ESERCIZI DI INFERENZA STATISTICA [1]

Fulvio De Santis - Marco Perone Pacifico - Luca Tardella - Isabella Verdinelli

adattamento

I PARTE: VARIABILI ALEATORIE, VEROSIMIGLIANZA E SUFFICIENZA

Variabili aleatorie

Esercizio 1. Un rappresentante percorre frequentemente in automobile il tratto tra New York e Boston. Si ipotizza che il tempo di percorrenza sia una variabile aleatoria con distribuzione normale di valore atteso 4.3 ore e varianza pari a 0.2^2 ore. Determinare la probabilità che un viaggio del rappresentante duri

a) più di 4.5 ore;

b) meno di 4 ore.

Esercizio 2. Un quiz è costituito da dieci domande a risposta multipla. Per ciascuna domanda sono previste 3 risposte, di cui solamente una è esatta. Per superare il test è necessario rispondere correttamente ad almeno sei domande. Supponendo di scegliere a caso le risposte a tutte le domande,

a) determinare la probabilità di superare il test;

b) stabilire se la probabilità di rispondere esattamente a tutte le domande del test è superiore alla probabilità di sbagliare tutte le risposte.

Esercizio 3. Si consideri un campione di 16 persone che soffrono di emicrania. Supponendo di utilizzare su questi pazienti un farmaco che, in base alla ricerca, si ritiene efficace nell' 80% dei casi, determinare:

a) la probabilità che abbia effetto su tutte le unità del campione;

b) la probabilità che il farmaco sia efficace su almeno 14 pazienti del campione;

c) il numero medio di pazienti del campione che ci si aspetta trovino giovamento dall'uso del farmaco.

d) Supponendo di somministrare il farmaco ad una popolazione di 1000 soggetti, determinare la probabilità che questo abbia effetto su almeno 750 individui.

Esercizio 4. Il tasso di guarigione garantito da un farmaco per una determinata malattia è pari al 70%.

a) Qual è la probabilità che, su 10 pazienti curati con il farmaco considerato, più di 8 guariscano?

b) Qual è il numero medio di pazienti per i quali ci si aspetta la guarigione?

Esercizio 5. Il 65% dei laureati di una facoltà viene assunto entro un anno dalla laurea. Supponendo di considerare $n = 9$ laureati della facoltà in esame, determinare:

a) il numero medio di laureati assunti entro un anno;

b) la probabilità che almeno cinque di questi trovino lavoro entro un anno;

c) la probabilità che al più due di questi trovi lavoro entro un anno.

Esercizio 6. Una variabile statistica X ha distribuzione normale di valore atteso μ e varianza σ^2 incognite. Determinare il valore di μ e σ sapendo che la probabilità che X assuma valori minori di 245 è pari a 0.33, e che la probabilità che X assuma valori superiori a 260 è pari a 0.48.

Esercizio 7. Il 2,35 per cento delle persone adulte di un collettivo è mancina. Determinare la probabilità che, su una scolaresca di 120 studenti, ne siano mancini

[1] *Gli esercizi contrassegnati dall'asterisco si riferiscono a prove di esame assegnate negli scorsi anni accademici.*

(a) 3;
(b) almeno 3;
(c) al massimo 3.

Esercizio 8*. In una indagine del 1994, il Census Bureau degli U.S.A. ha stabilito che il 70% dei cittadini americani aveva stipulato un contratto di assicurazione sanitaria privata. Sulla base di tale valutazione, qual era in quell'anno la probabilità che, su $n = 5$ cittadini scelti casualmente, al più 2 avessero il contratto?

Esercizio 9*. In una indagine del 1994, il Census Bureau degli U.S.A. ha stabilito che il 30% dei cittadini americani non aveva stipulato un contratto di assicurazione sanitaria privata. Sulla base di tale valutazione, qual era in quell'anno la probabilità che, su $n = 5$ cittadini scelti casualmente, almeno 3 persone non avessero il contratto?

