

Statistica II - Esercitazione 1

Stima puntuale: criteri di valutazione degli stimatori

Sia (X_1, \dots, X_n) un campione casuale a componenti i.i.d. ad una variabile casuale X avente legge di distribuzione $\varphi_X(x; \theta)$ con supporto \mathcal{X} ed indicizzata da un parametro θ di ignoto valore in uno spazio parametrico Θ .

Il nostro primo obiettivo è scegliere un valore di θ che 'meglio' di altri spieghi il motivo per cui sono stati osservati i valori (x_1, \dots, x_n) anziché altri; un tale valore di θ si chiamerà stima di θ .

Un procedimento di stima fa corrispondere ad ogni (x_1, \dots, x_n) un valore in Θ ; si tratta quindi di una funzione dallo spazio campionario $\mathcal{X} \times \dots \times \mathcal{X}$ in Θ , cioè di una statistica $T = T(X_1, \dots, X_n)$ che nella fattispecie prende il nome di stimatore di θ .

Ciò premesso, è desiderabile che uno stimatore per θ assuma valori prossimi al reale e ignoto valore di θ ; alla luce di ciò, un criterio atto alla valutazione dell'affidabilità di uno stimatore può essere intuitivamente basato sulla distanza di esso da θ , che deve essere opportunamente 'piccola'. Dato dunque uno stimatore T per θ , si consideri come distanza di T da θ il quadrato dello scarto tra i due:

$$(T - \theta)^2;$$

tale oggetto, poiché funzione del campione casuale X_1, \dots, X_n tramite T , è esso stesso una variabile aleatoria; per i fini all'oggetto, si rende quindi necessario considerare una caratteristica della relativa legge di distribuzione, ad esempio il valore atteso.

Errore Quadratico Medio

L'Errore Quadratico Medio di uno stimatore T per θ è definito come:

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

ove il valore atteso è inteso rispetto alla legge di distribuzione $\varphi_T(t; \theta)$ dello stimatore T .

Un'importante proprietà dell'Errore Quadratico Medio è la seguente:

$$\text{MSE}(T) = \text{Var}(T) + [E(T) - \theta]^2,$$

ove la quantità $B(T) = E(T) - \theta$ è detta distorsione (*bias* in inglese) dello stimatore T per θ ; pertanto:

$$\text{MSE}(T) = \text{Var}(T) + [B(T)]^2.$$

Correttezza

Uno stimatore T per θ è detto corretto o non distorto per θ se:

$$E(T) = \theta, \quad \forall \theta \in \Theta.$$

Più in generale, uno stimatore T per $\tau(\theta)$ è detto corretto o non distorto per $\tau(\theta)$ se:

$$E(T) = \tau(\theta), \quad \forall \theta \in \Theta,$$

dove $\tau(\cdot) : \Theta \mapsto \mathbb{R}$ è una funzione di θ .

Ai fini della minimizzazione dell'Errore Quadratico Medio di uno stimatore, teoricamente è necessario minimizzarne contemporaneamente varianza e distorsione; nella pratica, però, ciò non è affatto semplice. Il problema in questione è noto nella letteratura inglese come bias-variance trade-off e consiste nel fatto che minimizzando eccessivamente la distorsione si può avere come effetto un eccessivo aumento della varianza e viceversa. Tuttavia, in molti casi, un piccolo aumento nella distorsione di uno stimatore può comportare una grande diminuzione della variabilità dello stesso.

Efficienza relativa

Ai fini del confronto di due qualsivoglia stimatori di un medesimo parametro, è possibile ricorrere al rapporto dei rispettivi errori quadratici medi, ossia alla cosiddetta efficienza relativa; pertanto, considerati gli stimatori T_1 e T_2 per θ , T_1 è detto più efficiente di T_2 se:

$$\frac{\text{MSE}(T_1)}{\text{MSE}(T_2)} \leq 1, \quad \forall \theta \in \Theta$$

e

$$\exists \text{ almeno un } \theta \in \Theta : \frac{\text{MSE}(T_1)}{\text{MSE}(T_2)} < 1.$$

Non sempre risulta che uno stimatore sia più efficiente di un altro, e quindi preferibile. Infatti l'efficienza relativa può presentare al variare di $\theta \in \Theta$ valori sia inferiori a 1 sia superiori a 1, risultando $\text{MSE}(T_1) < \text{MSE}(T_2)$ per alcuni valori di $\theta \in \Theta$ e $\text{MSE}(T_1) > \text{MSE}(T_2)$ per altri.

Esercizio 1

Siano X_1, \dots, X_n i.i.d. ad una variabile casuale discreta X caratterizzata da legge di distribuzione:

$$\Pr(X = -1) = \frac{\theta}{2}, \quad \Pr(X = 0) = 1 - \theta, \quad \Pr(X = 1) = \frac{\theta}{2}.$$

1. Si determini lo spazio parametrico;
2. si calcolino media e varianza di X .

Si considerino ora i due seguenti stimatori di θ :

$$T_1 = \frac{1}{n} \sum_{i=1}^n |X_i|, \quad T_2 = X_n^2.$$

3. Si stabilisca se T_1 e T_2 sono distorti o meno per θ ;
4. si calcolino gli errori quadratici medi di T_1 e T_2 ;
5. studiando l'efficienza relativa dei due stimatori, si dica quale tra i due è da preferirsi.

Soluzione

Sia:

$$X \sim \varphi_X(x; \theta) = \Pr(X = x) = \begin{cases} \frac{\theta}{2} & \text{se } x \in \{-1, 1\} \\ 1 - \theta & \text{se } x = 0 \end{cases},$$

con $\theta \in \Theta$ ignoto e X_1, \dots, X_n i.i.d. a X . Trattasi di un problema di stima puntuale per θ .

1. Lo spazio parametrico Θ è costituito dai valori di θ tali da garantire che quanto fornito dal testo dell'esercizio sia effettivamente la funzione di probabilità di una variabile aleatoria discreta. Trattasi dunque di verificare per quali valori di θ siano soddisfatte le due seguenti condizioni:

$$\Pr(X = x) \geq 0, \quad \forall x \in \{-1, 0, 1\},$$

$$\sum_x \Pr(X = x) = \Pr(X = -1) + \Pr(X = 0) + \Pr(X = 1) = 1.$$

La prima condizione è verificata per $\{0 \leq \theta/2 \leq 1\} \cap \{0 \leq 1 - \theta \leq 1\}$, ovvero per $\{0 \leq \theta \leq 2\} \cap \{0 \leq \theta \leq 1\} = \{0 \leq \theta \leq 1\}$, mentre la seconda condizione è sempre verificata. In definitiva, lo spazio parametrico si identifica in $\Theta = [0, 1]$.

