Probabilità e Statistica
Esercitazioni

a.a. 2009/2010

C.d.L.: Ingegneria Elettronica e delle Telecomunicazioni, Ingegneria Informatica

Stima puntuale di parametri

Ines Campa

Campionamento

Esercizio 1. Durante un processo di produzione vengono estratti 300 campioni di un pezzo e ne viene misurato lo spessore in mm, fornendo una media campionaria di $2,6\ mm$. Si sa inoltre che la varianza della variabile casuale che dà lo spessore è $0,4\ mm^2$. Qual è la probabilità che il valore atteso dello spessore sia inferiore a $2,5\ mm$?

Risoluzione. Sia X la variabile casuale che denota lo spessore in mm, posto

$$Z = \frac{\overline{X}_n - \mu_X}{\sigma/\sqrt{n}} \quad \text{con} \quad \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

dal Teorema del limite centrale risulta che

$$Z \,\dot{\sim}\, \mathcal{N}(0;1) \quad \text{per} \quad n \geq 30.$$

Essendo $n = 300$ possiamo utilizzare l'approssimazione indicata.
Eseguiamo alcuni passaggi algebrici:

$$\mu_X < 2.5 \qquad -\mu_X > -2.5 \qquad 2.6 - \mu_X > 2.6 - 2.5$$

$$\frac{2.6 - \mu_X}{\sqrt{0.4/300}} > \frac{2.6 - 2.5}{\sqrt{0.4/300}} \qquad \frac{2.6 - \mu_X}{\sqrt{0.4/300}} > 2.74 \quad .$$

Ne segue che

$$\mu_X < 2.5 \Longleftrightarrow Z > 2.74$$

quindi

$$P\left[\mu_X < 2.5\right] = P\left[Z > 2.74\right] = 1 - 0.99693 = 0.00307.$$

–

Esercizio 2. In un processo di produzione vengono prodotte sbarre con una lunghezza data da una distribuzione normale, di cui si conosce la deviazione standard, pari a $0,2\ m$. Si esegue un campionamento per calcolare il valore atteso, misurando un numero n di sbarre e calcolando la media. Qual è la numerosità del campione che garantisce una probabilità superiore al 97% che la media campionaria disti dal valore atteso per meno di $1\ cm$?

Risoluzione. Sia X la v.c. che denota la lunghezza delle sbarre in m e sia $\overline{X}_n = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ bisogna determinare n tale che

$$P\left[\left|\overline{X}_n - \mu_X\right| < 0.01\right] > 0.97.$$

Essendo

$$\overline{X}_n \sim \mathcal{N}\left(\mu_X,\, \frac{\sigma_X^2}{n}\right) \Longrightarrow Z \sim \mathcal{N}(0,\, 1) \quad \text{con} \quad Z = \frac{\overline{X}_n - \mu_X}{\sigma_X/\sqrt{n}}$$

$$\left|\overline{X}_n - \mu_X\right| = \frac{\sigma_X}{\sqrt{n}}\left|Z\right| = \frac{0.2}{\sqrt{n}}\left|Z\right|.$$

–

Quindi il problema diventa determinare n:

$$P\left[|Z| < \frac{0.01}{0.2/\sqrt{n}}\right] > 0.97$$

$$P\left[|Z| < 0.05\sqrt{n}\right] > 0.97$$

$$2 \cdot P\left[0 \leq Z < 0.05\sqrt{n}\right] > 2 \cdot 0.485$$

$$P\left[Z < 0.05\sqrt{n}\right] - P\left[Z < 0\right] > 0.485$$

$$P\left[Z < 0.05\sqrt{n}\right] - 0.5 > 0.485$$

$$P\left[Z < 0.05\sqrt{n}\right] > 0.985$$

$$\Phi_Z\left(0.05\sqrt{n}\right) > \Phi_Z(2.17)$$

Quindi $0.05\sqrt{n} > 2,17 \Longrightarrow n > 1884$.

Esercizio 3 (Tema d'esame del 07/12/2004).
Si supponga che una distribuzione con media incognita abbia deviazione standard uguale a 1. Quanto deve essere grande un campione affinché si abbia una probabilità almeno del 90% che la media campionaria $\overline{X}_n$ disti meno di 0.5 dalla media della popolazione?

Risoluzione. Bisogna determinare n tale che

$$P\left[\left|\overline{X}_n - \mu_X\right| < 0.5\right] \geq 0.90.$$

Ricordiamo la *legge dei grandi numeri*: sia $f(\cdot)$ una densità con media μ e varianza σ^2, e sia $\overline{X}_n$ la media campionaria di un campione casuale di ampiezza n estratto da $f(\cdot)$, siano ϵ e δ due numeri fissati che soddisfano $\epsilon > 0$ e $0 < \delta < 1$, allora se $n > \frac{\sigma^2}{\epsilon^2 \delta}$:

$$P\left[\left|\overline{X}_n - \mu\right| < \epsilon\right] \geq 1 - \delta.$$

Abbiamo $\sigma^2 = 1$, $\epsilon = 0.5$ e $\delta = 1 - 0.90 = 0.10$, perciò

$$n > \frac{\sigma^2}{\epsilon^2 \delta} = \frac{1}{(0.5)^2 \cdot 0.10} = 40.$$

Quindi la probabilità che la media campionaria differisca dalla media incognita della popolazione di 0.5 è almeno del 90% se si prende un campione casuale di ampiezza superiore a 40.

