ESERCIZIO 5.1

Sia (X_1, X_2, …, X_n) un campione casuale generato da una v.c. $X \sim f(X; \theta)$ per la quale è noto che $E(X) = \theta$ e $Var(X) = 1$. Si considerino i 3 stimatori di θ:

$$T_1 = \frac{1}{n}\sum X_i \; ; \qquad T_2 = \frac{2}{n(n+1)}\sum iX_i \; ; \qquad T_3 = \sum (-1)^i X_i$$

a) si determini se sono corretti;
b) per quelli non corretti, si calcoli la distorsione d;
c) sapendo che $\sigma^2_{T_2} = \dfrac{2(2n+1)}{3n(n+1)}$, qual è, tra gli stimatori corretti, il migliore per θ?
d) e quale il migliore tra i 3?

<u>Soluzione</u>

a)
Per stabilire se T_1, T_2 e T_3 sono corretti occorre calcolare il loro valore atteso e confrontarlo con il parametro θ incognito.

Essendo uguale proprio alla media aritmetica, T_1 è senz'altro corretto:

$E(T_1) = \theta$

Per quanto riguarda T_2 si ha:

$$E(T_2) = \frac{2}{n(n+1)}\sum iE(X_i) = \frac{2}{n(n+1)}\theta\sum i = \frac{2}{n(n+1)}\theta\frac{n(n+1)}{2} = \theta$$

Infine per T_3:

$$E(T_3) = \sum (-1)^i E(X_i) = \theta\sum(-1)^i = \begin{cases} 0 & \text{se n è pari} \\ -\theta & \text{se n è dispari} \end{cases} \neq \theta$$

Quindi possiamo concludere che gli stimatori T_1 e T_2 sono corretti, mentre T_3 è distorto.

b)
La distorsione di T_3 può essere calcolata come:

$$d(T_3) = E(T_3) - \theta = \begin{cases} -\theta & \text{se n è pari} \\ -2\theta & \text{se n è dispari} \end{cases}$$

c)
La scelta è limitata ai due stimatori T_1 e T_2, di cui passiamo a considerare l'efficienza relativa:

$$\text{eff}\left(T_2 \mid T_1\right) = \frac{\text{EQM}\left(T_2\right)}{\text{EQM}\left(T_1\right)}$$

Essendo T_1 e T_2 entrambi corretti, il loro errore quadratico medio coinciderà con le loro varianze:

$$\sigma^2_{T_1} = \frac{1}{n}\sigma^2 = \frac{1}{n}$$

$$\sigma^2_{T_2} = \frac{2\left(2n+1\right)}{3n\left(n+1\right)}$$

Dal confronto tra le due quantità emerge che:

$$\text{eff}\left(T_2 \mid T_1\right) = \frac{\text{EQM}\left(T_2\right)}{\text{EQM}\left(T_1\right)} = \frac{\sigma^2_{T_2}}{\sigma^2_{T_1}} = \frac{\frac{2\left(2n+1\right)}{3n\left(n+1\right)}}{\frac{1}{n}} = \frac{2\left(2n+1\right)}{3\left(n+1\right)} = \frac{4n+2}{3n+3}$$

quantità che è uguale ad 1 se $n = 1$ ed è sempre maggiore di 1 quando $n > 1$.

Si può, dunque, concludere che, per quanto assurdo che possa sembrare, se estraessimo un campione di ampiezza $n = 1$ (cioè di una sola osservazione) sarebbe per noi indifferente utilizzare T_1 o T_2. Questo, d'altra parte, si verifica facilmente perché per $n = 1$ le due varianze $\sigma^2_{T_2} = \frac{2\left(2n+1\right)}{3n\left(n+1\right)}$ $\sigma^2_{T_1} = \frac{1}{n}$ sono uguali.

Se, invece, n è maggiore di 1 $\sigma^2_{T_1} < \sigma^2_{T_2}$, quindi T_1 è migliore di T_2. Si noti, inoltre, che eff($T_2|T_1$) tende a crescere all'aumentare di n, il che dimostra che più è grande il campione più *dobbiamo* preferire T_1 a T_2.

d)
Bisogna calcolare l'errore quadratico medio di T_3, quindi la sua varianza:

$$\text{Var}\left(T_3\right) = \text{Var}\left(\sum\left(-1\right)^i X_i\right) = \sum_i\left[\left(-1\right)^{2i}\text{Var}\left(X_i\right)\right] = n\sigma^2$$

Confrontiamo T_3 con il migliore tra T_1 e T_2, ossia T_1.

Tra i 2 stimatori sceglieremo quello con errore quadratico medio minore.

