Probabilità e Statistica

Docente: Prof. Elisabetta Scoppola

Foglio di esercizi 6

TA: Matteo Quattropani
matteo.quattropani@uniroma3.it;

22 gennaio 2018

Teorema del Limite Centrale, Media e Varianza Campionarie

Esercizio 12.1 Siano $Z_1, \dots, Z_n$ v.a.i.i.d. con distribuzione esponenziale di parametro 2. E sia

$$X_n = \sum_{i=1}^{n} Z_i.$$

Quanto vale, nell'approssimazione per n grande, la probabilità dell'evento $\{X_n \geq n/2\}$?
Soluzione. Notiamo che

$$\mathbb{E}[X_n] = n\mathbb{E}[Z_1] = \frac{n}{2}.$$

Per il TLC abbiamo

$$\mathbb{P}(X_n \geq n/2) = \mathbb{P}\left(X_n - \mathbb{E}[X_n] \geq 0\right) = \mathbb{P}\left(\frac{X_n - \mathbb{E}[X_n]}{\sqrt{\mathrm{Var}(X_n)}} \geq 0\right) \approx \frac{1}{2}.$$

Esercizio 12.2 Il numero di volte in cui un telefono squilla in un giorno è descritto da una v.a. di Poisson di parametro 2. Supponiamo che il numero di chiamate ricevute in un giorno sia indipendente dal numero di chiamate ricevute in un qualsiasi altro giorno. Sia $N(k)$ il numero di telefonate ricevute nei primi k giorni.

(1) Per k fissato, si stimi la probabilità dell'evento $\{N(k) \geq 3k\}$ (utilizza la disuguaglianza di Markov).
(2) Si stimi la stessa probabilità del punto precedente utilizzando la disuguaglianza di Chebyshev.
(3) Utilizzando il TLC si calcoli, nel limite $k \to \infty$ la probabilità dell'evento $\{N(k) \geq 2k + 3\sqrt{k}$.

Soluzione. $N(k)$ è la somma di variabili di Poisson indipendenti, pertanto è a sua volta una v.a. di Poisson di parametro $2k$. Ne segue che

$$\mathbb{E}[N(k)] = \mathrm{Var}(N(k)) = 2k.$$

(1) $\mathbb{P}(N(k) \geq 3k) \leq \frac{\mathbb{E}[N(k)]}{3k} = \frac{2}{3}$.
(2)

$$\begin{aligned}\mathbb{P}(N(k) \geq 3k) =& \mathbb{P}(N(k) - \mathbb{E}[N(k)] \geq k) \\ \leq& \mathbb{P}(|N(k) - \mathbb{E}[N(k)]| \geq k) \\ \leq& \frac{\mathrm{Var}(N(k))}{k^2} \\ =& \frac{2}{k}.\end{aligned}$$

(3) Sia

$$Z_k = \frac{N(k) - \mathbb{E}[N(k)]}{\sqrt{\mathrm{Var}(N(k))}}$$

per il TLC sappiamo che $Z_k \xrightarrow{d} \mathcal{N}(0,1)$ Allora

$$\begin{aligned}\mathbb{P}(N(k) \geq 2k + 3\sqrt{k}) =& \mathbb{P}\left(Z_k \geq \frac{3\sqrt{k}}{\sqrt{\mathrm{Var}(N(k))}}\right)\\ =& \mathbb{P}(Z_k \geq 3/\sqrt{2}) \to 1 - \Phi(3/\sqrt{2}) \approx 0.12\end{aligned}$$

Esercizio 12.3 Siano date $X_1, \ldots X_n, \ldots$ v.a.i.i.d. t.c.

$$\mathbb{P}(X_1 = 0) = \mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = \frac{1}{3}.$$

Si consideri

$$S_n = \sum_{i=1}^{n} X_i.$$

Stimare, prima con Chebyshev e poi con il TLC, la probabilità dell'evento $\{|S_{100}| \geq 10\}$.
Soluzione. Si noti che

$$\mathbb{E}[S_n] = 0, \qquad \mathrm{Var}(S_n) = n\mathbb{E}[X_1^2] = \frac{2}{3}n.$$

