Esercizi Unità 1

Analisi dei dati 2023/24

Cristiano Varin

1. Si consideri un campione casuale semplice di dimensione n da una variabile casuale di media μ e varianza pari a 3. Si consideri il seguente stimatore di μ:

$$T = \frac{1}{n-2} \sum_{i=1}^{n} X_i,$$

Si calcolino la distorsione e la varianza dello stimatore e si dica se si tratta di uno stimatore consistente.

Soluzione. La distorsione di T è

$$\text{Bias}(T) = \text{E}(T) - \mu = \frac{n}{n-2}\mu - \mu = \frac{2\mu}{n-2}.$$

La varianza di T è

$$\text{Var}(T) = \frac{3n}{(n-2)^2}.$$

Lo stimatore è consistente perché la sua distorsione e la sua varianza convergono a zero al crescere della dimensione campionaria.

2. Si consideri un campione casuale semplice di dimensione tre da una popolazione con valore atteso μ e varianza unitaria. Si considerino i seguenti stimatori di μ:

$$T_1 = \frac{1}{4}X_1 + \frac{1}{2}X_2 + \frac{1}{4}X_3 \quad \text{e} \quad T_2 = \bar{X},$$

Si dica quale dei due stimatori è preferibile (la risposta va motivata).

Soluzione. Entrambi gli stimatori hanno valore atteso pari a μ e quindi sono non distorti. Le varianze dei due stimatori sono

$$\text{Var}(T_1) = \frac{6}{16}, \quad \text{Var}(T_2) = \frac{1}{3},$$

quindi lo stimatore T_2 è preferibile.

3. Si consideri un campione casuale semplice (X_1, X_2, X_3) da una variabile di Poisson con media $\lambda > 0$. Si considerino i due stimatori di λ:

$$T_1 = \frac{2X_1 + X_2/2 + X_3}{5} \quad \text{e} \quad T_2 = \frac{X_1 + 2X_2 + 3X_3}{4}.$$

Si dica quale dei due stimatori è preferibile (la risposta va motivata).

Soluzione. I valori attesi dei due stimatori sono

$$\mathrm{E}(T_1) = \frac{7}{10}\lambda, \quad \mathrm{E}(T_2) = \frac{3}{2}\lambda.$$

Quindi, entrambi gli stimatori sono distorti

$$\mathrm{Bias}(T_1) = -\frac{3}{10}\lambda, \quad \mathrm{Bias}(T_2) = \frac{\lambda}{2}$$

Le varianze dei due stimatori sono

$$\mathrm{Var}(T_1) = \frac{21}{100}\lambda, \quad \mathrm{Var}(T_2) = \frac{14}{16}\lambda,$$

da cui segue che gli errori quadratici medi dei due stimatori sono

$$\mathrm{MSE}(T_1) = \frac{9}{100}\lambda^2 + \frac{21}{100}\lambda, \quad \mathrm{MSE}(T_2) = \frac{\lambda^2}{4} + \frac{14}{16}\lambda.$$

La differenza fra i due errori quadratici medi è

$$\begin{aligned}\mathrm{MSE}(T_2) - \mathrm{MSE}(T_1) &= \lambda\left\{\left(\frac{1}{4} - \frac{9}{100}\right)\lambda + \frac{14}{16} - \frac{21}{100}\right\} \\ &= \lambda(0.16\lambda + 0.665)\end{aligned}$$

Siccome la differenza dei due errori quadratici medi è positiva per qualsiasi valore ammissibile di λ, concludiamo che lo stimatore T_1 è preferibile.

4. Siano T_1 e T_2 due stimatori indipendenti tali che $\mathrm{E}(T_1) = \mathrm{E}(T_2) = \theta$, $\mathrm{Var}(T_1) = \sigma_1^2 > 0$ e $\mathrm{Var}(T_2) = \sigma_2^2 > 0$. Si consideri la combinazione lineare dei due stimatori

$$T_3 = aT_1 + (1-a)T_2, \quad a \in [0,1].$$

Si calcoli il valore di a per cui l'errore quadratico medio di T_3 è il più piccolo possibile.

