

TEST DI AUTOVALUTAZIONE VARIABILI ALEATORIE

I diritti d'autore sono riservati. Ogni sfruttamento commerciale non autorizzato sarà perseguito.

Metodi statistici per la biologia

1 Parte A

1.1

Una variabile casuale e':

- un sottoinsieme dello spazio degli esiti di un esperimento aleatorio;
- un elemento scelto a caso dello spazio degli esiti di un esperimento aleatorio;
- una quantità numerica il cui valore dipende dall'esito di un esperimento aleatorio;
- una successione $\{p(n) : n \in \mathbf{N}\}$ di numeri positivi la cui somma e' uguale a 1 .

1.2

Sia X una variabile aleatoria continua con densità di probabilità $f(x)$. Allora è *sempre* vero che

- $f(x) \leq 1$;
- $P(X = a) = f(a)$;
- $f(x) \geq 0$;
- nessuna delle precedenti.

1.3

Sia X una variabile casuale discreta tale che $P(X = 0) = 1/3$ e $P(X = 2) = 2/3$. Allora $Var(X)$ è uguale a

- 16/9;
- 4;
- 8/9;
- 8/3.

1.4

Sia X una variabile casuale continua con densità

$$f(x) = \begin{cases} \frac{1}{2} & \text{se } x \in [0, 2] \\ 0 & \text{altrimenti} \end{cases}$$

Allora $P(0.2 \leq X \leq 0.7)$ vale

- $\frac{1}{2}$;
- $\frac{1}{4}$;
- 0.7;
- 0.2.

1.5

Sia X una variabile aleatoria discreta tale che $P(X = 0) = p$, $P(X = -1) = P(X = 1) = (1 - p)/2$. Per quale valore di $p \in [0, 1]$, $Var(X)$ assume il valore massimo?

- $p = 0$;
- $p = 1$;
- $p = 1/2$;
- $Var(X)$ non dipende da p .

1.6

Sia X una variabile casuale discreta tale che $P(X = -1) = 1/3$ e $P(X = 2) = 2/3$. Allora $E(X)$ è uguale a

- 1;
- 1;
- 5/8;
- 1/8.

1.7

La media di una variabile casuale X continua con densità f è data dalla formula

- $\frac{X_1 + \dots + X_n}{n}$;
- $\int_{-\infty}^{+\infty} f(x) dx$;
- $\int_{-\infty}^{+\infty} x f(x) dx$;
- $\frac{f(X_1) + \dots + f(X_n)}{n}$.

1.8

Sia X una variabile casuale discreta. Allora $P(X \in \{x_1, x_2\})$ è uguale a

- $P(X = x_1)P(X = x_2)$;
- $P(X = x_1) + P(X = x_2)$;
- $x_1 + x_2$;
- $x_1 x_2$.

1.9

Sia X una variabile casuale continua, con densità f , e siano $x < y$ due numeri reali. Allora

- $P(X \in \{x, y\}) = 0$;
- $P(X \in \{x, y\}) = f(x) + f(y)$;
- $P(X \in \{x, y\}) = \int_x^y f(t) dt$;
- $P(X \in \{x, y\}) = f(x)f(y)$.

1.10

Sia $X \sim N(2, 4)$. Allora $E(X^2)$ è uguale a

- 0;
- 4;
- 6;
- 8.

1.11

Sia X una variabile casuale con distribuzione $B(n, p)$. Allora $P(X \leq 1)$ vale

- $np(1-p)^{n-1}$
- $(1-p)^n + np(1-p)^{n-1}$
- $(1-p)^{n-1}$
- i dati forniti non sono sufficienti a determinare la risposta

1.12

Una variabile casuale X è continua con densità data da

$$f(x) := \begin{cases} \frac{1}{x} & \text{se } 2 \leq x \leq 5 \\ 0 & \text{altrimenti} \end{cases}.$$

Allora $E(X)$ vale

- $\int_2^5 1 dx$
- $\int_2^5 \frac{1}{x} dx$
- $\int_2^5 x dx$
- 0

1.13

Sia $X \sim Po(\lambda)$. Sappiamo che $P(X \geq 1) = 2/3$. Allora λ è uguale a (ln denota il logaritmo in base e , o naturale)

- $2/3$
- $\ln(3/2)$
- $\ln(3)$
- non è possibile saperlo con i dati a disposizione

1.14

Siano X, Y variabili aleatorie indipendenti, ciascuna con distribuzione $N(0, 1)$. Quale delle seguenti variabili è standardizzata (cioè ha media zero e varianza uno)?

- $X + Y$
- $\frac{1}{2}(X + Y)$
- $\frac{1}{2}(X - Y)$
- $\frac{1}{\sqrt{2}}(X - Y)$

1.15

Sia X una variabile casuale discreta tale che

$$P(X = -1) = \frac{1}{2}, \quad P(X = 2) = \frac{1}{2}.$$

Allora $E(X^2)$ vale

- 1
- $\frac{1}{4}$
- $\frac{5}{2}$
- $\frac{1}{2}$

1.16

Una variabile casuale X è tale che $E(X) = -3$, $Var(X) = 2$. Se si pone $Y := -3X + 2$, quanto vale $Var(Y)$?