Sappiamo che: $EQM(T_1) = \sigma^2_{T_1} = \frac{1}{n}$. Dal momento che la distorsione di T_3 assume valori diversi a seconda se il campione sia di numerosità pari o dispari, $EQM(T_3)$ deve essere calcolato per entrambi i casi:

$$EQM(T_3) = \sigma^2_{T_3} + d^2_{T_3} = \begin{cases} n\sigma^2 + \theta^2 & \text{se n è pari} \\ n\sigma^2 + 4\theta^2 & \text{se n è dispari} \end{cases}$$

In entrambi i casi sceglieremo T_1, visto che $1/n$ è minore sia di $(n\sigma^2 + \theta^2)$ sia di $(n\sigma^2 + 4\theta^2)$; si noti, inoltre, che l'errore quadratico medio di T_1 tende a ridursi all'aumentare di n, mentre quello di T_3 all'aumentare di n cresce.

ESERCIZIO 5.2

Si considerino i due stimatori corretti T1 e T2 con varianze:

$$\sigma^2_{T_1} = \frac{1}{n}$$

e

$$\sigma^2_{T_2} = \frac{(0,7n_2 + 34,8)}{2n_2(n_2+1)}.$$

Si supponga di utilizzare T_1 e T_2 su due campioni di numerosià n_1 ed n_2.
Quanto deve essere ampio il campione su cui si utilizza T_2 affinchè la sua precisione sia equivalente a quella di T_1 relativamente ad un campione di n_1 = 10 unità?

<u>Soluzione</u>

Assumendo come misura di precisione di uno stimatore la sua varianza, utilizzare T_1 su un campione di 10 unità significa avere una precisione pari a:

$$\sigma^2_{T_1} = \frac{1}{n} = 0,1$$

Basta, dunque, porre anche la varianza di T_2 uguale a 0,1 e risolvere l'uguaglianza rispetto ad n_2:

$$\frac{(0,7n_2 + 34,8)}{2n_2(n_2+1)} = 0,1$$

da cui:

$$\frac{(0,7n_2 + 34,8)}{2n_2(n_2+1)} - \frac{1}{10} = 0 \Rightarrow \frac{10(0,7n_2 + 34,8) - 2n_2(n_2+1)}{20n_2(n_2+1)} = 0 \Rightarrow$$
$$\Rightarrow \frac{7n_2 + 348 - 2n_2^2 - 2n_2}{20n_2(n_2+1)} = 0$$

Dal momento che il denominatore è sempre positivo, l'uguaglianza va risolta rispetto al numeratore:

$$7n_2 + 348 - 2n_2^2 - 2n_2 = -2n_2^2 + 5n_2 + 348 = 0$$

Tale relazione è un'equazione di secondo grado nell'incognita n_2, che è soddisfatta per i due valori:

$$n_2 = \frac{-5 \pm \sqrt{25 + 4 \times 2 \times 348}}{2 \times (-2)} = \frac{-5 \pm \sqrt{2809}}{-4} = \frac{-5 \pm 53}{-4} = \begin{cases} 14,5 \\ -12 \end{cases}$$

di cui, naturalmente, quello negativo non è ammissibile.
L'ampiezza campionaria deve necessariamente essere un numero intero, quindi il valore 14,5 va arrotondato all'intero più grande (dato che i valori compresi tra 14 e 14,5 danno una precisione minore di 0,1).

L'ampiezza campionaria che consente a T_2 di assicurare la stessa efficienza che T_1 ha su un campione di 10 unità è, dunque, 15 unità.

ESERCIZIO 5.3

Sia (X_1, X_2, X_3, X_4, X_5) un campione casuale di ampiezza n = 5, generato da una v.c. $X \sim f(X)$ con media μ e varianza σ^2. Si considerino i due stimatori per μ:

$$T_1 = \bar{x} = \frac{1}{5}\sum_{i=1}^{5} X_i$$

e

$$T_2 = 0{,}1 * X_1 + 0{,}2 * X_2 + 0{,}4 * X_3 + 0{,}2 * X_4 + 0{,}1 * X_5 .$$

a) Si verifichi la correttezza di T_1 e T_2;
b) quale dei due è più efficiente?

Soluzione

a)

È noto che la media aritmetica campionaria è uno stimatore corretto della media della popolazione μ.
Per quanto riguarda T_2 si ha:

$$E(T_2) = E(0{,}1 * X_1 + 0{,}2 * X_2 + 0{,}4 * X_3 + 0{,}2 * X_4 + 0{,}1 * X_5) =$$

$$= 0{,}1 * E(X_1) + 0{,}2 * E(X_2) + 0{,}4 * E(X_3) + 0{,}2 * E(X_4) + 0{,}1 * E(X_5) =$$

$$= 0{,}1 * \mu + 0{,}2 * \mu + 0{,}4 * \mu + 0{,}2 * \mu + 0{,}1 * \mu = \mu$$

Possiamo, dunque, concludere che T_1 e T_2 sono entrambi stimatori corretti per μ.

b)

Rispondiamo a tale quesito calcolando l'efficienza relativa di T_1 e T_2:

$$\text{eff}(T_2 \mid T_1) = \frac{EQM(T_2)}{EQM(T_1)}$$

quindi calcolando $EQM(T_1) = \sigma^2_{T_1}$ ed $EQM(T_2) = \sigma^2_{T_2}$, essendo le loro distorsioni entrambe nulle.

