

# Soluzioni dei problemi proposti

## Capitolo 1

**1.9** Il numero telefonico può terminare con 100 differenti coppie ordinate ( $10 \times 10 = 100$ ). Per cui la coppia 99 (come del resto qualsiasi altra delle 100 possibili) ha probabilità  $1/100$ .

**1.10** La probabilità non può essere “misurata” come se fosse un parametro fisico. La probabilità di un evento è una valutazione umana, effettuata in base ad un insieme d’informazioni (non esaustive) disponibili e che crescono continuamente con l’esperienza. In quest’ambito, certamente i risultati di eventuali “prove sperimentali ripetute” trovano un posto importante, ma non costituiscono né uno strumento e né una procedura di misura della probabilità.

**1.11** Gli eventi  $a$  e  $b$  hanno la stessa probabilità. Se qualcuno rispondesse “è più probabile l’evento  $b$ ” lo farebbe certamente perché si riferirebbe non già a questa specifica successione (unica, come la  $a$ , nell’ambito delle  $2^{12}$  possibili) ma all’insieme delle 924 successioni caratterizzate da sei  $T$  e sei  $C$ .

**1.12** Ha ragione, infatti, essendo:

$$\begin{aligned} \Pr\{B\} &= \Pr\{B|A\} \Pr\{A\} + \Pr\{B|\bar{A}\} \Pr\{\bar{A}\} = \\ &= 0.09091 \times 0.10000 + 0.10101 \times 0.9 = 0.10000 = \Pr\{A\} \end{aligned}$$

l’eguaglianza  $\Pr\{A|B\} = \Pr\{B|A\}$  segue immediatamente. Il risultato vale in generale per le estrazioni senza rimessa da un’urna composta da  $d$  elementi difettosi e  $b$  buoni. Infatti abbiamo:

$$\Pr\{A\} = \frac{d}{d+b}; \quad \Pr\{B|A\} = \frac{d-1}{d+b-1}; \quad \Pr\{B|\bar{A}\} = \frac{d}{d+b-1}$$

da cui:

**542** Soluzioni dei problemi proposti

$$\begin{aligned}\Pr\{B\} &= \Pr\{B|A\} \Pr\{A\} + \Pr\{B|\bar{A}\} \Pr\{\bar{A}\} = \\ &= \frac{d-1}{d+b-1} \times \frac{d}{d+b} + \frac{d}{d+b-1} \times \frac{b}{d+b} = \frac{d}{d+b} = \Pr\{A\}.\end{aligned}$$

**1.13** L'apparente paradosso è conseguenza del fatto che si confondono due eventi diversi: il primo è "acquisto del biglietto vincente da parte di una specifica persona"; il secondo è semplicemente "acquisto del biglietto vincente". Il primo è raro mentre il secondo è certo.

**1.14** La probabilità che i 10 amici siano nati in giorni diversi è:

$$1 \times \frac{364}{365} \times \frac{363}{365} \times \dots \times \frac{356}{365} = \frac{365!}{365^{10} \times 355!} = 0.8831 \cong 88\%$$

quindi la probabilità che almeno due siano nati nello stesso giorno è pari a circa il 12%.

**1.15** Le probabilità richieste possono essere calcolate facilmente con l'aiuto dei diagrammi di Venn:

- $\Pr\{A \cup B\} = 0.5 + 0.4 - 0.3 = 0.6;$
- $\Pr\{\bar{A} \cup \bar{B}\} = 1 - \Pr\{A\} + \Pr\{B\} - \Pr\{B - (A \cap B)\} =$   
 $= (1 - 0.5) + 0.4 - (0.4 - 0.3) = 0.8;$
- $\Pr\{\bar{A} \cap \bar{B}\} = \Pr\{\overline{A \cup B}\} = 1 - 0.6 = 0.4;$
- $\Pr\{\bar{A} \cup \bar{B}\} = (1 - 0.5) + (1 - 0.4) - 0.4 = 0.7.$

**1.16** Dall'equazione  $\Pr\{A \cap B\} = \Pr\{A\} \times \Pr\{B|A\} = \Pr\{B\} \times \Pr\{A|B\}$  otteniamo:

$$\Pr\{B\} = \frac{\Pr\{A\} \times \Pr\{B|A\}}{\Pr\{A|B\}} = \frac{4}{15}.$$

**1.17** Tenendo conto delle ipotesi di incompatibilità ed  $s$ -indipendenza, le probabilità richieste sono:

- $\Pr\{B \cup C\} = 0.2 + 0.3 - 0.06 = 0.44;$
- $\Pr\{A \cup B \cup C\} = \Pr\{A\} + \Pr\{B \cup C\} - \Pr\{(A \cap B) \cup (A \cap C)\} =$   
 $= 0.1 + 0.44 = 0.54;$
- $\Pr\{A|B \cup C\} = \frac{\Pr\{A \cap (B \cup C)\}}{\Pr\{B \cup C\}} = \frac{\Pr\{(A \cap B) \cup (A \cap C)\}}{\Pr\{B \cup C\}} = 0;$
- $\Pr\{B \cap C|B \cup C\} = \frac{\Pr\{(B \cap C) \cap (B \cup C)\}}{\Pr\{B \cup C\}} = \frac{\Pr\{B \cap C\}}{\Pr\{B \cup C\}} = 0.136.$

**1.18** Utilizzando le soluzioni del problema precedente, e tenendo conto ancora una volta delle ipotesi di incompatibilità ed  $s$ -indipendenza, le probabilità richieste sono:

- $\Pr\{\bar{A} \cap \bar{B} \cap \bar{C}\} = 1 - \Pr\{A \cup B \cup C\} = 0.46;$

$$\begin{aligned}
 \text{b)} \quad & \Pr\{(\bar{A} \cap B \cap C) \cup (A \cap \bar{B} \cap C) \cup (A \cap B \cap \bar{C})\} = \\
 & = \Pr\{(\bar{A} \cap B \cap C)\} = \Pr\{(\bar{A} \cap B) \cap (\bar{A} \cap C)\} = \Pr\{B\} \times \Pr\{C\} = 0.06.
 \end{aligned}$$

Le asserzioni “nessuno dei tre si verifica” e “almeno uno si verifica” sono tra loro complementari. L’incompatibilità di  $A$  sia con  $B$  che con  $C$  implica le identità  $\bar{A} \cap B = B$  e  $\bar{A} \cap C = C$ .

**1.19** Conveniamo di indicare col primo numero il punteggio realizzato da Aldo e col secondo quello realizzato da Bruno.

a) Qualunque sia il punteggio realizzato da Aldo, Bruno ne realizza uno maggiore:

$$\begin{aligned}
 & \Pr\{[1 \cap (>1)] \cup [2 \cap (>2)] \cup [3 \cap (>3)] \cup [4 \cap (>4)] \cup [5 \cap (>5)]\} = \\
 & = \frac{1}{6} \times \frac{5}{6} + \frac{1}{6} \times \frac{4}{6} + \frac{1}{6} \times \frac{3}{6} + \frac{1}{6} \times \frac{2}{6} + \frac{1}{6} \times \frac{1}{6} = \frac{15}{36};
 \end{aligned}$$

b) Bruno realizza un punteggio maggiore oppure realizza lo stesso di Aldo e viene favorito dal successivo lancio della moneta:

$$\begin{aligned}
 & \Pr\{[1 \cap (>1)] \cup [2 \cap (>2)] \cup [3 \cap (>3)] \cup [4 \cap (>4)] \cup [5 \cap (>5)]\} + \\
 & + \Pr\{[(1 \cap 1) \cup (2 \cap 2) \cup (3 \cap 3) \cup (4 \cap 4) \cup (5 \cap 5) \cup (6 \cap 6)] \cap (\text{croce})\} = \\
 & = \frac{15}{36} + \frac{6}{36} \times \frac{1}{2} = \frac{18}{36} = \frac{1}{2};
 \end{aligned}$$

c) Bruno realizza un punteggio maggiore di 4 oppure realizza anche lui 4 e viene favorito dal successivo lancio della moneta:

$$\begin{aligned}
 & \Pr\{(>4)\} + \Pr\{(4) \cap (\text{croce})\} = \\
 & = \frac{2}{6} + \frac{1}{6} \times \frac{1}{2} = \frac{5}{12}.
 \end{aligned}$$

**1.20** L’insuccesso di un ciclo può avvenire in 3 maniere diverse a seconda che non riesca il primo intervento o non riesca il secondo o non riescano entrambi. Analogamente per il secondo ciclo. In totale ci sono  $3 \times 3$  maniere differenti in cui si può verificare l’insuccesso dei due cicli. Ogni maniera ha probabilità  $(1/2)^4$  per cui la probabilità d’insuccesso di entrambi i cicli è  $9 \times 1/16 = 0.56$ .

**1.21** Il successo di un ciclo può avvenire in una sola maniera, ossia quando entrambi gli interventi riescono. Un solo successo può verificarsi in 6 maniere diverse a seconda che l’insuccesso sia capitato al primo od al secondo ciclo (2) nonché a seconda che non sia riuscito il primo o il secondo o entrambi gli interventi (3). Ogni maniera ha probabilità  $(1/2)^4$  per cui la probabilità di un solo successo è  $6 \times 1/16 = 0.38$ .

**1.22** Due successi possono verificarsi in una sola maniera allorché entrambi i cicli sono coronati da successo ossia quando si verifica una successione di 4 interventi riusciti. La probabilità che ciò accada è  $(1/2)^4 = 0.06$ .

**1.23** La probabilità dell’evento  $A$  può essere calcolata come rapporto tra il numero di coppie di carte di denari (eventi favorevoli) ed il numero di coppie di carte (eventi

**544** Soluzioni dei problemi proposti

possibili) che possono essere estratte dal mazzo:

$$\Pr\{A\} = \frac{\binom{10}{2}}{\binom{40}{2}} = \frac{10 \times 9}{40 \times 39} = 0.058.$$

Analogamente, la probabilità dell'evento  $B$  può essere calcolata come rapporto tra il numero di coppie di carte di denari e di bastoni (eventi favorevoli) ed il numero di coppie di carte (eventi possibili) che possono essere estratte dal mazzo:

$$\Pr\{B\} = \frac{\binom{10}{1}\binom{10}{1}}{\binom{40}{2}} = 2 \times \frac{10 \times 10}{40 \times 39} = 0.128.$$

**1.24** Considerando ciascuno dei due eventi come concomitanze di specifiche estrazioni (ed utilizzando una simbologia intuitiva) abbiamo rispettivamente:

$$\Pr\{A\} = \Pr\{I_d \cap II_d\} = \Pr\{I_d\} \times \Pr\{II_d | I_d\} = \frac{10}{40} \times \frac{9}{39} = 0.058;$$

$$\begin{aligned} \Pr\{B\} &= \Pr\{(I_d \cap II_b) \cup (I_b \cap II_d)\} = \Pr\{I_d\} \times \Pr\{II_b | I_d\} + \Pr\{I_b\} \times \Pr\{II_d | I_b\} = \\ &= \frac{10}{40} \times \frac{10}{39} + \frac{10}{40} \times \frac{10}{39} = 2 \times \frac{10}{40} \times \frac{10}{39} = 0.128. \end{aligned}$$

**1.25** Entrambi gli eventi si possono presentare in  $C_{5,2}$  maniere diverse (incompatibili) ossia a seconda del numero di ordinamenti possibili di 5 pezzi di cui 2 di un tipo e 3 di un altro. Inoltre, considerando che ognuna delle maniere costituisce una concomitanza di eventi elementari la cui probabilità composta è sempre la stessa, otteniamo che:

$$\Pr\{2_A\} = \binom{5}{2} \times \frac{5}{15} \times \frac{4}{14} \times \frac{10}{13} \times \frac{9}{12} \times \frac{8}{11} = 0.400;$$

$$\Pr\{2_B\} = \binom{5}{2} \times \frac{500}{1500} \times \frac{499}{1499} \times \frac{1000}{1498} \times \frac{999}{1497} \times \frac{998}{1496} = 0.330.$$

Le stesse probabilità possono essere valutate come rapporto tra il numero di differenti gruppi di 5 pezzi di cui 2 difettosi (eventi favorevoli) ed il numero di differenti gruppi di 5 pezzi qualsiasi (eventi possibili) potenzialmente estraibili da entrambi i lotti:

$$\Pr\{2_A\} = \frac{\binom{5}{2} \times \binom{10}{3}}{\binom{15}{5}} = 0.400; \quad \Pr\{2_B\} = \frac{\binom{500}{2} \times \binom{1000}{3}}{\binom{1500}{5}} = 0.330$$

**1.26** Per il lotto  $A$  i risultati sono differenti; per il lotto  $B$  lo sono molto meno:

$$\Pr\{2_A\} = \binom{5}{2} \times \left(\frac{5}{15}\right)^2 \times \left(\frac{10}{15}\right)^3 = 0.329 \neq 0.400$$

$$\Pr\{2_B\} = \binom{5}{2} \times \left(\frac{500}{1500}\right)^2 \times \left(\frac{1000}{1500}\right)^3 = 0.329 \cong 0.330.$$

In generale, le valutazioni in caso di estrazione con e senza rimessa tendono a coincidere quanto più è inferiore a 0.10 il rapporto tra il numero di pezzi estratti ed il numero di pezzi presenti nel lotto.

**1.27** Nell'ipotesi che i lanci possano essere ritenuti eventi  $s$ -indipendenti (perché innanzitutto le condizioni di lancio sono omogenee, nonché sono assenti fenomeni di allenamento o affaticamento del concorrente) la probabilità  $p$  di centrare il bersaglio almeno una volta in  $n$  lanci è il complemento all'unità di non centrarlo mai. Pertanto, dovendo essere  $p = 1 - (1 - 0.75)^n \geq 0.99$ , troviamo per tentativi che  $n = 4$ . Infatti ad  $n = 3$  corrisponde  $p = 0.98$  e ad  $n = 4$  corrisponde  $p = 0.996$ .

**1.28** La probabilità di vincere un terno è:

$$p = \binom{5}{3} : \binom{90}{3} = \frac{5 \times 4 \times 3}{90 \times 89 \times 88} = 8.512 \cdot 10^{-5}$$

che in termini di scommessa coerente può essere espressa dal rapporto  $a/(a+b) = p$ , essendo "a" la somma che siamo disposti a scommettere per guadagnare "b" se l'evento  $T$  si verificherà. Da cui, tenendo conto che deve essere  $b/(a+b) = 1 - p$ , otteniamo che la scommessa coerente sarebbe di  $a = 1$  euro contro  $b = 11.747,00$  euro.

**1.29** Applicando il teorema di Bayes abbiamo:

$$\begin{aligned} & \Pr\{100 \text{ incontaminati} | 10 \text{ incontaminati}\} = \\ & = \frac{\Pr\{10 \text{ incontaminati} | 100 \text{ incontaminati}\} \times \Pr\{100 \text{ incontaminati}\}}{\Pr\{10 \text{ incontaminati}\}} = \\ & = \frac{1 \times 1/4}{\sum_{i=0}^3 (\Pr\{10 \text{ incontaminati} | i \text{ esemplari contaminati su } 100\} \times \Pr\{i\})} = \\ & = \frac{1/4}{\sum_{i=0}^3 \binom{100-i}{10} : \binom{100}{10} \times 1/4} = \frac{1/4}{1 \times 1/4 + 0.900 + 0.809 + 0.727} = 0.0931. \end{aligned}$$

**1.30** A priori possiamo dire che il lotto esaminato ha la stessa probabilità ( $1/3$ ) di essere quello indicato con  $A$ ,  $B$  o  $C$ . Inoltre le probabilità di estrarre un esemplare di 1<sup>a</sup> scelta dai lotti  $A$ ,  $B$  e  $C$  sono rispettivamente 1, 1 e  $1/3$ . L'evento cui siamo interessati implica la concomitanza che il lotto esaminato sia quello indicato con  $C$  (l'unico che ha esemplari di 2<sup>a</sup> scelta) e che da questo sia estratto un esemplare di 2<sup>a</sup> scelta. Applicando il teorema di Bayes, la probabilità di questo evento è:

**546 Soluzioni dei problemi proposti**


---

$$\begin{aligned} \Pr\left\{\left(C \cap 2^{\text{a scelta}}\right) \middle| 1^{\text{a scelta}}\right\} &= \Pr\{C \mid 1^{\text{a scelta}}\} \times \Pr\{2^{\text{a scelta}} \mid C\} = \\ &= \frac{\Pr\{1^{\text{a scelta}} \mid C\} \times \Pr\{C\}}{\Pr\{1^{\text{a scelta}}\}} \times \Pr\{2^{\text{a scelta}} \mid C\} = \\ &= \frac{1/3 \times 1/3}{1 \times 1/3 + 1 \times 1/3 + 1/3 \times 1/3} \times 2/3 = 0.0952. \end{aligned}$$

Si noti che questa probabilità è più che dimezzata rispetto a quella assoluta (grazie all'applicazione del teorema di Bayes che ha permesso l'utilizzo dell'informazione sperimentale "l'esemplare controllato è di 1<sup>a</sup> scelta"):

$$\Pr\{C \cap 2^{\text{a}}\} = \Pr\{C\} \times \Pr\{2^{\text{a scelta}} \mid C\} = 1/3 \times 2/3 = 0.222 > 2 \times 0.0952.$$

## Capitolo 2

**2.9** Indicando con  $b$  un pezzo buono e con  $c$  un pezzo cattivo, i possibili risultati della prima estrazione sono:  $\{b \cap b\}$ ,  $\{b \cap c\}$ ,  $\{c \cap b\}$ ,  $\{c \cap c\}$ , rispettivamente di probabilità  $p_1=30/50 \times 29/49$ ,  $p_2=30/50 \times 20/49$ ,  $p_3=20/50 \times 30/49$ ,  $p_4=20/50 \times 19/49$ . Pertanto, applicando la regola di Bayes, la richiesta probabilità è data complessivamente dalla somma:

$$\Pr\{b\} = \frac{15+2}{27} p_1 + \frac{15+1}{27} (p_2 + p_3) + \frac{15}{27} p_4.$$

**2.10** Applicando il Teorema di Bayes, la probabilità che l'esemplare sia non conforme ( $\bar{C}$ ), una volta che è stato scartato ( $SC$ ), è data da:

$$\begin{aligned} \Pr\{\bar{C} \mid SC\} &= \frac{\Pr\{\bar{C} \cap SC\}}{\Pr\{SC\}} = \frac{\Pr\{SC \mid \bar{C}\} \Pr\{\bar{C}\}}{\Pr\{SC \mid \bar{C}\} \Pr\{\bar{C}\} + \Pr\{SC \mid C\} \Pr\{C\}} = \\ &= \frac{0.90 \cdot 0.001}{0.90 \cdot 0.001 + 0.01 \cdot (1 - 0.001)} = 0.0826. \end{aligned}$$

**2.11** Grazie all'ipotesi di  $s$ -indipendenza non è necessario calcolare la probabilità di  $C$ . Infatti:

$$\Pr\{(A \cap B) \mid C\} = \frac{\Pr\{A \cap B \cap C\}}{\Pr\{C\}} = \Pr\{A \cap B\} = \frac{1}{5} \times \frac{1}{4} = \frac{1}{20}.$$

**2.12** Anche se risulta  $\Pr\{A \cap B \cap C\} = \Pr\{A\} \times \Pr\{B\} \times \Pr\{C\}$ , gli eventi  $A$ ,  $B$  e  $C$  non sono  $s$ -indipendenti perché  $\Pr\{A \cap B\} \neq \Pr\{A\} \times \Pr\{B\}$ ,  $\Pr\{A \cap C\} \neq \Pr\{A\} \times \Pr\{C\}$ , e  $\Pr\{B \cap C\} \neq \Pr\{B\} \times \Pr\{C\}$ .

**2.13** A differenza del problema precedente, ora risultano verificate le condizioni  $\Pr\{A \cap B\} = \Pr\{A\} \times \Pr\{B\}$ ,  $\Pr\{A \cap C\} = \Pr\{A\} \times \Pr\{C\}$ ,  $\Pr\{B \cap C\} = \Pr\{B\} \times \Pr\{C\}$ . Tuttavia, essendo  $\Pr\{A \cap B \cap C\} \neq \Pr\{A\} \times \Pr\{B\} \times \Pr\{C\}$ , anche questa volta gli eventi  $A$ ,  $B$  e  $C$  non

sono  $s$ -indipendenti.

**2.14** Indicando con  $d_I$  e  $d_{II}$  il numero dei pezzi non conformi estratti, rispettivamente dal I e dal II lotto (nell'ipotesi che il controllo dei lotti avvenga indipendentemente uno dall'altro) la probabilità di superare il controllo è:

$$\Pr\{(d_I = 0) \cap (d_{II} = 0)\} = \Pr\{d_I = 0\} \Pr\{d_{II} = 0\} = \frac{\binom{3}{0} \binom{17}{2}}{\binom{20}{2}} \times \frac{\binom{3}{0} \binom{17}{2}}{\binom{20}{2}} = 0.512.$$

Viceversa, se la fornitura avviene in un'unica soluzione, la probabilità di superare il controllo è lievemente inferiore e pari a:

$$\binom{34}{4} : \binom{40}{4} = 0.507.$$

**2.15** Dai dati in nostro possesso possiamo ricavare le seguenti valutazioni (indicando con  $C_i$  l'evento "il coperchio  $i$ -esimo copre la moneta", con  $P_1$  l'evento "abbiamo puntato sul coperchio n° 1" e con  $S_2$  l'evento "l'avversario ha scoperto il coperchio n° 2"):

$$\Pr\{C_i\} = \frac{1}{3}; \quad i=1, 2, 3$$

$$\Pr\{S_2|(C_1 \cap P_1)\} = \frac{1}{2}; \quad \Pr\{S_2|(C_2 \cap P_1)\} = 0; \quad \Pr\{S_2|(C_3 \cap P_1)\} = 1$$

e quindi, applicando la regola di Bayes, abbiamo le valutazioni *a posteriori*:

$$\begin{aligned} \Pr\{C_1|[S_2|(C_1 \cap P_1)]\} &= \frac{\Pr\{C_1\} \Pr\{S_2|(C_1 \cap P_1)\}}{\sum_{i=1}^3 \Pr\{C_i\} \Pr\{S_2|(C_i \cap P_1)\}} = \\ &= \frac{1/3 \times 1/2}{1/3 \times (1/2 + 0 + 1)} = \frac{1}{3} \end{aligned}$$

$$\begin{aligned} \Pr\{C_3|[S_2|(C_3 \cap P_1)]\} &= \frac{\Pr\{C_3\} \Pr\{S_2|(C_3 \cap P_1)\}}{\sum_{i=1}^3 \Pr\{C_i\} \Pr\{S_2|(C_i \cap P_1)\}} = \\ &= \frac{1/3 \times 1}{1/3 \times (1/2 + 0 + 1)} = \frac{2}{3}. \end{aligned}$$

Quindi, ancora una volta abbiamo la conferma che la strategia del cambiamento della puntata è quella che sfrutta meglio le informazioni di cui disponiamo *al momento*.

**2.16** I casi possibili sono  $6^3 = 216$ , mentre quelli favorevoli ad un incasso minore di 60 euro sono quelli riportati nella tabella seguente:

**548** Soluzioni dei problemi proposti

Caso	Vendita	Incasso	Probabilità
1	10,10,10	30	$(1/6)^3$
2	20,10,10	40	$(1/6)^3$
3	10,20,10	40	$(1/6)^3$
4	10,10,20	40	$(1/6)^3$
5	10,20,20	50	$(1/6)^3$
6	20,10,20	50	$(1/6)^3$
7	20,20,10	50	$(1/6)^3$
8	30,10,10	50	$(1/6)^3$
9	10,30,10	50	$(1/6)^3$
10	10,10,30	50	$(1/6)^3$

Conseguentemente:

$$\begin{aligned} \Pr\{\text{incasso} \geq 60\} &= 1 - \Pr\{\text{incasso} < 60\} = \\ &= 1 - \frac{10}{216} = 1 - 10 \times \left(\frac{1}{6}\right)^3 = 0.954. \end{aligned}$$

**2.17** L'evento si verifica se e solo se le prime 2 vendite hanno per oggetto altrettanti prodotti di costo inferiore a 50 euro ed il terzo no:

$$\Pr\{A\} = \left(\frac{4}{6}\right)^2 \times \left(1 - \frac{4}{6}\right) = 0.944.$$

**2.18** L'evento si verifica se e solo se le prime 3 vendite hanno per oggetto altrettanti prodotti di costo inferiore a 60 euro:

$$\Pr\{B\} = \left(\frac{5}{6}\right)^3 = 0.579.$$

**2.19** L'evento "almeno un successo" ( $n \geq 1$ ) è il complementare dell'evento "nessun successo" ( $n = 0$ ). Dovendo essere:

$$\Pr\{n \geq 1\} = 1 - \Pr\{n = 0\} = 1 - (1 - 0.60)^x \geq 0.90$$

ricaviamo che il numero minimo di tentativi da effettuare è:

$$x \geq \frac{\ln(0.10)}{\ln(0.40)} = 2.51.$$

**2.20** Le coppie distinte di lettere sono  $D_{24,2} = 24 \times 23$  mentre le triplette distinte di cifre sono  $D_{10,3} = 10 \times 9 \times 8$ . Quindi sono  $24 \times 23 \times 10 \times 9 \times 8 = 397'440$  i codici distintivi che possiamo realizzare. Pertanto, la probabilità di violarlo con un tentativo con-



dotto a caso è pari a  $2.52 \times 10^{-6}$ .

**2.21** Il numero di codici differenti è aumentato di un fattore  $P_5/(P_2 \times P_3) = 10$  pari al numero di possibili *ulteriori* ordinamenti. Il numero totale è quindi pari a 3'974'400. Ovviamente, giungiamo allo stesso risultato anche moltiplicando il numero di possibili scelte delle 2 lettere ( $C_{24,2} = 276$ ) e delle 3 cifre ( $C_{10,3} = 120$ ) per il numero delle possibili permutazioni di 5 oggetti distinti ( $P_5 = 120$ ). La corrispondente probabilità di violarlo con un tentativo condotto a caso è, quindi, 10 volte più bassa di prima.