Esercizio 10*. Da un'indagine demografica condotta dal *Census Bureau* statunitense, risulta che il 9.96 % dei cittadini americani di età superiore a 18 anni è di origine ispanica. Si consideri un campione casuale di $n = 1200$ cittadini americani e determinare

a) il numero medio di cittadini ispanici in un campione di tale ampiezza;

b) la probabilità che di un campione di tale ampiezza facciano parte meno di 100 cittadini ispanici.

Campioni Casuali, Statistiche Campionarie e Proprietà

Esercizio 1. Sia $X_1, \ldots, X_n$ un campione casuale da una popolazione di Poisson di parametro incognito θ. Determinare (in funzione di θ) la probabilità di osservare il seguente campione di dimensione $n = 5$:

$$\mathbf{x}_n = (2, 3, 1, 5, 5).$$

Che valore assume la probabilità considerata se $\theta = 2$? E se invece $\theta = 3$?

Esercizio 2. Sia $X_1, \ldots, X_n$ un campione casuale da una popolazione di esponenziale di parametro incognito θ. Determinare la funzione di densità congiunta (in funzione di θ) in corrispondenza del seguente campione di dimensione $n = 5$:

$$\mathbf{x}_n = (1.12, 0.88, 0.13, 0.42, 0.36).$$

Che valore assume la densità congiunta se si assume $\theta = 0.5$? E se invece si assume $\theta = 1$?

Esercizio 4. Il tempo di vita di un certo componente elettrico è una v.a. con valore atteso $\mu = 100$ e deviazione standard $\sigma = 20$. Se si provano $n = 16$ componenti di questo tipo (tra loro indipendenti), quanto vale approssimativamente la probabilità che la media campionaria delle loro durate di vita sia
a) minore di 104;
b) compresa tra 98 e 104.

Esercizio 5. In una azienda, si suppone che il numero di ore di straordinario degli impiegati in un mese sia una v.a. con valore atteso $\mu = 5.75$ ore e deviazione standard $\sigma = 0.48$ ore. Se si considera un campione casuale di $n = 36$ impiegati, qual è la probabilità che le ore complessive del loro lavoro straordinario sia compreso tra 202 e 210 ore?

Esercizio 6. Sia X una v.a. discreta con la seguente distribuzione:

$$P(X = 0) = 0.2, \quad P(X = 1) = 0.3, \quad P(X = 2) = 0.5.$$

Si determini, nel caso di un campione casuale di dimensione $n = 2$, la distribuzione della media campionaria, $\bar{X}$. Si calcoli quindi il valore atteso e la varianza della v.a. $\bar{X}$.

Esercizio 7*. Si suppone che il numero di telefonate che un operatore di un grande centralino riceve in un'ora del giorno sia una variabile aleatoria di Poisson di parametro $\theta = 12$. Considerato un campione casuale $X_1, \ldots, X_n$ di $n = 100$ operatori, determinare la probabilità (approssimazione) che il numero complessivo $Y = \sum_{i=1}^{100} X_i$ di telefonate a cui rispondono i 100 operatori in un'ora sia compreso tra 1150 e 1280 telefonate.
(Sugg.: ricordare che per una v.a. X con distribuzione di Poisson di parametro θ, si ha che $\mathrm{E}(X) = \mathrm{V}(X) = \theta$).

Esercizio 8*. Si suppone che il numero di clienti che si presentano a uno sportello di una grande banca in un giorno dell'anno sia una variabile aleatoria di Poisson di parametro $\theta = 30$. Considerato un campione casuale $X_1, \ldots, X_n$ di $n = 40$ sportelli, determinare la probabilità (approssimazione) che il numero complessivo $Y = \sum_{i=1}^{40} X_i$ di clienti serviti dai 40 sportelli in un giorno sia compreso tra 1180 e 1270.
(Sugg.: ricordare che per una v.a. X con distribuzione di Poisson di parametro θ, si ha che $\mathrm{E}(X) = \mathrm{V}(X) = \theta$).

