Esempio

Consideriamo un campione casuale $(X_1, \dots, X_n)$ estratto da una popolazione di densità Bernoulli di parametro θ. Il parametro θ è sconosciuto e vogliamo stimarlo. Consideriamo gli stimatori

$T_1 = \bar{X}_n$ (media campionaria) e $T_2 = \frac{1}{3}(X_2 + 2X_5)$.

(a) Gli stimatori T_1 e T_2 sono stimatori <u>corretti</u> del parametro θ.

Uno stimatore è corretto per la stima di un parametro (o di una funzione di quel parametro) se il suo valor medio coincide con quanto voglio stimare.

Nel nostro caso dobbiamo verificare $E_\theta(T_i) = \theta$, per $i = 1, 2$.

- $E_\theta(T_1) = E_\theta(\bar{X}_n) = \frac{1}{n}\sum_{i=1}^{n} E_\theta(X_i) = \theta$

linearità

$= \theta$ poiché le X_i sono i.i.d. $Be(\theta)$

- $E_\theta(T_2) = E_\theta\left[\frac{1}{3}(X_2 + 2X_5)\right] = \frac{1}{3}E_\theta(X_2) + \frac{2}{3}E_\theta(X_5) = \theta$

(b) T_1 è uno stimatore consistente, mentre T_2 non lo è.

Uno stimatore è consistente se la sua varianza tende a zero al crescere dell'ampiezza del campione. Calcoliamo la varianza degli stimatori proposti e vediamo se soddisfano questa proprietà.

$$\bullet\ Var_\theta(T_1) = \frac{1}{n^2}\sum_{i=1}^{n} Var_\theta(X_i) = \frac{\theta(1-\theta)}{n} \xrightarrow{n\to+\infty} 0$$

additività

$= \theta(1-\theta)$ poiché le X_i i.i.d. $Be(\theta)$

$$\bullet\ Var_\theta(T_2) = \frac{1}{9} Var_\theta(X_2) + \frac{4}{9} Var_\theta(X_5) = \frac{5}{9}\theta(1-\theta) \ \cancel{\xrightarrow{n\to+\infty}}\ 0$$

STIMA PUNTUALE DI MEDIA E VARIANZA

Sia $(X_1, \dots, X_n)$ un campione casuale estratto da una popolazione di densità f_θ, che ammette valor medio μ e varianza σ^2, entrambi incogniti che vogliamo stimare.

Stima del valor medio μ: si usa lo stimatore $\overline{X}_n$ perché è sempre uno stimatore corretto e consistente del valor medio. Infatti

$$E_\theta(\bar{X}_n) = \frac{1}{n}\sum_{i=1}^{n} E_\theta(X_i) = \mu \qquad \Rightarrow \text{correttezza}$$

$$Var_\theta(\bar{X}_n) = \frac{1}{n^2}\sum_{i=1}^{n} Var_\theta(X_i) = \frac{\sigma^2}{n} \xrightarrow{n\to+\infty} 0 \qquad \Rightarrow \text{consistenza}$$

<u>Stima della varianza σ^2</u> (la media della popolazione è incognita): lo stimatore "naturale" dato da

$$T = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

<u>non</u> è corretto, poiché $E_\theta(T) = \frac{(n-1)\sigma^2}{n}$. Andiamo a correggerlo. Si ha

$$E_\theta(T) = \frac{(n-1)\sigma^2}{n} \iff \frac{n}{n-1}E_\theta(T) = \sigma^2$$

$$\iff E_\theta\left[\frac{n}{n-1}T\right] = \sigma^2$$

Quindi $\frac{n}{n-1}T$ è uno stimatore corretto di σ^2. Se lo calcoliamo esplicitamente, otteniamo

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

ESERCIZI

1. Si consideri il modello statistico, dipendente dal parametro $\alpha > 0$, definito come

$$f_\alpha(x) = \begin{cases} c_\alpha x^{\alpha-1}(1-x) & \text{se } x \in (0,1) \\ 0 & \text{altrimenti.} \end{cases}$$

(a) Determinare il valore c_α per cui f_α risulta una densità.

(b) Calcolare il valor medio

(c) In base ad un campione di ampiezza n, usando la media campionaria, costruire uno stimatore del parametro α.