Stima parametrica

Esercizio 4. Si consideri per $\vartheta > 0$ la funzione definita da

$$f_X(x;\vartheta) = \begin{cases} \frac{4}{\vartheta^2}x & \text{se } 0 \le x \le \frac{\vartheta}{2} \\ \frac{4}{\vartheta^2} \cdot (\vartheta - x) & \text{se } \frac{\vartheta}{2} < x \le \vartheta \\ 0 & \text{altrimenti} \end{cases}$$

a) Verificare che per ogni $\vartheta > 0$, $f_X(\cdot;\vartheta)$ rappresenta una funzione di densità di probabilità.

b) Sia $X_1, X_2, \cdots, X_n$ un campione casuale estratto dalla popolazione di densità $f_X(\cdot, \vartheta)$, stabilire se $T = \frac{2}{n}\sum_{i=1}^{n} X_i$ è uno stimatore corretto di ϑ.

Risoluzione.

a) Risulta

$$\int_{-\infty}^{+\infty} f_X(x;\vartheta)\,dx = \frac{1}{2}\vartheta\frac{2}{\vartheta} = 1$$

quindi $f_X(\cdot;\vartheta)$ rappresenta una funzione di densità di probabilità $\forall \vartheta > 0$.

b) T è uno stimatore **corretto** o **non distorto** di ϑ se $E[T] = \vartheta$.
Nel nostro caso $E[X] = \frac{1}{2}\vartheta$ e

$$E[T] = E\left[\frac{2}{n}\sum_{i=1}^{n} X_i\right] = \frac{2}{n}\sum_{i=1}^{n} E[X_i] =$$

$$= \frac{2}{n} \cdot nE[X] = \frac{2}{n}n\frac{1}{2}\vartheta = \vartheta \Longrightarrow \text{lo stimatore è corretto.}$$

Esercizio 5. Sia X la variabile casuale di Poisson di parametro λ,

a) calcolare lo stimatore di massima verosimiglianza di λ;

b) verificare la correttezza dello stimatore trovato;

c) calcolare l'errore quadratico medio;

d) determinare uno stimatore con il metodo dei momenti;

e) supponendo poi di avere i seguenti dati campionati

$$2,\ 3,\ 2,\ 4,\ 3,\ 5,\ 4,\ 2,\ 3,\ 2,\ 3,\ 4,\ 2,\ 3,\ 4,\ 3,\ 2,\ 3,\ 4,\ 5,\ 4,\ 3,\ 3,\ 4$$

calcolare il valore stimato di λ.

Risoluzione.

a) X ha funzione di densità di probabilità

$$f_X(x;\lambda) = \begin{cases} \frac{\lambda^x}{x!}e^{-\lambda} & \text{se } x \in N \\ 0 & \text{altrove} \end{cases}.$$

La funzione di verosimiglianza è

$$L\left(x_1,\cdots,x_n,\lambda\right)=\frac{\lambda^{x_1}}{x_1!}e^{-\lambda}\cdot\frac{\lambda^{x_2}}{x_2!}e^{-\lambda}\cdot\ldots\cdot\frac{\lambda^{x_n}}{x_n!}e^{-\lambda}=\frac{\lambda^{x_1+x_2+\cdots+x_n}}{x_1!x_2!\cdots x_n!}e^{-n\lambda}=$$

$$=\frac{1}{x_1!x_2!\cdots x_n!}\lambda^{\sum\limits_{i=1}^{n}x_i}\cdot e^{-n\lambda}$$

con $x_i\in N$, $i=1,\cdots,n$. Il valore di λ, $\widehat{\lambda}$, che massimizza L è detto **stima di massima verosimiglianza** e $\widehat{\Lambda}\left(X_1,X_2,\cdots,X_n\right)$ è lo **stimatore di massima verosimiglianza**. Calcoliamo

$$\frac{d}{d\lambda}L\left(x_1,\cdots,x_n,\lambda\right)=\frac{e^{-n\lambda}}{x_1!x_2!\cdots x_n!}\left[\sum_{i=1}^{n}x_i\lambda^{\sum\limits_{i=1}^{n}x_i-1}-n\lambda^{\sum\limits_{i=1}^{n}x_i}\right].$$

Cerchiamo gli zeri della derivata prima di $L\left(x_1, \cdots, x_n, \lambda\right)$ rispetto a λ

$$\sum_{i=1}^{n} x_i \lambda^{\sum\limits_{i=1}^{n} x_i - 1} - n\lambda^{\sum\limits_{i=1}^{n} x_i} = 0$$

$$\frac{1}{\lambda} \sum_{i=1}^{n} x_i \lambda^{\sum\limits_{i=1}^{n} x_i} - n\lambda^{\sum\limits_{i=1}^{n} x_i} = 0$$

$$\frac{1}{\lambda} \sum_{i=1}^{n} x_i - n = 0 \Longrightarrow \widehat{\lambda}\left(x_1, x_2, \cdots, x_n\right) = \frac{\sum\limits_{i=1}^{n} x_i}{n} = \overline{x}_n .$$

Si verifica che $\widehat{\lambda}$ è un **punto di massimo**, anche se in modo piuttosto faticoso (fare, oppure..........).