Esercizio 9*. Si consideri un campione casuale di $n = 4$ osservazioni provenienti da una distribuzione normale di parametri $\mu = 8$ e $\sigma^2 = 8$. Date le tre statistiche campionarie:

$$T_1(\mathbf{X}_n) = \bar{X}_n, \qquad T_2(\mathbf{X}_n) = S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2, \qquad T_3(\mathbf{X}_n) = \frac{1}{4}\bar{X}_n + \frac{3}{4}S_n^2,$$

calcolarne il valore atteso e la varianza.

(**Sugg.** Ricordare che se una v.a. $Y \sim \chi^2_\nu$, si ha che $E[Y] = \nu$ e che $V[Y] = 2\nu$)

Esercizio 10* Sia $X_1, \ldots, X_n$ un campione casuale tale che, $\forall i = 1, \ldots, n$,

$$E[X_i] = \theta \qquad \text{e} \qquad V[X_i] = \frac{1}{2}\theta^2$$

a) Determinare valore atteso e varianza delle variabili aleatorie $\frac{1}{2}\bar{X}_n$ e $2X_1 + 2X_n - 3$.

b) Determinare l'approssimazione asintotica per la distribuzione di $\bar{X}_n$.

c) Supponendo che $\theta = 2$ e $n = 25$, determinare la probabilità (approssimazione) che la v.a. $\bar{X}_n$ assuma valori nell'intervallo $(2, 2.4)$.

Esercizio 11*. Sia $X_1, \ldots, X_n$ un campione casuale tale che, $\forall i = 1, \ldots, n$,

$$E[X_i] = \frac{\theta}{\theta + 1} \qquad \text{e} \qquad V[X_i] = \frac{\theta}{\theta^2(\theta + 2)}.$$

a) Determinare valore atteso e varianza delle variabili aleatorie $\bar{X}_n$ e $2\sum_{i=1}^{n} X_i - 3$.

b) Determinare l'approssimazione asintotica per la distribuzione di $\bar{X}_n$.

c) Supponendo che $\theta = 0.2$ e $n = 25$, determinare la probabilità (approssimazione) che la v.a. $\bar{X}_n$ assuma valori nell'intervallo $(1/2, 1)$.

SOLUZIONI

Variabili aleatorie normali e binomiali

Esercizio 1. $X \sim N(4.3, 0.2^2)$
(a) $Pr\{X > 4.5\} = Pr\{Z > \frac{4.5-4.3}{0.2}\} = 1 - F_Z(1) = 0.16$
(b) $Pr\{X < 4\} = \cdots = F_Z(-1.5) = 1 - F_Z(1.5) = 0.067$
Esercizio 2. $S = \sum_{i=1}^{10} X_i \sim \text{Binom}(10, \theta)$
a) $\mathbb{P}(S \geq 6) = \sum_{i=6}^{10} \binom{10}{i} \theta^i (1-\theta)^{10-i}$.
b) $\mathbb{P}(S = 10) = \theta^{10}$.
c) $\mathbb{P}(S = 0) = (1-\theta)^{10}$. Inoltre: $\qquad \theta^{10} > (1-\theta)^{10} \Leftrightarrow \theta > 1/2$.
Esercizio 3. $p_A = 0.8$; $p_B = 0.7$.
a) $S = \sum X_i \sim \text{Binom}(n = 16, \theta = 0.8)$; $\mathbb{P}(S = 10) = 0.8^{16} = 0.028$.
b) $\mathbb{P}(S \geq 14) = \sum_{i=14}^{16} \binom{16}{i} (0.8)^i (0.2)^{16-i}$.
c) $E(S) = np_A = 12.8$.

Esercizio 4. Se $X \sim \text{Bernoulli}(0.7)$; il numero di pazienti che guariscono è $S = \sum_{i=1}^{10} X_i \sim \text{Binom}(n = 10, 0.7)$.
(a) $Pr\{S > 8\} = \mathbb{P}\{S = 9\} + \mathbb{P}\{S = 10\} \simeq 0.121 + 0.028 = 0.149$.
(b) $E(S) = 10 \cdot 0.7 = 7$.