**2.22** Per quanto ne sappiamo, con la stessa probabilità il tecnico può aver sostituito una scheda di un qualsiasi tipo con una di un qualsiasi altro, purché differente dal primo. Applicando la regola di Bayes:

$$\begin{aligned} \Pr\{A\} &= \Pr\{A|(A \cap B)\} \times \Pr\{A \cap B\} + \Pr\{A|(A \cap C)\} \times \Pr\{A \cap C\} + \\ &+ \Pr\{A|(A \cap D)\} \times \Pr\{A \cap D\} + \Pr\{B|(B \cap A)\} \times \Pr\{B \cap A\} + \\ &+ \Pr\{B|(B \cap C)\} \times \Pr\{B \cap C\} + \Pr\{B|(B \cap D)\} \times \Pr\{B \cap D\} + \\ &+ \Pr\{C|(C \cap A)\} \times \Pr\{C \cap A\} + \Pr\{C|(C \cap B)\} \times \Pr\{C \cap B\} + \\ &+ \Pr\{C|(C \cap D)\} \times \Pr\{C \cap D\} + \Pr\{D|(D \cap A)\} \times \Pr\{D \cap A\} + \\ &+ \Pr\{D|(D \cap B)\} \times \Pr\{D \cap B\} + \Pr\{D|(D \cap C)\} \times \Pr\{D \cap C\} = \\ &= 3 \times \left( \frac{12}{52} \times \frac{1}{4} \times \frac{1}{3} \right) + 3 \times \left( \frac{14}{52} \times \frac{1}{4} \times \frac{1}{3} + 2 \times \frac{13}{52} \times \frac{1}{4} \times \frac{1}{3} \right) = \\ &= \frac{1}{12} \times \frac{3 \times 12 + 3 \times 14 + 6 \times 13}{52} = \frac{1}{12} \times \frac{156}{52} = 0.250. \end{aligned}$$

Il risultato è ovvio, essendoci simmetria tra i 4 casi possibili.

**2.23** Le probabilità richieste sono:

$$\Pr\{L\} = 86/100 = 0.86; \quad \Pr\{L|D\} = 39/47 = 0.83; \quad \Pr\{L|U\} = 47/53 = 0.89.$$

Il fatto che le probabilità condizionate siano diverse da quelle assolute indicano che l'occupazione dipende dal sesso.

**2.24** Applicando la regola di Bayes abbiamo:

$$\begin{aligned} \Pr\{L\} &= \Pr\{L|D\} \times \Pr\{D\} + \Pr\{L|U\} \times \Pr\{U\} = \\ &= 0.83 \times 0.47 + 0.89 \times 0.53 = 0.86. \end{aligned}$$

ossia la probabilità assoluta coincide con la "media pesata" di quelle condizionate.

**2.25** Applicando il teorema di Bayes abbiamo:

$$\begin{aligned} \Pr\{\bar{C}|S_1\} &= \frac{\Pr\{\bar{C}\} \times \Pr\{S_1|\bar{C}\}}{\Pr\{C\} \times \Pr\{S_1|C\} + \Pr\{\bar{C}\} \times \Pr\{S_1|\bar{C}\}} = \\ &= \frac{0.001 \times 0.90}{0.999 \times 0.01 + 0.001 \times 0.90} = 0.0826 \end{aligned}$$

essendo le probabilità di diagnosi (corretta o scorretta) delle probabilità condizionate per natura. La valutazione 0.0826 può essere interpretata come un aggiornamento del dato storico 0.001 alla luce del risultato sperimentale e delle qualità diagnostiche del

**550** Soluzioni dei problemi proposti

sistema adoperato.

**2.26** Applicando nuovamente il teorema di Bayes e tenendo conto della valutazione della probabilità di non conformità appena effettuata, abbiamo:

$$\begin{aligned} \Pr\{\bar{C}|(S_1 \cap S_2)\} &= \frac{\Pr\{\bar{C}|S_1\} \times \Pr\{S_2|\bar{C}\}}{\Pr\{C|S_1\} \times \Pr\{S_2|C\} + \Pr\{\bar{C}|S_1\} \times \Pr\{S_2|\bar{C}\}} = \\ &= \frac{0.0826 \times 0.90}{0.917 \times 0.01 + 0.0826 \times 0.90} = 0.890. \end{aligned}$$

Riscontriamo un evidente ed ulteriore “affievolimento della memoria” del dato storico assoluto  $\Pr\{\bar{C}\} = 0.001$  (valido per *tutta* la produzione) nella valutazione riguardante un esemplare *specifico*. Questa seconda applicazione del teorema di Bayes differisce dalla prima unicamente per la sostituzione dell’informazione a priori  $\Pr\{\bar{C}\} = 0.001$  con quella *aggiornata*  $\Pr\{\bar{C}|S_1\} = 0.0826$ .

**2.27** Basta applicare la regola di Bayes:

$$\begin{aligned} \Pr\{S\} &= \Pr\{S|\bar{C}\} \times \Pr\{\bar{C}\} + \Pr\{S|C\} \times \Pr\{C\} = \\ &= 0.90 \times 0.001 + 0.01 \times 0.999 = 0.011 \end{aligned}$$

avendo questa volta, ovviamente, utilizzato la valutazione della probabilità di non conformità relativa a tutti gli esemplari e non quella relativa all’esemplare esaminato.

**2.28** Ancora una volta basta applicare il teorema di Bayes tenendo conto che la coppia mista  $(A_1A_2)$  può verificarsi in due modalità diverse (la prima pompa di tipo  $A_1$  in coppia con la seconda di tipo  $A_2$  e viceversa):

$$\begin{aligned} \Pr\{A_1A_1|G\} &= \frac{\Pr\{A_1A_1\} \Pr\{G|A_1A_1\}}{\Pr\{A_1A_1\} \Pr\{G|A_1A_1\} + 2 \Pr\{A_1A_2\} \Pr\{G|A_1A_2\} + \Pr\{A_2A_2\} \Pr\{G|A_2A_2\}} = \\ &= \frac{C_{3,2}/C_{10,2} \times 0.04}{C_{3,2}/C_{10,2} \times 0.04 + 2 \frac{3 \times 7}{C_{10,2}} 0.006 + C_{7,2}/C_{10,2} \times 0.003} = \\ &= \frac{0.0027}{0.0027 + 0.0056 + 0.0014} = 0.2784. \end{aligned}$$

**2.29** In effetti sarebbe opportuno valutare almeno la probabilità delle altre tre ipotesi alternative  $(A_1A_2, A_2A_1$  e  $A_2A_2)$ .

Ripetendo l’applicazione del teorema di Bayes utilizzando gli stessi dati di prima, risulta  $\Pr\{A_1A_2|G\} = \Pr\{A_2A_1|G\} = 0.2887$  e  $\Pr\{A_2A_2|G\} = 0.1443$ . Pertanto l’ipotesi della fornitura conforme alle richieste  $(A_2A_2)$  risulta probabile solo al 14%, contro l’86% dell’unione degli altri tre casi di non conformità  $(A_1A_2, A_2A_1$  e  $A_1A_1)$ .

**2.30** È facile riconoscere che gli stati sono  $3^2$  e sono costituiti dai 3 stati (0-0), (1-1) e (2-2) e dalle 3 coppie di stati equivalenti (0-1) e (1-0), (0-2) e (2-0), (1-2) e (2-1).

Attesa l'ipotesi di  $s$ -indipendenza è facile calcolare la probabilità di ogni stato come prodotto delle due probabilità con i pedici corrispondenti. Tuttavia in casi più complessi risulta utile ricorrere alla sviluppo della potenza:

$$(p_0 + p_1 + p_2)^2 = p_0^2 + p_1^2 + p_2^2 + 2p_0p_1 + 2p_0p_2 + 2p_1p_2$$

che fornisce automaticamente le probabilità di tutti gli stati tenendo anche conto dell'esistenza di probabilità uguali.

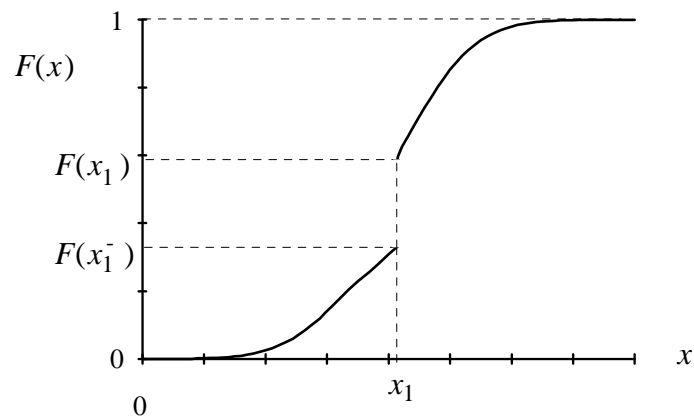
### Capitolo 3

**3.9** La differenza quadratica media rispetto ad un generico valore  $x$ , della v.a.  $X$ , è data da:

$$\begin{aligned} E\{[X - x]^2\} &= E\{[(X - E\{X\}) + (E\{X\} - x)]^2\} = \\ &= \text{Var}\{X\} + (E\{X\} - x)^2 \end{aligned}$$

che assume il valore minimo per  $x$  coincidente con la media (similmente al fatto che il momento d'inerzia è minimo rispetto al baricentro).

**3.10** La seguente Cdf è relativa ad una v.a. mista, definita sul semiasse positivo, con un accumulo di massa nel punto  $x_1$ , di valore  $F(x_1) - F(x_1^-)$ :



Questa v.a., pur essendo non continua, ha un insieme di definizione continuo.

**3.11** La differenza consiste nel fatto che un insieme di eventi a probabilità nulla *può verificarsi* contrariamente ad un insieme di eventi impossibili.

**3.12** La risposta è affermativa: una v.a. mista ha queste caratteristiche.

**3.13** La media della v.a. discreta è per definizione data da:

**552** Soluzioni dei problemi proposti

$$\begin{aligned} E\{y\} &= \sum_{y=0}^n y \binom{n}{y} p^y (1-p)^{n-y} = np \sum_{y=1}^{n-1} \binom{n-1}{y-1} p^{y-1} (1-p)^{(n-1)-(y-1)} = \\ &= np \sum_{x=0}^{n-1} \binom{n-1}{x} p^x (1-p)^{n-1-x} = np \end{aligned}$$

avendo posto  $n-1=m$  ed  $y-1=x$ .

**3.14** La richiesta varianza è:

$$\text{Var}\{y\} = \sum_{y=0}^n (y-np)^2 \binom{n}{y} p^y (1-p)^{n-y}$$

da cui, utilizzando i risultati del Problema 3.13, otteniamo:

$$\begin{aligned} \text{Var}\{y\} &= (np)^2 - 2np \times np + \sum_{y=1}^n y^2 \binom{n}{y} p^y (1-p)^{n-y} = \\ &= -(np)^2 + np \sum_{y=1}^n [(y-1)+1] \binom{n-1}{y-1} p^{y-1} (1-p)^{(n-1)-(y-1)} = \\ &= -(np)^2 + np \times (n-1)p + np = np(1-p). \end{aligned}$$

**3.15** La media della v.a. continua è:

$$\begin{aligned} E\{x\} &= \int_0^{\infty} x \lambda e^{-\lambda x} dx = -[x e^{-\lambda x}]_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx = \\ &= 0 - \frac{1}{\lambda} [e^{-\lambda x}]_0^{\infty} = \frac{1}{\lambda}. \end{aligned}$$

**3.16** Utilizzando la soluzione del Problema 3.15, otteniamo che la richiesta varianza è data da:

$$\begin{aligned} \text{Var}\{x\} &= \int_0^{\infty} (x-1/\lambda)^2 \lambda e^{-\lambda x} dx = -\frac{1}{\lambda^2} [x^2 e^{-\lambda x}]_0^{\infty} + \int_0^{\infty} 2x e^{-\lambda x} dx = \\ &= -\frac{1}{\lambda^2} + 0 + \frac{2}{\lambda} \times \frac{1}{\lambda} = \frac{1}{\lambda^2}. \end{aligned}$$

**3.17** Dalla trasformazione inversa  $X=(1-Y)^{1/3}$  deduciamo  $\left| \frac{dx}{dy} \right| = \frac{1}{3}(1-y)^{-2/3}$  e

quindi:

$$f_Y(y) = 1 / \left\{ \pi \left[ 1 + (1-y)^{2/3} \right] \right\} \frac{1}{3} (1-y)^{-2/3}; \quad -\infty < y < \infty.$$

**3.18** La funzione  $Y = \cos(X)$  non è biunivoca (essendo una curva convessa e simmetrica rispetto all'asse delle ordinate) nell'intervallo  $-\pi/2 \leq x \leq \pi/2$ , a cui corrisponde  $0 \leq y \leq 1$ . Dal suo diagramma deduciamo:

$$\begin{aligned} F_Y(y) &= [F_X(-x) - F_X(-\pi/2)] + [F_X(\pi/2) - F_X(x)] = \\ &= [F_X(-x) - 0] + [1 - F_X(x)] \end{aligned}$$

da cui:

$$\begin{aligned} f_Y(y) &= \frac{dF_Y(y)}{dy} = f_X(-x) \frac{dx}{dy} - f_X(x) \frac{dx}{dy} = \\ &= \{f_X[-\arccos(y)] + f_X[\arccos(y)]\} / \sqrt{1-y^2} = \\ &= (2/\pi) / \sqrt{1-y^2}; \quad \text{per } 0 \leq y \leq 1 \end{aligned}$$

essendo la derivata della trasformata inversa:

$$\frac{dx}{dy} = \frac{d\arccos(y)}{dy} = \frac{-1}{\sqrt{1-y^2}}.$$

**3.19** La funzione  $Y = X^2 + 1$  non è biunivoca per  $-1 \leq x \leq 1$  (a cui corrisponde  $1 \leq y \leq 2$ ). Invece è biunivoca per  $1 < x \leq 2$  (a cui corrisponde  $2 < y \leq 5$ ). Pertanto, essendo la trasformata inversa e la sua derivata uguali rispettivamente a:

$$X = \sqrt{Y-1}; \quad \left| \frac{dx}{dy} \right| = \frac{1}{2}(y-1)^{-1/2}$$

abbiamo:

$$f_Y(y) = \begin{cases} \frac{\sqrt{y-1}}{3} & \text{per } 1 \leq y \leq 2; \\ \frac{\sqrt{y-1}}{6} & \text{per } 2 < y \leq 5. \end{cases}$$

Per ulteriore esercizio, si verifichi che è pari all'unità l'integrale della  $f_Y(y)$  per  $1 \leq y \leq 5$ .

**3.20** Dalla trasformazione inversa  $X = (Y-1)^{1/3}$  deduciamo  $\left| \frac{dx}{dy} \right| = \frac{1}{3}(y-1)^{-2/3}$  e quindi:

$$f_Y(y) = \frac{1}{2} \times \frac{1}{3} (y-1)^{-2/3}; \quad 1 \leq y \leq 9.$$

**3.21** La pdf della vita residua all'età  $t_0$  non è altro che una pdf condizionata a  $T > t_0$ :

$$\begin{aligned} \Pr\{(t < T \leq t+dt) | T > t_0\} &= \frac{\Pr\{(t < T \leq t+dt) \cap (T > t_0)\}}{\Pr\{T > t_0\}} = \\ &= \frac{f_T(t) dt}{1 - F_T(t_0)} = \frac{2\lambda t e^{-\lambda t^2} dt}{e^{-\lambda t_0^2}} = 2\lambda t e^{-\lambda(t^2 - t_0^2)} dt; \quad t > t_0. \end{aligned}$$

È come dire che la massa di probabilità che compete a  $(t, t+dt)$  è rimasta quella originaria; invece, le alternative possibili sono diminuite del tratto  $(0, t_0)$ . Pertanto, solo dividendola per  $1 - F_T(t_0)$ , il suo integrale esteso da  $t_0$  ad  $\infty$  risulta essere uguale all'unità. Ovviamente, giungiamo allo stesso risultato se calcoliamo la derivata della

**554** Soluzioni dei problemi proposti

Cdf condizionata a  $T > t_0$  (3.7).

**3.22** Utilizzando l'operatore speranza matematica per il calcolo della media:

$$E\{Y\} = E\{X^2 + 1\} = 1 + \int_{-1}^2 \frac{x^4}{3} dx = 1 + \left[ \frac{x^5}{15} \right]_{-1}^2 = 1 + \frac{11}{5} = \frac{16}{5}$$

e per il calcolo del momento secondo:

$$E\{Y^2\} = E\{(X^2 + 1)^2\} = 1 + 2 \left[ \frac{x^5}{15} \right]_{-1}^2 + \left[ \frac{x^7}{21} \right]_{-1}^2 = 1 + 2 \frac{11}{5} + \frac{43}{7} = \frac{404}{35}$$

abbiamo anche i due addendi per calcolare la varianza:

$$\text{Var}\{Y\} = E\{Y^2\} - [E\{Y\}]^2 = \frac{404}{35} - \left(\frac{16}{5}\right)^2 = \frac{228}{175}.$$

Ovviamente otteniamo gli stessi risultati anche se invece della pdf della v.a.  $X$  utilizziamo la pdf della v.a.  $Y$  ottenuta come soluzione del problema 3.19. Ad esempio per la media abbiamo:

$$E\{Y\} = \int_{-\infty}^{+\infty} y f_Y(y) dy = \int_1^2 y \frac{(y-1)^{0.5}}{3} dy + \int_2^5 y \frac{(y-1)^{0.5}}{6} dy = \frac{16}{45} + \frac{128}{45} = \frac{16}{5}.$$

**3.23** Innanzitutto calcoliamo la Cdf della v.a.  $X$  (definita per  $-1 \leq x \leq 2$ ) integrando la pdf data:

$$F_X(x) = \frac{1+x^3}{9}; \quad \text{per } -1 \leq x \leq 2$$

essendo  $1/9$  la costante d'integrazione della pdf (necessaria affinché la Cdf sia positiva, non decrescente e tale che  $F_X(-1) = 0$  ed  $F_X(2) = 1$ ).

Per definizione di Cdf di  $Y$  è:

$$\begin{aligned} F_Y(y) &= \Pr\{Y \leq y\} = \Pr\{X^2 \leq y-1\} = \Pr\{-\sqrt{y-1} \leq X \leq \sqrt{y-1}\} = \\ &= \Pr\{X \leq \sqrt{y-1}\} - \Pr\{X \leq -\sqrt{y-1}\} \\ &= F_X(\sqrt{y-1}) - F_X(-\sqrt{y-1}). \end{aligned}$$

Considerando che la funzione di trasformazione è una parabola con vertice nel punto  $(0, 1)$ , deduciamo che all'intervallo di definizione  $[-1, 2]$  per  $X$  corrisponde l'intervallo  $[1, 5]$  per  $Y$ . Inoltre, poiché a valori  $y > 2$  corrispondono valori negativi  $x < -1$  per i quali  $F_X(x) = 0$ , possiamo scrivere:

$$F_Y(y) = \begin{cases} \frac{1+(y-1)^{3/2}}{9} - \frac{1+[-(y-1)^{3/2}]}{9} = \frac{2(y-1)^{3/2}}{9}; & \text{per } 1 \leq y \leq 2 \\ \frac{1+(y-1)^{3/2}}{9} - 0 = \frac{1+(y-1)^{3/2}}{9}; & \text{per } 2 < y \leq 5. \end{cases}$$

**3.24** Calcoliamo la Cdf della v.a.  $X$  integrando la pdf data

$$F_X(x) = \frac{x}{2}; \quad \text{per } 0 \leq x \leq 2.$$

Per definizione di Cdf di  $Y$  è:

$$\begin{aligned} F_Y(y) &= \Pr\{Y \leq y\} = \Pr\{X^3 \leq y-1\} = \Pr\{X \leq (y-1)^{1/3}\} = \\ &= F_X[(y-1)^{1/3}] = \frac{(y-1)^{1/3}}{2}; \quad \text{per } 1 \leq y \leq 9. \end{aligned}$$

**3.25** Imponiamo che sia pari a 0.5 il valore della Cdf di  $Y$ :

$$\begin{aligned} F_Y(y) &= \Pr\{Y \leq y\} = \Pr\{(X-1)^2 \leq y\} = \Pr\{1-\sqrt{y} \leq X \leq 1+\sqrt{y}\} = \\ &= F_X(1+\sqrt{y}) - F_X(1-\sqrt{y}) = \frac{1+\sqrt{y}}{16} - \frac{1-\sqrt{y}}{16} = 0.5 \end{aligned}$$

da cui risulta che la mediana di  $Y$  è  $y=2$ , la quale non corrisponde a quella di  $X$  (pari ad 8, per simmetria) in ragione della trasformazione non biunivoca. Se invece, ad esempio, la trasformazione fosse stata  $Y=(X-1)^3$ , avremmo avuto per  $Y$  la mediana  $y=7^3$ .

**3.26** Applicando la definizione di media abbiamo:

$$E\{X\} = x_1 P_X(x_1) + \int_0^2 x f_X(x) dx = \frac{1}{3}$$

inoltre, essendo:

$$E\{X^2\} = x_1^2 P_X(x_1) + \int_0^2 x^2 f_X(x) dx = \frac{8}{18}$$

abbiamo:

$$\text{Var}\{X\} = E\{X^2\} - [E\{X\}]^2 = \frac{1}{3}.$$

È interessante notare che otteniamo lo stesso risultato se (in analogia al teorema di Huygens) sommiamo le varianze delle due masse parziali (4/6 e 2/6), rispetto alle loro rispettive medie (0 ed 1), alla varianza rispetto alla media globale (1/3) delle due masse concentrate nelle loro rispettive medie:

$$\begin{aligned} &4/6 \times 0 + \int_0^2 (x-1)^2 1/6 dx + (4/6) \times (1/3)^2 + (2/6) \times (1-1/3)^2 = \\ &= 1/9 + 2/27 + 4/27 = 1/3. \end{aligned}$$

**3.27** Applicando la definizione di Mgf abbiamo:

$$\Phi_X(t) = e^{tx_1} P_X(x_1) + \int_0^2 e^{tx} f_X(x) dx = \frac{4}{6} + \frac{1}{6} \left[ \frac{e^{tx}}{t} \right]_0^2 = \frac{1}{6} \left( 4 + \frac{e^{2t}-1}{t} \right)$$

da cui, ad esempio, possiamo calcolare la media:

$$E\{X\} = \Phi_X'(0) = \frac{1}{6} \left( \frac{2te^{2t} - e^{2t} + 1}{t^2} \right) \Big|_{t=0} = \lim_{x \rightarrow \infty} \frac{e^{2t} + 2te^{2t}}{3} = \frac{1}{3}.$$

**3.28** La funzione è non decrescente e soddisfa le due condizioni  $F_X(-\infty)=0$  ed  $F_X(\infty)=1$ . Essendo a gradini caratterizza una v.a. discreta. I tre “salti” individuano i suoi tre possibili valori (0, 2, 3) cui competono rispettivamente le seguenti probabili-

**556** Soluzioni dei problemi proposti

tà:

$$\begin{aligned}
 P_X(0) &= F_X(0^+) - F_X(0^-) = \frac{1}{2} - 0 = \frac{1}{2}; \\
 P_X(2) &= F_X(2^+) - F_X(2^-) = \frac{5}{6} - \frac{1}{2} = \frac{1}{3}; \\
 P_X(3) &= F_X(3^+) - F_X(3^-) = 1 - \frac{5}{6} = \frac{1}{6}.
 \end{aligned}$$

La Mgf è:

$$\Phi_X(t) = \frac{1}{2} + \frac{1}{3}e^{2t} + \frac{1}{6}e^{3t}.$$

**3.29** Il cercato modello di pdf deve essere tale che:

$$\int_0^4 f_X(x) dx = \int_0^4 ax dx = 8a = 1$$

quindi la formulazione della pdf è semplicemente il prodotto  $8x$ .

**3.30** Imponendo che sia:

$$\int_{-\infty}^1 f_X(x) dx = a [\ln(1+x)]_{-\infty}^1 = a \ln(2) = 1$$

abbiamo  $a = 1/\ln(2)$ .

**3.31** Derivando la Cdf abbiamo:

$$f_X(x) = \begin{cases} 0; & \text{per } x < 0 \\ a; & \text{per } 0 \leq x \leq b; \quad a > 0 \\ 0; & \text{per } x > b \end{cases}$$

Essendo per continuità  $F_X(b) = 1$ , deduciamo che deve essere  $ab = 1$ .

**3.32** La formulazione della pmf ed i valori che essa assume sono:

$$P_X(x) = \frac{\binom{10}{x} \binom{90}{5-x}}{\binom{100}{5}}; \quad x = 0, \dots, 5;$$

$$0.583; \quad 0.339; \quad 0.702 \times 10^{-1}; \quad 0.683 \times 10^{-2}; \quad 0.251 \times 10^{-3}; \quad 0.335 \times 10^{-5}.$$

Il valore atteso non solo è diverso da quelli che può assumere la v.a. ma è anche sorprendentemente basso:

$$E\{X\} = \sum_{x=0}^5 x P_X(x) = 0.50.$$

**3.33** Per essere costretti ad effettuare  $x$  estrazioni, senza rimessa, per trovare anche il decimo esemplare difettoso, vuol dire che si verifica la seguente concomitanza: nelle prime  $x-1$  estrazioni sono trovati 9 esemplari ed all' $x$ -esima si trova il decimo. Quindi la probabilità di  $x$  è:



$$P_X(x) = \frac{\binom{10}{9} \binom{90}{x-10}}{\binom{100}{x-1}} \times \frac{1}{100-(x-1)}; \quad x=10, 11, 12, \dots, 100.$$

**3.34** Se dobbiamo effettuare  $x$  lanci per ottenere un numero inferiore a 3, vuol dire che si verifica la seguente concomitanza: nei primi  $x-1$  lanci compaiono tutti numeri uguali o superiori a 3 ed all' $x$ -esimo il primo inferiore a 3. Quindi la probabilità di  $x$  è:

$$P_X(x) = \left(\frac{2}{3}\right)^{x-1} \times \frac{1}{3}; \quad x=1, 2, \dots, \infty.$$

**3.35** La varianza della somma delle tre v.a. è:

$$\begin{aligned} \text{Var}\{X_1 + X_2 + X_3\} &= \text{Var}\{X_1\} + \text{Var}\{X_2\} + \text{Var}\{X_3\} + 2\text{Cov}\{X_1, X_2\} + \\ &+ 2\text{Cov}\{X_1, X_3\} + 2\text{Cov}\{X_2, X_3\} \end{aligned}$$

utilizzando le (3.61) e (3.81) abbiamo:

$$\begin{aligned} \text{Var}\{X_1\} &= 0.01; \quad \text{Var}\{X_2\} = 0.08; \quad \text{Var}\{X_3\} = 0.13; \\ \text{Cov}\{X_1, X_2\} &= 5.66 \times 10^{-3}; \quad \text{Cov}\{X_1, X_3\} = 18.03 \times 10^{-3}; \quad \text{Cov}\{X_2, X_3\} = 30.59 \times 10^{-3} \end{aligned}$$

quindi il richiesto scarto tipo è:

$$\sqrt{\text{Var}\{X_1 + X_2 + X_3\}} = \sqrt{0.3286} = 0.5732.$$

**3.36** Il numero di pezzi difettosi che l'operatore troverà è la somma  $Y$  di  $n=100$  v.a.  $X_i$  che, assumendo i valori 1 e 0 rispettivamente con probabilità  $p=0.01$  e  $q=1-p=0.99$ , hanno tutte media  $E\{X_i\}=1 \times p + 0 \times q = p$ . Per la linearità dell'operatore speranza matematica abbiamo:

$$E\{Y\} = E\left\{\sum_{i=1}^{100} X_i\right\} = \sum_{i=1}^{100} E\{X_i\} = n p = 100 \times 0.01 = 1$$

come, del resto, era intuitivo prevedere.

**3.37** È sufficiente ripetere lo stesso ragionamento impiegato per risolvere il problema precedente ottenendo che il numero atteso di guasti è  $E\{Y\} = P_1 + P_2 + P_3$ .