Utilizzando la disuguaglianza di Chebyshev otteniamo

$$\mathbb{P}(|S_n| \geq k) = \mathbb{P}(|S_n - \mathbb{E}[S_n]| \geq k) \leq \frac{\mathrm{Var}(S_n)}{k^2}$$

che nel nostro caso si traduce in $2/3$.
Si definisca

$$Z_n = \frac{S_n}{\sqrt{\mathrm{Var}(S_n)}}.$$

Allora

$$\mathbb{P}(|S_n| \geq k) = \mathbb{P}(|Z_n| \geq k\sqrt{\mathrm{Var}(S_n)})$$

dunque

$$\mathbb{P}(|S_{100}| \geq 10) = \mathbb{P}(|Z_{100}| \geq \sqrt{3/2}) \approx 2(1 - \Phi(\sqrt{3/2})) \approx 0.22.$$

Esercizio 12.4 Sia $\mathcal{Q}$ il quadrato di lato 2 centrato nell'origine, e sia $\mathcal{R}$ il rettangolo di base 2 ed altezza 1 centrato nell'origine. Siano $P_1, \ldots, P_n$ punti indipendenti scelti uniformemente a caso in $\mathcal{Q}$.

(1) Calcolare la probabilità dell'evento $\{P_1 \in \mathcal{R}\}$.
(2) Detto

$$\bar{P}_n = \frac{1}{\sqrt{n}}(P_1 + \cdots + P_n)$$

calcolare la probabilità dell'evento $\{\bar{P}_n \in \mathcal{R}\}$ nel limite $n \to \infty$.

Soluzione.

(1) Chiamiamo $P_i = (X_i, Y_i)$ dove X_i, Y_i i.i.d. $Unif[-1,1]$, $\forall i \in \mathbb{N}$. Ne segue che

$$\mathbb{P}(P_1 \in \mathcal{R}) = \mathbb{P}(X_1 \in [-1,1] \cap Y_1 \in [-1/2, 1/2]) = 1 \cdot \frac{1}{2} = \frac{1}{2}.$$

(2) Notiamo che

$$\mathbb{E}[X_i] = \mathbb{E}[Y_i] = 0, \qquad \mathrm{Var}[X_i] = \mathrm{Var}[Y_i] = \frac{1}{3}.$$

Dunque per il TLC $\sqrt{3}\bar{P}_n \xrightarrow{d} (Z_1, Z_2)$ dove $Z_1, Z_2 \sim \mathcal{N}(0,1)$ i.i.d., e quindi

$$\begin{aligned}\mathbb{P}(\bar{P}_n \in \mathcal{R}) \to& \mathbb{P}(Z_1 \in [-\sqrt{3}, \sqrt{3}])\mathbb{P}(Z_2 \in [-\sqrt{3}/2, \sqrt{3}/2])\\ =& (2\Phi(\sqrt{3}) - 1)(2\Phi(\sqrt{3}/2) - 1).\end{aligned}$$

Esercizio 12.5 Siano $X_1, \dots, X_n$ v.a.i.i.d. t.c.

$$\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = \frac{1}{8}, \qquad \mathbb{P}(X_1 = 0) = \frac{3}{4}.$$

Sia $S_n = \sum_{i=1}^n X_i$. Calcolare

(1) $\lim_{n\to\infty} \mathbb{P}(S_n/n > 0.1)$
(2) $\lim_{n\to\infty} \mathbb{P}(S_n/\sqrt{n} > 0.1)$

Soluzione. Notiamo che $\mathbb{E}[S_n] = 0$, e $\mathrm{Var}(S_n) = n/4$ quindi

(1) Per la LGN abbiamo

$$\lim_{n\to\infty} \mathbb{P}(S_n/n > 0.1) \le \lim_{n\to\infty} \mathbb{P}(|S_n/n| > 0.1) = 0$$

(2) Per il TLC

$$\lim_{n\to\infty} \mathbb{P}(S_n/\sqrt{n} > 0.1) = \lim_{n\to\infty} \mathbb{P}(S_n/\sqrt{\mathrm{Var}(S_n)} > 0.2) = 1 - \Phi(0.2).$$