2. La media di X risulta:

$$\begin{aligned} E(X) &= \sum_x x \cdot \Pr(X = x) = \\ &= -1 \cdot \Pr(X = -1) + 0 \cdot \Pr(X = 0) + 1 \cdot \Pr(X = 1) = \\ &= -\frac{\theta}{2} + \frac{\theta}{2} = 0; \end{aligned}$$

mentre, la varianza di X risulta:

$$\begin{aligned} \text{Var}(X) &= E(X^2) - [E(X)]^2 = E(X^2) = \sum_x x^2 \cdot \Pr(X = x) = \\ &= \frac{\theta}{2} + \frac{\theta}{2} = \theta. \end{aligned}$$

3. Essendo X_1, \dots, X_n i.i.d. a X , altresì $|X_1|, \dots, |X_n|$ sono i.i.d. a $|X|$ e, stante la formula del valore atteso di una trasformazione di una variabile casuale, risultano, $\forall \theta \in \Theta$:

$$E(T_1) \stackrel{\text{lin.}}{=} \frac{1}{n} \sum_{i=1}^n E(|X_i|) \stackrel{\text{i.d.}}{=} E(|X|) = \sum_x |x| \cdot \Pr(X = x) = \frac{\theta}{2} + \frac{\theta}{2} = \theta,$$

$$E(T_2) = E(X_n^2) \stackrel{\text{i.d.}}{=} E(X^2) = \theta;$$

i due stimatori in questione sono quindi ambedue non distorti per θ .

4. In virtù della non distorsione di T_1 per θ appena mostrata, risulta:

$$\begin{aligned} \text{MSE}(T_1) &= \text{Var}(T_1) \stackrel{\text{ind.}}{=} \frac{1}{n^2} \sum_{i=1}^n \text{Var}(|X_i|) \stackrel{\text{i.d.}}{=} \frac{1}{n} \text{Var}(|X|) = \\ &= \frac{1}{n} \{E(|X|^2) - [E(|X|)]^2\} = \frac{1}{n} [E(X^2) - \theta^2] = \frac{\theta(1-\theta)}{n}. \end{aligned}$$

In maniera assolutamente analoga, risulta:

$$\begin{aligned} \text{MSE}(T_2) &= \text{Var}(T_2) = \text{Var}(X_n^2) \stackrel{\text{i.d.}}{=} \text{Var}(X^2) = \\ &= E[(X^2)^2] - [E(X^2)]^2 = E(X^4) - \theta^2 = \theta(1-\theta), \end{aligned}$$

5. L'efficienza relativa di T_1 rispetto T_2 risulta:

$$\frac{\text{MSE}(T_1)}{\text{MSE}(T_2)} = \frac{\text{Var}(T_1)}{\text{Var}(T_2)} = \frac{1}{n} < 1 \quad \forall n > 1,$$

dunque T_1 è da preferirsi a T_2 .

Esercizio 2

Si supponga che una grandezza incognita μ sia stata misurata più volte tramite due strumenti caratterizzati da precisioni note e differenti. Le variabili casuali X_1, \dots, X_m , i.i.d. a $X \sim N(\mu, \sigma_X^2)$ con σ_X^2 nota e Y_1, \dots, Y_n , i.i.d. a $Y \sim N(\mu, \sigma_Y^2)$ con σ_Y^2 nota, interpretino rispettivamente le m e le n misurazioni della grandezza di interesse realizzate con i due strumenti in questione. Ai fini della stima di μ , si consideri dunque lo stimatore:

$$T = a\bar{X}_m + (1-a)\bar{Y}_n, \quad a \in \mathbb{R}.$$

1. Si calcolino media e varianza di T ;
2. si determini il valore di a che minimizzi l'errore quadratico medio di T ;
3. fissato a nel valore trovato al punto precedente, si identifichi la distribuzione di T .

Soluzione

Siano: $X \sim N(\mu, \sigma_X^2)$ con σ_X^2 nota e $Y \sim N(\mu, \sigma_Y^2)$ con σ_Y^2 nota, $\sigma_X^2 \neq \sigma_Y^2$, $\mu \in \mathbb{R}$ ignota. Dal testo dell'esercizio è ragionevole assumere che X e Y siano tra loro indipendenti. Siano inoltre: X_1, \dots, X_m i.i.d. a X e Y_1, \dots, Y_n i.i.d. a Y . Trattasi di un problema di stima puntuale per θ .

1. La media di T risulta:

$$\begin{aligned} E(T) &\stackrel{\text{lin.}}{=} aE(\bar{X}_m) + (1-a)E(\bar{Y}_n) \stackrel{\text{i.d.}}{=} aE(X) + (1-a)E(Y) = \\ &= a\mu + (1-a)\mu = \mu; \end{aligned}$$

essendo X e Y indipendenti, altresì \bar{X}_m e \bar{Y}_n sono indipendenti e la varianza di T risulta:

$$\begin{aligned} \text{Var}(T) &\stackrel{X,Y \text{ ind.}}{=} a^2 \text{Var}(\bar{X}_m) + (1-a)^2 \text{Var}(\bar{Y}_n) = \\ &\stackrel{\text{i.i.d.}}{=} a^2 \frac{\text{Var}(X)}{m} + (1-a)^2 \frac{\text{Var}(Y)}{n} = \\ &= a^2 \frac{\sigma_X^2}{m} + (1-a)^2 \frac{\sigma_Y^2}{n} = \\ &= a^2 \left(\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n} \right) - 2a \frac{\sigma_Y^2}{n} + \frac{\sigma_Y^2}{n}. \end{aligned}$$