- 6
- 6
- 8
- 18

1.17

Si considerino le variabili casuali $X \sim N(-1, 2)$, $Y \sim N(3, 2)$, $Z \sim N(0, 1)$, e assumiamo che siano indipendenti. Il valore di z per cui si ha $P(X + Y \geq 0) = P(Z \leq z)$ è

- 0
- 1
- 1
- 4

1.18

Sia $X \sim N(2, 1)$, e sia Φ la funzione di ripartizione della Normale standard. Allora per $x \in \mathbf{R}$, $P(X \leq x)$ è uguale a

- $\Phi(x)$
- $\Phi(x - 2)$
- $\Phi(x + 2)$
- $\Phi\left(\frac{x-1}{2}\right)$

1.19

Sia X una variabile casuale discreta tale che

$$P(X = 0) = \frac{1}{4} \quad P(X = 1) = \frac{1}{4} \quad P(X = 2) = \frac{1}{2}.$$

Allora $E(X^2)$ vale

- $\frac{9}{4}$
- 1
- $\frac{9}{16}$
- $\frac{11}{16}$

1.20

Sia X una variabile casuale, e si considerino le probabilità $P(X < 1)$ e $P(X < 2)$. Quale delle seguenti affermazioni è vera?

- Sicuramente $P(X < 1) \leq P(X < 2)$;
- Sicuramente $P(X < 1) \geq P(X < 2)$;
- il fatto che $P(X < 1)$ sia maggiore o minore di $P(X < 2)$ dipende dalla distribuzione di X ;
- almeno una delle due probabilità vale 0.

1.21

Sia $X \sim Po(\lambda)$, cioè X è una variabile di Poisson di parametro λ . Quale delle seguenti affermazioni è vera per ogni valore di λ ?

- $P(X = 0) = P(X = 1)$;
- $E(X^2) = \lambda^2$;
- $P(X = 1) = \lambda$;
- $E(X^2) = \lambda^2 + \lambda$.

1.22

Sia X una variabile casuale che assume i valori 1, 2, 3 con la stessa probabilità $\frac{1}{3}$. $Var(X)$ vale

- $\frac{2}{3}$;
- 4;
- $\frac{14}{3}$;
- 1.

1.23

Sia $X \sim B(1000, 0.005)$. La probabilità $P(X = 0)$ vale circa

- e^{-1000}
- $e^{-0.005}$
- e^{-5}
- $\frac{1}{1000}$

1.24

Sia X una variabile Normale di media 2 e varianza σ_1^2 , e Y una variabile normale di media 4 e varianza σ_2^2 . Si consideri l'affermazione: $P(X \geq 2) \leq P(Y \geq 3)$.

- è sicuramente falsa, a prescindere dai valori di σ_1 e σ_2
- è vera se $\sigma_1 > \sigma_2$, altrimenti è falsa
- è vera se $\sigma_2 > \sigma_1$, altrimenti è falsa
- è sicuramente vera, a prescindere dai valori di σ_1 e σ_2

1.25

Sia $X \sim B(1, 1/3)$ e sia $Y := 1 - X$. Qual è la distribuzione di Y ?

- $B(1, \frac{2}{3})$
- $B(1, \frac{1}{2})$
- $B(1, \frac{1}{3})$
- $Po(\frac{1}{3})$

1.26

Sia $X \sim N(\mu, \sigma^2)$ con $\mu = 11$ e σ^2 incognita. La probabilità $P(X \geq 11)$ vale

- $\frac{1}{2}$
- $\frac{1}{4}$
- $\Phi(\frac{11}{\sigma})$
- z_α con $\alpha = \frac{11}{\sigma}$

1.27

Sia $X \sim Po(3)$ e poniamo $Y := -X$. Qual è la distribuzione di Y ?

- $Po(-3)$
- $B(9, 1/3)$
- $N(3, 3)$
- nessuna delle precedenti

1.28

La distribuzione binomiale $B(n, p)$ può essere approssimata con la distribuzione di Poisson $Po(\lambda)$ purché siano soddisfatte le seguenti condizioni:

- n è grande, p è piccolo, $\lambda = np$.
- n è grande, p è piccolo, $\lambda = p$.
- np e $n(1 - p)$ sono grandi, $\lambda = np$.
- np e $n(1 - p)$ sono grandi, $\lambda = np(1 - p)$.

1.29

Siano $X \sim B(1, 1/4)$ e $Y \sim B(1, 1/2)$ variabili aleatorie indipendenti. Quale delle seguenti affermazioni è *falsa*?

- $E(X + Y) = 3/4$
- $Var(X + Y) = 7/16$
- $X + Y \sim B(2, 3/4)$
- $E(X - Y) = -1/4$

1.30

Sia $X \sim N(1, 1)$. Allora, per ogni $\alpha \in (0, 1)$, la probabilità $P(X \geq 1 + z_\alpha)$ vale

- α
- $1 - \alpha$
- $\alpha - 1$
- nessuna delle precedenti.