Esercizio 12.6 Sappiamo che il numero di biglietti acquistati da un tifoso in fila fuori dallo stadio è una qualche v.a. di media 2.4 e varianza 4. Supponiamo che prima di una partita ci siano 100 persone in fila al botteghino e che siano rimasti 250 biglietti disponibili. Dare un'approssimazione della probabilità dell'evento $A = \{$Ciascun tifoso in fila riesce ad acquistare i biglietti che vuole.$\}$.
Soluzione. Sia X_i, $i \le 100$ il numero di biglietti che l'i-esimo tifoso in fila intende acquistare. Sia $S = \sum_{i=1}^{100} X_i$. Allora $\mathbb{P}(A) = \mathbb{P}(S \le 250)$. Per il TLC

$$\mathbb{P}(S \le 250) \approx \mathbb{P}\left(Z \le \frac{250 - 100 \cdot (2.4)}{\sqrt{4 \cdot 100}} \right)$$

dove $Z \sim \mathcal{N}(0,1)$.
Eserizio 12.7 Dall'esperienza passata sappiamo che i voti dell'esame di statistica sono distributi approssimativamente come una v.a. normale con media $\mu = 24$ e varianza $\sigma^2 = 9$. Sapendo che la classe di quest'anno è composta da 81 persone, qual è la probabilità che la media campionaria sia compresa tra 23 e 25?
Soluzione. Sia X la media campionaria. Sappiamo che

$$\frac{X - 24}{3/9} \sim \mathcal{N}(0,1)$$

quindi

$$\mathbb{P}(23 \le X \le 25) = \mathbb{P}(-3 \le Z \le 3) = \Phi(3) - \Phi(-3),$$

dove $Z \sim \mathcal{N}(0,1)$.
Esercizio 12.8 La durata di un certo componente industriale è distribuita come una v.a. normale di media 100 ore e deviazione standard 6 ore. Qual è la probabilità che in un lotto di 100 pezzi la varianza campionaria dei componenti ecceda le 8 ore? Supponiamo ora che la varianza non sia nota, e che la radice della varianza campionaria del lotto sia di 6.4 ore. Qual è la probabilità che la media campionaria ecceda le 102 ore?
Soluzione. Sia S^2 la varianza campionaria. Sappiamo che

$$99\frac{S^2}{36} \sim \chi^2_{99}.$$

Ne segue che

$$\mathbb{P}(S^2 \ge 64) = \mathbb{P}(Y \ge \frac{64 \cdot 99}{36} = 176) \approx 3 \cdot 10^{-6},$$

dove $Y \sim \chi^2_{99}$.
Sia ora X la media campionaria, sappiamo che

$$\frac{X - 100}{6.4/10} \sim t_{99}.$$

Ne segue che

$$\mathbb{P}(X > 102) = \mathbb{P}(T > 2/(0.64) = 3.125) \approx 0.001,$$

dove $T \sim t_{99}$.

Verosimiglianza

Esercizio 13.1 Sia X una v.a. discrete tale che

$$\mathbb{P}(X=0) = \frac{2\theta}{3},\ \mathbb{P}(X=1) = \frac{\theta}{3},\ \mathbb{P}(X=2) = \frac{2(1-\theta)}{3},\ \mathbb{P}(X=3) = \frac{1-\theta}{3}.$$

Si noti che la legge qui sopra è ben definita per ogni scelta del parametro $\theta \in [0,1]$. Siano date 10 osservazioni indipendenti di tale v.a., i.e. $(3,0,2,1,3,2,1,0,2,1)$. Determinare lo stimatore di massima verosimiglianza per il parametro θ.

Soluzione. La funzione di verosimiglianza è data da

$$L(\theta) = \mathbb{P}(X=3)^2\mathbb{P}(X=0)^2\mathbb{P}(X=2)^3\mathbb{P}(X=1)^2$$

e quindi

$$L(\theta) = \left(\frac{2\theta}{3}\right)^2 \left(\frac{\theta}{3}\right)^3 \left(\frac{2(1-\theta)}{3}\right)^3 \left(\frac{1-\theta}{3}\right)^2.$$

Ne segue immediatamente che la log-verosimiglianza è data da

$$\begin{aligned} l(\theta) =& \log(L(\theta)) \\ =& 2\left(\log(2/3) + \log(\theta)\right) + 3(\log(1/3) + \log(\theta) + 3(\log(2/3) + \log(1-\theta) + 2(\log(1/3) + \log(1-\theta) \\ =& C + 5\log(\theta) + 5\log(1-\theta). \end{aligned}$$