Soluzione. Lo stimatore T_3 è ovviamente non distorto visto che

$$\mathrm{E}(T_3) = a\mathrm{E}(T_1) + (1-a)\mathrm{E}(T_2) = \theta.$$

L'errore quadratico medio di T_3 quindi coincide con la sua varianza

$$\mathrm{Var}(T_3) = a^2\mathrm{Var}(T_1) + (1-a)^2\mathrm{Var}(T_2) = a^2\sigma_1^2 + (1-a)^2\sigma_2^2.$$

Per cercare il valore di a che minimizza $\mathrm{Var}(T_3)$ ne calcoliamo la derivata prima

$$\frac{d}{da}\mathrm{Var}(T_3) = 2a\sigma_1^2 - 2(1-a)\sigma_2^2$$

e la poniamo pari a zero

$$\frac{d}{da}\mathrm{Var}(T_3) = 0 \longleftrightarrow \hat{a} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}.$$

Il valore $\hat{a}$ è il punto di minimo di $\mathrm{Var}(T_3)$ visto che la sua derivata seconda vale è una funzione positiva:

$$\frac{d^2}{da^2}\mathrm{Var}(T_3) = 2\sigma_1^2 + 2\sigma_2^2 > 0.$$

5. Sia X un campione casuale di dimensione uno da una variable casuale di Poisson con media $\lambda > 0$. Si considerano i due stimatori di λ :

$$T_1 = X \quad \text{e} \quad T_2 = 1.$$

Si calcolino i valori di λ per cui lo stimatore T_2 è preferibile allo stimatore T_1.

Soluzione. La distorsione dei due stimatori è

$$\text{Bias}(T_1) = 0 \quad \text{e} \quad \text{Bias}(T_2) = (1 - \lambda),$$

mentre la varianza è

$$\text{Var}(T_1) = \lambda \quad \text{e} \quad \text{Var}(T_2) = 0.$$

Quindi, l'errore quadratico medio dei due stimatori è

$$\text{MSE}(T_1) = \lambda \quad \text{e} \quad \text{MSE}(T_2) = (1 - \lambda)^2.$$

Lo stimatore T_2 è preferibile allo stimatore T_1 se

$$(1 - \lambda)^2 < \lambda,$$

ovvero se

$$\lambda^2 - 3\lambda + 1 < 0,$$

che corrisponde a

$$\frac{3 - \sqrt{5}}{2} < \lambda < \frac{3 + \sqrt{5}}{2}.$$

6. Sia $X_1, \ldots, X_n$ un campione casuale semplice da una variabile casuale con media μ e varianza $\sigma^2 > 0$. Si costruisca uno stimatore non distorto di $\gamma = \mu^2$.

Soluzione. Sappiamo che

$$\begin{aligned}\frac{\sigma^2}{n} &= \text{Var}(\bar{X}) \\ &= \text{E}(\bar{X}^2) - \{\text{E}(\bar{X})\}^2 \\ &= \text{E}(\bar{X}^2) + \mu^2,\end{aligned}$$

ovvero

$$\text{E}(\bar{X}^2) = \frac{\sigma^2}{n} - \mu^2.$$

Quindi, uno stimatore non distorto di $\gamma = \mu^2$ è

$$\hat{\gamma} = \frac{S}{n} - \bar{X}^2,$$

infatti

$$\begin{aligned}\text{E}(\hat{\gamma}) &= \text{E}\left(\frac{S}{n}\right) - \text{E}\left(\bar{X}^2\right) \\ &= \frac{\sigma^2}{n} - \left(\frac{\sigma^2}{n} - \mu^2\right) \\ &= \mu^2.\end{aligned}$$

7. Sia $X_1, \dots, X_n$ un campione casuale semplice da una variabile casuale uniforme nell'intervallo $(0, \theta)$, ovvero con densità

$$f(x;\theta) = \begin{cases} 1/\theta, & \text{se } 0 < x < \theta, \\ 0, & \text{altrimenti,} \end{cases}$$

per un qualche parametro ignoto $\theta > 0$. Si consideri lo stimatore $\hat{\theta} = 2\bar{X}$.