1.31

Sia X_1, X_2, \dots una successione di variabili casuali i.i.d. con distribuzione $B(m, p)$. Quale delle seguenti variabili casuali ha una distribuzione che, per n grande, approssima quella di una Normale Standard?

- $X_1 + \dots, X_n$;
- $\frac{\bar{X}_n - mp}{\sqrt{mp(1-p)}} \sqrt{n}$;
- $\frac{\bar{X}_n - mp}{\sqrt{n}}$;
- $\frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \sqrt{m}$.

1.32

Sia X_1, X_2, \dots una successione di variabili casuali i.i.d. con distribuzione $Po(\lambda)$. Quale delle seguenti variabili casuali ha una distribuzione che, per n grande, approssima quella di una Normale Standard?

- $\frac{\bar{X}_n - \lambda}{\sqrt{n}}$;
- $X_1 + \dots, X_n$;
- $\frac{\bar{X}_n}{\sqrt{\lambda}} \sqrt{n}$;
- $\frac{\bar{X}_n - \lambda}{\sqrt{\lambda}} \sqrt{n}$.

1.33

Sia X_1, X_2, \dots una successione di variabili casuali i.i.d. tali che $P(X_n = 1) = P(X_n = -1) = 1/2$. La variabile casuale $\frac{X_1 + X_2 + \dots + X_n}{\sqrt{n}}$

- ha varianza zero;
- ha media 1;
- ha distribuzione approssimativamente normale per n grande;
- ha distribuzione $N(0, 1)$ per ogni valore di n .

1.34

Siano X_1, X_2, X_3, \dots variabili i.i.d. con valore atteso μ e varianza 1. Sia inoltre $Z \sim N(0, 1)$. Fissato un arbitrario $x \in \mathbf{R}$, quale delle seguenti relazioni, per n sufficientemente grande, è una conseguenza del Teorema Centrale del Limite ?

- $P\left(\frac{X_1 + \dots + X_n}{\sqrt{n}} > \mu + x\right) \simeq P(Z \leq x)$
- $P\left(\frac{X_1 + \dots + X_n}{\sqrt{n}} < \mu\right) \simeq \frac{1}{2}$
- $P\left(\frac{X_1 + \dots + X_n}{n} < \mu + x\right) \simeq P(Z \leq x)$
- $P(X_1 + \dots + X_n \leq n\mu + x\sqrt{n}) \simeq P(Z \leq x)$

2 Parte B

2.1

Si sceglie “a caso” un campione di 5 oggetti da un lotto di 100 di cui 10 sono difettosi per effettuare un controllo di qualità. Sia X il numero di oggetti difettosi contenuti nel campione. Determinare la distribuzione di X .

Soluzione. La variabile X assume solo i valori $0, 1, \dots, 5$. $X = k$ significa che nel lotto di 5 oggetti k sono difettosi. Possiamo calcolare $\Pr(X = k)$ riconoscendo che X ha una distribuzione ipergeometrica oppure calcolare questa probabilità come casi favorevoli su casi possibili. *Casi possibili:* ci sono $\binom{100}{5}$ modi di scegliere 5 oggetti tra 100. *Casi favorevoli:* devo scegliere 5 oggetti di cui k difettosi. Scelgo prima i k difettosi tra i 10 difettosi in $\binom{10}{k}$ modi, poi scelgo i rimanenti $5 - k$ oggetti tra i rimanenti $100 - 10 = 90$ non difettosi in $\binom{90}{5-k}$ modi. In definitiva i casi favorevoli sono: $\binom{10}{k} \binom{90}{5-k}$ e quindi:

$$\Pr(X = k) = \frac{\binom{10}{k} \binom{90}{5-k}}{\binom{100}{5}}, \quad k = 0, 1, \dots, 5$$

2.2

In un'azienda agricola, si ritiene che la qualità dell'uva nel 2002 sarà ottima con probabilità 0.25 e buona con probabilità 0.55. Se la qualità sarà ottima, il vino prodotto sarà venduto a 5 Euro a bottiglia; se la qualità sarà buona, il vino prodotto sarà venduto a 3 Euro a bottiglia; se la qualità sarà inferiore a buona il vino prodotto sarà venduto a 1 Euro a bottiglia. Sia X il prezzo di una bottiglia di vino prodotto da quell'uva. Determinare la media e la varianza di X .

Soluzione. Sia ha:

$$P(X = 5) = 0.25, P(X = 3) = 0.55, P(X = 1) = 1 - 0.25 - 0.55 = 0.2.$$

Allora

$$\begin{aligned} E(X) &= 1 \times 0.2 + 3 \times 0.55 + 5 \times 0.25 = 3.1 \\ E(X^2) &= 1 \times 0.2 + 9 \times 0.55 + 25 \times 0.25 = 11.4, \end{aligned}$$

e quindi

$$\text{Var}(X) = E(X^2) - E(X)^2 = 1.79.$$

2.3

L'altezza media, in centimetri, di un bambino di 7 mesi è una variabile aleatoria normale di parametri $\mu = 71$ e $\sigma^2 = 6.25$. Qual è la percentuale di bambini di 7 mesi che superano i 74 centimetri di altezza?