2. Essendo T corretto per μ :

$$\text{MSE}(T) = \text{Var}(T) = a^2 \left(\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n} \right) - 2a \frac{\sigma_Y^2}{n} + \frac{\sigma_Y^2}{n} = f(a)$$

sia la funzione obiettivo di interesse. Condizione necessaria affinché \tilde{a} sia punto di minimo per $f(a)$ è che \tilde{a} sia punto di stazionarietà di f , ossia soluzione dell'equazione:

$$f'(a) = 0 \iff 2a \left(\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n} \right) - 2 \frac{\sigma_Y^2}{n} = 0,$$

da cui risulta

$$\tilde{a} = \frac{\frac{\sigma_Y^2}{n}}{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}.$$

Condizione sufficiente affinché \tilde{a} sia punto di minimo per f è che $f''(\tilde{a}) > 0$; poiché:

$$f''(a) = 2 \left(\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n} \right) > 0, \quad \forall a \in \mathbb{R},$$

\tilde{a} minimizza l'errore quadratico medio di T ed il relativo valore minimo risulta:

$$\begin{aligned} f(\tilde{a}) &= \min_a f(a) = \min_a \text{MSE}(T) = \\ &= \left(\frac{\frac{\sigma_Y^2}{n}}{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}} \right)^2 \left(\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n} \right) - 2 \frac{\frac{\sigma_Y^2}{n}}{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}} \cdot \frac{\sigma_Y^2}{n} + \frac{\sigma_Y^2}{n} = \\ &= \frac{\frac{\sigma_X^2}{m} \frac{\sigma_Y^2}{n}}{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}. \end{aligned}$$

Alternativamente, si noti che $f(a) = ka^2 + ba + c$, ove $k = \left(\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n} \right) > 0$, $b = -2 \frac{\sigma_Y^2}{n}$ e $c = \frac{\sigma_Y^2}{n}$, è l'equazione di una parabola avente concavità rivolta verso l'alto e vertice, ossia punto di minimo, di coordinate:

$$\left(-\frac{b}{2k}, -\frac{b^2 - 4kc}{4k} \right) = \left(\frac{\frac{\sigma_Y^2}{n}}{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}, \frac{\frac{\sigma_X^2}{m} \frac{\sigma_Y^2}{n}}{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}} \right),$$

con ascissa pari ad $\tilde{a} = \text{argmin}_a \text{MSE}(T)$ ed ordinata pari al valore minimo di $\text{MSE}(T)$.

3. Poiché combinazione lineare di variabili casuali Normali tra loro indipendenti, T ha anch'esso distribuzione Normale; nella fattispecie, per quanto determinato nei punti precedenti:

$$T \sim N \left(\mu, \frac{\frac{\sigma_X^2}{m} \frac{\sigma_Y^2}{n}}{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}} \right).$$

Esercizio 3

Sia X una variabile casuale di Poisson di parametro $\lambda \in \mathbb{R}^+$ e siano $T_1 = X$ e $T_2 = 1$ due stimatori per λ . Si dica per quali valori di λ T_2 è preferibile a T_1 .

Soluzione

Sia $X \sim \text{Poisson}(\lambda)$ con $\lambda \in \mathbb{R}^+$ ignoto. Trattasi di un problema di stima puntuale per λ .

Poichè $E(T_1) = E(X) = \lambda$, $\forall \lambda \in \mathbb{R}^+$, T_1 è stimatore non distorto per λ ; al contrario, essendo $E(T_2) = 1 \neq \lambda$, $\forall \lambda \in \mathbb{R}^+ - \{1\}$, T_2 è distorto per λ . Conseguentemente, gli errori quadratici medi di T_1 e T_2 risultano:

$$\begin{aligned} \text{MSE}(T_1) &= \text{Var}(T_1) = \text{Var}(X) = \lambda, \\ \text{MSE}(T_2) &= \text{Var}(T_2) + [B(T_2)]^2 = 0 + (1 - \lambda)^2 = 1 + \lambda^2 - 2\lambda; \end{aligned}$$

pertanto, T_2 è preferibile a T_1 se e solo se:

$$\frac{\text{MSE}(T_2)}{\text{MSE}(T_1)} < 1 \iff \lambda^2 - 3\lambda + 1 < 0.$$

In generale, la disequazione di II grado $a\lambda^2 + b\lambda + c < 0$ risulta soddisfatta $\forall \lambda \in \left(\frac{-b - \sqrt{b^2 - 4ac}}{2a}, \frac{-b + \sqrt{b^2 - 4ac}}{2a} \right)$, ossia nell'intervallo reale delimitato dalle radici dell'equazione di II grado $a\lambda^2 + b\lambda + c = 0$, che, nel presente caso, risultano:

$$\lambda_{1,2} = \frac{3 \mp \sqrt{5}}{2} = \begin{cases} 0.382 \\ 2.618 \end{cases}.$$

In definitiva, T_2 è da preferirsi a T_1 se e solo se $0.382 < \lambda < 2.618$.

Esercizio 4

Siano X_1, \dots, X_n i.i.d. secondo una qualche legge di distribuzione di media μ e varianza σ^2 :

1. si mostri che $\sum_{i=1}^n a_i X_i$ è stimatore non distorto per μ per ogni insieme di costanti $\{a_1, \dots, a_n\}$ tale che

$$\sum_{i=1}^n a_i = 1;$$

2. si mostri che, nel caso in cui $\sum_{i=1}^n a_i = 1$, $\text{Var} \left(\sum_{i=1}^n a_i X_i \right)$ è minima per $a_i = \frac{1}{n}$, $\forall i = 1, \dots, n$.

Soluzione

Sia X variabile casuale con varianza $\sigma^2 \in \mathbb{R}^+$ ignota e media $\mu \in \mathbb{R}$ ignota ed oggetto dell'interesse inferenziale. Siano inoltre: X_1, \dots, X_n i.i.d. a X . Trattasi di un problema di stima puntuale per μ .

1. Risulta:

$$\mathbb{E} \left(\sum_{i=1}^n a_i X_i \right) \stackrel{\text{lin.}}{=} \sum_{i=1}^n a_i \mathbb{E}(X_i) \stackrel{\text{i.d.}}{=} \sum_{i=1}^n a_i \mu = \mu \sum_{i=1}^n a_i, \quad \forall \mu \in \mathbb{R};$$

pertanto, lo stimatore in questione è non distorto per μ se e solo se l'insieme di costanti $\{a_1, \dots, a_n\}$ è tale che $\sum_{i=1}^n a_i = 1$.