Esercizio 5. Numero di laureati assunti $S \sim \text{Binom}(n = 9, 0.65)$
(a) $E(S) = n\theta = 9 \cdot 0.65 = 5.85$
(b) $Pr\{S \geq 5\} = \sum_{r=5}^{9} \binom{9}{r} (0.65)^r (0.35)^{9-r} \simeq 0.83$.
(c) $Pr\{S \leq 2\} = \cdots \simeq 0.011$.

Esercizio 6.
$Pr\{X < 245\} = Pr\{\frac{X-\mu}{\sigma} < \frac{245-\mu}{\sigma}\} = F_Z(\frac{245-\mu}{\sigma}) = 0.33 = F_Z(-0.44)$ quindi $\frac{245-\mu}{\sigma} = -0.44$
$Pr\{X > 260\} = Pr\{\frac{X-\mu}{\sigma} > \frac{260-\mu}{\sigma}\} = 1 - F_Z(\frac{260-\mu}{\sigma}) = 0.48$
quindi $F_Z(\frac{260-\mu}{\sigma}) = 0.52 = F_Z(0.06)$ e $\frac{260-\mu}{\sigma} = 0.06$.
Risolvendo otteniamo $\mu = 258.5$ e $\sigma = 30.5$.

Esercizio 7. Il numero di mancini nel campione ha distribuzione binomiale di parametri $n = 120, P = 0.0235$. Le probabilità richieste sono 0.23, 0.54, 0.69.

Esercizio 8. Il numero di cittadini nel campione che ha assicurazione sanitaria ha distribuzione binomiale di parametri $n = 5, P = 0.7$. La probabilità richiesta è 0.163.

Esercizio 9. È esattamente la stessa dell'esercizio precedente.

Esercizio 10. $X_i =$ etnia cittadino i-esimo campionario (1=ispanico, 0=non ispanico);

$Y_n = \sum_{i=1}^{n} X_i =$ numero di cittadini ispanici nel campione;

$\theta = 0.0996$

$X_i|\theta \sim \text{Ber}(\theta), \qquad Y_n|\theta \sim \text{Binom}(n, \theta)$.

a) $E_\theta[Y_n] = n\theta = (1200) * (0.096) = 119.52$.

b) Utilizzando l'approssimazione normale della distribuzione binomiale e ricordando che $V_\theta(Y_n) = n\theta(1-\theta)$ si ottiene

$$\mathbb{P}(Y_n < 100) \approx \mathbb{P}(Z < (100 - 119.52)/\sqrt{107.64}) = \Phi(-1.88) = 0.03,$$

dove $\Phi(\cdot)$ indica la funzione di ripartizione della v.a. normale standardizzata (Z).

Esercizio 1.

$$f_{X_1,\cdots,X_5}(x_1,\cdots,x_5)\theta) = \frac{\theta^{\sum_{i=1}^5 x_i} e^{-5\theta}}{\prod_{i=1}^5 x_i!} \quad \text{quindi} \quad f_{X_1,\cdots,X_5}(2,3,1,5,5;\theta) = \frac{\theta^{16} e^{-5\theta}}{172800}$$

per $\theta = 2$ la probabilità è $\frac{e^{-10}2^{16}}{172800} \simeq 0.000017$ mentre per $\theta = 3$ si ha $\frac{e^{-15}3^{16}}{172800} \simeq 0.000076$.

Esercizio 2.
$f_{X_1,\cdots,X_5}(x_1,\cdots,x_5;\theta) = \theta^5 e^{-\theta\sum_{i=1}^5 x_i}$ quindi $f_{X_1,\cdots,X_5}(1.12, 0.88, 0.13, 0.42, 0.36;\theta) = \theta^5 e^{-2.91\theta}$.

Per $\theta = 0.5$ vale 0.0073 mentre per $\theta = 1$ vale 0.054.