Quindi

$$\widehat{\lambda}\left(x_1, x_2, \cdots, x_n\right) = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x}_n$$

è la stima di massima verosimiglianza di λ e la media campionaria lo stimatore di massima verosimiglianza. Allo stesso risultato si perviene usando la **log-verosimiglianza**

$$\ln\left[L\left(x_1, \cdots, x_n, \lambda\right)\right] = -n\lambda + \sum_{i=1}^{n} x_i \ln\lambda - \sum_{i=1}^{n} \ln\left(x_i!\right).$$

Calcoliamo

$$\frac{d}{d\lambda} \ln\left[L\left(x_1, \cdots, x_n, \lambda\right)\right] = -n + \frac{\sum_{i=1}^{n} x_i}{\lambda}.$$

Poniamo uguale a zero la derivata prima

$$-n + \frac{\sum_{i=1}^{n} x_i}{\lambda} = 0 \Longrightarrow \widehat{\lambda}\left(x_1, x_2, \cdots, x_n\right) = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x}_n.$$

–

Lo stesso risultato di prima, ma si tratta di un punto di massimo?

$$\frac{d^2}{d\lambda^2} \ln \left[L \left(x_1, \cdots, x_n, \lambda \right) \right] \Bigg|_{\lambda = \widehat{\lambda}} = - \frac{\sum_{i=1}^{n} x_i}{\widehat{\lambda}^2} < 0 \Longrightarrow \widehat{\lambda} \text{ PUNTO DI MAX.}$$

b) Per determinare se lo stimatore è non distorto calcoliamo

$$E \left[\widehat{\Lambda} \left(X_1, X_2, \cdots, X_n \right) \right] = E \left[\overline{X}_n \right] = \frac{1}{n} E \left[\sum_{i=1}^{n} X_i \right] = \frac{1}{n} n E \left[X \right].$$

Essendo X v.c. di Poisson di parametro λ, si ha $E\left[X\right] = \lambda$, quindi

$$E \left[\widehat{\Lambda} \left(X_1, X_2, \cdots, X_n \right) \right] = \lambda \Longrightarrow \widehat{\Lambda} \text{ stimatore non distorto.}$$

c) Chiamiamo errore quadratico medio di uno stimatore $T = t\left(X_1, \cdots, X_n\right)$, la funzione di ϑ data da:

$$\mathrm{MSE}\left[T\right](\vartheta) := E\left[\left(T - \tau(\vartheta)\right)^2\right]$$

con $\tau(\vartheta)$ funzione del parametro da stimare. Ma

$$\mathrm{MSE}\left[T\right](\vartheta) := \mathrm{var}\left[T\right] + \left[D\left[T\right](\vartheta)\right]^2, \quad D\left[T\right](\vartheta) := \tau(\vartheta) - E\left[T\right]$$

Nel nostro caso $E\left[\widehat{\Lambda}\right] = \lambda \Longrightarrow D\left[\widehat{\Lambda}\right] = 0$, quindi

$$MSE\left[\widehat{\Lambda}\right] = \mathrm{var}\left[\widehat{\Lambda}\right] = \mathrm{var}\left[\overline{X}_n\right] \underbrace{=}_{X_i \ ind.} \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{var}\left[X_i\right] = \frac{1}{n^2} \cdot n\lambda = \frac{\lambda}{n}$$

d) Determiniamo, ora, uno stimatore **con il metodo dei momenti**. Ricordiamo che con

$$M_r' = \frac{1}{n}\sum_{i=1}^{n} X_i^r, \quad \mu_r' = E\left[X^r\right]$$

indicano rispettivamente il momento campionario assoluto di ordine r e il momento di ordine r della variabile casuale X.

Il metodo dei momenti consiste nel risolvere il sistema nelle incognite $\vartheta_1, \cdots, \vartheta_k$ di k, numero dei parametri incogniti, di equazioni

$$\begin{cases} \mu'_1 & = & M'_1 \\ \mu'_2 & = & M'_2 \\ . & & . \\ . & & . \\ . & & . \\ \mu'_k & = & M'_k \end{cases}$$

Nel nostro caso abbiamo solo un parametro da stimare, λ, quindi calcoliamo

$$\mu'_1 = E[X] = \lambda, \text{ e } M'_1 = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Posto $\mu'_1 = M'_1$, risulta $\overline{\Lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

e) Troviamo λ mediante i dati campionari

$$\widehat{\lambda} = \frac{2 \cdot 6 + 3 \cdot 9 + 4 \cdot 7 + 5 \cdot 2}{24} = \frac{77}{24} \approx 3.2083.$$

–

Esercizio 6. Sia $X_1,\ X_2\ \ldots X_n$ un campione casuale di ampiezza n estratto da una popolazione con distribuzione uniforme nell'intervallo $[a-b, a+b]$. Determinare gli stimatori di a e b con il metodo dei momenti.