2. Aggiungendo e togliendo $\frac{1}{n}$ ad a_i , $\forall i = 1, \dots, n$, risulta:

$$\begin{aligned} \text{Var} \left(\sum_{i=1}^n a_i X_i \right) &\stackrel{\text{ind.}}{=} \sum_{i=1}^n a_i^2 \text{Var}(X_i) \stackrel{\text{i.d.}}{=} \sum_{i=1}^n a_i^2 \sigma^2 = \sigma^2 \sum_{i=1}^n a_i^2 = \\ &\stackrel{\pm 1/n}{=} \sigma^2 \sum_{i=1}^n \left[\left(a_i - \frac{1}{n} \right) + \frac{1}{n} \right]^2 = \\ &= \sigma^2 \sum_{i=1}^n \left[\left(a_i - \frac{1}{n} \right)^2 + 2 \left(a_i - \frac{1}{n} \right) \frac{1}{n} + \frac{1}{n^2} \right] = \\ &= \sigma^2 \left[\sum_{i=1}^n \left(a_i - \frac{1}{n} \right)^2 + \frac{2}{n} \sum_{i=1}^n \left(a_i - \frac{1}{n} \right) + n \frac{1}{n^2} \right] = \\ &= \sigma^2 \left[\sum_{i=1}^n \left(a_i - \frac{1}{n} \right)^2 + \frac{2}{n} \underbrace{\sum_{i=1}^n a_i}_{=1} - \frac{2}{n} + \frac{1}{n} \right] = \\ &= \sigma^2 \left[\sum_{i=1}^n \left(a_i - \frac{1}{n} \right)^2 + \frac{1}{n} \right]; \end{aligned}$$

pertanto, lo stimatore in questione ha varianza minima e pari a $\frac{\sigma^2}{n}$ se e solo se $a_i = 1/n$, $\forall i = 1, \dots, n$.

Esercizio 5

Siano X_1, \dots, X_n i.i.d. a $X \sim \text{Bernoulli}(\theta)$. Ai fini della stima di θ , si considerino i due stimatori $T_1 = \bar{X}$ e $T_2 = \frac{\sum_{i=1}^n X_i + \sqrt{n/4}}{n + \sqrt{n}}$ e li si confronti in termini di errore quadratico medio per $n = 4$ e $n = 400$.

Soluzione

Sia $X \sim \text{Bernoulli}(\theta)$, $\theta \in [0, 1]$ ignoto; siano inoltre X_1, \dots, X_n i.i.d. a X . Trattasi di un problema di stima puntuale per θ .

Si noti che T_1 è stimatore non distorto per θ , infatti:

$$E(T_1) = E(\bar{X}) \stackrel{\text{lin.}}{=} \frac{1}{n} \sum_{i=1}^n E(X_i) \stackrel{\text{i.d.}}{=} \theta, \quad \forall \theta \in [0, 1];$$

pertanto:

$$\text{MSE}(T_1) = \text{Var}(T_1) \stackrel{\text{i.d.}}{=} \frac{\theta(1-\theta)}{n}.$$

Al contrario, T_2 è distorto per θ , infatti:

$$E(T_2) \stackrel{\text{lin.}}{=} \frac{\sum_{i=1}^n E(X_i) + \sqrt{n/4}}{n + \sqrt{n}} \stackrel{\text{i.d.}}{=} \frac{n\theta + \frac{\sqrt{n}}{2}}{n + \sqrt{n}} \neq \theta, \quad \forall \theta \in [0, 1],$$

con distorsione:

$$B(T_2) = \frac{n\theta + \frac{\sqrt{n}}{2}}{n + \sqrt{n}} - \theta = \frac{\sqrt{n}(1-2\theta)}{2(n + \sqrt{n})}.$$

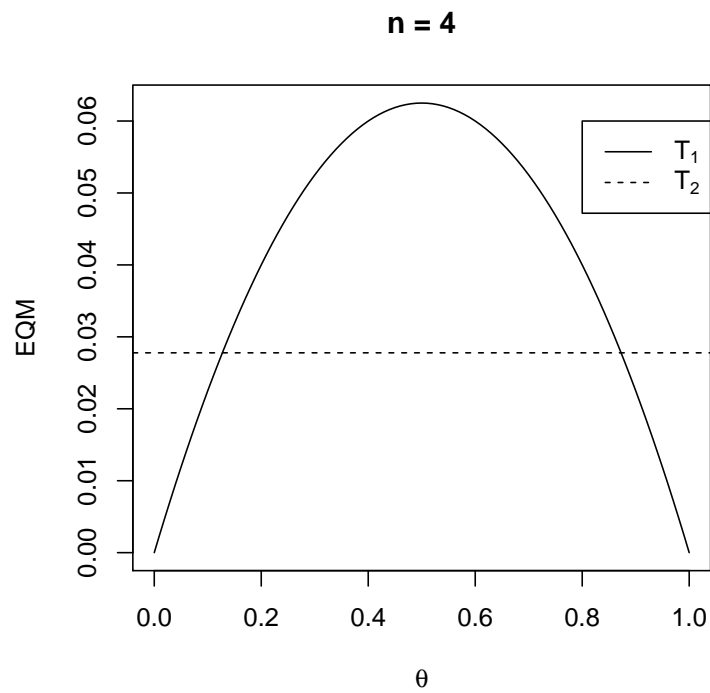
Essendo:

$$\text{Var}(T_2) \stackrel{\text{ind.}}{=} \frac{1}{(n + \sqrt{n})^2} \sum_{i=1}^n \text{Var}(X_i) \stackrel{\text{i.d.}}{=} \frac{1}{(n + \sqrt{n})^2} n\theta(1-\theta),$$