Dal momento che T_1 è la media aritmetica campionaria: $\sigma^2_{T_1} = \dfrac{\sigma^2}{n}$

$$EQM(T_1) = \sigma^2_{T_1} = \frac{\sigma^2}{5} = 0{,}2\sigma^2$$

Per quanto riguarda, invece, T_2:

$$EQM\left(T_2\right) = \sigma^2_{T_2} = Var\left(T_2\right) =$$
$$Var\left(0,1 * X_1 + 0,2 * X_2 + 0,4 * X_3 + 0,2 * X_4 + 0,1 * X_5\right) =$$

$$= 0,1^2 * Var\left(X_1\right) + 0,2^2 * Var\left(X_2\right) + 0,4^2 * Var\left(X_3\right) + 0,2^2 * Var\left(X_4\right) + 0,1^2 * Var\left(X_5\right) =$$

$$= 0,01^2\sigma^2 + 0,04^2\sigma^2 + 0,16^2\sigma^2 + 0,04^2\sigma^2 + 0,01^2\sigma^2 = 0,26\sigma^2$$

Quindi:

$$eff\left(T_2 \mid T_1\right) = \frac{\sigma^2_{T_2}}{\sigma^2_{T_1}} = \frac{0,26\sigma^2}{0,2\sigma^2} = 1,3 > 1$$

Questo vuol dire che T_1 è più efficiente di T_2.

ESERCIZIO 5.4

Sia $(X_1, X_2, \ldots, X_n)$ un campione casuale generato da una v.c. $X \sim f(X; \theta)$ per la quale è noto che $E(X) = \theta$ e $Var(X) = 1$. Si consideri lo stimatore:

$$T = \sum_i (X_i - i)$$

a) si verifichi se T è corretto per θ;
b) si calcoli l'errore quadratico medio di T.

<u>Soluzione</u>

a)

$$E(T) = E\left(\sum_i (X_i - i)\right) = E\left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} i\right) = E\left(\sum_{i=1}^{n} X_i\right) - E\left(\sum_{i=1}^{n} i\right) = n\theta - \frac{n(n+1)}{2}$$

T è distorto e la sua distorsione è uguale a:

$$d = E(T) - \theta = n\theta - \frac{n(n+1)}{2} - \theta = (n-1)\theta - \frac{n(n+1)}{2}$$

b)

Per il calcolo dell'EQM è necessaria la varianza di T:

$$Var(T) = Var\left(\sum_i (X_i - i)\right) = Var\left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} i\right) = Var\left(\sum_{i=1}^{n} X_i\right) - Var\left(\sum_{i=1}^{n} i\right) = n\sigma^2 - 0 = n$$

perché $\sigma^2 = 1$.
Per cui:

$$EQM(T) = Var(T) + d^2 = n + \left[(n-1)\theta - \frac{n(n+1)}{2}\right]^2$$

ESERCIZIO 5.5

Sia (X_1, X_2, …, X_n) un campione casuale generato da una v.c. $X \sim B(n; \pi)$. Si consideri lo stimatore di π:

$$T = \sum X_i$$

a) stabilire se T è uno stimatore corretto per π;
b) calcolare l'errore quadratico medio di T.

Soluzione

a)
Ricordando che la v.c. Binomiale ha valore atteso $E(X) = n\,\pi$ e varianza $Var(X) = n\,\pi(1-\pi)$, si ha che:

$$E(T) = E\left(\sum X_i\right) = \sum E(X_i) = n\pi$$

Bisogna dedurre che T non è uno stimatore corretto per π, e presenta una distorsione pari a:

$$d = \left[E(T) - \pi\right] = n\pi - \pi = \pi(n-1)$$

Inoltre considerando che $n\,\pi$ è proprio il valore atteso di X, si può dedurre che T sarebbe uno stimatore corretto della media di X (ossia del numero medio di successi della v.c. Binomiale).

b)
Per il calcolo dell'errore quadratico medio di T è necessario conoscere la sua varianza:

$$Var(T) = Var\left(\sum X_i\right) = \sum Var(X_i) = n\left[n\pi(1-\pi)\right] = n^2\pi(1-\pi)$$

Quindi:

$$EQM(T) = Var(T) + d^2 = n^2\pi(1-\pi) + \pi^2(n-1)^2 = n^2\pi(1-\pi) + \pi^2\left(n^2 + 1 - 2n\right) =$$

$$n^2\pi - n^2\pi^2 + n^2\pi^2 + \pi^2 - 2n\pi^2 = \pi^2 + n^2\pi - 2n\pi^2 = \pi^2\left(1 - 2n + \frac{n^2}{\pi}\right)$$