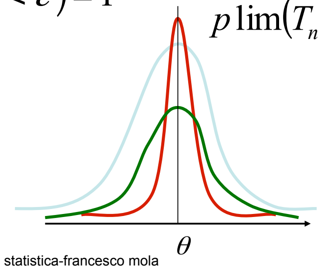
Consistenza

T_n é consistente in probabilità per θ se, fissato un $\varepsilon > 0$ si ha:

$$\lim_{n\to\infty} P\left(\left|T_n - \theta\right| < \varepsilon\right) = 1$$

$$T_n \xrightarrow{p} \theta$$

$$p\lim(T_n) = \theta$$



Metodi di costruzione degli stimatori (cenni)

- Metodo dei momenti
- Metodo della massima verosimiglianza
- Metodo dei minimi quadrati