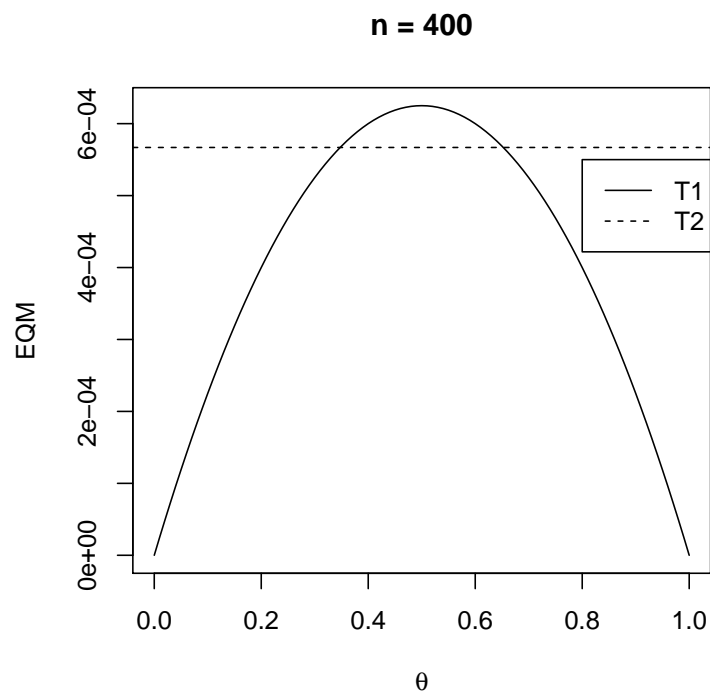
risulta:

$$\begin{aligned} \text{MSE}(T_2) &= \text{Var}(T_2) + [B(T_2)]^2 = \\ &= \frac{1}{(n + \sqrt{n})^2} n\theta(1-\theta) + \frac{n(1-2\theta)^2}{4(n + \sqrt{n})^2} = \\ &= \frac{n}{4(n + \sqrt{n})^2} [4\theta(1-\theta) + (1-2\theta)^2] = \\ &= \frac{n}{4(n + \sqrt{n})^2}. \end{aligned}$$

Fissato $n = 4$, risultano $\text{MSE}(T_1) = \frac{\theta(1-\theta)}{4}$ e $\text{MSE}(T_2) = \frac{1}{36}$ e dalla rappresentazione grafica dei relativi andamenti in funzione di θ , proposta nella Figura seguente, trae evidenza che per n piccolo T_2 è preferibile a T_1 a meno che θ sia prossimo a 0 o a 1.



Viceversa, fissato $n = 400$, risultano $\text{MSE}(T_1) = \frac{\theta(1-\theta)}{400}$ e $\text{MSE}(T_2) = 0.000567$ e dalla rappresentazione grafica dei relativi andamenti in funzione di θ , proposta nella Figura seguente, trae evidenza che per n grande T_1 è preferibile a T_2 a meno che θ sia prossimo a $\frac{1}{2}$.



Esercizio 6

Date X_1, \dots, X_n i.i.d. a $X \sim N(\mu, \sigma^2)$, $n > 1$, si consideri, ai fini della stima di σ^2 con μ ignota, lo stimatore

$$T_k = kS^2$$

con $k > 0$ ed $S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$ lo stimatore varianza campionaria corretta. Si identifichi il valore di b che minimizza $\text{MSE}(T_k)$.

Soluzione

Trattasi di un problema di stima puntuale per σ^2 . Abbiamo:

$$\begin{aligned} \text{MSE}(T_k) &= \text{Var}(kS^2) + [\text{E}(kS^2) - \sigma^2]^2 \\ &= k^2 \text{Var}(S^2) + [\text{E}(kS^2) - \sigma^2]^2 \\ &= \frac{k^2 2\sigma^4}{n-1} + (k-1)^2 \sigma^4 \\ &= \left[\frac{2k^2}{n-1} + (k-1)^2 \right] \sigma^4 \end{aligned}$$

che non dipende da μ ed è una funzione quadratica di σ^2 . Quindi dobbiamo trovare il valore di k che minimizza la funzione

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

Notando che $f(k)$ è equazione di una parabola avente concavità verso l'alto, il punto di minimo corrisponde a

$$k = \left(\frac{n-1}{n+1} \right)$$

quindi lo stimatore che minimizza $\text{MSE}(T_k)$ è

$$\frac{n-1}{n+1} S^2 = \frac{1}{n+1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Si noti infine che $\text{MSE}(S^2) = \frac{2}{n-1} \sigma^4 > \frac{2n-1}{n^2} \sigma^4 = \text{MSE} \left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right)$.

Esercizio 7

Siano X_1, \dots, X_n i.i.d. a $X \sim N(\mu, \sigma^2)$.

1. Sia $\sigma^2 = 1$. Ai fini della stima di $\tau(\mu) = \mu^2$, si consideri lo stimatore

$$T_1 = \bar{X}^2 - a.$$

Si determini il valore $a \in \mathbb{R}$ tale che lo stimatore T_1 sia corretto per $\tau(\mu) = \mu^2$.

2. Si consideri, ai fini della stima dello scarto quadratico medio σ con μ ignota, lo stimatore :

$$T_2 = a\sqrt{S^2} = a\sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}.$$

Si determini il valore di $a \in \mathbb{R}$ tale che lo stimatore T_2 sia corretto per σ .

Soluzione

1. Abbiamo

$$E(T_1) = E(\bar{X}^2 - a) = \text{Var}(\bar{X}) + [E(\bar{X})]^2 - a = \frac{1}{n} + \mu^2 - a$$

quindi lo stimatore T_1 è corretto per $\tau(\mu) = \mu^2$ per $a = \frac{1}{n}$.

2. Abbiamo

$$E(T_2) = E(a\sqrt{S^2}) = a\sqrt{\frac{\sigma^2}{n-1}} E\left(\sqrt{\frac{S^2(n-1)}{\sigma^2}}\right) = a\sqrt{\frac{\sigma^2}{n-1}} E(\sqrt{Q})$$

dove $Q = \frac{S^2(n-1)}{\sigma^2}$ è una v.c. χ_{n-1}^2 , quindi

$$E(\sqrt{Q}) = \int_0^\infty \sqrt{q} \varphi_Q(q; n-1) dq = \int_0^\infty \sqrt{q} \frac{q^{\frac{n-1}{2}-1} e^{-q/2} \left(\frac{1}{2}\right)^{(n-1)/2}}{\Gamma(\frac{n-1}{2})} dq$$

Riarrangiando i termini per ottenere la funzione di densità di una χ_n^2 , otteniamo

$$E(\sqrt{Q}) = \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2}) \left(\frac{1}{2}\right)^{1/2}} \underbrace{\int_0^\infty \frac{q^{\frac{n}{2}-1} e^{-q/2} \left(\frac{1}{2}\right)^{n/2}}{\Gamma(\frac{n}{2})} dq}_{=1}$$

quindi $a = \frac{\Gamma(\frac{n-1}{2})\sqrt{n-1}}{\sqrt{2}\Gamma(n/2)}$.