Soluzione. Sia X l'altezza, in centimetri, di un bambino di 7 mesi. $X \sim N(71, 6.25)$. Poiché

$$Z \equiv \frac{X - 71}{\sqrt{6.25}} \sim N(0, 1),$$

si ha

$$\begin{aligned} P(X \geq 74) &= P\left(Z \geq \frac{74 - 71}{\sqrt{6.25}}\right) = 1 - \Phi\left(\frac{74 - 71}{\sqrt{6.25}}\right) \\ &= 1 - \Phi(1.2) \simeq 1 - 0.8849 = 0.1151. \end{aligned}$$

Pertanto, circa l'11.5% dei bambini di 7 mesi supera i 74 centimetri di altezza.

2.4

In un canale di trasmissione vengono trasmessi simboli binari. I disturbi sul canale fanno sì che ogni simbolo trasmesso ha la probabilità del 2% di essere ricevuto errato, indipendentemente dagli altri simboli. I messaggi vengono trasmessi in "parole" composte da 50 simboli. Qual è la probabilità che una parola venga ricevuta con almeno due simboli errati? (Usare l'approssimazione di Poisson per una binomiale)

Soluzione. Sia Y il numero di simboli errati in una parola. Abbiamo che $Y \sim B(50, 0.02)$, e dobbiamo calcolare $P(Y \geq 2)$. Usando l'approssimazione di Poisson, $Y \approx Po(1)$. Dunque

$$P(Y \geq 2) = 1 - P(Y = 0) - P(Y = 1) \simeq 1 - 2e^{-1}.$$

2.5

Io prendo mediamente due raffreddori l'anno: il tempo tra la fine di un raffreddore e l'inizio del successivo ha distribuzione normale, con media 160 giorni e deviazione standard di 40 giorni.

- i. Qual è la probabilità che rimanga 200 giorni o più senza un raffreddore?

- ii. Qual è la probabilità che mi prenda un raffreddore entro 80 giorni dalla guarigione dal precedente?

Soluzione.

Sia X = tempo che intercorre tra due raffreddori successivi $\sim N(160, 1600)$, e $Z := \frac{X-160}{40} \sim N(0, 1)$.

i.

$$P(X \geq 200) = P(Z \geq 1) \simeq 0.158$$

ii.

$$P(X \leq 80) = P(Z \leq -2) \simeq 0.023$$

2.6

I semi di una pianta di piselli da giardino possono essere gialli o verdi. Un certo incrocio tra piante di piselli produce una progenie nel rapporto 3 gialli : 1 verde (cioè la probabilità che un seme prodotto sia giallo è $3/4$). Quattro semi così prodotti vengono scelti a caso.

- Qual è la probabilità che tre siano gialli e uno verde?
- Qual è la probabilità che siano tutti gialli?
- Qual è la probabilità che siano tutti dello stesso colore?

Soluzione. Sia X il numero di semi gialli tra i quattro scelti. Si ha che $X \sim B(4, 3/4)$. Pertanto:

a.

$$P(X = 3) = \binom{4}{3} \left(\frac{3}{4}\right)^3 \frac{1}{4} = 0.421875$$

b.

$$P(X = 4) = \left(\frac{3}{4}\right)^4 = 0.31640625$$

c.

$$P(X = 0 \text{ oppure } X = 4) = P(X = 0) + P(X = 4) = \left(\frac{1}{4}\right)^4 + \left(\frac{3}{4}\right)^4 = 0.3203125$$

2.7

Il quoziente di intelligenza (IQ) nella popolazione ha distribuzione normale con media 100 e deviazione standard 15. Quale valore minimo di IQ bisogna avere per appartenere al 5% della popolazione con maggiore IQ?

Soluzione. Sia $X \sim N(100, 15^2)$, e x il valore cercato. Dev'essere

$$P(X > x) = 0.05.$$

Ma

$$P(X > x) = P\left(\frac{X - 100}{15} > \frac{x - 100}{15}\right) = P\left(Z > \frac{x - 100}{15}\right),$$

dove $Z \sim N(0, 1)$. Perci dev'essere

$$\frac{x - 100}{15} = z_{0.05} \simeq 1.645 \Rightarrow x = 100 + 1.65 \times 15 = 124.67$$

2.8

L'altezza della popolazione femminile adulta degli Stati Uniti ha distribuzione normale di media 165 cm e deviazione standard 7.5 cm. È stato riscontrato che donne particolarmente alte vanno incontro a maggiori problemi dovuti all'osteoporosi, e quindi si propone l'adozione di uno *screening* preventivo.

- Supponiamo che si decida di eseguire lo screening su tutte le donne di altezza superiore a 175 cm. A quale percentuale di popolazione femminile corrispondono?
- Supponiamo che invece si decida di eseguire lo screening sul 5% più alto della popolazione femminile. Qual è l'altezza minima delle donne sottoposte a screening?

