Foglio di esercizi 7

Discussione soluzioni: 11.05.2022

1. Sia Y_i il valore ottenuto lanciando un dado a sei facce. Sia X_i l' *ultima* cifra di Y_i^2. (Per esempio, se $Y_i = 4$ allora $Y_i^2 = 16$ e $X_i = 6$). Definiamo $\bar{X}_n$ come

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n} \tag{1}$$

   (a) Calcolare il valore atteso e la varianza di X_i
   (b) Si predica il valore di $\bar{X}_n$ quando n è molto grande (a che numero sarà molto vicino?)
   (c) Si trovi il valore di n tale che il valore predetto al punto precendente per $\bar{X}_n$ sia corretto con una tolleranza di 0.01 con una probabilità di almeno 0.99. (Non bisogna usare la correzione di contiinuità in questo caso). (Suggerimento: ricorda che $\Phi(2.5) - \Phi(-2.5) \approx 0.99$, *i.e.,* the l'intervallo di confidenza bivariato al 99% è di circa 2.5σ.)
   (d) Sia $S_{100} = X_1 + X_2 + \cdots + X_{100}$. Se si dovesse predire la *prima* cifra di S_{100}, che cifra andrebbe scelta per mnassimizzare la probabilità di essere corretti? Quanto vale questa probabilità, approssimativamente? (Non c'è bisogno di usare la correzione di continuità).

2. Si consideri un campione $\{X_1, \ldots, X_n\}$ di variabili aleatorie indipendenti e identicamente distribuite (iid) con distribuzione di Bernoulli con parametro p: $X_i \sim Ber(p)$.

   (a) Verificare che la funzione di massa della distribuzione di Bernoulli può essere scritta come $p_X(x|p) = p^x(1-p)^{1-x}$ per $x \in \text{range}(X_i) = \{0, 1\}$.
   (b) Usando l'espressione ottenuta al punto precedente, calcolare lo stimatore di massima verosimiglianza per p.
   (c) Si tratta di uno stimatore corretto? Lo stimatore è consistente?

3. Si consideri un campione $\{X_1, \ldots, X_n\}$ di variabili aleatorie indipendenti e identicamente distribuite (iid) con dentsità

$$f_X(x|\lambda) := \begin{cases} \lambda e^{-\lambda x} & \text{se } x \geq 0 \\ 0 & \text{altrimenti} \end{cases}$$

   dove $\lambda > 0$ è il parametro della distribuzione. Questa distribuzione di probabilità è detta la distribuzione *esponenziale* con parametro λ.

   (a) Controllare che la funzione $f_X(x|\lambda)$ sia una densità di probabilità per ogni scelta di $\lambda > 0$
   (b) Sia X una variabile aleatoria con densità $f_X(x|\lambda)$. Si calcoli il valore atteso e la varianza di X
   (c) Si calcoli lo stimatore di massima verosimiglianza $\hat{\lambda}_n$ di λ per il campione $\{X_1, \ldots, X_n\}$
   (d) Quanto vale $\lim_{n\to\infty} \mathbb{P}(\frac{1}{\lambda} - \frac{1}{\hat{\lambda}_n} > \epsilon)$? A quale valore converge $\hat{\lambda}_n$?

4. Sia $X_1, \ldots, X_n$ un campione casuale estratto da una legge uniforme sull'intervallo (continuo) $[0, \theta]$ (il parametro della distribuzione è quindi θ). Determinare lo stimatore di massima verosimiglianza per θ. Suggerimento: Il massimo della funzione $L(x_1, \ldots,_n |\theta) = f(x_1, \ldots, x_n|\theta)$ in questo caso non si trova differenziando, ma facendo attenzione all'intervallo di θ dove $\mathcal{L}(x_1, \ldots, x_n|\theta) \neq 0$. Questo intervallo dipende da $\{x_1, \ldots, x_n\}$

5. Si consideri una popolazione di distribuzione (continua) $Unif(-\theta, 4\theta)$ e si proponga per θ uno stimatore $\hat{\theta}$ basato su un campione casuale $X_1, \ldots, X_{2000}$. Cliccando qui si trova un file csv contenente un campione di 2000 dati. Si dia una stima per θ usando $\hat{\theta}(x_1, \ldots, x_{2000})$ e i dati del file.