Esercizio 3. Se indichiamo con $W = \frac{(n-1)S^2}{\sigma^2}$, sappiamo che $W \sim \chi^2_{n-1}$. Poiché $n = 5$, $W = \frac{4S^2}{\sigma^2} \sim \chi^2_4$.

$Pr\{\frac{S^2}{\sigma^2} \leq 1.8\} = Pr\{\frac{4S^2}{\sigma^2} \leq 7.2\} = F_{\chi^2_4}(7.2) \simeq 0.87$.

Esercizio 4. Usiamo l'approssimazione normale alla distribuzione della media campionaria $\bar{X} \sim N(100, \frac{20^2}{16})$.

$Pr\{\bar{X} < 104\} = Pr\{Z < \frac{104-100}{20/4}\} = F_Z(0.8) \simeq 0.79$.

Esercizio 5. La popolazione ha distribuzione incognita, ma con valore atteso $\mu = 5.75$ e varianza $\sigma^2 = 0.48^2$. Per n sufficientemente grande, $\bar{X} \sim N(\mu, \sigma^2/n)$, quindi la somma $\sum X_i = n\bar{X} \simeq N(n\mu, n\sigma^2)$. Nel nostro caso $\sum X_i \simeq N(207, 8.2944)$.

$Pr\{202 \leq \sum X_i \leq 210\} = Pr\{\frac{202-207}{2.88} \leq Z \leq \frac{210-207}{2.88}\} = F_Z(1.04) - F_Z(-1.74) \simeq 0.809$.

Esercizio 6. $E(\bar{X}) = E(X) = 1.3$ e $V(\bar{X}) = \frac{V(X)}{n} = \frac{0.61}{2} = 0.305$.

La tabella di sinistra contiene tutti i campioni possibili, con la probabilità e la media campionaria corrispondente, la tabella di destra riassume la distribuzione della media campionaria.

(x_1, x_2)	Pr	$\bar{x}$
(0,0)	0.04	0
(0,1)	0.06	0.5
(0,2)	0.10	1
(1,0)	0.06	0.5
(1,1)	0.09	1
(1,2)	0.15	1.5
(2,0)	0.10	1
(2,1)	0.15	1.5
(2,2)	0.25	2

$\bar{x}$	Pr
0	0.04
0.5	0.12
1	0.29
0.5	0.30
2	0.25

Dalla seconda tabella si desumono i valori di $E(\bar{X})$ e $V(\bar{X})$.

Esercizio 7.

Usando l'approssimazione normale si può supporre $\sum X_i \sim N(1200, 1200)$.
$Pr\{1150 \leq \sum X_i \leq 1280\} = Pr\{\frac{1150-1200}{\sqrt{1200}} \leq Z \leq \frac{1280-1200}{\sqrt{1200}}\} = Pr\{-1.44 \leq Z \leq 2.31\} \simeq 0.91$.

Esercizio 8.

Anche qui usando l'approssimazione normale si può supporre $\sum X_i \sim N(1200, 1200)$.
$Pr\{1180 \leq \sum X_i \leq 1270\} = Pr\{\frac{1180-1200}{\sqrt{1200}} \leq Z \leq \frac{1270-1200}{\sqrt{1200}}\} = Pr\{-0.58 \leq Z \leq 2.02\} \simeq 0.70$.

Esercizio 9.
$E(T_1) = E(X) = 8 \qquad V(T_1) = \frac{V(X)}{n} = \frac{8}{4} = 2.$
$E(T_2) = V(X) = 8 \qquad V(T_2) = V(\frac{\sigma^2}{n-1} \cdot \frac{\sum(X_i-\bar{X})^2}{\sigma^2}) = \frac{\sigma^4}{(n-1)^2} \cdot 2(n-1) = 2\frac{\sigma^4}{n-1} = \frac{128}{3}.$
$E(T_3) = E(\frac{1}{4}T_1 + \frac{3}{4}T_2) = \frac{1}{4}E(T_1) + \frac{3}{4}E(T_2) = 8$
Per l'indipendenza tra $\bar{X}$ e S^2 si ha $V(T_3) = V(\frac{1}{4}T_1 + \frac{3}{4}T_2) = \frac{1}{16}V(T_1) + \frac{9}{16}V(T_2) = \frac{193}{8}$.