Risoluzione. L'intervallo ha ampiezza $a+b-(a-b)=2b$, ne segue che la funzione di densità di probabilità della v.c. X è

$$f_X(x; a, b) = \begin{cases} \frac{1}{2b} & \text{se } a-b \leq x \leq a+b \\ 0 & \text{altrove} \end{cases}$$

Per determinare gli stimatori di a e b con il metodo dei momenti, bisogna risolvere il sistema

$$\begin{cases} \mu_1' = M_1' \\ \mu_2' = M_2' \end{cases} \Longrightarrow \begin{cases} \mu_1' = \overline{X}_n \\ \mu_2' = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \end{cases}$$

Calcoliamo $\mu_1' = \frac{a+b+a-b}{2} = a$ e

–

$$\mu_2' = \mathrm{var}\,[X] + (\mu_1')^2 = \frac{[(a+b)-(a-b)]^2}{12} + a^2 = \frac{b^2}{3} + a^2$$

Sostituendo

$$\begin{cases} a = \overline{X}_n \\ \frac{b^2}{3} + a^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \end{cases}$$

ed eseguendo alcuni passaggi algebrici, ne segue che gli stimatori cercati sono

$$\widehat{a} = \frac{1}{n}\sum_{i=1}^{n} X_i \ \text{ e } \ \widehat{b} = \sqrt{\frac{3}{n}\sum_{i=1}^{n} X_i^2 - \frac{3}{n^2}\left(\sum_{i=1}^{n} X_i\right)^2}\,.$$

Esercizio 7. Sia data la variabile casuale unidimensionale X con funzione di densità di probabilità data da:

$$f_X(x;\vartheta) = \begin{cases} \frac{1}{18\vartheta^4} \cdot e^{-\frac{\sqrt{x}}{3\vartheta^2}} & \text{se } x \geq 0 \\ 0 & \text{altrimenti} \end{cases}$$

con $\vartheta > 0$.

a) Trovare uno stimatore di $\vartheta > 0$ con il metodo di massima verosimiglianza.

b) Calcolare la densità di probabilità della variabile casuale $Y = \sqrt{X}$.

Risoluzione.

a) Costruiamo la funzione di verosimiglianza

$$L(x_1, \cdots, x_n, \vartheta) = \frac{1}{18\vartheta^4} e^{-\frac{\sqrt{x_1}}{3\vartheta^2}} \cdot \frac{1}{18\vartheta^4} e^{-\frac{\sqrt{x_2}}{3\vartheta^2}} \cdot \ldots \cdot \frac{1}{18\vartheta^4} e^{-\frac{\sqrt{x_n}}{3\vartheta^2}} =$$

$$= \frac{1}{(18\vartheta^4)^n} e^{-\frac{(\sqrt{x_1}+\sqrt{x_2}+\ldots+\sqrt{x_n})}{3\vartheta^2}}$$

con $x_i \geq 0$, $i = 1, \ldots, n$.

Cerchiamo $\widehat{\vartheta}$ che massimizza $\ln L$ tra le soluzioni di $\frac{d}{d\vartheta}\ln L = 0$. Calcoliamo

$$\ln L = -n\ln 18\vartheta^4 + \ln e^{-\frac{\left(\sqrt{x_1}+\sqrt{x_2}+\ldots+\sqrt{x_n}\right)}{3\vartheta^2}}$$

$$= -n\ln 18 - 4n\ln\vartheta - \frac{\sqrt{x_1}+\sqrt{x_2}+\ldots+\sqrt{x_n}}{3\vartheta^2}$$

si ha

$$\frac{d}{d\vartheta}\ln L = -\frac{4n}{\vartheta} + \frac{2}{3\vartheta^3}\left(\sqrt{x_1}+\sqrt{x_2}+\ldots+\sqrt{x_n}\right) = 0.$$

Essendo $\vartheta > 0$ risulta

$$\widehat{\vartheta} = \sqrt{\frac{\sqrt{x_1}+\sqrt{x_2}+\ldots+\sqrt{x_n}}{6n}}$$

Vediamo se si tratta di un punto di massimo, calcoliamo

$$\frac{d^2}{d\vartheta^2}\ln L = +\frac{4n}{\vartheta^2} - \frac{2}{\vartheta^4}\left(\sqrt{x_1}+\sqrt{x_2}+\ldots+\sqrt{x_n}\right)$$

ne segue

$$\left.\frac{d^2}{d\vartheta^2}\ln L\right|_{\vartheta=\widehat{\vartheta}} = \frac{2}{\widehat{\vartheta}^2}\left(2n - \frac{\widehat{\vartheta}^2\cdot 6n}{\widehat{\vartheta}^2}\right) = -\frac{8n}{\widehat{\vartheta}^2} < 0 \Longrightarrow \widehat{\vartheta}\ \text{PUNTO DI MAX.}$$

Quindi

$$\widehat{\Theta} = \sqrt{\frac{\sum_{i=1}^{n} \sqrt{X_i}}{6n}}$$

è uno stimatore di ϑ di massima verosimiglianza.

b) Determiniamo la densità di probabilità della variabile casuale $Y = \sqrt{X}$.

Si ha $F_Y(y) = P[Y \leq y] = P\left[\sqrt{X} \leq y\right]$. Se $y < 0$ si ha $[\sqrt{x} \leq y] = \emptyset$.