Si poteva subito concludere che lo stimatore $\sqrt{S^2}$ sottostima σ (e quindi è stimatore distorto per σ) poiché segue dalla disuguaglianza di Jensen $E[f(X)] < f[E(X)]$ per funzioni strettamente concave, ovvero

$$E(\sqrt{S^2}) < \sqrt{E(S^2)} = \sqrt{\sigma^2} = \sigma$$

con $f(x) = \sqrt{x}$ funzione strettamente concava sul semiasse positivo.

Esercizio 8

Si supponga di effettuare $n = 3$ prove indipendenti nella variabile casuale X tale che $E(X) = \mu$ e $\text{Var}(X) = \mu^2$, $\mu \in \mathbb{R} - \{0\}$. Ai fini della stima di μ , si considerino i due stimatori:

$$T_1 = aX_1 + \frac{1}{2}X_2 + bX_3, \quad T_2 = bX_1 + \frac{1}{3}X_2 - aX_3$$

con $a, b \in \mathbb{R}$.

1. Si determinino i valori di a e b che garantiscano la non distorsione di T_1 e T_2 per μ ;
2. posti $a = \frac{1}{2}$ e $b = \frac{1}{3}$, si indichi quale tra i due stimatori è da preferirsi.

Soluzione

Sia X variabile casuale con varianza $\mu^2 \in \mathbb{R}^+$ ignota e media $\mu \in \mathbb{R} - \{0\}$ ignota ed oggetto dell'interesse inferenziale. Siano inoltre: X_1, X_2, X_3 i.i.d. a X . Trattasi di un problema di stima puntuale per μ .

1. Risultano:

$$\begin{aligned} E(T_1) &\stackrel{\text{lin.}}{=} aE(X_1) + \frac{1}{2}E(X_2) + bE(X_3) \stackrel{\text{i.d.}}{=} a\mu + \frac{1}{2}\mu + b\mu, \\ E(T_2) &\stackrel{\text{lin.}}{=} bE(X_1) + \frac{1}{3}E(X_2) - aE(X_3) \stackrel{\text{i.d.}}{=} b\mu + \frac{1}{3}\mu - a\mu; \end{aligned}$$

pertanto, al fine di garantire la proprietà di correttezza per μ di T_1 e T_2 , deve essere $\forall \mu \in \mathbb{R} - \{0\}$:

$$\begin{cases} E(T_1) = \mu \\ E(T_2) = \mu \end{cases} \iff \begin{cases} a\mu + \frac{1}{2}\mu + b\mu = \mu \\ b\mu + \frac{1}{3}\mu - a\mu = \mu \end{cases} \iff \begin{cases} a + b = \frac{1}{2} \\ b - a = \frac{2}{3} \end{cases} \iff \begin{cases} a = -\frac{1}{12} \\ b = \frac{7}{12} \end{cases}.$$

In definitiva:

$$T_1 = -\frac{1}{12}X_1 + \frac{1}{2}X_2 + \frac{7}{12}X_3, \quad T_2 = \frac{7}{12}X_1 + \frac{1}{3}X_2 + \frac{1}{12}X_3$$

sono gli stimatori corretti per μ richiesti.

2. Per quanto determinato al punto precedente, posti $a = \frac{1}{2}$ e $b = \frac{1}{3}$, i due stimatori per μ :

$$T_1 = \frac{1}{2}X_1 + \frac{1}{2}X_2 + \frac{1}{3}X_3, \quad T_2 = \frac{1}{3}X_1 + \frac{1}{3}X_2 - \frac{1}{2}X_3$$

risultano distorti per μ , infatti, $\forall \mu \in \mathbb{R} - \{0\}$, si ha che:

$$\begin{aligned} E(T_1) &= \left(\frac{1}{2} + \frac{1}{2} + \frac{1}{3}\right)\mu = \frac{4}{3}\mu \neq \mu, \\ E(T_2) &= \left(\frac{1}{3} + \frac{1}{3} - \frac{1}{2}\right)\mu = \frac{1}{6}\mu \neq \mu; \end{aligned}$$

sarà dunque da preferirsi lo stimatore, se esiste, caratterizzato da errore quadratico medio inferiore $\forall \mu \in \mathbb{R} - \{0\}$. A tal proposito, risultano:

$$[B(T_1)]^2 = \left(\frac{4}{3}\mu - \mu\right)^2 = \frac{1}{9}\mu^2, \quad [B(T_2)]^2 = \left(\frac{1}{6}\mu - \mu\right)^2 = \frac{25}{36}\mu^2$$

e:

$$\text{Var}(T_1) \stackrel{\text{ind.}}{=} \frac{1}{4}\text{Var}(X_1) + \frac{1}{4}\text{Var}(X_2) + \frac{1}{9}\text{Var}(X_3) \stackrel{\text{i.d.}}{=} \frac{11}{18}\mu^2,$$

$$\text{Var}(T_2) \stackrel{\text{i.d.}}{=} \frac{1}{9}\text{Var}(X_1) + \frac{1}{9}\text{Var}(X_2) + \frac{1}{4}\text{Var}(X_3) \stackrel{\text{i.d.}}{=} \frac{17}{36}\mu^2;$$

pertanto:

$$\text{MSE}(T_1) = \text{Var}(T_1) + [\text{B}(T_1)]^2 = \frac{11}{18}\mu^2 + \frac{1}{9}\mu^2 = \frac{13}{18}\mu^2,$$

$$\text{MSE}(T_2) = \text{Var}(T_2) + [\text{B}(T_2)]^2 = \frac{17}{36}\mu^2 + \frac{25}{36}\mu^2 = \frac{7}{6}\mu^2.$$

Evidentemente:

$$\text{MSE}(T_1) < \text{MSE}(T_2), \quad \forall \mu \in \mathbb{R} - \{0\},$$

dunque T_1 è da preferirsi a T_2 .

Esercizio 9

Il tempo di risposta di un calcolatore all'input di un terminale può essere modellizzato mediante una v.c. esponenziale negativa di ignota media θ . Si supponga che in un esperimento vengano misurati n tempi di risposta T_1, T_2, \dots, T_n .