(d) Dare una stima del parametro α, utilizzando lo stimatore trovato al punto precedente, nel caso in cui i dati campionari forniscano $\bar{x} = 0.52$.

<u>Soluzione</u>. (a) Imponiamo la condizione di normalizzazione. Otteniamo

$$\int_{\mathbb{R}} f_\alpha(x)\,dx = 1 \iff c_\alpha \int_0^1 x^{\alpha-1}(1-x)\,dx = 1$$

$$\iff c_\alpha \left[\frac{x^\alpha}{\alpha} - \frac{x^{\alpha+1}}{\alpha+1} \right]_0^1 = 1$$

$$\iff \quad c_\alpha = \alpha(\alpha+1)$$

e quindi la densità risulta

$$f_\alpha(x) = \begin{cases} \alpha(\alpha+1)\, x^{\alpha-1}(1-x) & \text{se } x \in (0,1) \\ 0 & \text{altrimenti.} \end{cases}$$

(b) Calcoliamo

$$\int_{\mathbb{R}} x f_\alpha(x)\,dx = \alpha(\alpha+1)\int_0^1 x^\alpha (1-x)\,dx$$

$$= \alpha(\alpha+1)\left[\frac{x^{\alpha+1}}{\alpha+1} - \frac{x^{\alpha+2}}{\alpha+2}\right]_0^1$$

$$= \frac{\alpha}{\alpha+2}.$$

(c) Sappiamo che $\overline{X}_n$ è uno stimatore corretto e consistente del valor medio. Quindi $\overline{X}_n$ è uno stimatore corretto e consistente di $\frac{\alpha}{\alpha+2}$ (cioè di una funzione del parametro α). Se n è grande, per la legge dei grandi numeri, si ha $\overline{X}_n \approx \frac{\alpha}{\alpha+2}$. Invertiamo questa relazione per trovare uno stimatore di α. Si ha

$$\overline{X}_n \approx \frac{\alpha}{\alpha+2} \iff \alpha \approx \frac{2\overline{X}_n}{1-\overline{X}_n}$$

Usiamo $T = \dfrac{2\overline{X}_n}{1-\overline{X}_n}$ come stimatore di α.

Oss. Lo stimatore di α così ottenuto è stato costruito a partire da uno stimatore corretto e consistente della media, non abbiamo però nessuna garanzia che sia uno stimatore corretto e consistente di α. In generale, non lo è e verificarlo è molto difficile.

(d) Se $\bar{x} = 0.52$, allora $\hat{\alpha} = \dfrac{2\bar{x}}{1-\bar{x}} = \dfrac{2(0.52)}{1-0.52} = 2.167$.

2. Siano $X_1, \dots, X_n$ variabili aleatorie i.i.d. con densità data da

$$f_L(x) = \begin{cases} 2 c_L x & \text{se } x \in [0, L] \\ 0 & \text{altrimenti;} \end{cases}$$

dove $L > 0$.

(a) Determinare il valore c_L per cui f_L risulta una densità.

(b) Calcolare il valor medio.

(c) Sia $T = \overline{X}_n$ uno stimatore per L. È corretto? Come può essere eventualmente modificato per diventare corretto?

<u>Soluzione.</u> (a) Imponiamo la normalizzazione. Si ha

$$\int_{\mathbb{R}} f_L(x)\,dx = 1 \iff 2c_L \int_0^L x\,dx = 1$$

$$\iff c_L \, x^2 \Big|_0^L = 1$$

$$\iff c_L = \frac{1}{L^2}$$

e quindi la densità risulta

$$f_L(x) = \begin{cases} \dfrac{2x}{L^2} & \text{se} \quad x \in [0,L] \\ 0 & \text{altrimenti.} \end{cases}$$

(b) Calcoliamo

$$\int_{\mathbb{R}} x\, f_L(x)\,dx = \frac{2}{L^2}\int_0^L x^2\,dx = \frac{2}{L^2}\cdot \frac{x^3}{3}\Big|_0^L = \frac{2}{3}L.$$

(c) T è uno stimatore corretto del valor medio, quindi $E_L(T) = \frac{2}{3}L$. Affinché sia uno stimatore corretto di L lo dobbiamo quindi moltiplicare per $\frac{3}{2}$. Pertanto, $S = \frac{3}{2}T$ è uno stimatore corretto di L.