6. Si consideri un campione $\{X_1, \ldots, X_n\}$ di variabili aleatorie indipendenti e identicamente distribuite (iid) con dentsità

$$f_X(x|\theta) := \begin{cases} \theta 2^\theta x^{-\theta-1} & \text{per } x \geq 2 \\ 0 & \text{altrimenti} \end{cases}$$

dove $\theta > 0$ è il parametro della distribuzione.

   (a) Controllare che la funzione $f_X(x|\theta)$ sia una densità di probabilità per ogni scelta di $\theta > 0$

   (b) Sia X una variabile aleatoria con densità $f_X(x|\theta)$. Si calcoli il valore atteso e la varianza di X

   (c) Si calcoli lo stimatore di massima verosimiglianza $\hat{\theta}_n$ di θ per il campione $\{X_1, \ldots, X_n\}$

7. (Da fare dopo la lezione di martedi) Si discuta se lo stimatore di massima verosimiglianza per il parametro λ di una distribuzione di poisson è corretto e/o consistente.

1. Sia Y_i il valore ottenuto lanciando un dado a sei facce. Sia X_i l' *ultima* cifra di Y_i^2. (Per esempio, se $Y_i = 4$ allora $Y_i^2 = 16$ e $X_i = 6$). Definiamo $\bar{X}_n$ come

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n} \tag{1}$$

   (a) Calcolare il valore atteso e la varianza di X_i

   (b) Si predica il valore di $\bar{X}_n$ quando n è molto grande (a che numero sarà molto vicino?)

   (c) Si trovi il valore di n tale che il valore predetto al punto precendente per $\bar{X}_n$ sia corretto con una tolleranza di 0.01 con una probabilità di almeno 0.99. (Non bisogna usare la correzione di contiinuità in questo caso). (Suggerimento: ricorda che $\Phi(2.5) - \Phi(-2.5) \approx 0.99$, *i.e.*, the l'intervallo di confidenza bivariato al 99% è di circa 2.5σ.)

   (d) Sia $S_{100} = X_1 + X_2 + \cdots + X_{100}$. Se si dovesse predire la *prima* cifra di S_{100}, che cifra andrebbe scelta per mnassimizzare la probabilità di essere corretti? Quanto vale questa probabilità, approssimativamente? (Non c'è bisogno di usare la correzione di continuità).

a) Ogni valore di Y è trasformato in uno di X

$$Y \qquad \begin{array}{ccc} 1 & \longrightarrow & 1 \\ 2 & \longrightarrow & 4 \\ 3 & \longrightarrow & 9 \\ 4 & \longrightarrow & 6 \\ 5 & \longrightarrow & 5 \\ 6 & \nearrow & \end{array} \qquad X$$

$\hookrightarrow \text{range}(X) = \{1,4,9,6,5\}$

Se la distribuzione di Y era uniforme, la pdm di Y è

$$p_X(1) = \mathbb{P}(Y=1) = \frac{1}{6}$$
$$p_X(4) = \mathbb{P}(Y=2) = \frac{1}{6}$$
$$p_X(9) = \mathbb{P}(Y=3) = \frac{1}{6}$$
$$p_X(6) = \mathbb{P}(Y=4) + \mathbb{P}(Y=6) = \frac{1}{3}$$
$$p_X(5) = \mathbb{P}(Y=5) = \frac{1}{6}$$

$$\Longrightarrow \quad \mathbb{E}(X) = \sum_{x \in \text{range}(X)} x \, p_X(x) = \frac{31}{6} = 5.1\bar{6}$$

$$\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = 32.5 - (4.5)^2 \doteq 5.805$$

b) Per la legge dei grandi numeri $\bar{X}_n \approx \mathbb{E}(X_1) = 5.1\bar{6}$

c) Usiamo il teorema del limite centrale nella forma generale

$$\overline{X}_n \underset{n \gg 1}{\approx} N(\mu^*, (\sigma_n^*)^2) \quad \text{con} \quad \begin{cases} \mu^* = E(X_j) = 5.16 \\ (\sigma_n^*)^2 = \frac{Var(X_j)}{n} = \frac{5.805}{n} \end{cases}$$