1. Si mostri che $\bar{T} = \frac{1}{n} \sum_{i=1}^n T_i$ è uno stimatore corretto per θ ;
2. si determini la legge di distribuzione di \bar{T} ;
3. si stabilisca se $\tilde{T} = 1/\bar{T}$ è uno stimatore corretto per $\tau(\theta) = \frac{1}{\theta}$;
4. si ricavi da \tilde{T} uno stimatore non distorto per $\tau(\theta) = \frac{1}{\theta}$ e se ne calcoli l'errore quadratico medio.

Soluzione

Sia $T \sim \text{Exp}\left(\frac{1}{\theta}\right)$, $\theta \in \mathbb{R}^+$ ignoto; siano inoltre T_1, \dots, T_n i.i.d. a T . Trattasi di un problema di stima puntuale per θ .

1. Lo stimatore $\bar{T} = \frac{1}{n} \sum_{i=1}^n T_i$ è sempre non distorto per la media della popolazione $E(T) = \theta$, infatti:

$$E(\bar{T}) \stackrel{\text{lin.}}{=} \frac{1}{n} \sum_{i=1}^n E(T_i) \stackrel{\text{i.d.}}{=} \frac{1}{n} n E(T) = \theta, \quad \forall \theta > 0.$$

2. Essendo $\text{Exp}\left(\frac{1}{\theta}\right) \equiv \text{Gamma}\left(1, \frac{1}{\theta}\right)$ ed in virtù della proprietà riproduttiva della variabile casuale Gamma rispetto al parametro di forma, si ha che:

$$\sum_{i=1}^n T_i \sim \text{Gamma}\left(n, \frac{1}{\theta}\right);$$

inoltre, essendo \bar{T} cambiamento di scala di $\sum_{i=1}^n T_i$ e $\frac{1}{\theta}$ parametro di scala per le variabili casuali in oggetto, si ha che:

$$\bar{T} = \frac{1}{n} \sum_{i=1}^n T_i \sim \text{Gamma}\left(n, \frac{1/\theta}{1/n}\right) \equiv \text{Gamma}\left(n, \frac{n}{\theta}\right).$$

3. Si noti che \tilde{T} è funzione di \bar{T} , di cui è nota dal punto precedente la legge di distribuzione. Pertanto:

$$\begin{aligned} E(\tilde{T}) &= E(1/\bar{T}) = \int_0^\infty \frac{1}{t} \frac{(n/\theta)^n}{\Gamma(n)} t^{n-1} e^{-\frac{n}{\theta}t} dt = \\ &= \left(\frac{n}{\theta}\right)^n \frac{1}{\Gamma(n)} \int_0^\infty t^{n-2} e^{-\frac{n}{\theta}t} dt; \end{aligned}$$

considerata ora la sostituzione $z = \frac{n}{\theta}t \Leftrightarrow t = \frac{\theta}{n}z$, che comporta $dt = \frac{\theta}{n}dz$, si ha che:

$$\begin{aligned} E(\tilde{T}) &= \left(\frac{n}{\theta}\right)^n \frac{1}{\Gamma(n)} \int_0^\infty \left(\frac{\theta}{n}z\right)^{n-2} e^{-z} \frac{\theta}{n} dz = \\ &= \left(\frac{n}{\theta}\right)^n \frac{1}{\Gamma(n)} \left(\frac{\theta}{n}\right)^{n-1} \int_0^\infty z^{(n-1)-1} e^{-z} dz = \\ &= \frac{n}{\theta} \frac{\Gamma(n-1)}{\Gamma(n)} = \frac{n}{(n-1)\theta} \neq \frac{1}{\theta}, \quad \forall \theta > 0, \end{aligned}$$

da cui risulta evidente che \tilde{T} è stimatore distorto per $\frac{1}{\theta}$.

4. Si consideri lo stimatore $a\tilde{T}$, $a \in \mathbb{R}^+$; si ha che:

$$E(a\tilde{T}) \stackrel{\text{lin.}}{=} aE(\tilde{T}) = a \frac{n}{n-1} \frac{1}{\theta};$$

ponendo $a = \frac{n-1}{n}$, lo stimatore:

$$U = \frac{n-1}{n} \tilde{T}$$

è non distorto per $1/\theta$. Si ha dunque che:

$$\begin{aligned} \text{MSE}(U) &= \text{Var}(U) = \left(\frac{n-1}{n}\right)^2 \text{Var}(\tilde{T}) = \\ &= \left(\frac{n-1}{n}\right)^2 \left\{ E(\tilde{T}^2) - [E(\tilde{T})]^2 \right\}, \end{aligned}$$

ove:

$$\begin{aligned} E(\tilde{T}^2) &= E(1/\tilde{T}^2) = \int_0^\infty \frac{1}{t^2} \frac{(n/\theta)^n}{\Gamma(n)} t^{n-1} e^{-\frac{n}{\theta}t} dt = \\ &= \left(\frac{n}{\theta}\right)^n \frac{1}{\Gamma(n)} \int_0^\infty t^{n-3} e^{-\frac{n}{\theta}t} dt; \end{aligned}$$

considerata nuovamente la sostituzione $z = \frac{n}{\theta}t \Leftrightarrow t = \frac{\theta}{n}z$, che comporta $dt = \frac{\theta}{n}dz$, si ha che:

$$\begin{aligned} E(\tilde{T}^2) &= \left(\frac{n}{\theta}\right)^n \frac{1}{\Gamma(n)} \int_0^\infty \left(\frac{\theta}{n}z\right)^{n-3} e^{-z} \frac{\theta}{n} dz = \\ &= \left(\frac{n}{\theta}\right)^n \frac{1}{\Gamma(n)} \left(\frac{\theta}{n}\right)^{n-2} \int_0^\infty z^{(n-2)-1} e^{-z} dz = \\ &= \left(\frac{n}{\theta}\right)^2 \frac{\Gamma(n-2)}{\Gamma(n)} = \frac{n^2}{\theta^2} \frac{\Gamma(n-2)}{(n-1)(n-2)\Gamma(n-2)} = \\ &= \frac{n^2}{(n-1)(n-2)\theta^2}. \end{aligned}$$