Per controllare questo risultato, è sufficiente notare che un guasto si verifica se a guastarsi è l'uno o l'altro o l'altro ancora; due guasti si verificano se continua a funzionare solo l'uno o solo l'altro o solo l'altro ancora; tre guasti si verificano solo se si guastano tutti. Pertanto la pmf del numero di guasti è:

$$\begin{aligned} P_Y(1) &= P_1 + P_2 + P_3 \\ P_Y(2) &= (1 - P_1)P_2P_3 + P_1(1 - P_2)P_3 + P_1P_2(1 - P_3) \\ P_Y(3) &= P_1P_2P_3 \end{aligned}$$

da cui è facile controllare che:

**558** Soluzioni dei problemi proposti

$$E\{Y\} = \sum_{i=1}^3 i P_Y(i) = P_1 + P_2 + P_3.$$

**3.38** Una successione di  $x$  pezzi conformi seguita da un pezzo difettoso (che determina l'immediato arresto della linea) è prodotta con probabilità  $p^x(1-p)$ . Pertanto il numero atteso di pezzi conformi prodotti, tra due arresti consecutivi della linea, è (cfr. Appendice A):

$$E\{X\} = \sum_{x=0}^{\infty} x P_X(x) = \sum_{x=0}^{\infty} x p^x (1-p) = \frac{p}{1-p}.$$

**3.39** È necessario seguire un ragionamento analogo a quello che ci ha permesso di risolvere il problema precedente, facendo però attenzione che questa volta il numero massimo di ispezioni "inutili" è finito e pari ad  $n-1$ :

$$E\{X\} = \sum_{x=1}^{n-1} x P_X(x) = \sum_{x=1}^{n-1} x p^x (1-p) = p \frac{(1-p^{n-1})^2}{1-p}$$

e, quindi, la media cercata è pari alla somma parziale della serie già considerata nel problema precedente (cfr. Appendice A).

**3.40** Le v.a.  $X$  ed  $Y$  hanno stessa media e varianza:

$$E\{X\} = E\{Y\} = \frac{1}{a} \int_0^a x dx = \frac{a}{2};$$

$$\text{Var}\{X\} = \text{Var}\{Y\} = \frac{1}{a} \int_0^a x^2 dx - \left(\frac{a}{2}\right)^2 = \frac{a^2}{3} - \frac{a^2}{4} = \frac{a^2}{12}$$

da cui essendo  $E\{\bullet\}$  un operatore lineare:

$$E\{Z\} = E\{X - Y\} = E\{X\} - E\{Y\} = 0$$

tenendo conto che  $\text{Var}\{\bullet\}$  è un operatore non lineare e che l' $s$ -indipendenza di  $X$  ed  $Y$  implica  $\text{Cov}\{X, Y\} = 0$ :

$$\text{Var}\{Z\} = \text{Var}\{X\} + \text{Var}\{Y\} = \frac{a^2}{6}.$$

**3.41** Il modello è valido, infatti la funzione è positiva ed inoltre:

$$\int_0^1 \int_0^1 f_{XY}(x, y) dx dy = \frac{1}{e-1} \int_0^1 [e^x]_0^1 dy = \frac{1}{e-1} \int_0^1 (e-1) dy = 1.$$

Le pdf marginali sono:

$$f_X(x) = \frac{1}{e-1} \int_0^1 e^x dy = \frac{e^x}{e-1}; \quad f_Y(y) = \frac{1}{e-1} \int_0^1 e^x dx = 1$$

con medie:

$$E\{X\} = \frac{[e^x(x-1)]_0^1}{e-1} = \frac{1}{e-1}; \quad E\{Y\} = \left[ \frac{y^2}{2} \right]_0^1 = \frac{1}{2}.$$

Le due v.a. sono  $s$ -indipendenti perché vale l'equazione  $f_{XY}(x, y) = f_X(x) f_Y(y)$ .

**3.42** Per calcolare la pdf della v.a.  $Z = X - Y$ , possiamo scegliere come variabile di comodo  $V = X$ :

$$f_{ZV}(z, v) = f_{XY}[x(z, v), y(z, v)] |J| = \frac{e^v}{e-1}$$

dove  $|J|=1$ . La v.a.  $V$  risulta limitata inferiormente dal valore della  $Z$ , ossia dalla differenza aleatoria e che separa  $X$  da  $Y$ . Inoltre  $V$  è limitata superiormente dal limite fisico costituito dalla capienza unitaria del serbatoio. Quindi è  $Z \leq V \leq 1$ . Integrando la  $f_{ZV}(z, v)$  tra  $z$  ed 1 rispetto a  $v$ , otteniamo la richiesta pdf:

$$f_Z(z) = \frac{1}{e-1} \int_z^1 e^v dv = \frac{1}{e-1} [e^v]_z^1 = \frac{e - e^z}{e-1}$$

il cui integrale tra 0 ed 1, ovviamente, è pari all'unità.

Il valore atteso del volume d'acqua residua è:

$$E\{Z\} = \int_0^1 z f_Z(z) dz = \frac{1}{e-1} \left[ \frac{e z^2}{2} - e^z (z-1) \right]_0^1 = \frac{\frac{e}{2} - 1}{e-1} = 0.209.$$

**3.43** La  $\Pr\{X > 3\} = 1 - \Pr\{X \leq 3\} = 1 - 0.16 \times 3 = 0.52$ . La Cdf di  $X$  è:

$$F_X(x) = \begin{cases} 0.16x; & \text{per } 0 < x \leq 5 \\ 0.80 + 0.02x; & \text{per } 5 < x \leq 15 \end{cases}$$

da cui possiamo verificare che  $\Pr\{X > 3\} = 1 - F_X(3) = 0.52$ .

La media di  $X$  è:

$$E\{X\} = \int_0^5 0.16x dx + \int_5^{15} 0.02x dx = 4.$$

La mediana è di poco inferiore alla media ed è pari a 3.125, essendo  $F_X(3.125) = 0.5$ . La varianza è:

$$\text{Var}\{X\} = E\{X^2\} - [E\{X\}]^2 = \int_0^5 0.16x^2 dx + \int_5^{15} 0.02x^2 dx - 16 = 28.33 - 16 = 12.33$$

quindi lo scarto tipo è 3.51.

**3.44** La Mgf è:

$$\Phi_X(t) = E\{e^{tX}\} = \int_0^5 0.16e^{tx} dx + \int_5^{15} 0.02e^{tx} dx = \frac{1}{t} (0.14e^{5t} + 0.02e^{15t} - 0.16)$$

da cui, ad esempio, possiamo calcolare la media:

$$\begin{aligned} E\{X\} &= \Phi_X'(0) = \frac{1}{t^2} [0.14e^{5t}(5t-1) + 0.02e^{15t}(15t-1) - 0.16] \Big|_{t=0} = \\ &= \lim_{t \rightarrow 0} \left\{ \frac{e^{5t}}{2} [7 + 3.5(5t-1)] + \frac{e^{15t}}{2} [9 + 4.5(15t-1)] \right\} = 4 \end{aligned}$$

avendo applicato la regola dell'Hôpital.

**3.45** Per tale v.a. la Mgf non esiste perché l'integrale diverge. Infatti, integrando per parti abbiamo:

**560** Soluzioni dei problemi proposti

$$\int_1^{\infty} e^{tx} \frac{1}{x^2} dx = \left[ -\frac{e^{tx}}{x} \right]_1^{\infty} + \int_1^{\infty} \frac{e^{tx}}{x} dx$$

essendo quello a secondo membro un noto integrale non convergente nell'intervallo  $[1, \infty)$ .

**3.46** La Cf di  $X$  è:

$$\Psi_X(t) = E\{e^{itX}\} = \int_0^{\infty} e^{-(1-it)x} dx = \frac{1}{1-it}$$

ed è legata alla Mgf dall'identità  $\Psi_X(t) = \Phi_X(it)$ .

**3.47** La Mgf di  $X$  è:

$$\Phi_X(t) = E\{e^{tX}\} = \frac{e^t + e^{-t}}{2}.$$

**3.48** Utilizzando la soluzione del problema precedente, abbiamo:

$$E\{X\} = \Phi_X'(0) = \frac{1}{(1-t)^2} \Big|_{t=0} = 1$$

$$\text{Var}\{X\} = E\{X^2\} - [E\{X\}]^2 = \Phi_X''(0) - 1 = \frac{2}{(1-t)^3} \Big|_{t=0} - 1 = 1.$$

**3.49** Essendo la pdf continua in  $x=0$ , è nulla la probabilità che il livello di emissioni sia pari esattamente al valore limite. Il valore atteso del danno è:

$$\int_0^{30} 3x^3 \frac{(900-x^2)}{36000} dx = \frac{1}{12000} \left[ 900 \frac{x^4}{4} - x^5 \right]_0^{30} = 5062.50.$$

**3.50** Essendo le v.a.  $X$  ed  $Y$   $s$ -indipendenti, abbiamo:

$$\Phi_Z(t) = E\{e^{t(X+Y)}\} = \Phi_X(t) \Phi_Y(t)$$

dove le due Mgf a secondo membro sono rispettivamente:

$$\Phi_X(t) = \int_0^2 x e^{tx} dx = \frac{2e^{2t}}{t} + \frac{1-e^{2t}}{t^2}; \quad t > 0$$

$$\Phi_Y(t) = \int_0^{\pi} e^{ty} \frac{\sin(y)}{2} dy = \left[ e^{ty} \frac{t \sin(y) - \cos(y)}{2(1+t^2)} \right]_0^{\pi} = \frac{e^{\pi t} + 1}{2(1+t^2)}.$$

Per il calcolo del secondo integrale si noti che valgono entrambe le seguenti equazioni:

$$\int e^{ty} \frac{\sin(y)}{2} dy = -\frac{1}{2} \int e^{ty} d \cos(y) = -\frac{1}{2} [e^{ty} \cos(y)] + \frac{t}{2} \int \cos(y) e^{ty} dy$$

$$\int e^{ty} \frac{\sin(y)}{2} dy = \frac{1}{2t} \int \sin(y) d e^{ty} = \frac{1}{2t} [e^{ty} \sin(y)] - \frac{1}{2t} \int e^{ty} \cos(y) dy$$

da cui:

$$\int \cos(y) e^{ty} dy = \frac{e^{ty} \sin(y) + t e^{ty} \cos(y)}{1+t^2}$$

che sostituito in una delle due equazioni precedenti fornisce l'integrale indefinito utilizzato nel calcolo di  $\Phi_Y(t)$ .

## Capitolo 4

**4.9** Le ipotesi rendono possibile l'applicazione del modello Binomiale. Fissando  $p=0.01$  ed  $n=20$ , la Cdf binomiale ci fornisce:

$$\Pr\{Y \leq 2\} = \sum_{y=0}^2 P_Y(y) = 0.8179 + 0.1652 + 0.0159 = 0.9990$$

quindi, solo una volta ogni 1000 prelievi rileveremo più di 2 provini inquinati oltre il limite tollerabile.

**4.10** Dalla definizione di periodo di ritorno, deduciamo che l'evento si verifica ogni  $1/0.083=12.05$  anni. Allo stesso risultato perveniamo considerando la v.a.  $Y = X + 1$  essendo  $X$  la v.a. Geometrica (con  $p=1-0.083$ ) che conta il numero di anni che precedono quello in cui si verifica l'eccesso di pioggia. Per cui il periodo di ritorno dell'allagamento risulta essere come prima  $T = 1/(1-p) = 12.05$  anni.

**4.11** Possiamo adottare il modello di Poisson con  $\mu = 0.82$  ottenendo:

$$\sum_{y=4}^{\infty} P_Y(y) = 1 - \sum_{y=0}^3 P_Y(y) = 1 - (0.4404 + 0.3611 + 0.1481 + 0.0405) = 0.0099$$

cioè è pari a circa l'1% la probabilità che la soglia venga superata (tuttavia, se le conseguenze fossero gravi il rischio sarebbe non trascurabile).

**4.12** Adottiamo il modello di Poisson con  $\mu = 15/30 = 0.5$ :

$$\sum_{y=2}^{\infty} P_Y(y) = 1 - \sum_{y=0}^1 P_Y(y) = 1 - (0.6065 + 0.3033) = 1 - 0.9098 = 0.0902$$

quindi, la probabilità non è trascurabile, essendo pari a circa il 10%.

**4.13** Preliminarmente, notiamo che essere costretti a somministrare il farmaco ad esattamente 6 pazienti (uno dopo l'altro) per osservarne 2 che manifestano effetti collaterali, equivale ad accumulare 4 osservazioni esenti da effetti prima delle 2 che invece ne sono caratterizzate. Pertanto, possiamo applicare il modello Binomiale Negativo con  $p=0.70$  e  $m=2$ . La richiesta probabilità è quindi:

$$\sum_{y=5}^{\infty} P_Y(y) = 1 - \sum_{y=0}^4 P_Y(y) = 1 - (0.0900 + 0.1260 + 0.1323 + 0.1235 + 0.1080) = 0.4202.$$

**4.14** Applicando il modello Ipergeometrico con  $N=80$  e  $D=30$ , otteniamo che per

**562** Soluzioni dei problemi proposti

$n=10$  il numero medio di confezioni avariate riscontrate è  $n D/N=3.75$ .

**4.15** Applichiamo il modello di v.a. Binomiale,  $Y$ , con  $p=0.95$ . Dobbiamo individuare il valore di  $n$  per cui  $\Pr\{Y \geq 3\}=0.98$  risolvendo per tentativi l'equazione:

$$\sum_{y=3}^n \binom{n}{y} p^y (1-p)^{n-y} > 0.98.$$

Il numero di esemplari da portare è 4; infatti per  $n=3$  abbiamo  $\Pr\{Y \geq 3\}=0.857$  e per  $n=4$  abbiamo  $\Pr\{Y \geq 3\}=0.986$ .

Applicando il modello Binomiale Negativo con  $p=0.05$ ,  $y=1$  ed  $m=3$  abbiamo:

$$\begin{aligned} P_Y(1) &= \Pr\{\text{uno solo guasto tra i primi tre}\} \times \Pr\{\text{il quarto funziona}\} = \\ &= 0.14 \times 0.95 = 0.13. \end{aligned}$$

**4.16** Al collaudo degli esemplari prelevati dalla linea  $A$  è applicabile il modello Binomiale. A quello degli esemplari prelevati dal lotto  $B$  (senza rimessa, come è prassi nei collaudi) è applicabile il modello Ipergeometrico. I risultati sono raccolti nella tabella seguente e sono indicativi del livello di approssimazione che comporta l'applicazione del modello Binomiale in caso di estrazione senza rimessa di un campione di dimensione non trascurabile rispetto a quella del lotto.

	linea $A$	lotto $B$
	prob. di superare il collaudo = 0.80	frazione di conformi = 0.80
$\Pr\{0 n=10\}$	$1.024 \times 10^{-7} >$	$7.6491 \times 10^{-9}$
$\Pr\{2 n=10\}$	$7.37 \times 10^{-5} >$	$1.96 \times 10^{-5}$
$\Pr\{5 n=10\}$	$0.0264 >$	$0.0210$
$\Pr\{10 n=10\}$	$0.1074 >$	$0.0937$

**4.17** In media arrivano  $20/60=0.333$  auto al secondo. Applicando il modello di Poisson abbiamo:

$$\sum_{y=2}^{\infty} P_Y(y) = 1 - \sum_{y=0}^1 P_Y(y) = 1 - (0.7168 + 0.2387) = 1 - 0.9555 = 0.0445.$$

**4.18** Possiamo applicare il modello Binomiale con  $p=0.94$ , avendo fissato il valore della Cdf  $= 1 - 0.999 = 0.001$ . Risolvendo per tentativi la seguente equazione:

$$\sum_{y=0}^9 \binom{n}{y} p^y (1-p)^{n-y} = 0.001$$

otteniamo che deve essere  $n=14$ , infatti, per  $n=13, 14$  e  $15$  abbiamo rispettivamente che la Cdf = 0.0060, 0.0010 e 0.0001.

**4.19** La formulazione della Mgf è:

$$\Phi_X(t) = E\{e^{tY}\} = E\{e^{t(4X+3)}\} = 0.3e^{3t} + 0.7e^{7t}.$$

**4.20** La v.a.  $Z$  “numero complessivo di richieste” è definita come somma di due v.a.  $s$ -indipendenti:  $Z = X + Y$ , dove  $X$  rappresenta la v.a. “numero di richieste di tipo tecnico” e  $Y$  la v.a. “numero di richieste di tipo commerciale”. Sia  $X$  che  $Y$  si possono ritenere distribuite secondo un modello di Poisson con parametri, rispettivamente,  $\lambda_X = 10$  richieste di tipo tecnico/minuto e  $\lambda_Y = 5$  richieste di tipo commerciale/minuto. Per la riproducibilità della v.a. di Poisson,  $Z$  è una v.a. di Poisson di parametro  $\lambda_Z = \lambda_X + \lambda_Y = 15$  richieste/minuto, per cui abbiamo:

$$\lambda_Z = 15; \quad z = 100; \quad t = 5; \quad \Pr\{Z = 100\} = \frac{(15 \times 5)^{100}}{100!} \cdot e^{-15 \times 5} = 0.0009.$$

**4.21** È plausibile ritenere che l'estrazione avvenga senza rimessa per cui possiamo utilizzare il modello Ipergeometrico con  $N=100$ ,  $D=3$ ,  $n=2$  e  $k=1$ . La richiesta probabilità è:

$$\Pr\{K=1\} = \frac{\binom{100-3}{1} \binom{3}{1}}{\binom{100}{2}} = 0.059.$$

Poiché la frazione di campionamento (rapporto tra il numero di pezzi estratti,  $n=2$ , ed il numero di pezzi presenti nel lotto,  $N=100$ ) è minore del 10%, è possibile ritenere che l'estrazione avvenga con rimessa e, quindi, approssimare il modello ipergeometrico con quello binomiale con  $p = D/N = 3/100 = 0.03$  ed  $n=2$ . Abbiamo:

$$\Pr\{K=1\} = \binom{2}{1} 0.03^1 (1-0.03)^{2-1} = 0.058.$$

**4.22** La v.a.  $X$  “numero giornaliero di autocarri inattivi” può essere ritenuta di tipo binomiale di parametri  $n=50$  e  $p=2/50=0.04$ . Il rischio richiesto è:

$$\Pr(X > 5) = 1 - \Pr(X \leq 5) = 1 - \sum_{x=0}^5 \binom{50}{x} 0.04^x (1-0.04)^{50-x} = 1 - 0.986 = 0.014.$$

**4.23** Possiamo adottare il modello Binomiale con  $p=1/3$  e  $n=10$  ottenendo:

**564** Soluzioni dei problemi proposti

$$\begin{aligned} \Pr\{Y \geq 8\} &= \sum_{y=8}^{10} P_Y(y) = \sum_{y=8}^{10} \binom{10}{y} \left(\frac{1}{3}\right)^y \left(1 - \frac{1}{3}\right)^{10-y} = \\ &= 0.00305 + 0.00034 + 0.00002 = 0.0034. \end{aligned}$$

**4.24** La probabilità che i tre amici totalizzino due vincite in un mese si calcola agevolmente tenendo conto dell'ipotesi di  $s$ -indipendenza degli esiti delle  $3 \times 4 = 12$  giocate complessive:

$$\binom{12}{2} p^2 (1-p)^{10} = \binom{12}{2} (24.97 \times 10^{-4})^2 (1 - 24.97 \times 10^{-4})^{10} = 4.013 \times 10^{-4}$$

dove  $p$  è la probabilità di vincere un ambo giocando due numeri su di una specifica ruota calcolata nell'Esempio 1.20.

**4.25** Il numero  $X$  di candidati da esaminare *prima di trovare* il primo idoneo è una v.a. Geometrica con parametro  $p = 0.6$ . Il numero totale di candidati da esaminare *per trovare* il primo idoneo è  $X + 1$ . Pertanto, per  $x + 1 = 10$ , abbiamo che la probabilità richiesta è:

$$p^9 (1-p) = 0.6^9 \times 0.4 = 4.03 \times 10^{-3}.$$

**4.26** Il numero di inserimenti a vuoto,  $Y$ , richiesto per raggiungere il credito necessario ad ottenere il caffè è una v.a. Binomiale Negativa con parametri  $p = 0.1$  ed  $m = 3$ . Pertanto la probabilità richiesta è:

$$P_Y(2) = \binom{4}{2} p^2 (1-p)^3 = 6 \times 0.1^2 \times 0.9^3 = 0.044.$$

**4.27** Il numero di allievi Gestionali che capitano negli 8 posti della prima fila è una v.a. Ipergeometrica  $K$  di parametri  $N = 25$ ,  $D = 13$  ed  $n = 8$ . Pertanto la probabilità richiesta è:

$$P_K(5) = \frac{\binom{13}{5} \binom{12}{3}}{\binom{25}{8}} = \frac{13!}{5!(13-5)!} \times \frac{12!}{3!(12-3)!} \times \frac{8!(25-8)!}{25!} = 0.262.$$

**4.28** Dalle sole informazioni in nostro possesso possiamo valutare la probabilità del manifestarsi della congestione in accordo ad una v.a. di Poisson con:

$$\lambda = 1/21 \text{ giorni}^{-1}$$

da cui otteniamo che la probabilità di congestione in una settimana è:

$$1 - \Pr\{\text{nessuna congestione in 7 giorni}\} = 1 - e^{-7\lambda} = 1 - e^{-1/3} = 0.284.$$

**4.29** Essendo  $\text{Cov}\{X_1, X_2\} = 0$ , abbiamo:

$$E[Z] = E[X_1] + E[X_2] = n_1 p + n_2 p = np = 3$$

$$\text{Var}[Z] = \text{Var}[X_1] + \text{Var}[X_2] = n_1 pq + n_2 pq = npq = 2.7.$$



Grazie all'ipotesi di  $s$ -indipendenza ed all'eguaglianza  $p_1 = p_2$ , possiamo affermare che sono soddisfatte le condizioni di riproducibilità della Binomiale. Pertanto la v.a.  $Z$  è binomiale con parametri  $p = p_1 = p_2$  ed  $n = n_1 + n_2$ .

**4.30** Adottando il modello di Poisson con  $\lambda = 2$  punti/set, per  $x = 5$  set, si ottiene:

$$\Pr\{Y \leq 5\} = \sum_{y=0}^5 \frac{(\lambda x)^y}{y!} e^{-\lambda x} = 0.0671;$$

$$\Pr\{Y \geq 7\} = 1 - \Pr\{Y \leq 5\} - \Pr\{Y = 6\} = 1 - 0.0671 - 0.0631 = 0.8698.$$

## Capitolo 5

**5.9** La richiesta pdf insiste sul segmento  $(0, 2)$  e, quindi, dovendo essere simmetrica e di area unitaria, deve essere un triangolo isoscele con vertice nel punto  $(1, 1)$ . Pertanto la sua pdf è costituita dai due segmenti seguenti:

$$f_Z(z) = \begin{cases} z & \text{per } 0 \leq z \leq 1 \\ 2 - z & \text{per } 1 < z \leq 2 \end{cases}$$

**5.10** Analogamente al problema precedente, la richiesta pdf insiste su un segmento di lunghezza unitaria (che parte dall'origine degli assi) e, quindi, deve essere un triangolo isoscele con vertice nel punto  $(1/2, 2)$  affinché la sua area sia unitaria. Pertanto la sua pdf è costituita dai due segmenti seguenti:

$$f(z) = \begin{cases} 4z & \text{per } 0 \leq z \leq 1/2; \\ 4 - 4z & \text{per } 1/2 < z \leq 1. \end{cases}$$

**5.11** Per la dimostrazione è sufficiente formulare la Mgf della v.a. Triangolare definita nel problema precedente e verificare che essa coincide con quella della v.a.  $Z$ , media di due v.a. Uniformi ridotte, già ricavata prima (cfr. Problema **Error! Reference source not found.**):

$$X = U^2; \quad U = X^{\frac{1}{2}}; \quad \left| \frac{du}{dx} \right| = \frac{1}{2} x^{\frac{1}{2}-1}$$

$$f_X(x) = [f_U(u) + f_U(-u)] \left| \frac{du}{dx} \right| = \frac{x^{\frac{1}{2}-1}}{\sqrt{2\pi}} e^{-\frac{x}{2}}; \quad 0 < x < \infty$$

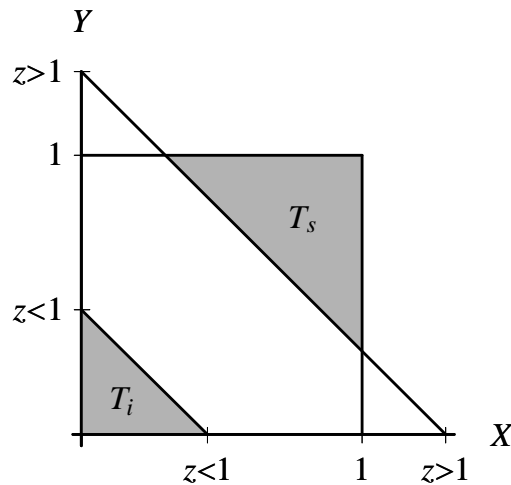
e traslando di  $1/2$  verso sinistra la funzione integranda del secondo integrale:

$$\int_0^{1/2} e^{tz} 4z dz + \int_0^{1/2} e^{tz+t/2} (2-4z) dz = \int_0^{1/2} e^{tz} [4z + e^{t/2} (2-4z)] dz =$$

$$= \frac{1}{t} \int_0^{1/2} [4z + e^{t/2} (2-4z)] de^{tz} = \frac{4}{t^2} (e^{t/2} - 1)^2.$$

**5.12** La v.a. bidimensionale  $(X, Y)$  è distribuita uniformemente sul quadrato  $Q$  di lato unitario.

## 566 Soluzioni dei problemi proposti



La retta  $X + Y = z$  interseca gli assi nei punti  $(0, z)$  e  $(z, 0)$ , essendo  $z$  uno specifico valore di  $Z$ . I punti  $(X, Y)$  di  $Q$ , sottostanti questa retta, rispettano la disuguaglianza  $X + Y < z$ . Se  $z \leq 1$ , l'intersezione con gli assi cade all'interno del lato di  $Q$  e detti punti appartengono al triangolo rettangolo  $T_i$  incluso tra la retta e gli assi. Il rapporto tra le aree di  $T_i$  e  $Q$  fornisce la probabilità di detti punti:

$$F_Z(z) = \Pr\{X + Y \leq z\} = \frac{\text{area di } T_i}{\text{area di } Q} = \frac{z^2}{2}; \quad \text{se } z \leq 1.$$

Se invece è  $z > 1$ , l'intersezione con gli assi cade all'esterno del lato di  $Q$  ed i punti, che rispettano la disuguaglianza  $X + Y < z$ , sono tutti quelli che appartengono a  $Q$  meno quelli che appartengono al triangolo rettangolo  $T_s$  incluso tra la retta e l'angolo del quadrato di coordinate  $(1, 1)$ . Il rapporto tra le aree di  $Q - T_s$  e  $Q$  fornisce la probabilità di questo secondo insieme di punti:

$$F_Z(z) = \Pr\{X + Y \leq z\} = \frac{\text{area di } Q - \text{area di } T_s}{\text{area di } Q} = 1 - \frac{(2-z)^2}{2}; \quad \text{se } z > 1.$$

La Cdf cercata è l'unione di queste due formulazioni e la sua derivata coincide con la pdf di cui al Problema (5.9).