Se $y \geq 0$, si ha

$$F_Y(y) = P\left[X \leq y^2\right] = \int_0^{y^2} \frac{1}{18\vartheta^4} e^{-\frac{\sqrt{x}}{3\vartheta^2}} dx.$$

Posto $t = \frac{\sqrt{x}}{3\vartheta^2}$, risulta $d\,t = \frac{1}{6\vartheta^2\sqrt{x}} d\,x \Longrightarrow 18\vartheta^4 t d\,t = d\,x$. Quindi

$$\int_0^{y^2} \frac{1}{18\vartheta^4} e^{-\frac{\sqrt{x}}{3\vartheta^2}} dx = \int_0^{\frac{y}{3\vartheta^2}} te^{-t} d\,t = 1 - e^{-\frac{y}{3\vartheta^2}} \left(1 + \frac{y}{3\vartheta^2}\right).$$

Ne segue

$$F_Y(y) = \begin{cases} 1 - e^{-\frac{y}{3\vartheta^2}}\left(1 + \frac{y}{3\vartheta^2}\right) & \text{se} \quad y \geq 0 \\ 0 & \text{se} \quad y < 0 \end{cases} ,$$

derivando si ottiene

$$f_Y(y) = \begin{cases} \frac{y}{9\vartheta^4} e^{-\frac{y}{3\vartheta^2}} & \text{se} \quad y \geq 0 \\ 0 & \text{se} \quad y < 0 \end{cases} .$$

Esercizio 8. Si supponga che X_1 , X_2, X_3 sia un campione casuale di ampiezza 3 estratto da una distribuzione esponenziale di media λ. Si considerino i seguenti stimatori del parametro λ:

$$\Lambda_1 = X_1,\ \Lambda_2 = \frac{X_1 + X_2}{2},\ \Lambda_3 = \frac{X_1 + 2X_2}{3},\ \Lambda_4 = \overline{X}_3$$

a) Indicare quali sono gli stimatori non distorti di λ.

b) Individuare tra gli stimatori non distorti quello con MSE minimo.

Risoluzione. Sia X una v.c. esponenziale di media λ, si ha

$$f_X(x;\lambda) = \begin{cases} \frac{1}{\lambda} e^{-\frac{x}{\lambda}} & \text{se} \quad x \geq 0 \\ 0 & \quad\quad \text{altrove} \end{cases}$$

con $E[X] = \lambda$ e $\operatorname{var}[X] = \lambda^2$.

a) Calcoliamo il valore atteso degli stimatori dati

$$E[\Lambda_1] = E[X_1] = \lambda,$$

$$E[\Lambda_2] = E\left[\frac{X_1 + X_2}{2}\right] = \frac{1}{2}(E[X_1] + E[X_2]) = \frac{1}{2} \cdot 2\lambda = \lambda,$$

– $$E[\Lambda_3] = E\left[\frac{X_1 + 2X_2}{3}\right] = \frac{1}{3}\left(E[X_1] + 2E[X_2]\right) = \frac{1}{3}\cdot(\lambda + 2\lambda) = \lambda,$$

$$E[\Lambda_4] = E\left[\overline{X}_3\right] = \lambda.$$

b) Tutti gli stimatori dati sono non distorti, per individuare il preferibile, calcoliamo la varianza di ciascuno:

$$\mathrm{var}[\Lambda_1] = \mathrm{var}[X_1] = \lambda^2,$$

$$\mathrm{var}[\Lambda_2] = \mathrm{var}\left[\frac{X_1 + X_2}{2}\right] = \frac{1}{4}\left(\mathrm{var}[X_1] + \mathrm{var}[X_2] + 2\mathrm{cov}[X_1, X_2]\right) =$$

$= \frac{1}{4}\cdot 2\lambda^2 = \frac{1}{2}\lambda^2$ (X_1 e X_2 sono indipendenti),

$$\mathrm{var}[\Lambda_3] = \mathrm{var}\left[\frac{X_1 + 2X_2}{3}\right] = \frac{1}{9}\left(\mathrm{var}[X_1] + 4\mathrm{var}[X_2]\right) = \frac{1}{9}\cdot\left(\lambda^2 + 4\lambda^2\right) = \frac{5}{9}\lambda^2,$$

$$\mathrm{var}[\Lambda_4] = \mathrm{var}\left[\overline{X}_3\right] = \frac{1}{3}\lambda^2.$$

Lo stimatore non distorto con varianza minima è la media campionaria.

Esercizio 9. Il numero di interruzioni (crash) di un personal computer segue una distribuzione di Poisson di parametro λ. Sia $X_1,\ X_2\ \ldots X_n$ un campione casuale di ampiezza n estratto dalla distribuzione di Poisson data. Se il costo di riparazione è dato da

$$Y_n = 3\overline{X}_n + \overline{X}_n^{\ 2}$$

dove $\overline{X}_n$ è la media campionaria su n crash, calcolare $E\left[Y_\infty\right]$.