Soluzione.

- Sia $X \sim N(165, 7.5^2)$.

$$P(X \geq 175) = P\left(\frac{X - 165}{7.5} \geq \frac{175 - 165}{7.5}\right) = P(Z \geq 1.33) = 1 - P(Z \leq 1.33) \simeq 0.092.$$

Pertanto le donne di altezza maggiore a 175 cm. corrispondono al 9.2% della popolazione.

- Sia x tale altezza minima. Dev'essere $P(X > c) = 0.05$, cioè

$$P\left(Z > \frac{c - 165}{7.5}\right) = 0.05 \Rightarrow \frac{c - 165}{7.5} = z_{0.05} = 1.65 \Rightarrow c \simeq 177.38.$$

2.9

Si ritiene che la probabilità che un individuo sia allergico al farmaco XXX sia $p = 0.002$. Si sceglie un campione casuale di 1500 persone.

- Determinare un'espressione *esatta* per la probabilità che 3 di queste persone siano allergiche al farmaco, mentre le altre 1497 non siano allergiche (non cercare di calcolarne il valore numerico)
- Determinare il valore numerico approssimato della probabilità al punto a), usando un'opportuna approssimazione.

Soluzione.

- Sia X il numero di individui allergici al farmaco. $X \sim B(1500, 0.002)$. Perciò

$$P(X = 3) = \binom{1500}{3} [0.002]^3 [0.998]^{1497}.$$

- Per l'approssimazione di Poisson, X ha approssimativamente la distribuzione di una variabile casuale di Poisson di parametro $1500 \times 0.002 = 3$. Perciò

$$P(X = 3) \simeq e^{-3} \frac{3^3}{3!} \simeq 0.224.$$

2.10

Il *peso corporeo* nella popolazione maschile adulta ha distribuzione normale con media 72 Kg e deviazione standard 13.2 Kg, mentre nella popolazione femminile ha distribuzione normale con media 64 Kg e deviazione standard 10.7 kg.

In un ascensore è espota la portata dichiarata di 5 persone e 400 Kg.

- Se nell'ascensore salgono 5 maschi, qual è la probabilità che il loro peso complessivo superi la portata dell'ascensore?

b) E se salgono 4 maschi e una femmina?

(Si assuma che i pesi dei diversi individui siano indipendenti.)

Soluzione.

a) Se $X_1, X_2, \dots, X_5 \sim N(72, 13.2^2)$ sono indipendenti, il peso complessivo è

$$X = \sum_{i=1}^5 X_i \sim N(5 \cdot 72, 5 \cdot 13.2^2) = N(360, 871.2).$$

Dunque

$$P(X > 400) = P\left(\frac{X - 360}{\sqrt{871.2}} > \frac{400 - 360}{\sqrt{871.2}}\right) = P(Z > 1.355) \simeq 0.088.$$

b) In questo caso il peso complessivo è

$$Y = \sum_{i=1}^4 X_i + Y_5,$$

dove $Y_5 \sim N(64, 10.7^2)$ è indipendente dalle altre variabili. Ne segue che

$$Y \sim N(4 \cdot 72 + 64, 4 \cdot 13.2^2 + 10.7^2) = N(352, 811.45).$$

Perciò

$$P(Y > 400) = P\left(\frac{Y - 352}{\sqrt{811.45}} > \frac{400 - 352}{\sqrt{811.45}}\right) = P(Z > 1.685) \simeq 0.046.$$

2.11

Sulla base delle osservazioni sul passato, si può ritenere che la probabilità che in un anno in Italia avvenga un terremoto di grande intensità (almeno 6.0 gradi nella scala Richter) sia pari a 0.13. Supponiamo che gli eventi che avvengano terremoti in anni diversi siano indipendenti. Trascuriamo inoltre la possibilità che in un singolo anno avvenga più di un terremoto. Sia X il numero di terremoti di grande intensità che avverranno in Italia nei prossimi 20 anni e sia A l'evento che nei prossimi 20 anni si verifichi più di 1 terremoto di grande intensità.

- Si determini la distribuzione della variabile X e si calcoli la probabilità dell'evento A .
- Si calcoli la probabilità dell'evento A usando l'approssimazione di Poisson.
- Si calcoli la probabilità dell'evento A usando l'approssimazione normale. È lecito farlo?