**5.13** Analogamente alla media di due v.a. Uniformi ridotte (cfr. Problema **Error! Reference source not found.**), la media  $Z$  di  $n$  v.a. Uniformi ridotte  $s$ -indipendenti,  $Y_i$ , ha la seguente Mgf:

$$\Phi_Z(t) = \left( \frac{e^{t/n} - 1}{t/n} \right)^n; \quad Z = \frac{\sum_{i=1}^n Y_i}{n}.$$

Lo sviluppo in serie di Mac Laurin di questa Mgf fornisce:

$$\begin{aligned} \left(\frac{e^{t/n} - 1}{t/n}\right)^n &= \left[\frac{n}{t} \left(1 + t/n + \frac{(t/n)^2}{2} + \frac{(t/n)^3}{3!} + \dots - 1\right)\right]^n = \\ &= \left[1 + \frac{1}{n} \left(t/2 + \frac{t^2/n}{3!} + \dots\right)\right]^n = \left[1 + \frac{\psi(n)}{n}\right]^n \end{aligned}$$

da cui, infine, otteniamo:

$$\lim_{n \rightarrow \infty} \left[1 + \frac{\psi(n)}{n}\right]^n = e^{\lim_{n \rightarrow \infty} \psi(n)} = e^{t/2}$$

che costituisce la Mgf di una v.a. ad un sol valore, 1/2 (con probabilità unitaria). Vale la pena notare che, in accordo col teorema del limite centrale, essa coincide pure con la Mgf di una v.a. Gaussiana (con media 1/2 e varianza nulla).

**5.14** È sufficiente esprimere in funzione del modello Esponenziale la probabilità condizionata assegnata:

$$\Pr\{x_1 < X \leq x_1 + dx | X > x_1\} = \frac{f(x_1) dx}{1 - F(x_1)} = \frac{\lambda e^{-\lambda x_1} dx}{e^{-\lambda x_1}} = \lambda dx.$$

**5.15** La richiesta vita media è:

$$\begin{aligned} \int_0^\infty x f_X(x) dx &= 2 \int_0^\infty \lambda x^2 e^{-\lambda x^2} dx = 2 \int_0^\infty y e^{-y} 1/(2\lambda)(y/\lambda)^{-1/2} dy \\ &= \frac{1}{\lambda^{1/2}} \int_0^\infty y^{3/2-1} e^{-y} dy = \frac{1}{\lambda^{1/2}} \Gamma(3/2) = \frac{1}{2} (\pi/\lambda)^{1/2} \end{aligned}$$

avendo utilizzato la trasformazione  $y = \lambda x^2$ .

**5.16** Supponendo che la sostituzione del componente critico avvenga in un tempo trascurabile, il numero di guasti segue una legge di probabilità di Poisson con  $\lambda = 1.3 \times 10^{-3}$ . Se  $Y$  è il numero di guasti ed  $n$  è il numero di ricambi, l'avaria coincide con l'evento  $\{Y > n\}$ . Quindi:

$$\Pr\{\text{avaria}\} = \Pr\{Y > n\} = 1 - \sum_{y=0}^n \frac{(\lambda x)^y}{y!} e^{-\lambda x} < 1 \times 10^{-6}.$$

Risolvendo per tentativi questa disequazione troviamo che solo per  $n = 2$  abbiamo un valore inferiore a  $1 \times 10^{-6}$  e pari a  $1.25 \times 10^{-7}$  (infatti per  $n = 0$  abbiamo  $9.06 \times 10^{-3}$  e per  $n = 1$  abbiamo  $4.12 \times 10^{-5}$ ).

**5.17** La pdf Weibull è  $f_W(w) = (\beta/\alpha)(w/\alpha)^{\beta-1} e^{-(w/\alpha)^\beta}$  da cui:

**568** Soluzioni dei problemi proposti

$$\begin{aligned}
 E\{W\} &= \beta \int_0^{\infty} (w/\alpha)^{\beta} e^{-(w/\alpha)^{\beta}} dw = \alpha \int_0^{\infty} z^{\left(\frac{1}{\beta}+1\right)^{-1}} e^{-z} dz = \\
 &= \alpha \Gamma\left(\frac{1}{\beta}+1\right); \quad z = (w/\alpha)^{\beta}; \quad w = \alpha z^{\frac{1}{\beta}}; \quad \frac{dw}{dz} = \frac{\alpha}{\beta} z^{\left(\frac{1}{\beta}-1\right)}.
 \end{aligned}$$

**5.18** Dalla Cdf relativa al gruppo  $A$ , abbiamo che la probabilità di mancato allarme è circa del 60%:

$$\Pr\{X < 2.80\} = \Pr\left\{\frac{X - \mu_A}{\sigma_A} < \frac{2.80 - 2.65}{0.59}\right\} = \Pr\{U < 0.25\} = 0.60.$$

Dalla Cdf relativa al gruppo  $B$ , abbiamo che la probabilità di falso allarme è circa del 20%:

$$\Pr\{X_B > 2.80\} = \Pr\left\{\frac{X_B - \mu_B}{\sigma_B} > \frac{2.80 - 2.39}{0.48}\right\} = \Pr\{U > 0.85\} = 0.20.$$

**5.19** La probabilità che il livello ematico di acido urico sia superiore a 7 mg/dl è:

$$\Pr\{X > 7\} = \Pr\left\{\frac{X - \mu_X}{\sigma_X} > \frac{7 - 5.4}{1}\right\} = \Pr\{U > 1.6\} = 0.0548$$

il cui inverso fornisce il “periodo di ritorno”  $T = 18.25$ , ossia, in media, ogni circa 18 soggetti controllati ne troviamo uno con livello ematico di acido urico superiore a 7 mg/dl.

**5.20** Condizioni sfavorevoli significa rapporto  $X/Y < 1$  che, a sua volta, implica che sia  $X - Y < 0$ . La v.a.  $Z = X - Y$  è anch'essa Normale con media  $\mu_Z = \mu_X - \mu_Y$  e scarto tipo  $\sigma_Z = \sqrt{\sigma_X^2 + \sigma_Y^2}$ , per cui:

$$\Pr\{Z < 0\} = \Pr\left\{\frac{Z - \mu_Z}{\sigma_Z} < \frac{0 - 4.6}{2.72}\right\} = \Pr\{U < -1.69\} = 0.045$$

quindi, in circa il 5% dei casi il dosaggio impiegato non risulta efficace.

**5.21** Se le prime 19 o 18 o 17 ... hanno peso complessivo  $\geq 130$  g l'evento  $\{n < 20\}$  si verifica. Conviene quindi calcolare la probabilità dell'evento complementare  $\{n \geq 20\}$  che si verifica solo quando le prime 20 hanno peso complessivo  $\leq 130$  g. Questo secondo evento si verifica il 50% delle volte, visto che il peso di 20 caramelle è una v.a. normale proprio di media  $\mu = 20 \times \mu_X = 20 \times 6.5 = 130$ .

**5.22** Imponendo che sia:

$$\int_0^1 \lambda e^{\lambda x} dx = \left[ e^{\lambda x} \right]_0^1 = e^{\lambda} - 1 = 1$$

abbiamo che per  $\lambda = \ln(2)$  la funzione può essere assunta come modello di pdf nell'intervallo  $[0, 1]$ . La Mgf di questo modello è:

$$\Phi_X(t) = E\{e^{tX}\} = \lambda \frac{e^{\lambda+t} - 1}{\lambda+t}$$

da cui, ad esempio, calcoliamo la media:

$$E\{X\} = \Phi'_X(0) = \lambda \left. \frac{e^{\lambda+t}(\lambda+t) - e^{\lambda+t} + 1}{(\lambda+t)^2} \right|_{t=0} = \frac{\lambda e^\lambda - e^\lambda + 1}{\lambda} = 0.557.$$

**5.23** Dai dati in nostro possesso riteniamo plausibile l'impiego del modello di Poisson per la pmf delle v.a.  $Y_A$  "numero di pezzi di tipo A" ed  $Y_B$  "numero di pezzi di tipo B" con, rispettivamente,  $\lambda_A = 6/15 = 0.4$  pezzi di tipo A/minuto e  $\lambda_B = 3/15 = 0.2$  pezzi di tipo B/minuto. Se riteniamo  $s$ -indipendenti le v.a.  $Y_A$  ed  $Y_B$ , la riproducibilità del modello di Poisson ci garantisce che la v.a.  $Y = Y_A + Y_B$  "numero di pezzi" è una v.a. di Poisson con  $\lambda = \lambda_A + \lambda_B = 0.4 + 0.2 = 0.6$  pezzi/minuto. La v.a.  $T$  "tempo intercorrente" tra gli eventi contati dalla v.a. di Poisson è di tipo Esponenziale, per cui risulta:

$$\Pr\{T \leq 3\} = 1 - e^{-0.6 \times 3} = 0.835.$$

**5.24** La v.a.  $Y$  "superficie complessiva ricoperta da  $n$  fogli di alluminio" è ottenuta come somma di  $n$  v.a. Gaussiane:  $Y = \sum_{i=1}^n X_i$ , essendo  $X_i$  la v.a. "superficie ricoperta dall' $i$ -esimo foglio di alluminio". La riproducibilità del modello Normale ci garantisce che  $Y$  è anch'essa una v.a. Normale di media  $\mu_Y = \sum_{i=1}^n \mu_{X_i} = 2n$  m<sup>2</sup> e scarto tipo  $\sigma_Y = \sqrt{\sum_{i=1}^n \sigma_i^2} = 0.1\sqrt{n}$  m<sup>2</sup>. Pertanto, il numero minimo di fogli di alluminio  $n$  è quello tale che:

$$\Pr\{Y > 50\} = \Pr\left\{\frac{Y - \mu_Y}{\sigma_Y} > \frac{50 - 2n}{0.1\sqrt{n}}\right\} = \Pr\left(U > \frac{50 - 2n}{0.1\sqrt{n}}\right) > 0.99.$$

Utilizzando le tabelle della Normale standard, si deduce che il primo percentile  $u_{0.01}$  è uguale a  $-2.33$ ; per cui il valore minimo di  $n$  è quello che soddisfa la seguente disuguaglianza:

$$\frac{50 - 2n}{0.1\sqrt{n}} < -2.33.$$

Il numero di fogli è 26; infatti per  $n = 25$  abbiamo  $\Pr\{Y > 50\} = 0.50$  e per  $n = 26$  abbiamo  $\Pr\{Y > 50\} = 0.99996$ .

**5.25** Per la riproducibilità della v.a. Normale, la v.a.  $Z = R - S$  è anch'essa una v.a. Normale con media  $\mu_Z = \mu_R - \mu_S = 637 - 450 = 187$  e scarto tipo  $\sigma_Z = \sqrt{\sigma_R^2 + \sigma_S^2} = \sqrt{30^2 + 90^2} = 94.87$ . Poiché:

$$\Pr\{R > S\} = \Pr\{Z = R - S > 0\}$$

**570** Soluzioni dei problemi proposti

la richiesta probabilità è:

$$\Pr\{Z > 0\} = 1 - \Pr\{Z \leq 0\} = 1 - \Pr\left\{\frac{Z - \mu_Z}{\sigma_Z} \leq \frac{0 - 187}{94.87}\right\} = 1 - \Pr\{U \leq -1.971\} = 0.98.$$

**5.26** Ritenendo plausibile l'adozione del modello di Poisson, abbiamo:

$$\lambda = 0.5; \quad x = 50; \quad \Pr\{Y \leq 20\} = \sum_{y=0}^{20} \frac{(0.5 \times 50)^y}{y!} e^{-0.5 \times 50} = 0.185.$$

Il calcolo di questa sommatoria risulta alquanto laborioso. Diventa più semplice se approssimiamo il modello di Poisson di parametro  $\mu_Y = 0.5 \times 50 = 25$  con un modello Normale di parametri:  $\mu = \mu_Y = 25$  e  $\sigma = \sigma_Y = \sqrt{\mu_Y} = \sqrt{25} = 5$ . Pertanto, una buona approssimazione della probabilità richiesta è:

$$\Pr\{Y \leq 20\} = \Pr\left\{\frac{Y - \mu}{\sigma} \leq \frac{20 + 0.5 - 25}{5}\right\} = \Pr\{U \leq -0.9\} = 0.184.$$

L'addizione del valore 0.5 migliora l'approssimazione, poiché la densità "gaussiana" che si estende da  $20 - 0.5$  a  $20 + 0.5$  viene assegnata al valore "poissoniano" 20. Analogamente, qualora avessimo voluto calcolare  $\Pr\{Y \geq 20\}$  avremmo dovuto sottrarre il valore 0.5. L'approssimazione gaussiana adottata è valida quanto più la pmf della v.a. di Poisson ha forma simmetrica. In pratica ciò accade quando  $\mu > 10$ .

**5.27** Calcoliamo innanzitutto la probabilità che un prodotto sia conforme:

$$\begin{aligned} \Pr\{119.7 < X \leq 120.3\} &= \Pr\left\{\frac{119.7 - 120.0}{0.2} < \frac{X - \mu_X}{\sigma_X} \leq \frac{120.3 - 120.0}{0.2}\right\} = \\ &= \Pr\{-1.5 < U \leq 1.5\} = F_U(1.5) - F_U(-1.5) = 0.9332 - 0.0668 = 0.8664. \end{aligned}$$

Applicando il modello di v.a. Binomiale,  $Y$ , con  $n = 100$  e  $p = 0.8664$ , abbiamo:

$$\Pr\{Y \geq 75\} = \sum_{y=75}^{100} P_Y(y) = \sum_{y=75}^{100} \binom{100}{y} 0.8664^y (1 - 0.8664)^{100-y} = 0.9995.$$

Il calcolo di questa sommatoria risulta alquanto laborioso. Diventa più semplice se approssimiamo il modello Binomiale di parametri  $n = 100$  e  $p = 0.8664$  con un modello Normale di parametri:  $\mu = \mu_Y = np = 86.64$  e  $\sigma = \sigma_Y = \sqrt{np(1-p)} = 3.40$ .

Una buona approssimazione della probabilità richiesta è:

$$\Pr\{Y \geq 75\} = \Pr\left\{\frac{Y - \mu}{\sigma} \geq \frac{75 - 0.5 - 86.64}{3.40}\right\} = \Pr\{U \geq -3.57\} = 0.9998.$$

La sottrazione del valore 0.5 migliora l'approssimazione. Qualora avessimo voluto calcolare  $\Pr\{Y \leq 75\}$  avremmo dovuto, invece, aggiungere il valore 0.5.

L'approssimazione gaussiana adottata è tanto più valida quanto più la pmf della v.a. Binomiale ha forma simmetrica. In pratica ciò accade quando  $np > 10$  e  $n(1-p) > 10$ .

**5.28** Sulla base delle informazioni in nostro possesso è ragionevole assumere un mo-

dello Gumbel (dei massimi) per la v.a.  $X$ , assegnando ai parametri di posizione e scala,  $a$  e  $b$ , valori tali da conferire alla distribuzione la media e la deviazione standard fornite (cfr. Appendice A):

$$E\{X\} = a - b\Gamma'(1) = a + 0.57721b = 770$$

$$\text{Var}\{X\} = b^2\pi^2/6 = 370^2$$

da cui si ricavano i valori  $a = 603.5$  e  $b = 288.5$ . Quindi:

$$\Pr\{X > 1000\} = 1 - F_X(1000) = 1 - e^{-e^{-\frac{1000-603.5}{288.5}}} \cong 0.22; \quad T(1000) \cong \frac{1}{0.22} \cong 4.5 \text{ anni.}$$

**5.29** Indicando con  $T$  la durata complessiva e con  $T_j$  quella della generica batteria, abbiamo:

$$\mu_T = E\{T_1 + T_2 + T_3\} = 3 \times 1.5 = 4.5 \text{ h}$$

$$\sigma_T = \sqrt{\text{Var}\{T\}} = \sqrt{\text{Var}\{T_1 + T_2 + T_3\}} = \sqrt{3} \cdot 0.15 = 0.26 \text{ h.}$$

Generalizziamo ad un  $n$  qualsiasi il risultato precedente e consideriamo il quantile  $u_{0.90}$  della Normale Standard. Otteniamo il valore corrispondente al numero di batterie cercato maggiorando le soluzioni della seguente equazione di 2° grado:

$$\frac{5 - n\mu}{\sqrt{n}\sigma} = u_{0.90}; \quad (5 - n\mu)^2 - u_{0.90}^2 n\sigma^2 = 0$$

la cui soluzione è  $n = 3.33$ ; quindi, assumiamo pari ad 4 il numero richiesto.

**5.30** Per la v.a.  $X$  Normale di parametri  $\mu = 50$  cm e  $\sigma = 6$  cm, è immediato calcolare:

$$\begin{aligned} \Pr\{49 < X \leq 54\} &= F_X(54) - F_X(49) = \\ &= F_U\left(\frac{54-50}{6}\right) - F_U\left(\frac{40-50}{6}\right) = (1 - 0.2525) - 0.4338 \cong 0.31. \end{aligned}$$

L'intervallo, simmetrico rispetto alla media  $\mu$ , si ottiene imponendo:

$$\begin{aligned} \Pr\{\mu - \delta < X \leq \mu + \delta\} &= F_X(\mu + \delta) - F_X(\mu - \delta) = \\ &= F_U\left(\frac{\delta}{\sigma}\right) - F_U\left(\frac{-\delta}{\sigma}\right) = 2F_U\left(\frac{\delta}{\sigma}\right) - 1 = 0.25 \end{aligned}$$

da cui, consultando la tabella della coda destra della Gaussiana standard, si ricava:

$$\frac{\delta}{\sigma} = u_{0.375} \cong 0.32; \quad \delta = 0.32\sigma = 1.92$$

e quindi l'intervallo di misure ricercato è  $50 \pm 1.92$ , compreso tra circa 48 e 52 cm.

**Capitolo 6**

**6.9** Non avremmo nessun motivo per giudicare anomala la successione di tutte croci (o tutte teste), per il semplice motivo che ogni successione ha la stessa probabilità  $(1/2^6)$  di verificarsi con lanci *ritenuti* corretti di una moneta *ritenuta* perfetta. L'istintiva sensazione di sorpresa che proveremmo è dovuta semplicemente al fatto che tra tutte le  $2^6$  possibili successioni ce ne sono:

una sola	con 6 croci
sei	con 1 testa
quindici	con 2 teste
venti	con 3 teste
quindici	con 4 teste
sei	con 5 teste
una sola	con 6 teste

Per cui, è naturale essere abituati ad osservare uno dei 63 risultati (con almeno 1 testa) tra i 64 possibili. Tale abitudine c'induce pure a confondere casuale con disordinato: casuale significa solo che è stato scelto a caso tra tutte le alternative possibili, e tra queste può pure essere estratta una molto "ordinata", come la suddetta successione di 6 croci.

**6.10** Le frequenze competenti a ciascun intervallo d'età sono:

$$0.3030, 0.4545, 0.2424$$

l'età media e quella mediana sono rispettivamente:

$$54.389, 54.334.$$

**6.11** Essendo  $\mu = 6$  e  $\sigma = 2$ , dalla disuguaglianza di Chebyshev abbiamo:

$$\Pr\{|X - 6| \geq 2k\} \leq \frac{1}{k^2}; \quad 2k = 3$$

per cui il richiesto limite inferiore di probabilità è:

$$\Pr\{|X - 6| < 3\} > 1 - (2/3)^2 = 0.555.$$

**6.12** Media e varianza, calcolati utilizzando la funzione di probabilità empirica, sono:

$$\bar{x} = \sum_{i=1}^6 x_i P_i = 7.583; \quad s^2 = \sum_{i=1}^6 (x_i - \bar{x})^2 P_i = 2.576.$$

Media e varianza, calcolati dal campione di dati, sono:

$$\bar{x} = \frac{1}{12} \sum_{j=1}^{12} x_j = 7.583; \quad s^2 = \frac{1}{11} \sum_{j=1}^{12} (x_j - \bar{x})^2 = 2.811.$$

**6.13** Ordinati in senso crescente i due campioni di determinazioni:

	1	2	3	4	5	6	7	8	9
NA	16.4	27.1	35.6	49.6	66.0	71.7	72.1	72.8	94.2
MI	58.1	62.1	71.6	72.8	80.6	82.0	89.0	101.2	

nel primo caso, essendo dispari il numero di determinazioni, la mediana è subito indi-



viduata nella 5° determinazione, 66.0 mm. Nel secondo caso, il numero di determinazioni è pari e quindi la mediana cade tra la 4° e la 5° determinazione ed è uguale a  $(72.8+80.6)/2=76.7$ .

**6.14** Dalla diseuguaglianza di Chebyshev e dalle tabelle della distribuzione Normale — per  $k = 2, 2.5, 4$  — si ottengono rispettivamente i seguenti valori delle code:

$k$	2	2.5	3
Chebyshev	0.250	0.160	0.111
Normale	0.046	0.012	0.003

**6.15** Associamo alla generica ripetizione  $i$ -esima la v.a. bernoulliana  $X_i$  che assume i valori 1 e 0 a seconda che l'evento  $A$  si verifichi o meno. La media delle  $X_i$  è proprio la frequenza di  $A$  e converge alla media delle medie delle  $X_i$ ,  $\sum_{i=1}^n p/n = p$ , perché la varianza della somma delle  $X_i$  è  $\text{Var}\left\{\sum_{i=1}^n X_i\right\} = \sum_{i=1}^n \text{Var}\{X_i\} = n p q$  e, quindi, rapportata ad  $n^2$  soddisfa la condizione richiesta dalla legge dei grandi numeri.

**6.16** La misura della grandezza  $X$  è data dalla media aritmetica di tutte le sue valutazioni disponibili. I prodotti  $\bar{X}_n \times n$  e  $\bar{X}_m \times m$  ci forniscono rispettivamente la somma delle  $n$  ed  $m$  valutazioni disponibili; pertanto la misura di  $X$  è:

$$(\bar{X}_n \times n + \bar{X}_m \times m)/(n + m)$$

che equivale alla media pesata delle due misure disponibili.

**6.17** Le misure  $\bar{X}_n$  ed  $\bar{X}_m$  sono affette rispettivamente dagli errori medi  $S_n = S/\sqrt{n}$  ed  $S_m = S/\sqrt{m}$  (essendo  $S$  l'errore medio delle  $n + m$  valutazioni) da cui risulta:

$$n = S^2/S_n^2; \quad m = S^2/S_m^2.$$

Pertanto, i pesi attribuiti alle misure  $\bar{X}_n$  ed  $\bar{X}_m$ , nel problema precedente, sono inversamente proporzionali alle stime delle loro rispettive varianze.

**6.18** Imponendo la condizione richiesta ed utilizzando il valore 1.28 del 90° percentile della Cdf Normale si ottiene:

$$\Pr\{\bar{x} \leq +5\} = \Pr\left\{\frac{\bar{X}}{\sigma/\sqrt{n}} \leq \frac{+5}{20/\sqrt{n}}\right\} = 0.90; \quad \frac{+5}{20/\sqrt{n}} = 1.28; \quad n = \left(\frac{1.28 \times 20}{5}\right)^2 \cong 26.$$

**6.19** Trasformando la limitazione imposta all'errore assoluto di  $S$  in una corrispondente limitazione alla v.a. Chi-Quadrato, si ottiene

$$\begin{aligned} & \Pr\{-0.25 \times S < S - \sigma < 0.25 \times S\} = \Pr\{0.75 \times S < \sigma < 1.25 \times S\} = \\ & = \Pr\left\{10 \times (0.75)^2 < (n-1) \frac{\sigma^2}{S^2} < 10 \times (1.25)^2\right\} = \\ & = \Pr\{5.62 < \chi^2 < 15.62\} \cong 0.9 - 0.1 = 0.8; \quad \nu = 10. \end{aligned}$$

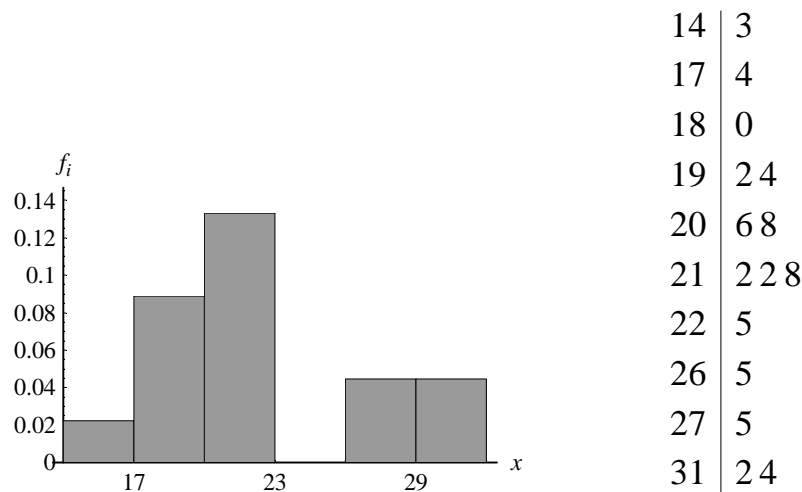
**582** Soluzioni dei problemi proposti

**6.20** Media, mediana, varianza (corretta), deviazione standard, quartili e range sono rispettivamente:

$$\begin{aligned} \bar{x} &= \sum_{i=1}^n x_i / n = 22.2, & \hat{x}_{0.5} &= 21.2, & s^2 &= \sum_{i=1}^n (x_i - \bar{x})^2 / (n-1) = 24.2, \\ s &= 4.92, & \hat{x}_{0.25} &= 19.3, & \hat{x}_{0.75} &= 24.5, \\ \text{Max}(x_1, \dots, x_n) - \text{Min}(x_1, \dots, x_n) &= 31.4 - 14.3 = 17.1. \end{aligned}$$

**6.21** Arrotondando il valore  $m = 1 + 3.3 \log_{10}(n) = 4.88$  abbiamo  $m = 5$  da cui, arrotondando  $[\text{Max}(x_1, \dots, x_n) - \text{Min}(x_1, \dots, x_n)] / m = 3.42$ , segue un'ampiezza dell'intervallo  $\Delta x = 3$ . Arrotondando per difetto  $\text{Min}(x_1, \dots, x_n) = 14.3$  abbiamo che l'estremo inferiore del primo intervallo è uguale a 14. Arrotondando per eccesso  $\text{Max}(x_1, \dots, x_n) = 31.4 \cong 32$  ed essendo intero il rapporto  $(32 - 14) / 3 = 6$  fissiamo al valore 32 l'estremo superiore dell'ultimo intervallo. Ai 6 intervalli così individuati corrispondono le frequenze assolute  $n_i = \{1, 4, 6, 0, 2, 2\}$  e le densità di frequenza  $f_i = n_i / (n \Delta x) = \{1/45, 4/45, 6/45, 0, 2/45, 2/45\}$ .