$$\longrightarrow \frac{\overline{X}_n - \mu^*}{\sigma_n^*} \approx Z \sim N(0,1)$$

quindi: $99\% = P(|Z| \leq 2.5) \approx P\left(\left|\frac{\overline{X}_n - \mu^*}{\sigma_n^*}\right| \leq 2.5\right)$

$$= P(\underbrace{|\overline{X}_n - \mu^*|}_{\text{errore}} \leq \underbrace{2.5\,\sigma_n^*}_{\text{tolleranza}})$$

vogliamo che la tolleranza sia di 0.01 $\longrightarrow 2.5\,\sigma_n^* = 0.01$

$$\Longleftrightarrow 2.5\sqrt{\frac{5.805}{n}} = 0.01$$

$$\Longleftrightarrow n = \left(\frac{2.5\sqrt{5.805}}{0.01}\right)^2$$

$$= 362487$$

n minimo

d) Per la legge dei grandi numeri
ci aspettiamo

$$S_n = n\overline{X}_n = 100\,\overline{X}_{100} \approx 100 \cdot 5{,}16 = 516$$

$\longrightarrow \overline{X}_n$ si concentra intorno a questo # quindi la prima cifra dovrebbe essere 5.

Usiamo il CLT per approssimare la P che questo capiti

$$\overline{X}_n \approx N(\mu^*, (\sigma_n^*)^2) \longrightarrow S_n = n\overline{X}_n \approx N(n\mu^*, \underbrace{n^2(\sigma_n^*)^2}_{n(\sigma^*)^2})$$

$$\xrightarrow{\text{std.}} \frac{S_n - n\mu^*}{\sqrt{n}\,\sigma^*} \approx Z \sim N(0,1)$$

$$\mathbb{P}(\text{prima cifra} = 5) \approx \mathbb{P}(S_n \in [500, 599]) + \underbrace{\mathbb{P}(S_n \in [5000, 5999]) + \dots}_{\text{molto piccoli}}$$

$$\approx \mathbb{P}(500 \le S_n \le 599) = \mathbb{P}\Big(\underbrace{\frac{500 - n\mu^*}{\sqrt{n}\,\sigma^*}}_{-0.6917} \le Z \le \underbrace{\frac{599 - n\mu^*}{\sqrt{n}\,\sigma^*}}_{3.4170}\Big)$$

$$= \Phi(3.4170) - \Phi(-0.6917) = 0.7551$$

$\longrightarrow$ $\mathbb{P}$ di aver ragione è circa 75%.

2. Si consideri un campione $\{X_1, \dots, X_n\}$ di variabili aleatorie indipendenti e identicamente distribuite (iid) con dentsità

$$f_X(x|\lambda) := \begin{cases} \lambda e^{-\lambda x} & \text{se } x \geq 0 \\ 0 & \text{altrimenti} \end{cases}$$

dove $\lambda > 0$ è il parametro della distribuzione. Questa distribuzione di probabilità è detta la distribuzione *esponenziale* con parametro λ.

(a) Controllare che la funzione $f_X(x|\lambda)$ sia una densità di probabilità per ogni scelta di $\lambda > 0$

(b) Sia X una variabile aleatoria con densità $f_X(x|\lambda)$. Si calcoli il valore atteso e la varianza di X

(c) Si calcoli lo stimatore di massima verosimiglianza $\hat{\lambda}_n$ di λ per il campione $\{X_1, \dots, X_n\}$

(d) Quanto vale $\lim_{n\to\infty} \mathbb{P}(\frac{1}{\lambda} - \frac{1}{\hat{\lambda}_n} > \epsilon)$? A quale valore converge $\hat{\lambda}_n$?

a) • $f_X(x|\lambda) \geq 0$ ✓

• $\int_{-\infty}^{\infty} f_X(x|\lambda)\,dx = \int_0^{\infty} \lambda e^{-\lambda x}\,dx = \lambda \int_0^{\infty} e^{-\lambda x}\,dx = \lambda \cdot \frac{1}{\lambda} = 1$ ✓

b) $\mathbb{E}(X) = \int_0^{\infty} x \lambda e^{-\lambda x}\,dx = \frac{1}{\lambda}$

$Var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \int_0^{\infty} x^2 \lambda e^{-\lambda x}\,dx - \frac{1}{\lambda^2} = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$