(a) Si calcoli la distorsione di $\hat{\theta}$.

(b) Si valuti la consistenza di $\hat{\theta}$.

Soluzione.

(a) Il valore atteso dello stimatore è $\mathrm{E}(\hat{\theta}) = \mathrm{E}(2\bar{X}_n) = 2\mathrm{E}(X_1)$. Siccome

$$\mathrm{E}(X_1) = \int_0^\theta \frac{x}{\theta}\, dx = \frac{\theta}{2}$$

allora abbiamo che $\mathrm{E}(\hat{\theta}) = \theta$, cioè lo stimatore è non distorto.

(b) Secondo la 'Legge dei grandi numeri' abbiamo che

$$\bar{X} \xrightarrow{p} \mathrm{E}(X_1) = \frac{\theta}{2}, \quad \text{per } n \to \infty.$$

Siccome $\hat{\theta}$ è una trasformazione continua di $\bar{X}$ abbiamo che

$$\hat{\theta} = g(\bar{X}) \xrightarrow{p} g\{\mathrm{E}(X_1)\} = 2\frac{\theta}{2} = \theta, \quad \text{per } n \to \infty,$$

ovvero $\hat{\theta}$ è uno stimatore consistente di θ.

Alternativamente, la consistenza dello stimatore si può verificare controllando che la varianza dello stimatore converge a zero al crescere della dimensione campionaria, come mostrato nel seguito. La varianza dello stimatore è

$$\mathrm{Var}(\hat{\theta}) = \mathrm{Var}(2\bar{X}_n) = \frac{4}{n}\mathrm{Var}(X_1).$$

Resta da calcolare $\mathrm{Var}(X_1)$:

$$\begin{aligned} \mathrm{Var}(X_1) &= \mathrm{E}(X_1^2) - \mathrm{E}(X_1)^2 \\ &= \mathrm{E}(X_1^2) - \frac{\theta^2}{4} \\ &= \int_0^\theta \frac{x^2}{\theta}\, \mathrm{d}x - \frac{\theta^2}{4} \\ &= \frac{\theta^2}{3} - \frac{\theta^2}{4} \\ &= \frac{\theta^2}{12}. \end{aligned}$$

Quindi la varianza dello stimatore è

$$\mathrm{Var}(\hat{\theta}) = \frac{4}{n}\frac{\theta^2}{12} = \frac{\theta^2}{3n}$$

quantità che converge a zero al divergere di n.

8. Sia $X_1, \ldots, X_n$ un campione casuale semplice da una variabile casuale discreta con funzione di probabilità:

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

per un qualche parametro ignoto $\theta \in [0, 1]$. Si consideri lo stimatore $\hat{\theta} = 2\bar{X} + 1$.

(a) Si calcoli la distorsione di $\hat{\theta}$.

(b) Si valuti la consistenza di $\hat{\theta}$.

Soluzione.

(a) Il valore atteso di X_1 è

$$\mathrm{E}(X_1) = -1\left(\frac{1}{2}\right) + 0\left(\frac{1-\theta}{2}\right) + 1\left(\frac{\theta}{2}\right) = \frac{\theta - 1}{2}.$$

Quindi, il valor atteso di $\hat{\theta}$ è

$$\mathrm{E}(\hat{\theta}) = 2\mathrm{E}(X_1) + 1 = \theta,$$

ovvero $\hat{\theta}$ è non distorto.