**6.22** Sulla base dei calcoli già effettuati, l'istogramma ed il diagramma *Stem-and-Leaf*, del campione della v.a.  $X$  esaminato nei due problemi precedenti, sono rispettivamente:



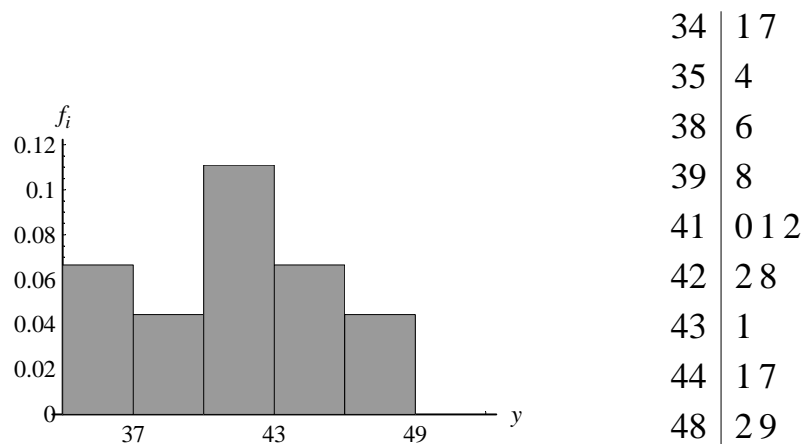
**6.23** Media, mediana, varianza (corretta), deviazione standard, quartili e range sono rispettivamente:

$$\begin{aligned} \bar{y} &= \sum_{i=1}^n y_i / n = 41.3, & \hat{y}_{0.5} &= 41.2, & s^2 &= \sum_{i=1}^n (y_i - \bar{y})^2 / (n-1) = 19.4, \\ s &= 4.40, & \hat{y}_{0.25} &= 39.2, & \hat{y}_{0.75} &= 43.6, \\ \text{Max}(y_1, \dots, y_n) - \text{Min}(y_1, \dots, y_n) &= 48.9 - 34.1 = 14.8. \end{aligned}$$

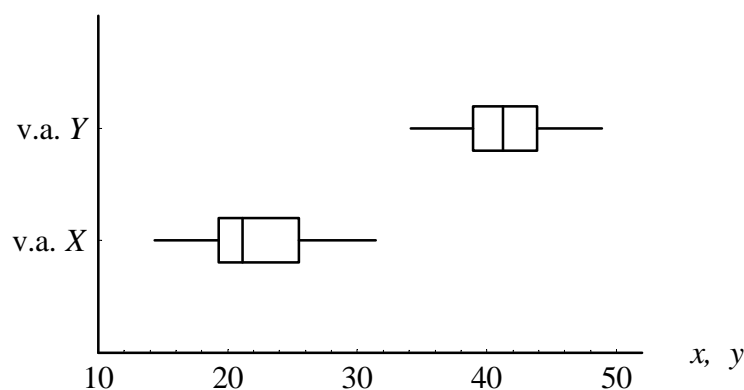
**6.24** Arrotondando il valore  $m = 1 + 3.3 \log_{10}(n) = 4.88$  abbiamo  $m = 5$  da cui, arrotondando  $[\text{Max}(y_1, \dots, y_n) - \text{Min}(y_1, \dots, y_n)] / m = 2.96$ , segue un'ampiezza dell'in-

tervallo  $\Delta y = 3$ . Arrotondando per difetto  $\text{Min}(y_1, \dots, y_n) = 34.1$  abbiamo che l'estremo inferiore del primo intervallo è uguale a 34. Arrotondando per eccesso  $\text{Max}(y_1, \dots, y_n) = 48.9 \cong 50$  ma essendo non intero il rapporto  $(50 - 34)/3 = 5.33$  fissiamo al valore 52 l'estremo superiore dell'ultimo intervallo. Ai 6 intervalli così individuati corrispondono le frequenze assolute  $n_i = \{3, 2, 5, 3, 2, 0\}$  e le densità di frequenza  $f_i = n_i / (n \Delta y) = \{3/45, 2/45, 5/45, 3/45, 2/45, 0\}$ .

**6.25** Sulla base dei calcoli già effettuati, l'istogramma ed il diagramma *Stem-and-Leaf*, del campione della v.a.  $Y$  esaminato nei due problemi precedenti, sono rispettivamente:



**6.26** Utilizzando i valori  $\text{Min}(x_1, \dots, x_n) = 14.3$ ,  $\text{Max}(x_1, \dots, x_n) = 31.4$ ,  $\hat{x}_{0.25} = 19.2$ ,  $\hat{x}_{0.75} = 26.5$ , nonché  $\text{Min}(y_1, \dots, y_n) = 34.1$ ,  $\text{Max}(y_1, \dots, y_n) = 48.9$ ,  $\hat{y}_{0.25} = 38.6$  e  $\hat{y}_{0.75} = 44.1$ , già calcolati in precedenza, i richiesti grafici *Box-Plot*, rispettivamente del campione della v.a.  $X$  e di quello della v.a.  $Y$ , sono:



**6.27** La media, mediana, varianza (corretta), deviazione standard, quartili e range sono rispettivamente:

**584** Soluzioni dei problemi proposti

$$\bar{x} = \sum_{i=1}^n x_i / n = 28.96, \quad \hat{x}_{0.5} = 23, \quad s^2 = \sum_{i=1}^n (x_i - \bar{x})^2 / (n-1) = 730.71,$$

$$s = 27.03, \quad \hat{x}_{0.25} = 14, \quad \hat{x}_{0.75} = 39,$$

$$\text{Max}(x_1, \dots, x_n) - \text{Min}(x_1, \dots, x_n) = 123 - 0 = 123.$$

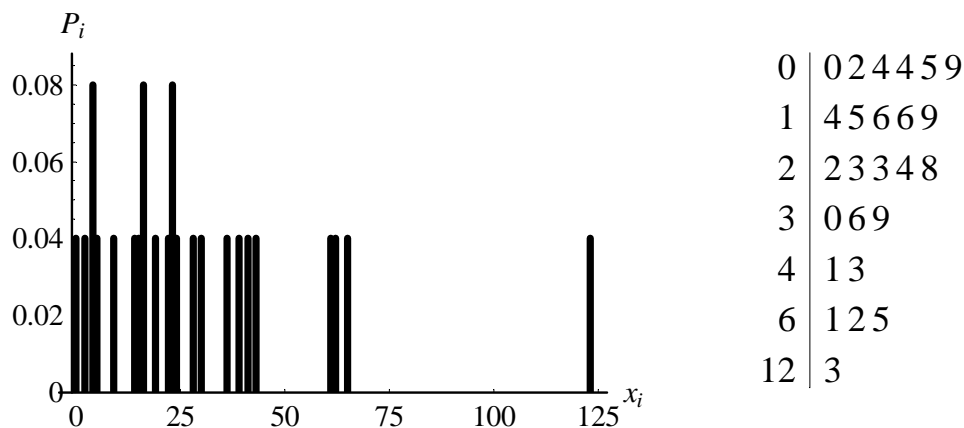
**6.28** Il vettore dei dati ordinati  $x_i$ , quello delle corrispondenti frequenze assolute  $n_i$  e quello delle frequenze relative  $P_i$  sono:

$$x_i = \{0, 2, 4, 5, 9, 14, 15, 16, 19, 22, 23, 24, 28, 30, 36, 39, 41, 43, 61, 62, 65, 123\};$$

$$n_i = \{1, 1, 2, 1, 1, 1, 1, 2, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1\};$$

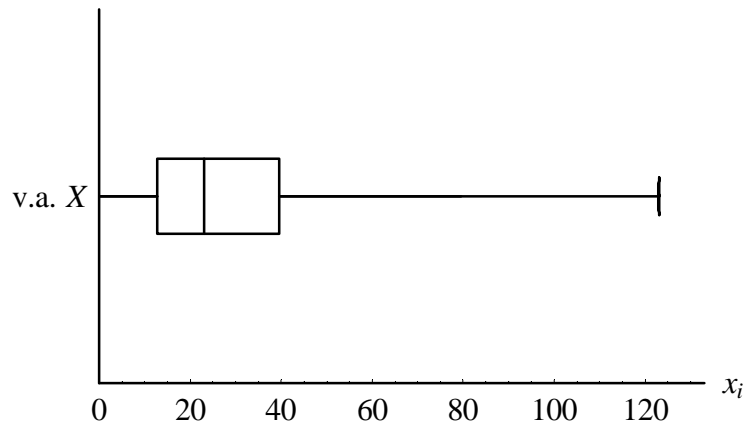
$$P_i = \{0.04, 0.04, 0.08, 0.04, 0.04, 0.04, 0.04, 0.08, 0.04, 0.04, 0.08, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 0.04\}.$$

**6.29** Sulla base dei calcoli già effettuati, il diagramma a barre ed il diagramma *Stem-and-Leaf*, del campione della v.a.  $X$  esaminato nei due problemi precedenti, sono rispettivamente:



È evidente che il diagramma a barre essendo alquanto “piatto” risulta meno informativo dello *Stem-and-Leaf* che, invece, sembra far intravedere meglio la forma della legge di probabilità.

**6.30** In questo caso il Box-Plot evidenzia che un dato è addirittura distante dalla mediana più di 3 volte la distanza interquartilica pari a  $\hat{x}_{0.75} - \hat{x}_{0.25} = 39 - 14 = 25$ .



## Capitolo 7

**7.9** La relazione  $\Pr\{S^2 - \sigma^2 < 1.5 \sigma^2\} > 0.95$  può essere facilmente trasformata nella  $\Pr\{(n-1)S^2/\sigma^2 < (n-1)(1.5+1)\} > 0.95$ , essendo  $(n-1)S^2/\sigma^2$  una v.a.  $\chi^2$  con  $\nu = n-1$  gradi di libertà.

Dalle tabelle deduciamo che deve essere almeno  $n=5$ . Infatti per  $\nu = n-1=3$  la probabilità è inferiore a 0.95 essendo  $2.5(n-1) = 7.50 < 7.81 = \chi_{3,0.95}^2$ . Invece per  $\nu = n-1=4$  abbiamo  $2.5(n-1) = 10.00 > 9.49 = \chi_{4,0.95}^2$  e, quindi, la probabilità è appena superiore a 0.95.

**7.10** La condizione imposta  $\Pr\{|\bar{X} - \mu| < 0.5 S\} > 0.90$  è facilmente trasformata nella  $\Pr\{-0.5 \sqrt{n} < (\bar{X} - \mu)/(S/\sqrt{n}) < 0.5 \sqrt{n}\} > 0.90$ , da cui ricaviamo che deve essere  $\Pr\{(\bar{X} - \mu)/(S/\sqrt{n}) < 0.5 \sqrt{n}\} > 0.95$  essendo  $(\bar{X} - \mu)/(S/\sqrt{n})$  una v.a.  $T$  con  $\nu = n-1$  gradi di libertà.

Dalle tabelle deduciamo che deve essere almeno  $n=13$ . Infatti per  $\nu = n-1=11$  la probabilità è inferiore a 0.95 essendo  $0.5 \sqrt{n} = 1.732 < 1.796 = T_{11,0.95}$ . Invece per  $\nu = n-1=12$  abbiamo  $0.5 \sqrt{n} = 1.803 > 1.782 = T_{12,0.95}$  e, quindi, la probabilità è appena superiore a 0.95.

**7.11** L'ellisse considerata ha semiassi di lunghezza  $k\sigma_X$  e  $k\sigma_Y$ . Un punto cade all'interno di tale ellisse se e solo se le sue coordinate verificano l'equazione di un'ellisse contenuta nella prima:

$$\frac{x^2}{\sigma_X^2} + \frac{y^2}{\sigma_Y^2} < k^2$$

**586** Soluzioni dei problemi proposti

e quest'evento si verifica con probabilità  $\Pr\{\chi^2 < k^2\}$  perché il primo membro della disequaglianza — essendo la somma di 2 quadrati di v.a. Normali standard  $s$ -indipendenti — costituisce una v.a.  $\chi^2$  con 2 gradi di libertà.

**7.12** Poiché una semplice trasformazione delle variabili  $X$  ed  $Y$  può far ruotare quest'ellisse facendo sovrapporre i suoi assi agli assi coordinati, il problema viene ridotto a quello precedente.

**7.13** Per verificare l'affermazione calcoliamo la Mgf del quadrato di una v.a. normale standard:

$$\begin{aligned}\Phi_{U^2}(t) &= E\left\{e^{tU^2}\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{tu^2} e^{-u^2/2} du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{u^2}{2}(1-2t)} du = \\ &= \frac{1}{(1-2t)^{1/2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}} dy = \frac{1}{(1-2t)^{1/2}}; \quad y = u(1-2t)^{1/2}\end{aligned}$$

da cui deduciamo che la Mgf della  $\sum_{i=1}^v U_i^2$ , essendo ovviamente uguale a:

$$\left[\Phi_{U^2}(t)\right]^v = \frac{1}{(1-2t)^{v/2}}$$

risulta effettivamente coincidente con quella della v.a.  $\chi^2$ .

**7.14** La pdf del quadrato di una v.a. Normale standard  $X = U^2$  è (cfr. Problema **Error! Reference source not found.**):

$$f_X(x) = \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{x}{2}}; \quad x \geq 0. \quad (1)$$

Quindi, la pdf della somma  $Y$  di due  $X$   $s$ -indipendenti può essere ottenuta a partire dalla pdf bidimensionale:

$$f(x_1, x_2) = f_X(x_1) f_X(x_2) = \frac{1}{2\pi} x_1^{-\frac{1}{2}} x_2^{-\frac{1}{2}} e^{-\frac{x_1+x_2}{2}}$$

operando la trasformazione:

$$y = x_1 + x_2; \quad z = \frac{x_2}{y}; \quad |J| = y; \quad 0 \leq z = \frac{x_2}{x_1 + x_2} \leq 1$$

otteniamo:

$i$	1	2	3	4	5	6	7
$x_i$	1.45	1.75	2.05	2.35	2.65	2.95	3.25
$f_i \times \Delta x = \frac{n_i}{n}$	0.020	0.060	0.340	0.220	0.240	0.100	0.020

che, integrata rispetto a  $z$ , tenendo conto che  $\int (z - z^2)^{-\frac{1}{2}} dz = -\arcsen(1 - 2z)$ , fornisce:

$$x = n / \sum_{i=1}^n (1/x_i). \quad (2)$$

Analogamente, possiamo ricavare la pdf della somma  $Z$ , di  $Y$  e di un'ulteriore  $X$ , a partire dalla pdf bidimensionale:

$$f(y, x) = f_Y(y) f_X(x) = \frac{1}{2\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{y+x}{2}}$$

che può essere trasformata:

$$z = y + x; \quad t = \frac{x}{z}; \quad |J| = z; \quad 0 \leq t = \frac{x}{y+x} \leq 1; \quad f(z, t) = \frac{1}{2\sqrt{\pi}} \left(\frac{z}{2}\right)^{\frac{1}{2}} e^{-\frac{z}{2}} t^{-\frac{1}{2}}$$

e quindi integrata rispetto a  $t$ :

$$f_Z(z) = \frac{1}{\sqrt{\pi}} \left(\frac{z}{2}\right)^{\frac{1}{2}} e^{-\frac{z}{2}}; \quad z \geq 0. \quad (3)$$

Le (1) (2) e (3), tenendo rispettivamente conto delle identità:

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}; \quad \Gamma(1) = 1; \quad \Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right)$$

possono essere poste nella seguente forma:

$$f_X(x) = \frac{1}{2\Gamma\left(\frac{1}{2}\right)} \left(\frac{x}{2}\right)^{\frac{1}{2}-1} e^{-\frac{x}{2}}; \quad f_Y(y) = \frac{1}{2\Gamma\left(\frac{2}{2}\right)} \left(\frac{y}{2}\right)^{\frac{2}{2}-1} e^{-\frac{y}{2}}$$

$$f_Z(z) = \frac{1}{2\Gamma\left(\frac{3}{2}\right)} \left(\frac{z}{2}\right)^{\frac{3}{2}-1} e^{-\frac{z}{2}}$$

da cui appare evidente la ricorsività che conduce direttamente alla pdf della Chi-Quadrato. Inoltre possiamo verificare facilmente che:

$$\int_0^{\infty} f(\kappa) d\kappa = \frac{1}{2\Gamma\left(\frac{\nu}{2}\right)} \int_0^{\infty} \left(\frac{\kappa}{2}\right)^{\frac{\nu}{2}-1} e^{-\frac{\kappa}{2}} d\kappa = \frac{2}{2\Gamma\left(\frac{\nu}{2}\right)} \int_0^{\infty} (x)^{\frac{\nu}{2}-1} e^{-x} dx = 1; \quad x = \frac{\kappa}{2}.$$

**7.15** I valori di una v.a. che delimitano una massa di probabilità centrata sono quelli che definiscono code di probabilità simmetriche. Nel nostro caso sono i percentili  $z_{0.05}$  e  $z_{0.95}$ . Il primo è immediatamente individuabile nella tabella del 95° percentile della  $Z_{5,8}$ :

$$\Pr\{Z_{5,8} < 3.69\} = 0.95.$$

Anche il secondo può essere dedotto dalla stessa tabella tenendo però conto che è pari al reciproco del valore 4.82 del 95° percentile della  $Z_{8,5}$  (si noti l'inversione dei pe-

**588** Soluzioni dei problemi proposti

dici):

$$\Pr\left\{Z_{5,8} < \frac{1}{4.82} = 0.21\right\} = 0.05.$$

Quindi:

$$\Pr\{0.21 \leq Z_{5,8} \leq 3.69\} = 0.90.$$

**7.16** Dovendo valutare la seguente probabilità:

$$\Pr\{|\bar{X} - \mu| < 2\}$$

basterà dividere entrambi i membri della disuguaglianza per  $s/\sqrt{n}$ , così da ottenere:

$$\Pr\left\{\frac{2}{5/\sqrt{25}} < \frac{\bar{X} - \mu}{S/\sqrt{n}} < \frac{2}{5/\sqrt{25}}\right\}$$

e considerare che nell'ipotesi di campionamento da una popolazione Gaussiana, la quantità  $(\bar{X} - \mu)/(S/\sqrt{n})$  è una v.a.  $T$  di Student con  $\nu = n - 1 = 24$  gradi di libertà. Dalla tabella della  $T$  di Student deduciamo:

$$\Pr\{|\bar{X} - \mu| < 2\} = \Pr\{-2 < T_{24} < 2\} = 1 - 2[\Pr\{T_{24} > 2\}] \cong 1 - 2(0.025) = 0.95$$

avendo utilizzato il valore presente in tabella più vicino a 2.0 (cioè 2.064).

**7.17** Il valore dell'errore,  $\varepsilon$ , che soddisfa la condizione imposta  $\Pr(\bar{X} - \mu > \varepsilon) = 0.01$ , può essere individuato considerando la seguente trasformazione:

$$\Pr\left\{\frac{\bar{X} - \mu}{S/\sqrt{n}} > \frac{\varepsilon}{5/\sqrt{25}}\right\} = 0.01.$$

Nell'ipotesi di campionamento da una popolazione Gaussiana,  $(\bar{X} - \mu)/(S/\sqrt{n})$  è una v.a.  $T$  di Student con  $\nu = n - 1 = 24$  gradi di libertà. Dalla tabella della  $T$  di Student troviamo che esso è pari a 2.492. Pertanto:

$$\frac{\varepsilon}{5/\sqrt{25}} = 2.492; \quad \varepsilon = 2.492.$$

**7.18** Dovendo valutare la seguente probabilità  $\Pr\{S_A^2/S_B^2 > 2\}$ , basterà moltiplicare entrambi i membri della disuguaglianza per  $\sigma_B^2/\sigma_A^2$ , ottenendo:

$$\Pr\left\{\left(S_A^2\sigma_B^2\right)/\left(S_B^2\sigma_A^2\right) > 2 (1.8/2)\right\}.$$

Nell'ipotesi di campionamento da una popolazione Gaussiana, il primo membro della disequazione è una v.a.  $Z$  di Fisher con  $\nu_1 = \nu_2 = n - 1 = 20$  gradi di libertà. Dalla tabella del 90° percentile della pdf di Fisher, deduciamo che la probabilità richiesta è:

$$\Pr\{Z_{20,20} > 1.8\} = 0.10.$$

**7.19** La v.a.  $X_1^2/36$ , essendo una Normale standard al quadrato, è una  $\chi^2$  con  $\nu = 1$ . Quindi la probabilità richiesta è  $\Pr\{X_1^2/36 \leq 7.88\} = 0.995$ .



**7.20** La v.a.  $Y$  è una  $T$  di Student con  $\nu = 1$ . Poiché il valore 15 non è presente in tabella, procediamo per interpolazione lineare. La probabilità richiesta è:

$$\Pr\{Y > 15\} = 0.010 + \frac{31.821 - 15}{31.821 - 12.706}(0.025 - 0.010) = 0.023.$$

**7.21** La v.a.  $Z$  è la radice quadrata di una v.a. Chi-Quadrato,  $K$ , con  $\nu = 3$ . Dalla tabella della  $\chi^2$ , abbiamo:

$$\Pr\{Z \leq 3.06\} = \Pr\{Z^2 = K \leq 9.35\} = 1 - 0.025 = 0.975.$$

**7.22** È sufficiente eguagliare la Mgf della v.a. Chi-Quadrato con  $\nu$  gradi di libertà con la Mgf della v.a. Esponenziale con  $\lambda = 1/2$ :

$$\left(\frac{1}{1-2t}\right)^{\frac{\nu}{2}} = \left(\frac{1/2}{1/2-t}\right)^{\frac{\nu}{2}} = \frac{\lambda}{\lambda-t}; \quad \lambda = \frac{1}{2}$$

ottenendo che deve essere  $\nu = 2$ .

**7.23** Indicando con  $Z$  la v.a. “distanza del proiettile dal bersaglio”, la probabilità cercata è formalmente esprimibile in termini di Cdf della v.a. Chi-Quadrato con  $\nu = 2$ :

$$\Pr\{Z \leq 20\} = \Pr\{X^2 + Y^2 \leq 400\} = \Pr\left\{\left(\frac{X-0}{10}\right)^2 + \left(\frac{Y-0}{10}\right)^2 \leq 4\right\} = \Pr\{K \leq 4\}$$

e quindi, in base alla soluzione del problema precedente, può essere calcolata utilizzando la Cdf della v.a. Esponenziale con  $\lambda = 1/2$ :

$$\Pr\{K \leq 4\} = 1 - e^{-2} = 0.86.$$

**7.24** Applicando la trasformazione di v.a., otteniamo:

$$f_Y(y) = \lambda e^{-\frac{y}{2}} \frac{\left(\frac{y}{2}\right)^{n-1}}{\Gamma(n)} \frac{1}{2\lambda} = \frac{1}{2\Gamma(n)} \left(\frac{y}{2}\right)^{n-1} e^{-\frac{y}{2}}$$

che è la pdf di una v.a. Chi-Quadrato con  $\nu = 2n$  gradi di libertà. Si giunge allo stesso risultato considerando che la Mgf della v.a.  $Y$  è:

$$\Phi_Y(t) = E\{e^{tY}\} = E\{e^{2t\lambda X}\} = \left(\frac{\lambda}{\lambda - 2t\lambda}\right)^n = \frac{1}{(1-2t)^n}.$$

**7.25** Partendo dal numeratore della quantità a secondo membro si ha:

**590** Soluzioni dei problemi proposti

$$\begin{aligned}
 \sum_{i=1}^n \sum_{j=1}^n (X_i - X_j)^2 &= \sum_{i=1}^n \sum_{j=1}^n (X_i - X_j - \bar{X} + \bar{X})^2 = \sum_{i=1}^n \sum_{j=1}^n [(X_i - \bar{X}) - (X_j - \bar{X})]^2 = \\
 &= \sum_{i=1}^n \sum_{j=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^n \sum_{j=1}^n (X_j - \bar{X})^2 - 2 \sum_{i=1}^n \sum_{j=1}^n [(X_i - \bar{X})(X_j - \bar{X})] = \\
 &= n \sum_{i=1}^n (X_i - \bar{X})^2 + n \sum_{j=1}^n (X_j - \bar{X})^2 - 2 \left[ \sum_{i=1}^n (X_i - \bar{X}) \right] \left[ \sum_{j=1}^n (X_j - \bar{X}) \right] = \\
 &= 2n \sum_{i=1}^n (X_i - \bar{X})^2 - 2 \left[ \left( \sum_{i=1}^n X_i \right) - n\bar{X} \right] \left[ \left( \sum_{j=1}^n X_j \right) - n\bar{X} \right] = 2n \sum_{i=1}^n (X_i - \bar{X})^2
 \end{aligned}$$

da cui l'uguaglianza resta dimostrata.

**7.26** Nel caso di un campione di dimensione  $n=2$  la media e la varianza campionaria possono essere riformulate come segue (cfr. problema precedente):

$$\bar{X} = \frac{X_1 + X_2}{2}; \quad S^2 = \sum_{i=1}^2 \frac{(X_i - \bar{X})^2}{2-1} = \frac{(X_1 - X_2)^2}{2}.$$

Per dimostrare che  $\bar{X}$  ed  $S^2$  sono  $s$ -indipendenti sarà sufficiente dimostrare che lo sono le v.a.  $Y_1 = X_1 + X_2$  ed  $Y_2 = X_1 - X_2$ . Indichiamo con  $\mu_X$  e  $\sigma_X^2$  rispettivamente la media e la varianza della popolazione gaussiana da cui è estratto il campione. In virtù della riproducibilità della v.a. Gaussiana (§5.3) abbiamo:

$$\begin{aligned}
 \Phi_{Y_1}(t) &= \Phi_{X_1+X_2}(t) = \exp(2\mu_X t + \sigma_X^2 t^2); \\
 \Phi_{Y_2}(t) &= \Phi_{X_1-X_2}(t) = \exp(\sigma_X^2 t^2).
 \end{aligned}$$

Indicando con  $Y_3$  la v.a. somma delle v.a.  $Y_1$  ed  $Y_2$ , abbiamo:

$$\begin{aligned}
 Y_3 &= Y_1 + Y_2 = X_1 + X_2 + X_1 - X_2 = 2X_1 \\
 \Phi_{Y_3}(t) &= \Phi_{2X_1}(t) = \exp(2\mu_X t + 2\sigma_X^2 t^2)
 \end{aligned}$$

da cui, risultando verificata l'uguaglianza:

$$\Phi_{Y_1}(t) \cdot \Phi_{Y_2}(t) = \Phi_{Y_3}(t)$$

si evince che le v.a.  $Y_1$  ed  $Y_2$  sono  $s$ -indipendenti.