Soluzione.

a) Per lo schema delle prove ripetute e indipendenti, $X \sim B(20, 0.13)$. Di conseguenza

$$\begin{aligned} P(A) &= P(X > 1) = 1 - P(X = 0) - P(X = 1) = 1 - \binom{20}{0} 0.13^0 0.87^{20} - \binom{20}{1} 0.13^1 0.87^{19} \\ &= 1 - 0.87^{20} - 2 \cdot 0.13 \cdot 0.87^{19} = 1 - 0.062 - 0.184 \simeq 0.754. \end{aligned}$$

b) Per l'approssimazione di Poisson $X \approx Y := Po(20 \cdot 0.13) = Po(2.6)$, quindi

$$P(A) \simeq P(Y > 1) = 1 - P(Y = 0) - P(Y = 1) = 1 - e^{-2.6} - e^{-2.6} \cdot 2.6 = 0.732.$$

c) Secondo le regole viste a lezione non sarebbe lecito usare l'approssimazione normale, perché $np = 20 \cdot 0.13 = 2.6 < 5$. Tuttavia il risultato è comunque molto buono: usando l'approssimazione di continuità, dato che $E(X) = 20 \cdot 0.13 = 2.6$ e $Var(X) = 20 \cdot 0.13 \cdot (1 - 0.13) = 2.262$, si ottiene

$$P(X > 1) = P(X > 1.5) = P\left(\frac{X - 2.6}{\sqrt{2.262}} > \frac{1.5 - 2.6}{\sqrt{2.262}}\right) = P(Z > -0.73) = \Phi(0.73) = 0.767.$$

2.12

Per le donne con un'età compresa tra i 18 e i 24 anni, la pressione sanguigna sistolica è distribuita normalmente con media 114.8 mmHg, e deviazione standard 13.1 mmHg.

- Se si seleziona a caso una donna con un'età compresa tra i 18 e i 24 anni, qual è la probabilità che la sua pressione arteriosa sia maggiore di 140 mmHg?
- Se si selezionano a caso 3 donne con un'età compresa tra i 18 e i 24 anni, qual è la probabilità che la pressione arteriosa di *tutte e tre* sia maggiore di 140 mmHg?

Soluzione.

- a) Sia $X \sim N(114.8, (13.1)^2)$, e $Z = \frac{X-114.8}{13.1} \sim N(0,1)$.

$$P(X > 140) = P\left(Z > \frac{140 - 114.8}{13.1}\right) = P(Z > 1.9237) = 1 - P(Z \leq 1.9237) \simeq 0.0272.$$

- b) Trattandosi di tre prove indipendenti, la probabilità richiesta è $P(X > 140)^3 = (0.0272)^3 \simeq 0.00002$

2.13

In un corso di laurea a numero chiuso si ritiene che il numero di studenti non debba essere superiore a 80. È noto, però, che il 25% degli studenti che superano l'esame di ammissione cambiano idea, non confermando l'iscrizione. Se 100 studenti superano l'esame di ammissione, qual è la probabilità che più di 80 decidano di iscriversi?

Soluzione. Per $i = 1, \dots, 100$, sia

$$X_i = \begin{cases} 1 & \text{se l}'i\text{-mo studente che ha superato l'esame di ammissione si iscrive} \\ 0 & \text{altrimenti} \end{cases}$$

Notare che $X_i \sim Be(0.75)$. Allora, usando anche la correzione di continuità,

$$\begin{aligned} P\left(\sum_{i=1}^{100} X_i > 80\right) &= P\left(\sum_{i=1}^{100} X_i > 80.5\right) = P\left(\frac{\bar{X}_{100} - 0.75}{\sqrt{0.25 \cdot 0.75}} \geq \frac{0.805 - 0.75}{\sqrt{0.25 \cdot 0.75}}\right) \\ &\simeq P(N(0, 1) \geq 2.93) = 1 - \Phi(2.93) \simeq 0.0017. \end{aligned}$$

2.14

Nella battitura di un testo, ogni carattere viene sbagliato con probabilità 0.005, indipendentemente dagli altri. Un articolo contiene 2500 caratteri. Usando l'approssimazione normale, calcolare la probabilità che nell'articolo ci siano oltre 15 errori.

Soluzione. Per $i = 1, \dots, 2500$, sia

$$X_i = \begin{cases} 1 & \text{se l}'i\text{-mo carattere viene sbagliato} \\ 0 & \text{altrimenti} \end{cases}$$

Notare che $X_i \sim Be(2500, 0.005)$. Essendo $2500 \times 0.005 > 5$, è lecito usare l'approssimazione normale. Allora, usando anche la correzione di continuità,

$$\begin{aligned} P\left(\sum_{i=1}^{2500} X_i > 15\right) &= P\left(\frac{\bar{X}_{2500} - 0.005}{\sqrt{0.005 \cdot 0.995}} \sqrt{2500} \geq \frac{15.5 - 0.005}{\sqrt{0.005 \cdot 0.995}} \sqrt{2500}\right) \\ &\simeq P(N(0, 1) \geq 0.85) = 1 - \Phi(0.85) \simeq 0.197. \end{aligned}$$

2.15

In un paese di provincia, i guadagni medi annui di un gruppo familiare sono di 20.000 Euro, con una varianza di 9.000.000 Euro². Se vengono scelti a caso 40 gruppi familiari, qual è la probabilità che la media aritmetica dei loro redditi sia superiore a 20.750 Euro?