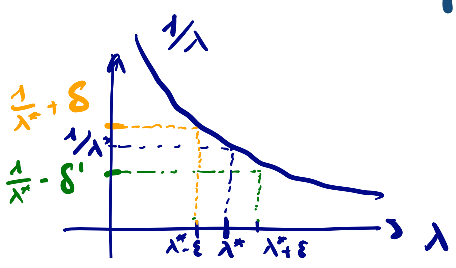
c) $\mathcal{L}(x_1 \dots x_n; \lambda) = \prod_{i=1}^{n} f_X(x_i|\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^n x_i} = e^{n \log\lambda - \lambda \sum_{i=1}^n x_i}$

$$\frac{d}{d\lambda}\left(n \log\lambda - \lambda \sum_i x_i\right) = 0$$

$$\Longleftrightarrow \quad \frac{n}{\hat{\lambda}_n} = \sum_{i=1}^n x_i \quad \Longleftrightarrow \quad \hat{\lambda}_n = \frac{1}{\frac{1}{n}\sum_{i=1}^n x_i}$$

d) $\mathbb{P}\left(\left|\frac{1}{\lambda^*} - \frac{1}{\hat{\lambda}_n}\right| > \varepsilon\right) = \mathbb{P}\left(\left|\frac{1}{\lambda^*} - \frac{1}{n}\sum x_i\right| > \varepsilon\right) \xrightarrow[n\to\infty]{LLN} 0$

Siccome $\lambda^* > 0$ fisso ed $\varepsilon > 0$ piccolo abbiamo



$$\{\lambda : |\lambda^* - \lambda| < \varepsilon\} \subseteq \{\lambda : |\frac{1}{\lambda} - \frac{1}{\lambda^*}| < \max(\delta, \delta')\}$$

$$\Longrightarrow \mathbb{P}\left(\left|\lambda^* - \frac{1}{\frac{1}{n}\sum_i X_i}\right| < \varepsilon\right) \leq \mathbb{P}\left(\left|\frac{1}{\lambda^*} - \frac{1}{n}\sum X_i\right| \leq \underbrace{\max(\delta, \delta')}_{\varepsilon'}\right)$$

$$= \mathbb{P}\left(\left|\frac{1}{\lambda^*} - \frac{1}{n}\sum_i X_i\right| \leq \varepsilon'\right) \xrightarrow[LLN]{n \to \infty} 0$$

$\forall \varepsilon' > 0$

Si consideri un campione $\{X_1, \ldots, X_n\}$ di variabili aleatorie indipendenti e identicamente distribuite (iid) con distribuzione di Bernoulli con parametro p: $X_i \sim Ber(p)$.

(a) Verificare che la funzione di massa della distribuzione di Bernoulli può essere scritta come $p_X(x|p) = p^x(1-p)^{1-x}$ per $x \in \text{range}(X_i) = \{0, 1\}$.

(b) Usando l'espressione ottenuta al punto precedente, calcolare lo stimatore di massima verosimiglianza per p.

(c) Si tratta di uno stimatore corretto? Lo stimatore è consistente?

a) Sappiamo che $p_X(x) = \begin{cases} p & \text{se } x = 1 \\ 1-p & \text{se } x = 0 \end{cases}$

Verifichiamo $\begin{cases} p_X(1|p) = p^1(1-p)^{1-1} = p & x = 1 \\ p_X(0|p) = p^0(1-p)^{1-0} = 1-p & x = 0 \end{cases}$ ✓

b) $$\mathcal{L}(x_1, \ldots x_n | p) = \prod_{j=1}^{n} p_X(x_j|p) = \prod_{j=1}^{n} p^{x_j}(1-p)^{1-x_j}$$

$$= \left(\prod_{j=1}^{n} p^{x_j}\right)\left(\prod_{j=1}^{n}(1-p)^{1-x_j}\right)$$

$$= p^{\sum_{j=1}^{n} x_j} \cdot (1-p)^{\sum_{j=1}^{n}(1-x_j)} =$$

$$= \exp\left(\log p \sum_{j=1}^{n} x_j + \log(1-p)\left(n - \sum_{j=1}^{n} x_j\right)\right)$$