**7.27** Essendo una v.a. riproducibile, la  $\chi^2$  con  $\nu=50$  coincide con la somma di 50  $\chi^2$  con  $\nu=1$ . Se ne deduce che, per la convergenza in distribuzione dimostrata dal teorema del limite centrale, i percentili di una  $\chi^2$  con  $\nu=50$  gradi di libertà tendono a quelli di una Gaussiana con media  $\nu=50$  e varianza  $2\nu=100$  (rispettivamente pari ad  $n$  volte la media e la varianza di una  $\chi^2$  con  $\nu=1$ ). Indicato con  $\kappa_{0.95,\nu}$  il percentile cercato, abbiamo:

$$\kappa_{0.95,\nu} = F_U^{-1}(0.95)\sqrt{2\nu} + \nu = 1.645 \cdot \sqrt{2\nu} + \nu$$

essendo  $F_U^{-1}(0.95)$  il 95° percentile della v.a. Gaussiana standard. Pertanto risulta:

$$\kappa_{0.95, \nu=50} = 66.45; \quad \kappa_{0.95, \nu=100} = 123.26,$$

mentre la tabella della  $\chi^2$  fornisce rispettivamente i valori esatti:

$$\kappa_{0.95, \nu=50} = 67.51; \quad \kappa_{0.95, \nu=100} = 124.34$$

indicando che l'errore relativo connesso alla procedura diminuisce al crescere di  $\nu$ .

**7.28** Indicando con  $K$  la v.a. Chi-Quadrato, per definizione abbiamo:

$$Z_1 = \frac{K_1/10}{K_2/6},$$

da ciò si deduce che essendo:

$$Z_2 = \frac{1}{Z_1} = \frac{K_2/6}{K_1/10}$$

il percentile cercato è quello indicato in tabella per  $\nu_1 = 6$  e  $\nu_2 = 10$ , cioè 3.22.

**7.29** Per definizione abbiamo:

$$T = \frac{U}{\sqrt{K_\nu/\nu}}$$

dove  $U$  e  $K_\nu$  indicano le v.a.,  $s$ -indipendenti, Gaussiana standard e Chi-Quadrato con  $\nu$  gradi di libertà. Vale la seguente identità:

$$T^2 = \frac{U^2}{K_\nu/\nu} = \frac{K_1/1}{K_\nu/\nu} = Z$$

da cui si evince che il quadrato di una  $T$  di Student con  $\nu$  gradi di libertà è una v.a. di Fisher con  $\nu_1 = 1$  e  $\nu_2 = \nu$  gradi di libertà.

**7.30** Indicati con  $t_{0.05}$  e  $t_{0.95}$  rispettivamente il 5° ed il 95° percentile della  $T$  con  $\nu = 10$  gradi di libertà e con  $z_{0.90}$  il 90° percentile di una v.a.  $Z$  di Fisher con  $\nu_1 = 1$  e  $\nu_2 = \nu = 10$  gradi di libertà, utilizzando la soluzione del problema precedente, abbiamo:

$$\Pr\{t_{0.05} \leq T \leq t_{0.95}\} = 0.90 = \Pr\{T^2 \leq t_{0.95}^2\} = \Pr\{Z \leq z_{0.90}\}$$

da cui:

$$t_{0.95}^2 = z_{0.90}; \quad t_{0.95} = \sqrt{z_{0.90}}.$$

Infatti, dalle tabelle delle due v.a. abbiamo:

$$z_{0.90} = 3.29; \quad t_{0.95} = 1.81 = \sqrt{3.29}.$$

## Capitolo 8

**8.9** La risposta è affermativa. Infatti la media dei primi due costituisce uno stimatore con le caratteristiche richieste. Esso è corretto come i primi due condividendo la speranza matematica, essendo:

$$E\{(\hat{p}_1 + \hat{p}_2)/2\} = (E\{\hat{p}_1\} + E\{\hat{p}_2\})/2 = p.$$

**592** Soluzioni dei problemi proposti

Inoltre esso ha varianza:

$$\text{Var}\{(\hat{p}_1 + \hat{p}_2)/2\} = (\text{Var}\{\hat{p}_1\} + \text{Var}\{\hat{p}_2\})/4 + 2\text{Cov}\{\hat{p}_1, \hat{p}_2\}/4$$

che è notevolmente inferiore, essendo pari alla metà della media delle varianze dei primi due stimatori meno un'ulteriore aliquota dovuta alla covarianza negativa.

**8.10** È sufficiente utilizzare il primo stimatore per i dati pseudo-sperimentali generati a partire da una serie di determinazioni  $u_i$  della v.a. Uniforme definita in  $(0, 1)$  ed il secondo stimatore a partire dalla serie di determinazioni  $u'_i = 1 - u_i$ , ottenute come complemento all'unità delle prime.

**8.11** Per effettuare una verifica rapida, si può generare un numero alto di determinazioni (ad esempio  $n=20\cdot000$ ), calcolarne media e varianza campionaria e controllare che queste siano prossime rispettivamente a 0.5 ed a  $1/12$ , valori della media e della varianza teorica della Uniforme definita nell'intervallo  $(0, 1)$ . Una verifica più rigorosa è un test d'ipotesi di adattamento (cfr. 10.4) effettuato sui dati generati.

**8.12** Generate  $n$  determinazioni,  $u_i$ , di una v.a. uniforme definita in  $(0, 1)$ , per ottenere  $n$  determinazioni  $x_i$  di  $X$ , può essere utilizzata la trasformazione:

$$x_i = F_X^{-1}(u_i) = \ln[u_i(e-1) + 1].$$

**8.13** Una possibile procedura è: a) estrarre  $n$  valori da una tabella di numeri pseudo-casuali; b) ridurli all'intervallo  $(0, 1)$  ottenendo  $n$  determinazioni,  $u_i$ , di una v.a. Uniforme definita in  $(0, 1)$ ; utilizzare quest'ultimi come valori della Cdf Gaussiana standard,  $F_X(x_i)$ , per ottenere da una tabella di quest'ultima i corrispondenti percentili. Questi ultimi costituiscono le determinazioni cercate.

**8.14** Una volta generate  $n$  determinazioni,  $x_i$ , di una v.a. Gaussiana standard, anche senza l'uso di strumenti di calcolo come nel caso del problema precedente, le corrispondenti determinazioni,  $y_i$ , di una Gaussiana di media 5 e varianza 9 sono ottenute con la trasformazione  $y_i = 3x_i + 5$ .

**8.15** Si possono generare  $n$  successioni di  $k$  determinazioni di una v.a. uniforme definita nell'intervallo  $(0, 1)$ ,  $u_{i,j}$  ( $i=1, \dots, n$ ;  $j=1, \dots, k$ ), e ricavare le  $n$  determinazioni,  $x_i$ , come medie normalizzate dei dati delle rispettive  $n$  successioni:

$$x_i = \left( \sum_{j=1}^k \frac{u_{i,j}}{k} - \frac{1}{2} \right) / \sqrt{1/(k \cdot 12)}.$$

Per il teorema del limite centrale l'approssimazione migliora al crescere di  $k$ .

**8.16** La variabile  $Z = (Y - a)/(b - a) = F_X(x)$  è uniforme in  $(0, 1)$  per cui la sua trasformata lineare  $Y = (b - a)Z + a$  resta uniforme ma risulta definita in  $(a, b)$ .

**8.17** Si possono generare  $n$  successioni di  $k$  determinazioni di una v.a. esponenziale, di parametro  $\lambda$ ,  $x_{i,j}$  ( $i=1, \dots, n$ ;  $j=1, \dots, k$ ), e ricavare  $n$  determinazioni dello stima-

tore del parametro  $\theta = 1/\lambda$  come medie dei  $k$  dati delle rispettive  $n$  successioni:

$$\hat{\theta}_i = \sum_{j=1}^k \frac{x_{i,j}}{k}$$

e verificare che la differenza tra la media  $\sum_{i=1}^n \hat{\theta}_i/n$  ed il valore vero  $\theta = 1/\lambda$  diminuisce all'aumentare del prodotto  $kn$ .

**8.18** Come per il problema precedente, si possono generare  $n$  successioni di  $k$  determinazioni di una v.a. esponenziale, di parametro  $\lambda$ ,  $x_{i,j}$  ( $i=1, \dots, n$ ;  $j=1, \dots, k$ ), e ricavare  $n$  determinazioni dello stimatore del parametro  $\theta = 1/\lambda$  come medie campionarie dei  $k$  dati delle rispettive  $n$  successioni. Quindi, calcolare la varianza delle  $n$  determinazioni:

$$\sum_{i=1}^n \frac{(\hat{\theta}_i - \theta)^2}{n}$$

e verificare che la differenza tra questa ed il valore vero  $1/(k\lambda^2)$  diminuisce all'aumentare del prodotto  $kn$ .

**8.19** Fissato, nel primo quadrante di un sistema di assi cartesiani, un quadrato con vertice nell'origine e con lati di lunghezza  $a$  (sovrapposti agli assi) tale da includere la figura piana, è sufficiente generare un numero elevato  $n$  di coppie di determinazioni  $s$ -indipendenti di v.a. uniformi definite nell'intervallo  $(0, a)$ . Il rapporto tra il numero,  $n_I$ , di punti che cadono all'interno della figura piana ed il numero totale,  $n$ , fornisce una stima della sua area.

**8.20** Utilizzando l'equazione (8.8) con  $p = \Pr\{\bar{X} \leq \mu_X\} = 0.5$  ed  $u_{0.975} = 1.96$ , la soluzione del problema è il valore più piccolo di  $n$  che verifica la disequazione:

$$1.96 \sqrt{(1-0.5)/(0.5n)} \times 100 < 10.$$

Risolvendo si ricava:

$$n > \left( \frac{1.96 \times 100}{10} \right)^2 = 384.2.$$

Pertanto è necessario generare almeno 385 campioni pseudo-casuali.

**8.21** In base ai parametri  $a$  e  $b$  adottati per il generatore, per ogni seme  $x_0$ , la lunghezza massima è  $m = 32$ . Infatti, ad esempio, per  $x_0 = 1$  ed  $x_0 = 2$  abbiamo rispettivamente:

8, 15, 22, 29, 4, 11, 18, 25, 0, 7, 14, 21, 28, 3, 10, 17, 24, 31, 6, 13, 20, 27, 2, 9, 16, 23, 30, 5, 12, 19, 26, 1, 8, 15, 22, 29, 4, 11, 18, 25, ... ;

9, 16, 23, 30, 5, 12, 19, 26, 1, 8, 15, 22, 29, 4, 11, 18, 25, 0, 7, 14, 21, 28, 3, 10, 17, 24, 31, 6, 13, 20, 27, 2, 9, 16, 23, 30, 5, 12, 19, 26, ... .

**8.22** No, in quanto essendo  $3 \neq k \times 4 + 1$  (con  $k = 0, 1, 2, \dots$ ) non sarebbe rispettata una delle condizioni necessarie per il raggiungimento della lunghezza massima del ciclo.

**594** Soluzioni dei problemi proposti

Infatti, fissando  $a=3$  e mantenendo inalterate le altre ipotesi, invece delle due precedenti sequenze di lunghezza massima  $m=32$  abbiamo rispettivamente due sequenze entrambe di lunghezza 16:

10, 5, 22, 9, 2, 13, 14, 17, 26, 21, 6, 25, 18, 29, 30, 1, 10, 5, 22, 9, 2, 13, 14, 17, 26, 21, 6, 25, 18, 29, 30, 1, 10, 5, 22, 9, 2, 13, 14, 17, ... ;

13, 14, 17, 26, 21, 6, 25, 18, 29, 30, 1, 10, 5, 22, 9, 2, 13, 14, 17, 26, 21, 6, 25, 18, 29, 30, 1, 10, 5, 22, 9, 2, 13, 14, 17, 26, 21, 6, 25, 18, ... .

**8.23** Generando 100'000 punti pseudo-casuali nel rettangolo di base  $3\pi$  ed altezza 1.1, 38'531 sono risultati compresi tra la funzione  $\sin x$  e l'asse delle ascisse. Invece, generando 100'000 punti pseudo-casuali nel rettangolo di base  $3\pi$  ed altezza  $-1.1$ , 18'989 sono risultati compresi tra la funzione  $\sin x$  e l'asse delle ascisse. Pertanto la stima dell'integrale è:  $3\pi \times 1.1 \times (38'531 - 18'989) / 100'000 = 2.026$ .

**8.24** Risulta  $A \times f_W(x) > f_K(x)$  per  $x \leq 10$  ed  $A \times f_W(x) \cong f_K(x)$  per  $x > 10$ . Essendo la Cdf della v.a. Weibull invertibile, è facile generare una sequenza di determinazioni  $w_i$  della v.a.  $W$ , nonché un uguale numero di determinazioni  $u_i$  della v.a. Uniforme definita in  $(0, 1)$ . Ogni  $w_i$  è accettata come determinazione della v.a.  $K$  se risulta  $u_i \leq f_K(w_i) / [A \times f_W(w_i)]$ . Visto che il rapporto  $f_K(w_i) / [A \times f_W(w_i)]$  è prossimo all'unità per  $w_i > 10$ , è per questi valori di  $w_i$  che il metodo raggiunge la massima efficienza, risultando tutti accettati come determinazioni pseudo-casuali della v.a. Chi-Quadrato  $K$ .

**8.25** La funzione di trasformazione richiesta è:

$$y = \varphi(u) = \begin{cases} 0; & 0 \leq u < 1/3; \\ 3u - 1; & 1/3 \leq u < 2/3; \\ 1; & 2/3 \leq u \leq 1. \end{cases}$$

Si osservi che la  $Y$  è una v.a. mista, in quanto è continua in  $(0, 1)$  ed ha due masse concentrate rispettivamente in 0 ed 1 essendo  $\Pr\{Y=0\} = \Pr\{Y=1\} = 1/3$ .

**8.26** Consideriamo le due seguenti trasformazioni:

$$\varphi_1(w) = F_W(w) = 1 - e^{-\left(\frac{w}{\alpha}\right)^\beta}; \quad \varphi_2(u) = F_{Z_M}^{-1}(u) = a - b \ln \ln \left(\frac{1}{u}\right).$$

La prima può essere utilizzata per trasformare le determinazioni pseudo-casuali della v.a. Weibull in quelle della v.a. Uniforme definita in  $(0, 1)$ ; mentre la seconda per trasformare quest'ultime in determinazioni della v.a. Gumbel.

**8.27** Fissato un sistema di assi cartesiani, si può considerare il quadrato di vertici  $(0, 0)$ ,  $(1, 0)$ ,  $(1, 1)$  e  $(0, 1)$  e il settore circolare inscritto in esso. Il rapporto tra l'area del quadrante e quella del quadrato è pari a  $p = \pi/4$ . Pertanto, generando punti con ascisse ed ordinate ottenute come determinazioni pseudo-casuali  $s$ -indipendenti di due v.a. uniformi, entrambe definite in  $(0, 1)$ , la frazione di punti che cade nel settore circolare è una stima della probabilità  $p$  che, moltiplicata per 4, fornisce a sua volta una stima

di  $\pi$ .

**8.28** Dalla condizione richiesta  $|\hat{\pi} - \pi| < \pi/100$  ricaviamo  $|(\hat{p} - p)/p| \times 100 < 1$  essendo  $\hat{p} = \hat{\pi}/4$  e  $p = \pi/4$ . Utilizzando l'equazione (8.8):

$$1.64 \sqrt{\left(1 - \frac{\pi}{4}\right) / \left(\frac{n\pi}{4}\right)} \times 100 < 1; \quad u_{0.975} = 1.64$$

si ricava che il numero  $n$  di pseudo-esperimenti da effettuare deve essere superiore a  $164^2 (4 - \pi) / \pi = 7349.05$ .

**8.29** Essendo la v.a. Gamma la somma di  $n$  v.a. Esponenziali  $s$ -indipendenti ed equidistribuite, è sufficiente sommare 3 determinazioni  $s$ -indipendenti della v.a. Esponenziale di parametro  $\lambda$  per ottenerne una della v.a.  $Z$ .

**8.30** Dopo aver generato una determinazione pseudo casuale  $x_i$  della v.a. Gaussiana di parametri  $\mu_X$  e  $\sigma_X$ , possiamo ottenere la corrispondente determinazione  $y_i$  generandola dalla Cdf gaussiana condizionata  $F(y|x)$  di parametri  $\sigma^2 = \sigma_Y^2(1 - \rho^2)$  e  $\mu = \mu_Y + \rho x_i \sigma_Y / \sigma_X$ .

## Capitolo 9

**9.9** A rigore, il metodo della massima verosimiglianza non è applicabile in questo caso come in tutti quelli in cui l'insieme di definizione della v.a. dipende dai parametri incogniti.

Tuttavia, disponendo di un campione casuale  $x_1, x_2, \dots, x_n$ , notiamo che la funzione di verosimiglianza  $L(a; x_1, x_2, \dots, x_n) = a^{-n}$  è massima, per valori finiti di  $a$ , in corrispondenza di  $a = \max\{x_1, x_2, \dots, x_n\}$ . Infatti la pdf Uniforme caratterizzata da questo valore del parametro è quella che presenta la massima pdf sui valori che sono stati osservati.

**9.10** Per un campione casuale, tratto da una popolazione poissoniana, la funzione di verosimiglianza ed il suo logaritmo naturale sono:

$$L(\mu; y_1, y_2, \dots, y_n) = \prod_{i=1}^n \frac{\mu^{y_i}}{y_i!} e^{-\mu} = e^{-n\mu} \prod_{i=1}^n \frac{1}{y_i!} \mu^{\sum_{i=1}^n y_i};$$

$$\ln(L) = -n\mu + \ln(\mu) \sum_{i=1}^n y_i + \text{costante}$$

da cui:

$$\frac{d \ln(L)}{d \mu} = -n + \frac{1}{\mu} \sum_{i=1}^n y_i = 0; \quad \hat{\mu} = \sum_{i=1}^n y_i / n.$$

**9.11** Possiamo ipotizzare che i dati ottenuti sono determinazioni di una v.a. Binomiale con  $n=1000$ . La stima puntuale della proporzione incognita  $p$  è  $\hat{p} = 30/1000 = 0.03$ . Il limite superiore di confidenza è:

**596** Soluzioni dei problemi proposti

$$p_2 = \frac{u_\alpha^2 + 2n\hat{p} + \sqrt{u_\alpha^4 + 4nu_\alpha^2\hat{p} - 4nu_\alpha^2\hat{p}^2}}{2(n + u_\alpha^2)} = 0.0403; \quad u_\alpha = 1.96.$$

Tuttavia considerando che  $n$  è elevato, possiamo trascurare i termini che *aggiungerebbero* (a quella funzione di  $n$ ) quantità certamente trascurabili:

$$p_2 = \hat{p} + u_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.0407; \quad u_\alpha = 1.96.$$

È evidente che l'approssimazione è accettabile e che il limite ottenuto corrisponde all'utilizzo della funzione ancillare  $U$  con  $\bar{x} = \hat{p}$  e  $\sigma = \hat{p}(1-\hat{p})$ .

**9.12** Calcolando due volte il logaritmo della Cdf Weibull otteniamo:

$$\ln \left[ \frac{1}{1 - F_W(w)} \right] = (w/\alpha)^\beta; \quad Y(w) = \ln \ln \left[ \frac{1}{1 - F_W(w)} \right] = \beta \ln(w) - \beta \ln(\alpha).$$

Stimata la retta di regressione di  $Y(w_i) = \ln \ln [1 - i/(n+1)]^{-1}$  in  $\ln(w_i)$ , i valori  $w_0$  e  $w_1$  corrispondenti a  $Y=0$  ed  $Y=1$  sono tali che  $\hat{\alpha} = w_0$  e  $\hat{\beta} = [\ln(w_1) - \ln(\hat{\alpha})]^{-1}$ .

**9.13** Il numero medio di guasti per ora di lavoro è  $\hat{\mu} = 3/7500 = 0.0004$ . Utilizzando la funzione ancillare  $U$  abbiamo:

$$\Pr \left\{ -u_{\alpha/2} < \frac{\hat{\mu} - \mu}{\sqrt{\mu/n}} < u_{\alpha/2} \right\} = 1 - \alpha; \quad u_{\alpha/2} = 1.96$$

che ci permette d'impostare l'equazione  $(\hat{\mu} - \mu)^2 = u_{\alpha/2}^2 \frac{\mu}{n}$  le cui soluzioni costituiscono i limiti cercati:  $\mu_1 = 1.36 \times 10^{-4}$  e  $\mu_2 = 11.76 \times 10^{-4}$ .

**9.14** La stima della frazione di morti è  $\hat{p} = 22/143 = 0.154$ . Utilizzando la funzione ancillare  $U$  possiamo scrivere:

$$\Pr \left\{ -u_{\alpha/2} < \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} < u_{\alpha/2} \right\} = 1 - \alpha; \quad u_{\alpha/2} = 1.96$$

cui corrisponde l'equazione quadratica  $n(\hat{p} - p)^2 = u_{\alpha/2}^2 p(1-p)$  le cui soluzioni forniscono i limiti  $p_1 = 0.104$  e  $p_2 = 0.222$  ossia, al livello di fiducia del 95%, l'intervallo 10%÷22% include la percentuale effettiva di mortalità. Tale valutazione concorda con quanto l'OMS affermò negli stessi giorni: "il tasso di fatalità della SARS va dallo 0 al 50 per cento a seconda dell'età, con una stima complessiva di fatalità del 14-15 per cento".



**9.15** La v.a.  $X$ , numero di esemplari difettosi giornalieri, è assimilabile ad una v.a. di Poisson. Per la stima del parametro incognito  $\mu$ , si può adottare il metodo della massima verosimiglianza (si otterrebbe lo stesso risultato utilizzando il metodo dei momenti). La funzione di verosimiglianza ed il suo logaritmo naturale sono:

$$L(\mu; y_1, y_2, \dots, y_{30}) = \prod_{i=1}^{30} \frac{\mu^{y_i}}{y_i!} e^{-\mu};$$

$$\ln(L) = \sum_{i=1}^{30} \ln \left( \frac{\mu^{y_i}}{y_i!} e^{-\mu} \right) = \sum_{i=1}^{30} [y_i \ln \mu - \mu - \ln(y_i!)] = \ln \mu \sum_{i=1}^{30} y_i - n\mu - \sum_{i=1}^{30} \ln(y_i!)$$

da cui:

$$\frac{\partial \ln(L)}{\partial \mu} = \frac{1}{\mu} \sum_{i=1}^{30} y_i - n; \quad \hat{\mu} = \frac{1}{30} \sum_{i=1}^{30} y_i = 5.833.$$

Quindi, utilizzando la Cdf della v.a. di Poisson, è possibile valutare la probabilità richiesta:

$$F_Y(3) = \sum_{y=0}^3 \frac{\hat{\mu}^y}{y!} e^{-\hat{\mu}} = 0.166.$$

**9.16** La funzione di verosimiglianza ed il suo logaritmo naturale sono:

$$L(p; y_1, y_2, \dots, y_{10}) = \prod_{i=1}^{10} \binom{n_i}{y_i} p^{y_i} (1-p)^{n_i - y_i};$$

$$\ln(L) = \sum_{i=1}^{10} \left[ \ln \binom{n_i}{y_i} + y_i \ln p + (n_i - y_i) \ln(1-p) \right]$$

da cui:

$$\frac{d \ln(L)}{dp} = \sum_{i=1}^{10} \left[ y_i \frac{1}{p} - \frac{(n_i - y_i)}{1-p} \right] = 0; \quad \hat{p} = \frac{\sum_{i=1}^{10} y_i}{\sum_{i=1}^{10} n_i} = \frac{83}{147} = 0.56.$$

In pratica, lo stimatore di massima verosimiglianza di  $p$  è uguale al rapporto tra il numero totale di elementi difettosi ed il numero complessivo di elementi estratti. È bene evidenziare che potremmo ottenere un risultato anche molto diverso considerando la media delle singole stime  $\hat{p}_i = y_i/n_i$ . Invece, i risultati coincidono quando tutti i campioni hanno la stessa ampiezza,  $n_1 = n_2 = \dots = n_{10}$ . Infine, notiamo che il numero complessivo di elementi non conformi,  $Y$ , può essere considerato come somma  $Y = \sum_{i=1}^{10} Y_i$  del numero di elementi non conformi estratti ogni giorno,  $Y_i$ . Ritenendo  $p$  costante, per la riproducibilità della v.a. Binomiale,  $Y$  è una v.a. Binomiale, di parametri  $n = \sum_{i=1}^{10} n_i$  e  $p$ . Pertanto è come se avessimo un unico campione di  $n = \sum_{i=1}^{10} n_i$  elementi di cui  $y = \sum_{i=1}^{10} y_i$  sono non conformi, da cui deduciamo la stessa stima  $\hat{p}$  di prima.

**9.17** La funzione di verosimiglianza ed il suo logaritmo naturale sono:

**598** Soluzioni dei problemi proposti

$$L(\lambda; y_1, y_2, \dots, y_{10}) = \prod_{i=1}^{10} \frac{(\lambda x_i)^{y_i}}{y_i!} e^{-\lambda x_i};$$

$$\ln(L) = \sum_{i=1}^{10} [y_i \ln \lambda + y_i \ln x_i - \ln(y_i!) - \lambda x_i]$$

da cui:

$$\frac{d \ln(L)}{d \lambda} = \sum_{i=1}^{10} \left( y_i \frac{1}{\lambda} - x_i \right) = 0; \quad \hat{\lambda} = \frac{\sum_{i=1}^{10} y_i}{\sum_{i=1}^{10} x_i} = \frac{24}{30.5} = 0.787.$$

In pratica, lo stimatore di massima verosimiglianza di  $\lambda$  è uguale al rapporto tra il numero totale di difetti e la superficie complessiva delle lamiera. È bene evidenziare che potremmo ottenere un risultato anche molto diverso considerando la media delle singole stime  $\hat{\lambda}_i = y_i/x_i$ . Invece i risultati coincidono quando tutte le lamiera hanno la stessa dimensione,  $x_1 = x_2 = \dots = x_{10}$ . Infine, notiamo che il numero complessivo di difetti,  $Y$ , può essere considerato come somma  $Y = \sum_{i=1}^{10} Y_i$  del numero di difetti riscontrati su ogni lamiera,  $Y_i$ . Ritenendo  $\lambda$  costante, per la riproducibilità della v.a. di Poisson,  $Y$  è una v.a. di Poisson, di parametro  $\mu = \sum_{i=1}^{10} \mu_i = \sum_{i=1}^{10} \lambda x_i = \lambda_i \sum_{i=1}^{10} x_i$ . Pertanto è come se avessimo una lamiera di  $\sum_{i=1}^{10} x_i$  m<sup>2</sup> in cui sono presenti  $y = \sum_{i=1}^{10} y_i$  difetti, da cui deduciamo la stessa stima  $\hat{\lambda}$  di prima.