Soluzione. Sia X_i il reddito dell' i -esimo gruppo familiare. Dobbiamo calcolare

$$P(\bar{X}_{40} \geq 20750).$$

Usando l'approssimazione normale

$$\begin{aligned} P(\bar{X}_{40} \geq 20750) &= P\left(\frac{\bar{X}_{40} - 20000}{\sqrt{9000000}}\sqrt{40} \geq \frac{20750 - 20000}{\sqrt{9000000}}\sqrt{40}\right) \\ &\simeq P(N(0,1) \geq 1.58) = 1 - \Phi(1.58) = 0.0575. \end{aligned}$$

2.16

Un test a risposta multipla consiste di 100 domande, che prevedono 4 possibili risposte delle quali solo una è corretta (come nella prima parte di questo esame). Uno studente risponde a caso a tutte le domande del test. Si calcoli

(a) la media e la varianza del numero di risposte corrette che egli fornisce;

(b) la probabilità che il numero di risposte corrette sia compreso tra 20 e 30 (impiegando l'approssimazione normale).

Soluzione.

(a). Sia X il numero di risposte corrette. Si ha $X \sim B(100, 0.25)$, per cui $E(X) = 25$, $Var(X) = 100 \times 0.25 \times 0.75 = 18.75$.

(b). Per l'approssimazione normale, $\frac{X-25}{\sqrt{18.75}}$ ha distribuzione approssimativamente $N(0, 1)$. Pertanto, usando anche la correzione di continuità,

$$\begin{aligned} P(20 \leq X \leq 30) &= P(X \leq 30) - P(X < 20) = P(X \leq 30.5) - P(X \leq 19.5) \\ &\simeq P\left(N(0,1) \leq \frac{30.5 - 25}{\sqrt{18.75}}\right) - P\left(N(0,1) \leq \frac{19.5 - 25}{\sqrt{18.75}}\right) = \Phi(1.27) - \Phi(-1.27) = 2\Phi(1.27) - 1 = 0,796 \end{aligned}$$

2.17

Una Compagnia di Assicurazioni vende 10000 assicurazioni sulla vita al costo di 1000 Euro ciascuna. Nel caso di decesso dell'assicurato entro cinque anni dall'acquisto della polizza, la Compagnia paga 100000 Euro alla famiglia. Le statistiche demografiche indicano che ognuno degli assicurati ha probabilità 0.0098 di morire entro cinque anni, indipendentemente dagli altri. Si noti che in bilancio della Compagnia al termine dei cinque anni sarà in attivo se i decessi sono al massimo 99.

a. Qual è la probabilità che il bilancio della Compagnia risulti in attivo?

b. Quanto vale la probabilità al punto a. se il costo della polizza viene aumentato a 1100 Euro?

Soluzione. a. Sia X il numero di decessi tra i 10000 assicurati. Per l'approssimazione normale (che è valida essendo $np = 98 > 5$ e $n(1-p) = 9902 > 5$), $\frac{X-98}{\sqrt{98(1-0.0098)}}$ ha distribuzione approssimativamente $N(0, 1)$. Pertanto, usando anche la correzione di continuità,

$$P(X \leq 99) = P(X \leq 99.5) = P\left(\frac{X - 98}{\sqrt{98(1 - 0.0098)}} \leq \frac{99.5 - 98}{\sqrt{98(1 - 0.0098)}}\right) \simeq \Phi(0.15227) \simeq 0.56.$$

b. In questo caso la Compagnia resta in attivo se il numero di decessi è non più di 109. Pertanto, come sopra,

$$P(X \leq 109) = P(X \leq 109.5) = P\left(\frac{X - 98}{\sqrt{98(1 - 0.0098)}} \leq \frac{109.5 - 98}{\sqrt{98(1 - 0.0098)}}\right) \simeq \Phi(1.1674) \simeq 0.8785.$$

2.18

Il numero giornaliero di passeggeri sui treni da Milano a Firenze è una variabile aleatoria di distribuzione incognita. Supponendo che il valore atteso sia pari a 3000 e la varianza pari a 10^6 , si calcoli approssimativamente la probabilità che in 30 giorni il numero complessivo di viaggiatori sia almeno 10^5 .

Soluzione. Sia X_i il numero di passeggeri de giorno i -mo.

$$\begin{aligned} P(X_1 + \dots + X_{30} > 10^5) &= P(\bar{X}_{30} > \frac{10^5}{30}) \\ &= P\left(\frac{\bar{X}_{30} - 3000}{1000} \sqrt{30} > \frac{333}{1000} \sqrt{30}\right) \simeq 1 - \Phi(1.825) \simeq 0.034. \end{aligned}$$

2.19

Il numero di accessi al giorno alla pagina web della Facoltà di Scienze MM.FF.NN. ha distribuzione di Poisson, il numero medio di accessi al giorno è 74 e i numeri di accessi in giorni distinti sono indipendenti. Qualè la probabilità che il numero di accessi in un anno (= 365 giorni) superi 27200?