Risulta quindi:

$$\begin{aligned} \text{Var}(U) &= \left(\frac{n-1}{n}\right)^2 \text{Var}(\tilde{T}) = \\ &= \left(\frac{n-1}{n}\right)^2 \left\{ E(\tilde{T}^2) - [E(\tilde{T})]^2 \right\} = \\ &= \left(\frac{n-1}{n}\right)^2 \left\{ \frac{n^2}{(n-1)(n-2)\theta^2} - \left[\frac{n}{(n-1)\theta}\right]^2 \right\} = \\ &= \left(\frac{n-1}{n}\right)^2 \underbrace{\frac{n^2}{(n-1)^2(n-2)\theta^2}}_{\text{Var}(\tilde{T})} = \frac{1}{(n-2)\theta^2}. \end{aligned}$$

Esercizio 10

Sia (X_1, \dots, X_n) un campione casuale a componenti i.i.d. a X caratterizzata da funzione di densità di probabilità:

$$\varphi_X(x; \theta) = \theta x^{\theta-1},$$

$0 < x < 1$, $\theta > 0$ e sia:

$$T = -\frac{n}{\sum_{i=1}^n \log X_i}$$

uno stimatore per θ . Si verifichi se:

1. T è stimatore corretto per θ (suggerimento: si determini la legge di distribuzione di $Y = -\log X$);
2. il relativo errore quadratico medio.

Soluzione

Sia $X \sim \varphi_X(x; \theta) = \theta x^{\theta-1}$, $0 < x < 1$, $\theta \in \mathbb{R}^+$ ignoto; siano inoltre X_1, \dots, X_n i.i.d. a X . Trattasi di un problema di stima puntuale per θ .

1. Come suggerito dal testo dell'esercizio, si consideri la trasformazione $Y = f(X) = -\log X$ di X : essa è biunivoca con inversa $X = f^{-1}(Y) = e^{-Y}$; in virtù del metodo basato sulla densità, Y risulta dunque caratterizzata da funzione di densità di probabilità:

$$\varphi_Y(y; \theta) = \varphi_X(f^{-1}(y); \theta) \left| \frac{dx}{dy} \right| = \theta e^{-(\theta-1)y} e^{-y} = \theta e^{-\theta y}, \quad y > 0,$$

ossia $Y \sim \text{Exp}(\theta)$, o, equivalentemente, $Y \sim \text{Gamma}(1, \theta)$.

Essendo X_1, \dots, X_n i.i.d. a X , altresì Y_1, \dots, Y_n sono i.i.d. a Y e, in virtù della proprietà riproduttiva della variabile casuale Gamma rispetto il parametro di forma, risulta:

$$W = \sum_{i=1}^n Y_i = \sum_{i=1}^n (-\log X_i) = -\sum_{i=1}^n \log X_i \sim \text{Gamma}(n, \theta);$$

pertanto:

$$T = -\frac{n}{\sum_{i=1}^n \log X_i} = \frac{n}{W}$$

si identifica in una trasformazione di W ed il relativo valore atteso può essere determinato come segue:

$$\begin{aligned} E(T) &= E\left(\frac{n}{W}\right) = \\ &= \int_0^\infty \frac{n}{w} \frac{\theta^n}{\Gamma(n)} w^{n-1} e^{-\theta w} dw = \\ &= n \int_0^\infty \frac{\theta^n}{(n-1)\Gamma(n-1)} w^{(n-1)-1} e^{-\theta w} dw = \\ &= \frac{n}{n-1} \theta \underbrace{\int_0^\infty \frac{\theta^{n-1}}{\Gamma(n-1)} w^{(n-1)-1} e^{-\theta w} dw}_{=1} = \\ &= \frac{n}{n-1} \theta, \end{aligned}$$

ove l'ultimo integrale riportato ha valore pari a 1 poiché relativo alla funzione di densità di probabilità della variabile casuale $\text{Gamma}(n-1, \theta)$ integrata sul proprio supporto.

In definitiva, poiché:

$$E(T) = \frac{n}{n-1} \theta \neq \theta, \quad \forall \theta > 0,$$

T è stimatore distorto per θ .

2. Essendo T distorto per θ , si ha che:

$$\begin{aligned}\text{MSE}(T) &= \text{Var}(T) + [\text{B}(T)]^2 = \\ &= \{E(T^2) - [E(T)]^2\} + [\text{B}(T)]^2,\end{aligned}$$

ove:

$$[\text{B}(T)]^2 = \left(\frac{n}{n-1} \theta - \theta \right)^2 = \frac{1}{(n-1)^2} \theta^2$$

e:

$$\begin{aligned}E(T^2) &= E\left(\frac{n}{W}\right)^2 = \\ &= \int_0^\infty \left(\frac{n}{w}\right)^2 \frac{\theta^n}{\Gamma(n)} w^{n-1} e^{-\theta w} dw = \\ &= n^2 \int_0^\infty \frac{\theta^n}{(n-1)(n-2)\Gamma(n-2)} w^{(n-2)-1} e^{-\theta w} dw = \\ &= \frac{n^2}{(n-1)(n-2)} \theta^2 \underbrace{\int_0^\infty \frac{\theta^{n-2}}{\Gamma(n-2)} w^{(n-2)-1} e^{-\theta w} dw}_{=1} = \\ &= \frac{n^2}{(n-1)(n-2)} \theta^2,\end{aligned}$$

ove l'ultimo integrale riportato ha valore pari a 1 poiché relativo alla funzione di densità di probabilità della variabile casuale $\text{Gamma}(n-2, \theta)$ integrata sul proprio supporto.

Si ha dunque:

$$\begin{aligned}\text{Var}(T) &= E(T^2) - [E(T)]^2 = \\ &= \frac{n^2}{(n-1)(n-2)} \theta^2 - \frac{n^2}{(n-1)^2} \theta^2 = \\ &= \frac{n^2}{(n-1)^2(n-2)} \theta^2\end{aligned}$$

e, in definitiva:

$$\begin{aligned}\text{MSE}(T) &= \text{Var}(T) + [\text{B}(T)]^2 = \\ &= \frac{n^2}{(n-1)^2(n-2)} \theta^2 + \frac{1}{(n-1)^2} \theta^2 = \\ &= \frac{n^2 + n - 2}{(n-1)^2(n-2)} \theta^2 = \frac{(n-1)(n+2)}{(n-1)^2(n-2)} \theta^2 = \\ &= \frac{n+2}{(n-1)(n-2)} \theta^2.\end{aligned}$$