**9.18** Lo stimatore dei momenti del parametro  $a$  si ottiene eguagliando il momento primo del campione casuale, ossia la media campionaria,  $\bar{X}$ , alla formulazione del corrispondente momento della v.a.  $X$ , ossia il valore atteso,  $E\{X\}$ . Quindi:

$$E\{X\} = \int_1^{+\infty} x \frac{a}{x^{a+1}} dx = \frac{a}{a-1} = \bar{X}; \quad a = \frac{\bar{X}}{\bar{X}-1}.$$

**9.19** La funzione di verosimiglianza ed il suo logaritmo naturale sono:

$$L(a; x_1, \dots, x_n) = \prod_{i=1}^n \frac{a}{x_i^{a+1}} = \frac{a^n}{\prod_{i=1}^n x_i^{a+1}};$$

$$\ln(L) = n \ln(a) - (a+1) \sum_{i=1}^n \ln(x_i)$$

da cui:

$$\frac{\partial \ln(L)}{\partial a} = \frac{n}{a} - \sum_{i=1}^n \ln(x_i) = 0; \quad \hat{a} = \frac{n}{\sum_{i=1}^n \ln(x_i)}.$$

**9.20** Dalla relazione:

$$Y(x) = u = \frac{x - \mu}{\sigma}$$

si deduce che l'intercetta con l'asse delle  $x$  ( $u=0$ ) ed il reciproco del coefficiente angolare coincidono, rispettivamente, con le stime dei parametri incogniti  $\mu$  e  $\sigma$ :

$$\hat{\mu} = 99.56; \quad \hat{\sigma} = 1/0.097 = 10.31.$$

**9.21** Calcolando due volte il logaritmo della Cdf Gumbel “dei valori massimi” otteniamo:

$$\ln[F_{Z_M}(z)]^{-1} = e^{-\frac{z-a}{b}}; \quad Y(z) = -\ln \ln[F_{Z_M}(z)]^{-1} = \frac{z-a}{b}.$$

Disponendo di un campione casuale  $z_1, z_2, \dots, z_n$ , è possibile stimare la retta di regressione dei punti di ordinata  $Y(z_i) = -\ln \ln[i/(n+1)]^{-1}$  ed ascissa  $z_i$ . I valori  $z_0$  e  $z_1$  corrispondenti a  $Y=0$  ed  $Y=1$  sono tali che  $\hat{a} = z_0$  e  $\hat{b} = z_1 - \hat{a}$ .

**9.22** Dalla relazione:

$$Y(x) = \ln[1 - F_X(x)]^{-1} = \lambda x$$

si deduce che il coefficiente angolare coincide con la stima  $\hat{\lambda} = 10.25$  del parametro incognito  $\lambda$ .

**9.23** Uno stimatore è detto più efficiente di un altro se presenta un errore quadratico medio minore. Poiché:

$$E\{[T_n - \mu]^2\} = \text{Var}\{T_n\} + (E\{T_n\} - \mu)^2 = \text{Var}\left\{\frac{x_{(1)} + x_{(n)}}{2}\right\} + 0 = \frac{\sigma^2}{2};$$

$$E\{[\bar{X} - \mu]^2\} = \text{Var}\{\bar{X}\} + (E\{\bar{X}\} - \mu)^2 = \frac{\sigma^2}{n} + 0 = \frac{\sigma^2}{n}$$

per  $n > 2$  lo stimatore  $\bar{X}$  risulta sempre più efficiente dello stimatore  $T_n$ . Per  $n=2$  i due stimatori sono, ovviamente, equivalenti in quanto  $\bar{X}$  coincide con  $T_n$ . È utile notare che entrambi gli stimatori sono corretti e, quindi, hanno errore quadratico medio coincidente con la varianza.

**9.24** Sfruttando il risultato ottenuto nell'Esempio 9.7 per lo stimatore del parametro  $\lambda$ , si ha:

$$\hat{\lambda} = \frac{9}{2+1+8+4+3+9+6+1+4+12} = 0.18.$$

Indicando con  $X$  la v.a. ritardo, si ha:

$$\Pr\{X > 10\} = e^{-0.18 \times 10} = 0.165.$$

**9.25** Sfruttando il risultato ottenuto nel problema precedente è possibile, per la proprietà d'invarianza dello stimatore di massima verosimiglianza, calcolare una stima puntuale del tempo medio di attesa:  $\bar{x} = 1/\hat{\lambda} = 5.56$  minuti. Utilizzando la funzione

**600** Soluzioni dei problemi proposti

ancillare  $Z_n = 2\lambda n\bar{X}$  otteniamo l'intervallo di confidenza del tempo medio di attesa:

$$\Pr\left\{\frac{2n\bar{x}}{\kappa_2} < \frac{1}{\lambda} < \frac{2n\bar{x}}{\kappa_1}\right\} = 1 - \alpha.$$

Fissando  $1 - \alpha = 0.90$ , individuiamo i percentili  $\kappa_1$  e  $\kappa_2$  della v.a.  $\chi^2$  con  $\nu = 2 \times n = 2 \times 10 = 20$  gradi di libertà:

$$F_K(\kappa_1) = 1 - F_K(\kappa_2) = 0.05; \quad \kappa_1 = 10.85; \quad \kappa_2 = 31.41$$

ed otteniamo l'intervallo di confidenza del tempo medio di attesa (3.54, 10.25).

**9.26** In generale, per un assegnato livello di confidenza  $1 - \alpha$ , l'intervallo di confidenza della media con  $\sigma$  incognita:

$$\left(\bar{x} - t_2 s / \sqrt{n}, \bar{x} + t_1 s / \sqrt{n}\right)$$

risulta sempre più ampio dell'intervallo di confidenza con  $\sigma$  nota:

$$\left(\bar{x} - u_2 \sigma / \sqrt{n}, \bar{x} + u_1 \sigma / \sqrt{n}\right)$$

in quanto, a parità di valori numerici di  $\sigma$  ed  $s$ , i percentili  $t_1$  e  $t_2$  della v.a. di Student con  $\nu = n - 1$  gradi di libertà risultano sempre maggiori dei corrispondenti percentili  $u_1$  e  $u_2$  della v.a. Normale standard. I percentili coincidono solo quando  $n \rightarrow \infty$ , ossia quando l'ampiezza del campione diventa tanto grande che la varianza campionaria coincide con la varianza della popolazione.

In termini numerici, risulta:

$$\begin{aligned} \bar{x} &= 9.74, \quad s^2 = 3.66; \\ \nu &= 9, \quad F_T(t_1) = 1 - F_T(t_2) = 0.025, \quad t_1 = -2.262, \quad t_2 = 2.262; \\ F_U(u_1) &= 1 - F_U(u_2) = 0.025, \quad u_1 = -1.96, \quad u_2 = 1.96. \end{aligned}$$

Da cui otteniamo gli intervalli di confidenza con  $\sigma$  incognita (8.37, 11.10) e con  $\sigma$  nota (8.55, 10.92). Euristicamente, possiamo affermare che il primo intervallo è più ampio del secondo perché essendo la varianza incognita l'incertezza è maggiore.

**9.27** Partendo dal presupposto che la media  $\mu$  della popolazione debba essere 99 volte su 100 contenuta in un intervallo del tipo:

$$\bar{x} - u_2 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + u_1 \frac{\sigma}{\sqrt{n}}$$

con  $u_1 = u_2 = u_{\alpha/2} = 2.576$  ( $\alpha = 0.01$ ) si può imporre che  $|\bar{x} - \mu| \leq \varepsilon$ . Mediante la disuguaglianza  $u_{\alpha/2} \sigma / \sqrt{n} \leq \varepsilon$ , si ricava  $n \geq (u_{\alpha/2} \sigma / \varepsilon)^2$  ed essendo  $\varepsilon = 0.1$  e  $\sigma = 0.5$  si ha  $n \geq 165.9$ , ossia  $n \geq 166$ .

**9.28** Le stime campionarie puntuali di media e varianza sono:  $\bar{x} = 7.96$ ,  $s^2 = 0.12$ . Se il campione è casuale ed è estratto da una popolazione gaussiana, mediante l'utilizzo della funzione ancillare  $K = (n-1)S^2 / \sigma^2$  otteniamo l'intervallo di confidenza della varianza:

$$\Pr \left\{ \frac{(n-1)s^2}{\kappa_2} < \sigma^2 < \frac{(n-1)s^2}{\kappa_1} \right\} = 1 - \alpha.$$

Fissando  $1 - \alpha = 0.90$ , individuamo i percentili  $\kappa_1$  e  $\kappa_2$  della v.a.  $\chi^2$  con  $n-1=7$  gradi di libertà:

$$F_K(\kappa_1) = 1 - F_K(\kappa_2) = 0.05; \quad \kappa_1 = 2.17; \quad \kappa_2 = 14.07$$

ed otteniamo l'intervallo di confidenza della varianza (0.06, 0.38) e dello scarto tipo (0.24, 0.61). L'ipotesi che la popolazione sia gaussiana è importante. Infatti, se fosse violata, la v.a.  $(n-1)S^2/\sigma^2$  potrebbe avere una Cdf molto diversa dalla  $\chi^2$ , anche se  $n$  fosse molto grande.

**9.29** Le stime campionarie puntuali di media e varianza per i gruppi A e B sono:  $\bar{x}_A = 90.75$ ,  $s_A^2 = 8.54$ ,  $\bar{x}_B = 88.14$ ,  $s_B^2 = 9.99$ . Se i campioni sono casuali e sono estratti da popolazioni gaussiane, mediante l'utilizzo della funzione ancillare  $Z = (S_A^2/S_B^2)(\sigma_B^2/\sigma_A^2)$  otteniamo l'intervallo di confidenza del rapporto di varianze ( $\sigma_B^2/\sigma_A^2$ ):

$$\Pr \left\{ z_1 \frac{s_A^2}{s_B^2} < \frac{\sigma_B^2}{\sigma_A^2} < z_2 \frac{s_A^2}{s_B^2} \right\} = 1 - \alpha.$$

Fissando  $1 - \alpha = 0.90$ , individuamo i percentili  $z_1$  e  $z_2$  della v.a.  $Z$  di Fisher con  $n_A - 1 = 9$  e  $n_B - 1 = 7$  gradi di libertà:

$$F_Z(z_1) = 1 - F_Z(z_2) = 0.05; \quad z_1 = 1/3.29 = 0.303; \quad z_2 = 3.68$$

ed otteniamo l'intervallo di confidenza del rapporto di varianze ( $\sigma_B^2/\sigma_A^2$ ) (0.36, 4.31). Vale la pena sottolineare che se la popolazione non fosse gaussiana le v.a.  $(n_A - 1)S_A^2/\sigma_A^2$  e  $(n_B - 1)S_B^2/\sigma_B^2$  potrebbero avere una Cdf molto diversa dalla  $\chi^2$ , anche se  $n$  fosse molto grande, per cui anche la v.a.  $(S_A^2/S_B^2)(\sigma_B^2/\sigma_A^2)$  potrebbe avere una Cdf molto diversa da quella di una  $Z$  di Fisher.

**9.30** La funzione di verosimiglianza ed il suo logaritmo naturale sono:

$$\begin{aligned} L(\mu; y_1, y_2, \dots, y_{50}) &= (\Pr\{0 \text{ difetti}|\mu\})^{32} (\Pr\{1 \text{ difetti}|\mu\})^{12} (\Pr\{2 \text{ difetti}|\mu\})^6 = \\ &= \left( \frac{\mu^0}{0!} e^{-\mu} \right)^{32} \left( \frac{\mu^1}{1!} e^{-\mu} \right)^{12} \left( \frac{\mu^2}{2} e^{-\mu} \right)^6; \end{aligned}$$

$$\ln[L(\mu; y_1, y_2, \dots, y_{50})] = -32\mu - 12\mu + 12\ln \mu - 6\mu + 12\ln \mu - 6\ln 2$$

da cui:

$$\frac{\partial \ln(L)}{\partial \mu} = -32 - 12 + \frac{12}{\mu} - 6 + \frac{12}{\mu} = 0; \quad \hat{\mu} = \frac{12 + 2 \times 6}{32 + 12 + 6} = \frac{24}{50} = 0.48.$$

Sfruttando il metodo utilizzato nell'Esempio 9.16, stimiamo, al livello di confidenza  $1 - \alpha = 0.95$ , i due intervalli (0.38, 0.66) e (0.29, 0.67). Il primo intervallo è stato

**602 Soluzioni dei problemi proposti**


---

ottenuto ritenendo la statistica  $\hat{\mu}$  distribuita come una v.a. Normale di media  $\mu$  e varianza  $\mu/n$ ; il secondo intervallo è stato ottenuto adottando l'approssimazione consistente nel sostituire il valore stimato  $\hat{\mu} = 24/50 = 0.48$  a  $\mu$  nell'espressione della varianza.

**Capitolo 10**

**10.9** Alla registrazione di 4 risultati  $\oplus$ , sul totale di 6, corrisponde una coda pari ad  $\alpha = 0.34$ . Pertanto solo accettando un rischio di errore di I specie superiore al 34% potremmo ritenere significativo il test.

**10.10** Confrontando le frequenze osservate con quelle attese (in parentesi), il test risulta non significativo perché la statistica  $X^2$  è uguale a 5.36 ed è inferiore anche al 95° percentile, 5.99, della distribuzione  $\chi^2$  con  $(3-1) \times (2-1) = 2$  gradi di libertà:

	a	b	c	
I)	20 (26.67)	60 (53.33)	80 (80)	160
II)	30 (23.33)	40 (46.67)	70 (70)	140
	50	100	150	300

**10.11** Il test di Fisher risulta negativo, avendo l'evento osservato (caratterizzato da 2 risultati “-” per il II campione) ed i due più estremi una probabilità complessiva elevata, uguale a:  $0.137 + 0.020 + 0.001 = 0.158$ .

**10.12** L'intervallo di confidenza, al livello  $1 - \alpha = 0.90$ , dell'inquinamento acustico medio, è:

$$\bar{x} \pm t_{0.95} \frac{s}{\sqrt{n}} = 95.56 \pm 5.19;$$

$$90.37 < \mu < 100.75$$

per cui il test risulta significativo risultando il valore “storico” di 90 dB esterno a tale intervallo. Se si ammette che il livello di rumore non sarebbe potuto mai essere diminuito, il livello di significatività non è 0.90 bensì 0.95 (rimanendo esclusa la coda inferiore di rigetto). Ovviamente la statistica di Student:

$$\frac{\bar{x} - 90}{s/\sqrt{n}} = 1.99 > t_{0.95} = 1.86$$

non può che confermare questo risultato.

**10.13** Risulta che:

$$(n-1) \frac{s^2}{\sigma^2} = 41.40 > \kappa_{0.95} = 28.87$$

pertanto il test è significativo e, quindi, l'ipotesi  $\mathcal{H}_0 = \{\sigma^2 = 0.00010\}$  è rigettata.

**10.14** Indicando col simbolo  $\oplus$  la prevalenza chilometrica dell'autovettura appartenente al secondo gruppo, i risultati possono essere "ricodificati" come segue:

$\oplus \quad \ominus \quad \oplus \quad \oplus \quad \oplus \quad \oplus \quad \ominus \quad \oplus \quad \oplus \quad \oplus$

La Cdf Binomiale con  $p=0.5$  ed  $n=10$  presenta in corrispondenza del valore 2 una coda pari a 0.055. Pertanto, accettando un livello di rischio di I specie di poco superiore a 0.05, il test risulta significativo.

**10.15** La distribuzione di campionamento Binomiale ( $p=1/4$ ,  $n=4$ ) presenta una probabilità  $\alpha=(1/4)^4=0.004$ . La pmf binomiale ( $p=3/4$ ,  $n=4$ ), corrispondente all'ipotesi alternativa, ha una massa di probabilità pari a  $\beta=1-(3/4)^4=0.68$  che grava sull'intervallo di accettazione.

**10.16** Impiegando la statistica  $Z$  di Fisher, con  $\nu_1=12$  e  $\nu_2=14$  gradi di libertà, il test risulta non significativo. Infatti abbiamo:

$$z = \frac{s_1^2}{s_2^2} = 1.42 < z_{0.95} = 2.53.$$

**10.17** La statistica (10.11) che si ottiene con le frequenze osservate, (6, 10, 1, 2, 1) rispettivamente per i cinque valori (0, 1, 2, 3, 4), e le corrispondenti frequenze attese  $20 P_Y(y_i)$ , (4.97, 7.42, 4.99, 1.99, 0.52) essendo:

$$P_Y(y_i) = \binom{10}{y_i} 0.13^{y_i} (1-0.13)^{10-y_i}$$

fornisce il valore  $\kappa=4.74$  che è minore del 95° percentile della v.a.  $\chi^2$  con 4 gradi di libertà (pari a 9.488). Pertanto, al livello di significatività  $1-\alpha=0.95$ , non possiamo rigettare l'ipotesi di Cdf Binomiale.

**10.18** La statistica (10.11) che si ottiene con le frequenze osservate (11, 12, 6, 0, 1), rispettivamente per i cinque valori (0, 1, 2, 3, 4), e le corrispondenti frequenze attese  $20 P_Y(y_i)$ , (11.04, 11.04, 5.52, 1.84, 0.460) essendo:

$$P_Y(y_i) = \frac{1^{y_i}}{y_i!} e^{-1}$$

fornisce il valore  $\kappa=2.60$  che è minore del 95° percentile della  $\chi^2$  con 4 gradi di libertà (pari a 9.488). Pertanto, non possiamo rigettare l'ipotesi di Cdf di Poisson al livello di significatività  $1-\alpha=0.95$ .

**10.19** La statistica (10.14) fornisce il valore 0.113 che è minore del valore critico al livello di significatività  $1-\alpha=0.95$ , pari a  $D_n=1.36/\sqrt{n}=0.192$ . Pertanto non possiamo rigettare l'ipotesi di Cdf Gaussiana.

**10.20** Il rischio di I specie,  $\alpha$ , è  $2.1 \times 10^{-7}$  perché tale è l'area sottesa dalla v.a. Nor-

**604** Soluzioni dei problemi proposti

male di media 50 e scarto tipo  $5/\sqrt{10}=1.58$  nella regione di rigetto ( $\bar{X} \leq 42$ ). Il rischio di II specie,  $\beta$ , è 0.103 perché tale è l'area sottesa dalla v.a. Normale di media 40 e scarto tipo 1.58 nella regione di accettazione ( $\bar{X} > 42$ ).

**10.21** Affermare che la probabilità di falso allarme sia più bassa consiste nel rigettare l'ipotesi nulla, correndo un rischio  $\alpha$  pari a:

$$\alpha = \sum_{i=0}^2 \binom{16}{i} 0.3^i (1-0.3)^{16-i} = 0.09936.$$

**10.22** Per una generica ampiezza del campione,  $n$ , l'intervallo di accettazione è:

$$\left( 25 - 1.645 \frac{1.5}{\sqrt{n}}; \quad 25 + 1.645 \frac{1.5}{\sqrt{n}} \right).$$

Il valore della potenza nei confronti di  $\mathcal{H}_1 = \{\mu = 26.5\}$ ,  $\Psi(\mathcal{H}_1)$ , è dato dall'area sottesa alla pdf Gaussiana, di media 26.5 e scarto tipo  $\sigma/\sqrt{n}$ , all'esterno dell'intervallo di accettazione. Il valore minimo di  $n$  è quello che soddisfa la seguente disuguaglianza:

$$\Psi(\mathcal{H}_1) = F_U \left( \frac{25 - 1.645 \frac{1.5}{\sqrt{n}} - 26.5}{\frac{1.5}{\sqrt{n}}} \right) + \left[ 1 - F_U \left( \frac{25 + 1.645 \frac{1.5}{\sqrt{n}} - 26.5}{\frac{1.5}{\sqrt{n}}} \right) \right] \geq 0.99.$$

Risolvendo per tentativi si ricava  $n=16$ ; infatti, per  $n=15$  e per  $n=16$  otteniamo, rispettivamente,  $\Psi(\mathcal{H}_1)=0.987$  e  $\Psi(\mathcal{H}_1)=0.991$ .

**10.23** Il test è direzionale e la regione di rigetto è costituita dall'insieme di valori  $\{8, 9, 10\}$ . Il valore della potenza nei confronti di  $\mathcal{H}_1 = \{p=0.8\}$ ,  $\Psi(\mathcal{H}_1)$ , è dato dalla massa di probabilità della pmf binomiale, di parametri  $n=10$  e  $p=0.8$ , che grava sulla regione di rigetto:

$$\begin{aligned} \Psi(\mathcal{H}_1) &= \binom{10}{8} 0.8^8 (1-0.8)^{10-8} + \binom{10}{9} 0.8^9 (1-0.8)^{10-9} + \\ &+ \binom{10}{10} 0.8^{10} (1-0.8)^{10-10} = 0.6778. \end{aligned}$$

**10.24** Utilizzando la statistica di Student, troviamo il seguente valore di  $t$ :

$$t = \frac{120.87 - 120.00}{\frac{3.15}{\sqrt{15}}} = 1.070$$

cui corrisponde, in corrispondenza di 14 gradi di libertà, una significatività del test pari a  $1 - \alpha = 0.849$ , avendo fissato l'ipotesi alternativa unidirezionale.

**10.25** Con riferimento alla distribuzione Ipergeometrica con  $N=100$ ,  $D=25$  e  $n=20$ , il rischio di II specie,  $\beta$ , è:



$$\beta = \sum_{i=0}^1 \frac{\binom{25}{i} \binom{75}{20-i}}{\binom{100}{20}} = 0.0015 + 0.0134 = 0.0149.$$

**10.26** Adottando classi di ampiezza  $\Delta x = 10$ , individuate dai valori centrali (45, 55, 65, 75, 85, 95, 105), otteniamo il vettore (4, 10, 20, 29, 24, 8, 5) di frequenze osservate,  $O_i$ , cui corrisponde il seguente vettore di frequenze attese,  $E_i = 100 \times P_i$ , (3.80, 11.09, 21.08, 26.11, 21.08, 11.09, 3.80), essendo:

$$P_i = F_U \left( \frac{x_i + \Delta x/2 - 75}{15} \right) - F_U \left( \frac{x_i - \Delta x/2 - 75}{15} \right).$$

Con le suddette frequenze, la statistica (10.11) fornisce il valore  $\kappa = 2.14$ , che è minore del 90° percentile della v.a.  $\chi^2$ , con  $\nu = 6$ , pari a 10.64. Pertanto, al livello di significatività  $1 - \alpha = 0.90$ , non possiamo rigettare l'ipotesi di Cdf gaussiana.

**10.27** In questo caso le frequenze attese,  $E_i$ , assumono un valore costante uguale a  $100/7 = 14.29$  e la statistica (10.11) fornisce il valore  $\kappa = 41.53$  che è maggiore del 90° percentile della v.a.  $\chi^2$ , con  $\nu = 6$ , pari a 10.64. Pertanto, il test è significativo ( $p\text{-value} = 2.27 \times 10^{-7}$ ) e quindi l'ipotesi che la Cdf sia di tipo uniforme è rigettata. Ovviamente il risultato era prevedibile in quanto l'istogramma delle densità di frequenza riportato nell'Esempio 6.5 ha una forma molto diversa da quella di una pdf uniforme.

**10.28** L'intervallo di confidenza, (0.36, 4.31), del rapporto di varianze ( $\sigma_B^2 / \sigma_A^2$ ), determinato nel Problema 9.29, include il valore 1, per cui, al livello di confidenza  $1 - \alpha = 0.90$ , non può essere rigettata l'ipotesi di uguaglianza delle varianze. Verificata l'ipotesi di uguaglianza delle varianze, per confrontare le medie dei due campioni possiamo utilizzare la stima *pooled* della varianza campionaria e la statistica di Student. Fissato il livello di significatività  $1 - \alpha = 0.90$  e le ipotesi  $\mathcal{H}_0 = \{\mu_A = \mu_B\}$  ed  $\mathcal{H}_1 = \{\mu_A \neq \mu_B\}$ , abbiamo:

$$t = \frac{90.75 - 88.14}{\sqrt{(8.54 \times 9 + 9.99 \times 7) / 16} \sqrt{1/10 + 1/8}} = 1.82 > t_{0.95} = 1.746; \quad \nu = 16$$

per cui è possibile rigettare l'ipotesi nulla essendo significativa la differenza riscontrata.

**10.29** Adottando classi di ampiezze individuate dai valori degli estremi superiori (120, 240, 360, 480), otteniamo il vettore (22, 12, 7, 6) di frequenze osservate,  $O_i$ , cui corrisponde il vettore (21.20, 11.64, 6.38, 3.51) di frequenze attese,  $E_i = 47 \times P_i$ , essendo:

$$P_i = e^{-0.005 x_{i-1}} - e^{-0.005 x_i}; \quad x_0 = 0.$$

## 606 Soluzioni dei problemi proposti

Con le suddette frequenze, la statistica (10.11) fornisce il valore  $\kappa = 1.88$ , per il quale la Cdf della v.a.  $\chi^2$ , con  $\nu = 3$ , fornisce un  $p\text{-value} = 0.60$ . Conseguentemente  $\mathcal{H}_0$  verrebbe rigettata solo con un rischio di I specie almeno pari ad  $\alpha = 0.60$ .

**10.30** Il rischio di I specie è  $\alpha \cong 0.90$ , essendo  $1 - \alpha \cong 0.10$  l'area sottesa dalla pdf gaussiana con  $\mu = 2$  e  $\sigma = 2/\sqrt{25} = 0.40$ , nella regione di accettazione:

$$\begin{aligned} F_U\left(\frac{2.048-2}{0.40}\right) - F_U\left(\frac{1.952-2}{0.40}\right) &= F_U(0.12) - F_U(-0.12) = \\ &= 1 - 0.4522 - 0.4522 = 0.0956. \end{aligned}$$

## Capitolo 11

**11.9** Non necessariamente. Per rispondere al secondo quesito si può eseguire un test di Student multiplo, valutando gli intervalli di confidenza delle quattro medie, con l'uso della varianza *pooled*, e controllando quale intervallo include la media del primo gruppo:

livello del fattore	osservazioni	media e varianza <i>pooled</i>	intervallo di confidenza della media al livello $1 - \alpha = 0.95$
1	8.99, 9.89, 9.61, 9.31, 9.35	9.43 0.8929	8.60, 10.26
2	11.62, 11.40, 11.11, 10.80, 11.72	11.33 0.8929	10.50, 12.16
3	11.85, 11.33, 11.84, 13.83, 11.05	11.98 0.8929	11.15, 12.81
4	7.57, 11.27, 8.94, 7.89, 9.28	8.99 0.8929	8.16, 9.82

da cui risulta che solo il quarto gruppo non permette di rigettare l'ipotesi di media uguale a quella del primo, con un livello di confidenza di 0.95. Per il calcolo della semi ampiezza degli intervalli di confidenza, si può utilizzare la funzione *CONFIDENCE* di Microsoft Excel che, in questo caso, per  $\alpha = 0.05$ ,  $\sigma = \sqrt{0.8929} = 0.9449$  ed  $n = 5$ , fornisce il valore 0.828.