Questa funzione è continua in p, quindi il massimo si trova con $\frac{d}{dp} = 0$ (consideriamo soltanto l'esponente come al solito)

$$\frac{d}{dp}\left(\log p \sum_{j=1}^{n} x_j + \log(1-p)\left(n - \sum_{j=1}^{n} x_j\right)\right) = 0$$

$$\Longleftrightarrow \frac{1}{p}\sum_{j=1}^{n} x_j - \frac{1}{1-p}\left(n - \sum_{j=1}^{n} x_j\right) = 0$$

$$\Longleftrightarrow \frac{1}{p}\sum_{j=1}^{n} x_j = \frac{1}{1-p}\left(n - \sum_{j=1}^{n} x_j\right) \Longleftrightarrow \frac{1-p}{p} = \frac{n - \sum_{j=1}^{n} x_j}{\sum_{j=1}^{n} x_j}$$

$$\Longleftrightarrow \frac{1}{p} - \cancel{1} = \frac{n}{\sum_{j=1}^{n} x_j} - \cancel{1} \Longleftrightarrow \hat{p} = \frac{1}{n}\sum_{j=1}^{n} x_j$$

c) $$\mathbb{E}_p(\hat{p}) = \mathbb{E}_p\left(\frac{1}{n}\sum_{j=1}^{n} X_j\right) = \frac{1}{n}\sum_{j=1}^{n} \underbrace{\mathbb{E}_p(X_j)}_{p} = \frac{1}{n}\cdot n\cdot p = p$$

$\Longrightarrow$ corretto

$$\lim_{n\to\infty} \mathbb{P}_p(|\hat{p}_n - p| > \varepsilon) = \lim \mathbb{P}_p\left(\left|\frac{1}{n}\sum_{j=1}^{n} x_j - \mathbb{E}_p(X_j)\right| > \varepsilon\right) = 0$$

legge grandi n

$\Longrightarrow$ consistente

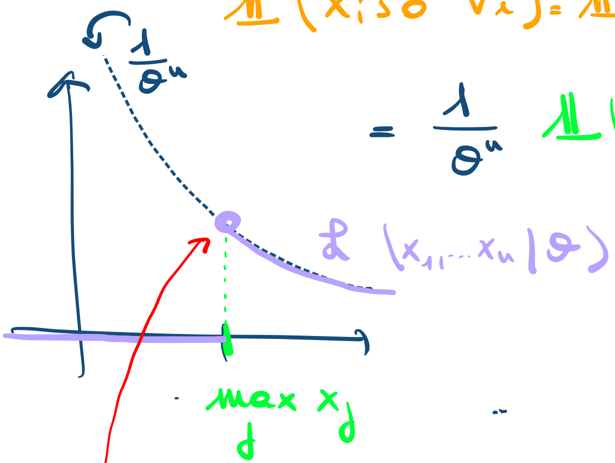
Sia $X_1, \ldots, X_n$ un campione casuale estratto da una legge uniforme sull'intervallo (continuo) $[0, \theta]$ (il parametro della distribuzione è quindi θ). Determinare lo stimatore di massima verosimiglianza per θ. Suggerimento: Il massimo della funzione $L(x_1, \ldots, _n|\theta) = f(x_1, \ldots, x_n|\theta)$ in questo caso non si trova differenziando, ma facendo attenzione all'intervallo di θ dove $\mathcal{L}(x_1, \ldots, x_n|\theta) \neq 0$. Questo intervallo dipende da $\{x_1, \ldots, x_n\}$

$$f_X(x|\theta) = c_\theta \, \mathbb{1}(x \in (0,\theta)) \xRightarrow{\int_{-\infty}^{\infty} f_X(x|\theta)\,dx = 1} c_\theta = \frac{1}{\theta}$$

$$\mathcal{L}(x_1 \ldots x_n; \theta) = \prod_{i=1}^{n} f_X(x_i|\theta) = \left(\frac{1}{\theta}\right)^n \prod_{i=1}^{n} \underbrace{\mathbb{1}(x_i \in (0,\theta))}_{\mathbb{1}(x_i > 0) \cdot \mathbb{1}(x_i < \theta)}$$