Risoluzione. Sia X una v.c. di Poisson di parametro λ, si ha

$$f_X\left(x;\lambda\right) = \begin{cases} \frac{\lambda^x}{x!}e^{-\lambda} & \text{se} \quad x = 0, 1, 2, 3 \ldots \\ 0 & \qquad\quad \text{altrove} \end{cases}$$

con $E\left[X\right] = \lambda$ e $\mathrm{var}\left[X\right] = \lambda$. Calcoliamo il valore atteso dello stimatore dato

$$E\left[Y_n\right] = E\left[3\overline{X}_n + \overline{X}_n^2\right] = 3E\left[\overline{X}_n\right] + E\left[\overline{X}_n^2\right] = 3\lambda + E\left[\overline{X}_n^2\right].$$

Ma

$$E\left[\overline{X}_n^2\right] - \left(E\left[\overline{X}_n\right]\right)^2 = \mathrm{var}\left[\overline{X}_n\right], \quad \mathrm{var}\left[\overline{X}_n\right] = \frac{\mathrm{var}\left[X\right]}{n} = \frac{\lambda}{n},$$

quindi

$$E\left[Y_n\right] = 3\lambda + \frac{\lambda}{n} + \lambda^2$$

e passando al limite per $n \longrightarrow \infty$ si ha

$$E\left[Y_\infty\right] = 3\lambda + \lambda^2.$$

Esercizio 10. Per un corso la prova d'accertamento del profitto è costituita da una batteria di 30 quesiti con risposta dicotomica. Si può ipotizzare che chi risponde a caso ad un quesito abbia probabilità $\frac{1}{2}$ di indovinare la risposta esatta. Supponiamo che ogni risposta esatta sia valutata $\frac{1}{30}$, si dica qual è per una persona perfettamente ignorante, la probabilità di ottenere esattamente la sufficienza in una prova d'esame.
Sia X la variabile casuale che conteggia il numero di risposte esatte in una prova. È ragionevole pensare che $X = S + I$ dove S indica la variabile casuale che conteggia il numero dei quesiti di cui il candidato conosce la risposta esatta, mentre I è la variabile casuale che conteggia il numero di risposte indovinate. Sia S la v.c. binomiale di parametri $n = 30$ e p e sia $I/S = s$ la v.c. binomiale di parametri $n = 30 - s$ e $p = \frac{1}{2}$.

a) Si determini la funzione di densità di probabilità congiunta di S e I, $p_{S,I}(s, i)$.

b) Si mostri che $p_X(x) = \sum_{s=0}^{x} p_{S,I}(s, x - s)$ con $x = 0, \cdots, 30$ e dunque X è una variabile casuale binomiale di parametri $n = 30$ e $\frac{1+p}{2}$.

c) Si determini uno stimatore non distorto di p.

Risoluzione. Per ipotesi X è la v.c. che conteggia il numero di risposte esatte, ha distribuzione $Bi\left(30, \frac{1}{2}\right)$ con

$$p_X(x) = \left(\begin{array}{c} 30 \\ x \end{array} \right) p^x (1-p)^{30-x} \quad x = 0, \cdots, 30.$$

Nel caso ricerchiamo la probabilità di ottenere esattamente la sufficienza si ha

$$\overline{p} = p_X(18) = \left(\begin{array}{c} 30 \\ 18 \end{array} \right) \left(\frac{1}{2}\right)^{18} \left(\frac{1}{2}\right)^{12} \approx 0.08055.$$

a) Sia $0 \leq s \leq 30$ e $0 \leq i \leq 30 - s$ allora

$$p_{S,I}(s,i) = \left(\begin{array}{c} 30 \\ s \end{array} \right) p^s (1-p)^{30-s} \cdot \left(\begin{array}{c} 30-s \\ i \end{array} \right) \left(\frac{1}{2}\right)^{i} \left(\frac{1}{2}\right)^{30-s-i} =$$

$$= \left(\begin{array}{c} 30 \\ s \end{array} \right) \cdot \left(\begin{array}{c} 30-s \\ i \end{array} \right) p^s \left[\frac{1}{2}(1-p)\right]^{30-s}.$$

b) Scelto $0 \leq s \leq 30$ e $0 \leq i \leq 30 - s$ risulta

$$p_X(x) = \sum_{(s,i):\ s+i=x} p_{S,I}(s,i) = \sum_{(s,i):\ i=x-s} p_{S,I}(s, x-s) =$$

$$= \sum_{s=0}^{x} \binom{30}{x} \cdot \binom{x}{s} p^s \left[\frac{1-p}{2}\right]^{30-s} =$$

$$= \binom{30}{x} \cdot \left(\frac{1-p}{2}\right)^{30-x} \sum_{s=0}^{x} \binom{x}{s} p^s \left(\frac{1-p}{2}\right)^{x-s} .$$

Ricordando il binomio di Newton

$$\sum_{s=0}^{x} \binom{x}{s} p^s \left(\frac{1-p}{2}\right)^{x-s} = \left(\frac{1+p}{2}\right)^{x}$$

Si ha

$$p_X(x) = \binom{30}{x} \left(\frac{1-p}{2}\right)^{30-x} \left(\frac{1+p}{2}\right)^{x} ,$$

con $x = 0, 1, \ldots,\ 30$.