Soluzione. Sia X_i il numero di accessi nel giorno i . $X_i \sim Po(74)$. Usiamo l'approssimazione normale. Dato che i numeri dono molto elevati, la correzione di continuità non è rilevante.

$$P(X_1 + \dots + X_{365} > 27200) = P\left(\frac{\bar{X} - 74}{\sqrt{74}/\sqrt{365}} > \frac{27200 - 74}{\sqrt{74}/\sqrt{365}}\right) \simeq 1 - \Phi(1.156) = 0.1238.$$

2.20

I portafogli obbligazionari nei quali investono i grandi istituti finanziari, sono costituiti da centinaia di diverse obbligazioni. Consideriamo il caso semplificato in cui una banca acquista oggi 100 diverse obbligazioni al prezzo di 10 Euro l'una. Si ritiene che il valore di domani delle stesse obbligazioni abbia media 10 Euro, e deviazione standard 3 Euro, e che i valori di domani di obbligazioni diverse siano indipendenti. Calcolare, usando l'approssimazione normale, la probabilità che tra oggi e domani la banca abbia una perdita superiore ai 50 Euro.

Soluzione. Sia X_i la differenza tra il prezzo di domani e quello di oggi dell' i -ma obbligazione. Sappiamo che $E(X_i) = 0$, $Var(X_i) = 9$. Usando l'approssimazione normale:

$$P(X_1 + \dots + X_{100} < -50) = P\left(\frac{\bar{X}}{3/10} < -\frac{0.5}{3/10}\right) \simeq \Phi\left(-\frac{0.5}{3/10}\right) = 1 - \Phi\left(\frac{0.5}{3/10}\right) = 0.0478.$$

2.21

Un certo incrocio tra piante di piselli produce una progenie che può avere fiori rossi oppure fiori bianchi. La probabilità di ottenere una pianta con i fiori rossi è $\frac{9}{16}$. Se vengono esaminate 320 piante nella progenie, qual è la probabilità che non più della metà di esse abbiano i fiori rossi? (Usare l'approssimazione normale)

Soluzione. Sia X_i la variabile che vale 1 se l' i -ma pianta ha i fiori rossi, e 0 altrimenti.

$$\begin{aligned} P(X_1 + \dots + X_{320} \leq 160) &= P(X_1 + \dots + X_{320} \leq 160.5) \\ &= P\left(\frac{\bar{X} - \frac{9}{16}}{\sqrt{\frac{9}{16} \frac{7}{16}}/\sqrt{320}} \leq \frac{\frac{160.5}{320} - \frac{9}{16}}{\sqrt{\frac{9}{16} \frac{7}{16}}/\sqrt{320}}\right) = \Phi\left(\frac{\frac{160.5}{320} - \frac{9}{16}}{\sqrt{\frac{9}{16} \frac{7}{16}}/\sqrt{320}}\right) \simeq 0.014 \end{aligned}$$

2.22

Un generatore di numeri casuali genera numeri compresi tra 0 e 1, secondo una distribuzione di media 0.5 e deviazione standard 0.289. Se vengono generati 100 numeri, qual è la probabilità (approssimata) che la loro media campionaria sia compresa tra 0.47 e 0.53?

Soluzione. Indicando con X_1, \dots, X_{100} i numeri generati, si ha che le variabili X_i sono i.i.d. con $\mu = E(X_i) = 0.5$ e $\sigma = \sqrt{Var(X_i)} = 0.289$. Detta $\bar{X} = (X_1 + \dots + X_{100})/n$ la media campionaria, dobbiamo calcolare $P(0.47 \leq \bar{X} \leq 0.53)$. Usando l'approssimazione normale si ha

$$\begin{aligned} P(0.47 \leq \bar{X} \leq 0.53) &= P\left(\frac{0.47 - 100 \cdot \mu}{\sigma/\sqrt{100}} \leq \frac{\bar{X} - 100 \cdot \mu}{\sigma/\sqrt{100}} \leq \frac{0.53 - 100 \cdot \mu}{\sigma/\sqrt{100}}\right) \\ &\approx P(-1.03 \leq Z \leq 1.03) = 2\Phi(1.03) - 1 = 2 \cdot 0.85 - 1 = 0.7. \end{aligned}$$

2.23

In una elezione votano un milione di persone, che devono scegliere tra i due candidati A e B . Il voto di un certo numero n di elettori è sotto il controllo di una organizzazione malavitosa, che garantisce che essi votino per il candidato A . Tutti gli altri elettori votano "a caso", scegliendo con uguale probabilità i candidati, ognuno indipendentemente dagli altri. Supponiamo che l'organizzazione malavitosa controlli $n = 2000$ voti. Qual è la probabilità (approssimata) che il candidato A vinca le elezioni?

Soluzione. Sia X il numero di voti ricevuti da A tra i 998 000 elettori non controllati. Notare che $X \sim B(998\,000, 1/2)$. Il candidato A vince se $X > 498\,000$. Usando l'approssimazione normale (dati i numeri elevati la correzione di continuità non è rilevante)

$$P(X > 498\,000) = P\left(\frac{X - 499\,000}{\frac{1}{2}\sqrt{998\,000}} > -\frac{1\,000}{\frac{1}{2}\sqrt{998\,000}}\right) \simeq \Phi(2) = 0.977.$$