$$= \frac{1}{\theta^n} \underbrace{\left(\prod_{i=1}^{n} \mathbb{1}(x_i > 0)\right)}_{\mathbb{1}(x_i > 0 \ \forall i) = \mathbb{1}(\min_i x_i > 0)} \underbrace{\left(\prod_{j=1}^{n} \mathbb{1}(\theta > x_j)\right)}_{\mathbb{1}(\theta > x_j \ \forall j) = \mathbb{1}(\theta > \max_j x_j)}$$

$$= \frac{1}{\theta^n} \mathbb{1}(\theta > \max_j x_j) \, \mathbb{1}(0 < \min_i x_i)$$



$$\Longrightarrow \hat{\theta}(x_1 \ldots x_n) = \max_j (x_j) = \max(x_1, \ldots, x_n)$$

il massimo della funzione $\mathcal{L}(x_1 \ldots x_n|\theta)$ è raggiunto al suo punto di discontinuità

Si consideri un campione $\{X_1, \dots, X_n\}$ di variabili aleatorie indipendenti e identicamente distribuite (iid) con dentsità

$$f_X(x|\theta) := \begin{cases} \theta 2^\theta x^{-\theta-1} & \text{per } x \geq 2 \\ 0 & \text{altrimenti} \end{cases}$$

dove $\theta > 0$ è il parametro della distribuzione.

(a) Controllare che la funzione $f_X(x|\theta)$ sia una densità di probabilità per ogni scelta di $\theta > 0$

(b) Sia X una variabile aleatoria con densità $f_X(x|\theta)$. Si calcoli il valore atteso e la varianza di X

(c) Si calcoli lo stimatore di massima verosimiglianza $\hat{\theta}_n$ di θ per il campione $\{X_1, \dots, X_n\}$

a) • $f_X(x|\theta) = \begin{cases} \theta 2^\theta x^{-\theta-1} & x \geq 2 \\ 0 & \text{altrimenti} \end{cases} \geq 0$ ✓

• $\int_{-\infty}^{\infty} f_X(x|\theta)\,dx = \int_2^\infty \theta 2^\theta x^{-\theta-1}\,dx = 2^\theta \int_2^\infty \theta x^{-\theta-1}\,dx = 2^\theta \cdot \frac{1}{2^\theta} = 1$ ✓

b) $\mathbb{E}(X) = \int_2^\infty x\,\theta 2^\theta x^{-\theta-1}\,dx = \frac{\theta 2^\theta}{\theta-1} \int_2^\infty \theta-1\, x^{-(\theta-1)-1}\,d\alpha$

$= \frac{\theta}{\theta-1} 2^\theta \cdot \frac{1}{2^{\theta-1}} = 2\frac{\theta}{\theta-1}$

$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \int x^2 \theta 2^\theta x^{-\theta-1}\,dx - \left(\frac{2\theta}{\theta-1}\right)^2$

$= 4\left(\frac{\theta}{\theta-2} - \left(\frac{\theta}{\theta-1}\right)^2\right)$

c) $\mathcal{L}(x_{\dots} x_n, \theta) = \theta^n 2^{n\theta} \prod x_i^{-\theta-1}$

$= e^{n\log\theta + n\theta\log 2 - (\theta+1)\sum_i \log(x_i)}$

$\frac{d}{d\theta}\left(n\log\theta + n\theta\log 2 - (\theta+1)\sum_i \log x_i\right) = 0$

$\Longleftrightarrow \frac{n}{\theta} = \sum_i \log x_i - n\log 2 \Longleftrightarrow \hat{\theta} = \frac{1}{\frac{1}{n}\sum_i \log x_i - \log 2}$

. (Da fare dopo la lezione di martedi) Si discuta se lo stimatore di massima verosimiglianza per il parametro λ di una distribuzione di poisson è corretto e/o consistente.

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{(visto in classe)}$$

$$\mathbb{E}(\hat{\lambda}) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \frac{1}{n}\sum_{i=1}^{n}\lambda = \lambda \quad \checkmark$$

$\Rightarrow$ corretto

$$\lim_{n\to\infty} \mathbb{P}(|\hat{\lambda} - \lambda| > \varepsilon) = \lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum X_i - \mathbb{E}(X_i)\right| > \varepsilon\right)$$

$$= 0$$

legge dei grandi numeri (LLN)

$\Rightarrow$ consistente