Pertanto X è una binomiale di parametri 30 e $\frac{1+p}{2}$, di media

$$E\left[X\right] = 30 \cdot \frac{1+p}{2} = 15\left(1+p\right).$$

c) Determiniamo, inoltre, uno stimatore di p con il metodo dei momenti:

$$\overline{X}_n = 15(1+p) \Longrightarrow \widehat{p} = \frac{\overline{X}_n}{15} - 1 \quad \text{stimatore di } \ p \ .$$

Calcoliamo $E\left[\widehat{p}\right]$

$$E\left[\widehat{p}\right] = \frac{1}{15} \cdot 15(1+p) - 1 = p,$$

quindi $\widehat{p}$ è corretto.

Esercizio 11. Sia $X_1, \ X_2 \ \ldots X_n$ un campione casuale proveniente da una distribuzione rettangolare nell'intervallo $[0, \ 2\vartheta]$ con $\vartheta > 0$.

1) Si determini lo stimatore di ϑ con il metodo dei momenti.

2) Si verifichi se tale stimatore è consistente.

Risoluzione.

1) L'intervallo ha ampiezza $2\vartheta - 0 = 2\vartheta$, ne segue che la funzione di densità di probabilità della v.c. rettangolare X è

$$f_X(x; \vartheta) = \begin{cases} \frac{1}{2\vartheta} & \text{se } 0 \leq x \leq 2\vartheta \\ 0 & \text{altrove} \end{cases}$$

Per determinare lo stimatore di ϑ con il metodo dei momenti, bisogna risolvere l'equazione

$$\mu_1' = M_1' \Longrightarrow \mu_1' = \overline{X}_n,$$

rispetto a ϑ.
Calcoliamo $\mu_1' = \frac{0+2\vartheta}{2} = \vartheta$, sostituendo nell'equazione precedente risulta $\overline{X}_n$ uno stimatore di ϑ. Essendo $E\left[\overline{X}_n\right] = \vartheta$, lo stimatore è non distorto.

2) Calcoliamo

$$\operatorname{var}\left[\overline{X}_n\right] \underbrace{=}_{X_i \text{ indip.}} \frac{1}{n^2}\sum_{i=1}^{n} \operatorname{var}\left[X_i\right] = \frac{1}{n^2} \cdot n \cdot \frac{\vartheta^2}{3} = \frac{\vartheta^2}{3n}.$$

Ne segue

$$\lim_{n \longrightarrow +\infty} E\left[\left(\overline{X}_n - \vartheta\right)^2\right] = \lim_{n \longrightarrow +\infty} \frac{\vartheta^2}{3n} = 0,$$

quindi lo stimatore è consistente.

Esercizio 12 (Tema d'esame del 10/01/2006).

Sia $X_1,\ X_2\ \ldots X_n$ un campione casuale di ampiezza n estratto da una popolazione con densità di probabilità

$$f_X(x,\vartheta) = \begin{cases} \frac{2}{\vartheta}\left(1 - \frac{x}{\vartheta}\right) & \text{se } 0 \leq x \leq \vartheta \\ 0 & \text{altrove} \end{cases}$$

con $\vartheta > 0$.

1. Determinare uno stimatore T di ϑ con il metodo dei momenti.
2. Stabilire se T è distorto e calcolarne l'errore quadratico medio.

Risoluzione. Individuiamo lo stimatore T del parametro ϑ, risolvendo rispetto a ϑ

$$\overline{X}_n = E[X].$$

Calcoliamo

$$E[X] = \int_0^{\vartheta} \frac{2}{\vartheta} x \left(1 - \frac{x}{\vartheta}\right) d\,x = \frac{1}{3}\vartheta$$

quindi

–

$$\overline{X}_n = \frac{1}{3}\vartheta \Longrightarrow T = 3\overline{X}_n.$$

Lo stimatore è non distorto, infatti, $E\left[3\overline{X}_n\right] = 3E\left[\overline{X}_n\right] = 3\frac{\vartheta}{3} = \vartheta$
Calcoliamo, infine, l'errore quadratico medio

$$\mathrm{MSE}\left[T\right] = \mathrm{var}\left[T\right] = \mathrm{var}\left[3\overline{X}_n\right] = 9\frac{\mathrm{var}\left[X\right]}{n}.$$

$$E\left[X^2\right] = \int_0^{\vartheta} \frac{2}{\vartheta}x^2\left(1 - \frac{x}{\vartheta}\right)d\,x = \frac{\vartheta^2}{6} \Longrightarrow \mathrm{var}\left[X\right] = \frac{\vartheta^2}{6} - \frac{\vartheta^2}{9} = \frac{\vartheta^2}{18}.$$

Ne segue $\mathrm{MSE}\left[T\right] = \frac{\vartheta^2}{2n}$.

Esercizio 13 (Tema d'esame del 12/07/2005).
Sia $X_1,\ X_2\ \ldots X_n$ un campione casuale di ampiezza n estratto da una popolazione con densità di probabilità

$$f_X(x, \vartheta) = \begin{cases} 5^{-\vartheta}\vartheta\, x^{\vartheta-1} & \text{se } 0 < x < 5 \\ 0 & \text{altrove} \end{cases}$$

con $\vartheta > 0$. Determinare lo stimatore di massima verosimiglianza $\widehat{\Theta}$ di ϑ.