Esercizio 10. $E(\frac{1}{2}\bar{X}) = \frac{\theta}{2}$ $\quad V(\frac{1}{2}\bar{X}) = \frac{\theta^2}{8n}$
$E(2X_1 + 2X_n - 3) = 4\theta - 3 \quad V(2X_1 + 2X_n - 3) = 4\theta^2$.
$\bar{X} \sim N(\theta, \frac{\theta^2}{2n})$ quindi, nel caso $\theta = 2$, $n = 25$, $\bar{X} \sim N(2, \frac{4}{50})$.
$Pr\{2 < \bar{X} < 2.4\} = Pr\{0 < Z < 1.41\} \simeq 0.42$.

Esercizio 11. $E(\bar{X}) = \frac{\theta}{\theta+1} \quad V(\bar{X}) = \frac{\theta}{n\theta^2(\theta+2)}$
$E(2\sum X_i - 3) = 2n\frac{\theta}{\theta+1} - 3 \quad V(2\sum X_i - 3) = 4n\frac{\theta}{\theta^2(\theta+2)}$.
$\bar{X} \sim N(\frac{\theta}{\theta+1}, \frac{\theta}{n\theta^2(\theta+2)})$ quindi, nel caso $\theta = 0.2$, $n = 25$, $\bar{X} \sim N(0.16, 0.09)$.
$Pr\{\frac{1}{2} < \bar{X} < 1\} = \cdots \simeq 0.13$.

Esercizio 12.

a) La distribuzione esatta di $Y_n = \sum_{i=1}^{n} X_i$ è la distribuzione di Poisson di parametro $n\theta$. Quindi Y_n può assumere i valori $y = 0, 1, \dots, y \dots$ con probabilità

$$f_{Y_n}(y;\theta) = \mathbb{P}(Y_n = y;\theta) = \frac{e^{-n\theta}(n\theta)^y}{y!}.$$

La variabile aleatoria $U_n = 2Y_n - 3$ assume di conseguenza i valori $u = 0, -3, -1, 1, 3, 5, \dots 2y - 3, \dots$ con probabilità

$$f_{U_n}(u;\theta) = \mathbb{P}(U_n = u;\theta) = \mathbb{P}(2Y_n - 3 = u;\theta) =$$

$$\mathbb{P}(Y_n = \frac{u+3}{2};\theta) = f_{Y_n}((u+3)/2;\theta) = \frac{e^{-n\theta}(n\theta)^{(u+3)/2}}{[(u+3)/2]!}.$$

b) Sono verificate le ipotesi del teorema del limite centrale: le v.a. $X_1, \dots, X_n$ sono i.i.d., e valore atteso e varianza di X_i sono finiti. Pertanto,

$$Y_n = \sum_{i=1}^{n} X_i \,\dot\sim\, N(n\theta, n\theta) = N(2, 2).$$

Per le proprietà delle v.a. normali, una combinazione lineare di una v.a. normale ha ancora distribuzione normale. Poichè

$$E\theta[U_n] = 2E\theta[Y_n] - 3 = 2n\theta - 3, \qquad \mathbb{V}_\theta[U_n] = 4\mathbb{V}_\theta[Y_n] = 4n\theta,$$

si ha che

$$U_n = 2Y_n - 3 \,\dot\sim\, N(2n\theta - 3, 4n\theta) = N(1, 8),$$

e quindi che

$$\mathbb{P}(U_n > 0) \approx \mathbb{P}\left(\frac{U_n - 1}{\sqrt{8}} > \frac{-1}{\sqrt{8}}\right) = \mathbb{P}\left(Z > -0.36\right) = 1 - \Phi(-0.36) = \Phi(0.36) = 0.64.$$