Dove C è una costante che non dipende da θ. Derivando otteniamo

$$\frac{dl(\theta)}{d\theta} = \frac{5}{\theta} - \frac{5}{1-\theta}$$

e ponendo la derivata uguale a zero otteniamo $\hat{\theta} = 1/2$. Verifichiamo che anche le condizioni del secondo ordine siano soddisfatte:

$$\frac{d^2 l(\theta)}{d\theta^2} = -5\left(1/\theta^2 + 1/(1-\theta)^2\right) < 0, \qquad \forall \theta \in (0,1).$$

Esercizio 13.2 Sia X una v.a. Geometrica di parametro θ (incognito). Ricordiamo che questo significa che

$$\mathbb{P}(X = x) = \theta(1-\theta)^{x-1},\ \forall x \in \mathbb{N}.$$

Date le osservazioni indipendenti $x_1, \ldots, x_n$ di X si determini lo stimatore di massima verosimiglianza per il parametro θ.

Soluzione. La funzione di verosimiglianza è data da

$$L(\theta) = \theta^n (1-\theta)^{\sum_{i=1}^n x_i - 1}$$

ne segue che

$$l(\theta) = n\log(\theta) + \log(1-\theta)\sum_{i=1}^n (x_i - 1).$$

Derivando la log-verosimiglianza otteniamo

$$\frac{dl(\theta)}{d\theta} = N/\theta - \frac{1}{1-\theta}\sum_{i=1}^n (x_i - 1)$$

e ponendo quest'ultima uguale a zero otteniamo

$$\frac{1}{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

In conclusione $\hat{\theta} = \frac{1}{\bar{x}_n}$, i.e. lo stimatore di m.v. non è altro che l'invero della media campionaria della nostra v.a.. Ci resta solamente da verificare che tale valore sia effettivamente un massimo per la log-verosimiglianza. Calcoliamo la derivata seconda

$$\frac{d^2 l(\theta)}{d\theta^2} = -\frac{n}{\theta^2} - \left(\frac{1}{1-\theta}\right)^2 \sum_{i=1}^{n}(x_i - 1).$$

Utilizziamo ora il fatto che

$$\sum_{i=1}^{n}(x_i - 1) = n\bar{x}_n - n = n\left(\frac{1-\hat{\theta}}{\hat{\theta}}\right)$$

pertanto, se valutata in $\hat{\theta}$, la derivata seconda risulta

$$\left.\frac{d^2 l(\theta)}{d\theta^2}\right|_{\theta=\hat{\theta}} = -\frac{n}{\hat{\theta}^2} - \left(\frac{1}{1-\hat{\theta}}\right)^2 n\left(\frac{1-\hat{\theta}}{\hat{\theta}}\right) = -n\frac{1}{\hat{\theta}(1-\hat{\theta})} < 0.$$

Esercizio 13.3 Una v.a. è distribuita secondo la legge di Pareto di parametri $a > 0$ e $\theta > 0$ se ha densità

$$f(x) = \theta z^{\theta} x^{-\theta-1}\mathbf{1}_{x \geq a}.$$

Supponiamo che a sia noto, e che possiamo osservare n realizzazioni indipendenti di una v.a. con tale legge. Dare un'espressione per lo stimatore di massima verosimiglianza per θ.

Soluzione. Calcoliamo la funzione di log-verosimiglianza

$$\begin{aligned} l(\theta) =& \sum_{i=1}^{n}[\log(\theta) + \theta\log(a) - (\theta+1)\log(x_i)] \\ =& n\log(\theta) + n\theta\log(a) - (\theta+1)\sum_{i=1}^{n}\log(x_i). \end{aligned}$$

Derivando otteniamo

$$\frac{dl(\theta)}{d\theta} = \frac{n}{\theta} + n\log(a) - \sum_{i=1}^{n}\log(x_i)$$

e questa è uguale a zero se e solo se

$$\theta = \hat{\theta} := \frac{1}{\bar{\log(x)}_n - \log(a)}.$$

Notiamo che il denominatore risulta essere maggiore di 0 dalla definizione della densità. La condizione di secondo ordine è banale da verificare in questo caso.