Risoluzione. Calcoliamo la funzione di verosimiglianza

$$L(x_1, x_2, \cdots, x_n, \vartheta) = \prod_{i=1}^{n} f(x_i, \vartheta) = 5^{-n\vartheta}\vartheta^n \prod_{i=1}^{n} x_i^{\vartheta-1}$$

con $0 < x_i < 5,\ i = 1, 2, \ldots, , n$. Quindi

$$\ln L = -n\vartheta \ln 5 + n \ln \vartheta + (\vartheta - 1) \sum_{i=1}^{n} \ln x_i \quad \text{e}$$

–

$$\frac{d}{d\vartheta}\ln L = -n\ln 5 + \frac{n}{\vartheta} + \sum_{i=1}^{n}\ln x_i.$$

Poniamo

$$\frac{d}{d\vartheta}\ln L = 0 \Longrightarrow \widehat{\vartheta} = \frac{n}{n\ln 5 - \sum\limits_{i=1}^{n}\ln x_i}.$$

Inoltre

$$\left.\frac{d^2}{d\vartheta^2}\ln L\right|_{\vartheta = \widehat{\vartheta}} = -\frac{n}{\widehat{\vartheta}^2} < 0,$$

quindi

$$\widehat{\Theta} = \frac{n}{n\ln 5 - \sum\limits_{i=1}^{n}\ln x_i}.$$

Esercizio 14 (Tema d'esame del 28/06/2005).

Sia $X_1, \ldots, X_n$ un campione casuale, di dimensione n, estratto da una distribuzione rettangolare uniforme sull'intervallo $[a, 2a]$.

1. Determinare uno stimatore T_1 di a con il metodo dei momenti. Verificare se lo stimatore T_1 è distorto e calcolarne l'errore quadratico medio $MSE[T_1]$.
2. Considerato poi lo stimatore $T_2 = \dfrac{1}{2}X_1 + \dfrac{1}{6}X_2$, verificare se T_2 è distorto e calcolarne l'errore quadratico medio $MSE[T_2]$.
3. Supposto $n = 3$, quale dei due stimatori T_1 e T_2 di a è preferibile (giustificare la risposta)?

$$\left[\begin{array}{l} T_1 = \frac{2}{3}\overline{X}_n \\ T_1 \text{ non distorto} \\ MSE[T_1] = \frac{a^2}{27n} \\ T_2 \text{ non distorto} \\ MSE[T_2] = \frac{5a^2}{216} \\ T_1 \text{ preferibile} \end{array}\right]$$

Esercizio 15 (Tema d'esame del 11/04/2006).

Determinare lo stimatore di massima verosimiglianza del parametro λ per un campione estratto dalla densità

$$f_X(x) = \begin{cases} \lambda^2 (x-1) e^{-\lambda(x-1)} & \text{se } x > 1, \\ 0 & \text{altrove.} \end{cases}$$

$$\left[T = \frac{2}{\bar{X}_n - 1} \right]$$

Esercizio 16 (Tema d'esame del 04/07/2006).

Sia $X_1, \ldots, X_8$ un campione aleatorio, di dimensione 8, estratto da una distribuzione rettangolare uniforme sull'intervallo $[-1, b]$, con $b > -1$.

Si chiede:

1. determinare uno stimatore T_1 di b con il metodo dei momenti;
2. determinare se lo stimatore T_1 sia distorto;
3. calcolare l'errore quadratico medio $\mathrm{MSE}[T_1]$;
4. considerato poi lo stimatore $T_2 = 4\overline{X}_8 - X_3 - X_5 + 1$, calcolarne l'errore quadratico medio $\mathrm{MSE}[T_2]$;
5. determinare quale dei due stimatori T_1 e T_2 di b sia preferibile, giustificando la risposta.

$$\begin{bmatrix} T_1 = 2\overline{X}_8 + 1 \\ T_1 \text{ non distorto} \\ MSE[T_1] = \frac{(b+1)^2}{24} \\ MSE[T_2] = \frac{(b+1)^2}{6} \\ T_1 \text{ preferibile} \end{bmatrix}$$

Esercizio 17 (Tema d'esame del 05/04/2005).

Sia X una variabile casuale normale con media μ e varianza σ^2. Siano X_1, X_2, X_3 le variabili casuali indipendenti descritte dalle tre determinazioni x_1, x_2, x_3 di un campione casuale estratto da essa. Per stimare il parametro μ si considerano i due seguenti stimatori:

$$\overline{X}_3 = \frac{X_1 + X_2 + X_3}{3}\,, \quad T = \frac{1}{5}X_1 + \frac{1}{5}X_2 + \frac{3}{5}X_3\,.$$

1. Dire se $\overline{X}_3$ e T sono stimatori non distorti di μ e motivare la risposta.
2. Calcolare $\mathrm{MSE}[\overline{X}_3]$ e $\mathrm{MSE}[T]$ e stabilire quale tra i due stimatori $\overline{X}_3$ e T di μ sia preferibile, motivando la risposta.

$$\left[\begin{array}{l} \overline{X}_3 \text{ non distorto} \\ T \text{ non distorto} \\ MSE[\overline{X}_3] = \frac{\sigma^2}{3} \\ MSE[T] = \frac{11\sigma^2}{25} \\ \overline{X}_3 \text{ preferibile} \end{array}\right]$$