(b) Utilizzando la 'Legge dei grandi numeri' e le proprietà delle trasformazioni continue delle variabili casuali verifichiamo immediatamente la consistenza di $\hat{\theta}$:

$$\hat{\theta} = g(\bar{X}) \xrightarrow{p} g\{\mathrm{E}(X_1)\} = 2\mathrm{E}(X_1) + 1 = \theta, \quad \text{per } n \to \infty.$$

Alternativamente possiamo provare la consistenza dello stimatore non distorto $\hat{\theta}$ verificando che la sua varianza converge a zero. La varianza di X_1 è

$$\mathrm{Var}(X_1) = \mathrm{E}(X_1^2) - \mathrm{E}(X_1)^2 = \mathrm{E}(X_1^2) - \frac{(\theta-1)^2}{4}.$$

Siccome

$$\mathrm{E}(X_1^2) = -1^2\left(\frac{1}{2}\right) + 0^2\left(\frac{1-\theta}{2}\right) + 1^2\left(\frac{\theta}{2}\right) = \frac{\theta+1}{2},$$

abbiamo che

$$\mathrm{Var}(X_1) = \frac{\theta+1}{2} - \frac{(\theta-1)^2}{4} = \frac{-\theta^2 + 4\theta + 1}{4}.$$

Quindi la varianza di $\hat{\theta}$ è

$$\mathrm{var}(\hat{\theta}) = \frac{4}{n}\mathrm{var}(X_1) = \frac{-\theta^2 + 4\theta + 1}{n},$$

quantità che converge a zero al crescere della dimensione campionaria.

9. Si risolva l'esercizio 8.3 del libro di testo Baron (2014, pagina 234).

Soluzione. Abbiamo che

$$\begin{aligned}
\Pr(X \geq 127.78) &= \Pr(X - 48.23 \geq 127.78 - 48.23) \\
&\leq \Pr\left(\frac{|X - 48.23|}{26.52} \geq \frac{79.55}{26.52}\right) \\
&= \Pr\left(\frac{|X - 48.23|}{26.52} \geq 3\right) \\
&\leq \frac{1}{9}
\end{aligned}$$

10. Sia $X_1, \ldots, X_n$ un campione casuale semplice da una variabile casuale con funzione di densità
$$f(x;\theta) = 2\theta^2 x^{-3}, \quad \text{per } x > \theta \text{ e } \theta > 0.$$
Si consideri lo stimatore $\hat{\theta} = \bar{X}/2$.

(a) Si calcoli la distorsione di $\hat{\theta}$.
(b) Si calcoli l'errore standard di $\hat{\theta}$.
(c) Si valuti la consistenza di $\hat{\theta}$.

Soluzione.

(a) Il valore atteso di X_1 è
$$\mathrm{E}(X_1) = \int_\theta^\infty 2\theta^2 x^{-2}\, dx = 2\theta$$
Quindi lo stimatore $\hat{\theta}$ è non distorto poiché
$$\mathrm{E}(\hat{\theta}) = \mathrm{E}(\bar{X}/2) = \mathrm{E}(X_1/2) = \theta.$$

(b) La varianza dello stimatore è pari a
$$\mathrm{Var}(\hat{\theta}) = \frac{1}{4n}\mathrm{Var}(X_1).$$
Dobbiamo valutare $\mathrm{Var}(X_1)$:
$$\begin{aligned}
\mathrm{Var}(X_1) &= \mathrm{E}(X_1^2) - \mathrm{E}(X_1)^2 \\
&= \mathrm{E}(X_1^2) - 4\theta^2 \\
&= \int_\theta^\infty 2\theta^2 x^{-1}\, dx - 4\theta^2
\end{aligned}$$
Siccome l'integrale è divergente, la varianza non è finita e quindi l'errore standard di $\hat{\theta}$ non esiste finito.

(c) Siccome la varianza non è finita, lo stimatore non è consistente.