**11.10** Basta considerare l'identità elementare:

$$\sum_{i=1}^n \sum_{j=1}^b (x_{ij} - \bar{x})^2 = \sum_{i=1}^n \sum_{j=1}^b [(x_{ij} - \bar{x}_{i\bullet} - \bar{x}_{\bullet j} + \bar{x}) + (\bar{x}_{i\bullet} - \bar{x}) + (\bar{x}_{\bullet j} - \bar{x})]^2$$

e tenere conto che sono nulli i tre prodotti incrociati dei tre termini del quadrato del trinomio considerato. Per esempio:

$$\begin{aligned}
 & \sum_{i=1}^n \sum_{j=1}^b \{[(x_{ij} - \bar{x}_{i\bullet}) - (\bar{x}_{\bullet j} - \bar{x})](\bar{x}_{i\bullet} - \bar{x})\} = \sum_{i=1}^n \sum_{j=1}^b [(x_{ij} - \bar{x}_{i\bullet})(\bar{x}_{i\bullet} - \bar{x})] + \\
 & - \sum_{i=1}^n \sum_{j=1}^b [(\bar{x}_{\bullet j} - \bar{x})(\bar{x}_{i\bullet} - \bar{x})] = \sum_{i=1}^n (\bar{x}_{i\bullet} - \bar{x}) \sum_{j=1}^b (x_{ij} - \bar{x}_{i\bullet}) - \sum_{i=1}^n (\bar{x}_{i\bullet} - \bar{x}) \sum_{j=1}^b (\bar{x}_{\bullet j} - \bar{x}) = \\
 & = \sum_{i=1}^n (\bar{x}_{i\bullet} - \bar{x}) \left[ \sum_{j=1}^b x_{ij} - \sum_{j=1}^b \bar{x}_{i\bullet} \right] - \sum_{i=1}^n (\bar{x}_{i\bullet} - \bar{x}) \left[ \sum_{j=1}^b (\bar{x}_{\bullet j} - \bar{x}) \right] = 0.
 \end{aligned}$$

**11.11** La tabella ANOVA completa è:

orig. della varianza	devianza	gradi di libertà	varianza, $S^2$	$z^*$	$p$ -value
fattore A	15.1078	2	7.55391	1.91	0.2613
fattore B	4.58149	2	2.29074	0.58	0.6009
errore	15.7958	4	3.94894		
totale	35.4851	8			

per cui nessuno dei due fattori risulta significativo, anche se fissiamo un livello minimo di significatività  $1 - \alpha = 0.9000$ .

**11.12** Essendo  $1/z^* = 400$ , abbiamo:

$$p = \Pr\{Z_{6,2} > 1/z^*\} = 0.0025$$

per cui il test è significativo al livello  $1 - p = 0.9975$ . In questa condizione è il modello interpretativo a dover essere rigettato o almeno una delle sue ipotesi. L'ipotesi che più plausibilmente non è verificata è quella di normalità dei rilievi sperimentali, visto che — per ogni trattamento — le misure sono state ottenute su tre vetture (differenti anche per il modello). Evidentemente la somma dell'errore sperimentale e dell'effetto dovuto allo specifico modello di autovettura non costituisce una v.a. normale. Pertanto, si dovrebbe depurare la varianza intraclasse dell'aliquota dovuta al fattore autovettura-giorno, come richiesto dal Problema 11.13.

**11.13** La richiesta ANOVA a due vie, riportata nella tabella seguente, indica che l'effetto del fattore autovettura-giorno è significativo — come previsto nella risposta al problema precedente — al livello  $1 - p = 0.9989$ . Permane invece non significativo l'effetto del fattore pilota.

**608** Soluzioni dei problemi proposti

origine della varianza	devianza	gradi di libertà	varianza, $S^2$	$Z^*$	$p$ -value
fattore principale	16.576	2	8.2878	0.05	0.9515
fattore secondario	19036	2	9518.4	57.86	0.0011
errore	658.04	4	164.51		
totale	19711	8			

**11.14** Dal modello di analisi  $x_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$ , i valori attesi cercati sono facilmente dedotti ricordando le ipotesi di  $s$ -indipendenza ed omogeneità delle varianze:

orig. della varianza	valore atteso della varianza, $E\{S^2\}$
fattore principale	$\sigma^2 + \frac{bc}{(n-1)} \sum_{i=1}^n \alpha_i^2$
fattore secondario	$\sigma^2 + \frac{nc}{(b-1)} \sum_{j=1}^b \beta_j^2$
interazione	$\sigma^2 + \frac{c}{(n-1)(b-1)} \sum_{i=1}^n \sum_{j=1}^b (\alpha\beta)_{ij}^2$
errore	$\sigma^2$

**11.15** La produzione complessiva è caratterizzata da una varianza pari a  $\sigma^2 + (1/3) \sum_{i=1}^3 (\mu_i - \mu)^2$  e da una media  $\mu = (\mu_1 + \mu_2 + \mu_3) / 3$ . Infatti in ogni “baricentro”  $\mu_i$  è concentrato  $1/3$  della “massa” totale e ciascuna di queste contribuisce alla varianza complessiva con la somma di due aliquote,  $1/3\sigma^2$  e  $1/3(\mu_i - \mu)^2$ , dovute rispettivamente alla varianza intorno alle singole medie ed alla varianza delle medie parziali intorno a quella complessiva. I valori attesi dell’ANOVA sono quelli riportati nella seguente tabella.

origine della varianza	valore atteso della varianza $E\{S^2\}$
fattore $L$	$\sigma^2 + \frac{m}{2} \sum_{i=1}^3 (\mu_i - \mu)^2$
errore	$\sigma^2$

**11.16** La prima eguaglianza segue dal seguente sviluppo:

$$\begin{aligned}
 \Sigma_t &= \sum_{i=1}^n \sum_{j=1}^m (x_{ij} - \bar{x})^2 = \sum_{i=1}^n \sum_{j=1}^m (x_{ij}^2 - 2\bar{x}x_{ij} + \bar{x}^2) = \\
 &= \sum_{i=1}^n \sum_{j=1}^m x_{ij}^2 - 2\bar{x} \sum_{i=1}^n \sum_{j=1}^m x_{ij} + nm\bar{x}^2 = \sum_{i=1}^n \sum_{j=1}^m x_{ij}^2 - 2\bar{x}(nm\bar{x}) + nm\bar{x}^2 = \\
 &= \sum_{i=1}^n \sum_{j=1}^m x_{ij}^2 - nm\bar{x}^2 = \sum_{i=1}^n \sum_{j=1}^m x_{ij}^2 - \frac{T^2}{nm}.
 \end{aligned}$$

Analogamente, per la seconda eguaglianza si ha:

$$\begin{aligned}
 m\Sigma_f &= m \sum_{i=1}^n (\bar{x}_i - \bar{x})^2 = m \sum_{i=1}^n (\bar{x}_i^2 - 2\bar{x}\bar{x}_i + \bar{x}^2) = \\
 &= m \left( \sum_{i=1}^n \bar{x}_i^2 - 2\bar{x} \sum_{i=1}^n \bar{x}_i + n\bar{x}^2 \right) = m \left( \sum_{i=1}^n \left( \frac{T_i}{m} \right)^2 - 2\bar{x} \sum_{i=1}^n \frac{T_i}{m} + n\bar{x}^2 \right) = \\
 &= \frac{1}{m} \sum_{i=1}^n T_i^2 - 2\bar{x}nm\bar{x} + nm\bar{x}^2 = \frac{1}{m} \sum_{i=1}^n T_i^2 - nm\bar{x}^2 = \frac{1}{m} \sum_{i=1}^n T_i^2 - \frac{T^2}{nm}.
 \end{aligned}$$

La terza eguaglianza segue dalla (11.4).

**11.17** Nel caso di un piano non bilanciato partendo dalla (11.4) e sviluppando i calcoli si ottiene:

origine della varianza	devianza	gradi di libertà	varianza, $S^2$
Fattore (var. interclassi)	$\sum_{i=1}^n m_i (\bar{x}_i - \bar{x})^2$	$n - 1$	$\frac{\sum_{i=1}^n m_i (\bar{x}_i - \bar{x})^2}{n - 1}$
Errore (var. intraclassi)	$\sum_{i=1}^n \sum_{j=1}^{m_i} (x_{ij} - \bar{x}_i)^2$	$\sum_{i=1}^n (m_i - 1)$	$\frac{\sum_{i=1}^n \sum_{j=1}^{m_i} (x_{ij} - \bar{x}_i)^2}{\sum_{i=1}^n (m_i - 1)}$
Totale	$\sum_{i=1}^n \sum_{j=1}^{m_i} (x_{ij} - \bar{x})^2$	$(\sum_{i=1}^n m_i) - 1$	

**11.18** Eseguendo l'analisi della varianza ad una via dei dati sperimentali assegnati, si ottiene la seguente tabella:

**610** Soluzioni dei problemi proposti

origine della varianza	devianza	gradi di libertà	varianza, $S^2$	$z^*$	$z_{0.95}$
Fattore (var. interclassi)	78.44	5	15.69	7.43	3.11
Errore (var. intraclassi)	25.33	12	2.11		
Totale	103.78	17			

Poiché risulta  $z^* > z_{0.95}$  il test è significativo al livello 0.95. L'ipotesi nulla di inefficacia del fattore "tipologia di confetti" va rigettata.

**11.19** Essendo:

$$\hat{\mu}_i = \bar{x}_{i\bullet} = \frac{\sum_{j=1}^m x_{ij}}{m}; \quad \hat{\mu} = \bar{x} = \frac{\sum_{i=1}^n \bar{x}_{i\bullet}}{n} = \frac{\sum_{i=1}^n \sum_{j=1}^m x_{ij}}{nm}; \quad \hat{\alpha}_i = \hat{\mu}_i - \hat{\mu} = \bar{x}_{i\bullet} - \bar{x}$$

risulta:

$$\hat{\alpha}_1 = -1.22; \quad \hat{\alpha}_2 = 1.11; \quad \hat{\alpha}_3 = -2.55; \\ \hat{\alpha}_4 = 2.44; \quad \hat{\alpha}_5 = -2.22; \quad \hat{\alpha}_6 = 2.44.$$

Si osservi che (§11.3):

$$\sum_{i=1}^n \hat{\alpha}_i = 0.$$

**11.20** In questo caso, essendo il piano non bilanciato (cioè il numero di prove  $m_i$  varia al variare del livello del fattore) si ha:

$$\hat{\mu}_i = \bar{x}_{i\bullet} = \frac{\sum_{j=1}^{m_i} x_{ij}}{m_i}; \quad \hat{\mu} = \bar{x} = \frac{\sum_{i=1}^n \sum_{j=1}^{m_i} x_{ij}}{\sum_{i=1}^n m_i} = \frac{\sum_{i=1}^n m_i \bar{x}_{i\bullet}}{\sum_{i=1}^n m_i}; \quad \hat{\alpha}_i = \hat{\mu}_i - \hat{\mu} = \bar{x}_{i\bullet} - \bar{x}.$$

Risulta pertanto:

$$\hat{\mu} = \bar{x} = \frac{\sum_{i=1}^3 m_i \bar{x}_{i\bullet}}{\sum_{i=1}^3 m_i} = \frac{10 \times 1 + 20 \times 2 + 30 \times 3}{10 + 20 + 30} = 2.33; \\ \hat{\alpha}_1 = -1.333, \quad \hat{\alpha}_2 = -0.333, \quad \hat{\alpha}_3 = 0.667.$$

Si osservi che in questo caso, trattandosi di piano non bilanciato, la condizione che devono soddisfare le  $\hat{\alpha}_i$  non è:

$$\sum_{i=1}^n \hat{\alpha}_i = 0$$

ma la seguente:

$$\sum_{i=1}^n m_i \hat{\alpha}_i = 0.$$

**11.21** Eseguendo l'analisi della varianza ad una via dei dati sperimentali disponibili, si ottiene la seguente tabella:

origine della varianza	devianza	gradi di libertà	varianza, $S^2$	$z^*$	$z_{0.95}$
Fattore (var. interclassi)	0.0198	4	0.0094	5.99	3.06
Errore (var. intraclassi)	0.0124	15	0.0008		
Totale	0.0322	19			

Risultando  $z^* > z_{0.95}$  il test è significativo al livello 0.95. L'ipotesi nulla di inefficacia del fattore "fonderia" va rigettata.

**11.22** Eseguendo l'analisi della varianza a due vie con interazione si ottiene la seguente tabella:

origine della varianza	devianza	gradi di libertà	varianza, $S^2$	$z^*$	$z_{0.95}$
Modalità di prelievo (B)	$3.47 \times 10^{-6}$	2	$1.733 \times 10^{-6}$	0.131	3.68
Zona di prelievo (A)	$3.833 \times 10^{-5}$	4	$9.583 \times 10^{-6}$	0.726	3.05
Interazione (AB)	$7.687 \times 10^{-5}$	8	$9.608 \times 10^{-6}$	0.728	2.64
Errore	$1.98 \times 10^{-4}$	15	$1.320 \times 10^{-5}$		
Totale	$3.17 \times 10^{-4}$	29			

Per entrambi i fattori, risulta  $z^* < z_{0.95}$ , per cui non è possibile rigettare, al livello di significatività 0.95, l'ipotesi nulla di inefficacia degli stessi fattori. Alla stessa conclusione si giunge anche per l'interazione.

**11.23** Eseguendo l'analisi della varianza a due vie dei dati sperimentali assegnati, si ottiene la seguente tabella:

**612** Soluzioni dei problemi proposti

origine della varianza	devianza	gradi di libertà	varianza, $S^2$	$z^*$	$z_{0.99}$	$p$ -value
Fattore $A$ (sito di estrazione)	$3.393 \times 10^{-3}$	4	$8.48 \times 10^{-4}$	3.096	4.18	0.033
Fattore $B$ (profondità di estrazione)	$5.72 \times 10^{-6}$	4	$1.43 \times 10^{-6}$	0.005	4.18	0.999
Interazione ( $AB$ )	$1.197 \times 10^{-4}$	16	$7.48 \times 10^{-6}$	0.027	2.81	0.999
Errore	$6.85 \times 10^{-3}$	25	$2.73 \times 10^{-4}$			
Totale	0.01034	49				

Pertanto, al livello di significatività 0.99, il test non permette di rigettare l'ipotesi di inefficacia dei fattori e della loro interazione.

**11.24** Eseguendo l'analisi della varianza ad una via si ottiene la seguente tabella:

origine della varianza	devianza	gradi di libertà	varianza, $S^2$	$z^*$	$z_{0.99}$	$p$ -value
Fattore $A$ (sito di estrazione)	$3.393 \times 10^{-3}$	4	$8.48 \times 10^{-4}$	5.47	3.77	0.001
Errore	$6.98 \times 10^{-3}$	45	$1.55 \times 10^{-4}$			
Totale	$1.04 \times 10^{-2}$	49				

Risultando  $z^* > z_{0.99}$  è possibile rigettare l'ipotesi di inefficacia del fattore  $A$  al livello 0.99. Si vede che, nel caso in cui  $B$  ed  $AB$  non producono effetti significativi, inglobando le quantità  $\Sigma_B$  e  $\Sigma_I$  nell'errore, si può riuscire ad evidenziare l'efficacia di un trattamento che in prima analisi (problema 11.23) era stato considerato inefficace. Per ulteriori considerazioni al riguardo, si veda anche il problema 11.26.

**11.25** In accordo con la (11.14) si ha:

$$\begin{aligned}
 (\widehat{\alpha\beta})_{12} &= \bar{x}_{12\cdot} - \bar{x}_{\cdot 2\cdot} - \bar{x}_{1\cdot\cdot} + \bar{x} = \\
 &= 2.50 - 4.27 - 3.44 + 4.48 = -0.73.
 \end{aligned}$$

**11.26** In generale la potenza del test cresce al crescere del numero di gradi di libertà dell'errore. Pertanto, se si rileva che uno o più fattori non sono efficaci, si può effettuare nuovamente il test inglobando il loro effetto nell'errore. Così operando si ottiene la seguente tabella:



origine della varianza	devian- za	gradi di libertà	varian- za, $s^2$	$z^*$	$p$ -value
Commissione	60.04	2	30.02	2.96	0.0628
Errore	426.27	42	10.15		
Totale	486.31	44			

Il numero di gradi di libertà dell'errore passa da 30 a 42. Si vede (cfr. Tabella 11.25) che operando questa nuova scomposizione il valore di  $p$ -value si riduce da circa 0.10 a circa 0.06.

**11.27** Il modello di tabella ANOVA usata nell'Esempio 11.7 è quello a due vie con interazione riportato in Tabella 11.23 cui corrisponde un insieme di dati raccolti secondo lo schema di Tabella 11.24. Invece, trascurando la presenza del fattore  $B$ , i dati di Tabella 11.24 sarebbero stati riaggregati nella maniera seguente:

Livello del fattore	osservazioni
$A_1$	24, 27, 25, 30, 26, 30, 23, 25, 28, 29, 29, 29, 25, 21, 26
$A_2$	23, 26, 26, 18, 30, 22, 25, 22, 25, 23, 21, 24, 26, 21, 23
$A_3$	24, 26, 30, 24, 26, 21, 30, 26, 30, 24, 30, 26, 18, 20, 30

e il modello di tabella ANOVA da utilizzare sarebbe stato quello ad una via riportata in Tabella 11.11. Non è difficile verificare che risultano valide le seguenti uguaglianze:

$$m \sum_{i=1}^n (\bar{x}_i - \bar{x})^2 = bc \sum_{i=1}^n (\bar{x}_{i\bullet\bullet} - \bar{x})^2$$

$$\sum_{i=1}^n \sum_{j=1}^m (x_{ij} - \bar{x})^2 = \sum_{i=1}^n \sum_{j=1}^b \sum_{k=1}^c (x_{ijk} - \bar{x})^2$$

(essendo i primi membri tratti dalla Tabella 11.11 ed i secondi dalla Tabella 11.23) che indicano che in questo caso la  $\Sigma_e$  calcolata usando la scomposizione dell'ANOVA ad una via coincide con la somma delle quantità  $nc\Sigma_b$ ,  $c\Sigma_i$  e  $\Sigma_e$  relative alla scomposizione dell'ANOVA a due vie con interazione.

**11.28** Il piano è non bilanciato, pertanto si ha:

**614** Soluzioni dei problemi proposti

origine della varianza	devianza	gradi di libertà	varianza, $S^2$	$z^*$	$z_{0.95}$
Fattore (var. interclas- si)	0.192	1	0.192	3.77	4.75
Errore (var. intraclas- si)	0.610	12	0.051		
Totale	0.802				

Il test è non significativo. Non si può rigettare l'ipotesi che il livello medio di creatività sia lo stesso.

**11.29** Dallo svolgimento dell'Esempio 10.5 risulta:

$$\frac{\bar{x}_2 - \bar{x}_1}{s\sqrt{1/n + 1/m}} = 1.94$$

da cui si ottiene:

$$t_{0.025} = -2.179 \quad \text{e} \quad t_{0.975} = 2.179.$$

Pertanto, risultando 1.94 interno all'intervallo si deduce che non è possibile rigettare l'ipotesi nulla.

**11.30** Il test di Student e l'ANOVA nel caso di un fattore a due livelli e test bilaterali sono totalmente equivalenti. Infatti, la variabile  $Z^*$  usata per l'ANOVA:

$$\frac{\sum_{i=1}^n m_i (\bar{x}_i - \bar{x})^2}{n-1} \left( \frac{\sum_{i=1}^n \sum_{j=1}^{m_i} (x_{ij} - \bar{x}_i)^2}{\sum_{i=1}^n (m_i - 1)} \right)^{-1}$$

nel caso in esame ( $n = 2$ ) può essere particolarizzata come segue:

$$\begin{aligned} & \frac{\sum_{i=1}^2 m_i [\bar{x}_i - (m_1 \bar{x}_1 + m_2 \bar{x}_2) / (m_1 + m_2)]^2}{2-1} \frac{m_1 m_2^2 (\bar{x}_1 - \bar{x}_2)^2 + m_2 m_1^2 (\bar{x}_1 - \bar{x}_2)^2}{(m_1 + m_2)^2} \\ &= \frac{\sum_{i=1}^2 \sum_{j=1}^{m_i} (x_{ij} - \bar{x}_i)^2}{m_1 + m_2 - 2} = \frac{\sum_{i=1}^2 \sum_{j=1}^{m_i} (x_{ij} - \bar{x}_i)^2}{m_1 + m_2 - 2} = \\ &= \frac{m_1 m_2 (m_1 + m_2) (\bar{x}_1 - \bar{x}_2)^2}{(m_1 + m_2)^2} = \frac{m_1 m_2}{m_1 + m_2} (\bar{x}_1 - \bar{x}_2)^2 \\ &= \frac{\sum_{i=1}^2 \sum_{j=1}^{m_i} (x_{ij} - \bar{x}_i)^2}{m_1 + m_2 - 2} = \frac{(m_1 - 1) s_1^2 + (m_2 - 1) s_2^2}{m_1 + m_2 - 2} \end{aligned}$$

che coincide col quadrato della variabile  $T$  usata per il test di Student:

$$\frac{\bar{x}_2 - \bar{x}_1}{s\sqrt{1/m_1 + 1/m_2}}; \quad s = \sqrt{\frac{(m_1 - 1) s_1^2 + (m_2 - 1) s_2^2}{m_1 + m_2 - 2}}$$

## Capitolo 12

**12.9** Trattandosi del caso in cui la prestazione ideale è  $x_0 = 0$  (LB), il valore atteso della perdita è:

$$E\{L(x)\} = \frac{C}{\Delta^2} (\mu^2 + \sigma^2) = \frac{15,00}{0,08^2} (0,04^2 + 0,004^2) = 3,79 \text{ euro.}$$

**12.10** I valori assunti dalla funzione  $SN$  suggeriscono di adottare la configurazione contraddistinta col n° 2:

combinazione $i$	$SN = -10 \log_{10} \left[ \sum_{i=1}^n (x_i - x_0)^2 / n \right]$
1	-15.2
2*	-1.6
3	-16.4

**12.11** I valori assunti dalla funzione  $SN$ , valida per il caso LB, suggeriscono di adottare la configurazione n° 1:

combinazione $i$	$SN = -10 \log_{10} \left( \sum_i x_i^2 / n \right)$
1*	-19.6
2	-23.5
3	-26.7

**12.12** In corrispondenza della configurazione n° 3 è massimo il valore assunto dalla funzione  $SN$  valida per il caso HB:

combinazione $i$	$SN = -10 \log_{10} \left[ \sum_i (1/x_i^2) / n \right]$
1	19.2
2	23.4
3*	26.7

**12.13** Il numero minimo di trattamenti coincide con il numero totale di gradi di libertà degli effetti dei fattori e delle loro interazioni che dobbiamo stimare:

$$1 + 5 \times (2 - 1) + 10 \times (2 - 1) \times (2 - 1) = 16.$$

**12.14** Il piano  $2^{5-2}$  non è ortogonale in quanto non sono nulli tutti i prodotti interni dei vettori dei codici (-1 e +1) tra tutte le combinazioni di colonne. Infatti, per le coppie di colonne  $BD$ ,  $CD$  e  $DE$  il prodotto è uguale a -4.

**12.15** Per  $k = 4$ , il piano fattoriale  $2^k$  può essere esteso, nel rispetto dei vincoli dell'ortogonalità, ad un numero di fattori pari a:

**626** Soluzioni dei problemi proposti

$$p = \sum_{i=2}^4 \binom{4}{i} = 11.$$

Pertanto, un piano fattoriale  $2^4$  può essere impiegato per un numero di fattori al più uguale a  $k' = k + p = 4 + 11 = 15$ , realizzando un piano  $2^{15-11}$ .

**12.16** In base alle regole che definiscono l'operatore  $I$  possiamo scrivere:  $EE = I = EABCD$ . Pertanto, dall'equazione simbolica  $I = ABCDE$  ricaviamo che il piano è di risoluzione V in quanto il numero di lettere al secondo membro dell'equazione simbolica è uguale a 5.

**12.17** Dall'equazione simbolica  $I = ABE$ , poiché  $E = EI = EABE = AB$  si deduce che i codici ( $-1$  e  $+1$ ) della colonna corrispondente al fattore  $E$  sono stati ottenuti moltiplicando quelli delle colonne corrispondenti ai fattori  $A$  e  $B$ .

**12.18** Il piano ha risoluzione III in quanto il numero di lettere al secondo membro dell'equazione simbolica è 3 ( $I = ABE$ ). Inoltre, dall'elenco dei confondimenti relativi a ciascun effetto principale:

$$A = AI = AABE = BE$$

$$B = BI = BABE = AE$$

$$C = CI = ABCE$$

$$D = DI = ABDE$$

$$E = EI = EABE = AB$$

si deduce che ciascun effetto principale è confuso con l'interazione di almeno due fattori ma non con altri fattori principali.

**12.19** A partire dai generatori del piano è possibile dedurre:

$$I = FF = ABEF$$

$$I = GG = ACDFG$$

da cui:

$$I = ABEFACDFG = BCDEG$$

e quindi l'equazione simbolica del piano è:

$$I = ABEF = ACDFG = BCDEG.$$

Il piano è di risoluzione IV in quanto il numero *minimo* di lettere che compaiono nell'equazione simbolica è 4. Inoltre, nessun effetto principale è confuso con altri fattori principali o con l'interazione di due fattori principali, mentre alcune interazioni di due fattori possono essere confuse con quelle di altri due.

**12.20** L'elenco dei confondimenti richiesti si può ricavare dall'equazione simbolica del piano  $I = ABD = ACE = BCDE$ :

$$A = BD = CE = ABCDE; \quad B = AD = CDE = ABCE; \quad C = AE = BDE = ABCD;$$

$$D = AB = BCE = ACDE; \quad E = AC = BCD = ABDE; \quad BC = DE = ABE = ACD;$$

$$CD = BE = ADE = ABC.$$

**12.21** L'elenco dei confondimenti che scaturiscono dalla riduzione è:

<i>a)</i>	<i>b)</i>
$I = DEF = ABCF = ABCDE$	$I = ABCE = ADEF = BCDF$
$A = BCF = ADEF = BCDE$	$A = BCE = DEF = ABCDF$
$B = ACF = ACDE = BDEF$	$B = ACE = CDF = ABDEF$
$C = ABF = ABDE = CDEF$	$C = ABE = BDF = ACDEF$
$D = EF = ABCE = ABCDF$	$D = AEF = BCF = ABCDE$
$E = DF = ABCD = ABCEF$	$E = ABC = ADF = BCDEF$
$AB = CF = CDE = ABDEF$	$F = ADE = BCD = ABCEF$
$AC = BF = BDE = ACDEF$	$AB = CE = ACDF = BDEF$
$AD = AEF = BCE = BCDF$	$AD = EF = ABCF = BCDE$
$AE = ABF = BCD = BCEF$	$AE = BC = DF = ABCDEF$
$AF = BC = ABE = BCDEF$	$AF = DE = ABCD = BCEF$
$BD = ACE = BEF = ACDF$	$BD = CF = ABEF = ACDE$
$DE = ACD = BDF = ACEF$	$BF = CD = ABDE = ACEF$
$CD = ABE = CEF = ABDF$	$ABD = ACF = BEF = CDE$
$CE = ABD = CDF = ABEF$	$ABF = ACD = BDE = CEF$

Il piano che scaturisce dalla coppia di generatori *a*) è di risoluzione III; infatti, alcune interazioni di due fattori sono confuse con alcuni effetti principali. Il piano che scaturisce dalla coppia di generatori *b*) è di risoluzione IV; infatti, nessun effetto principale è confuso con altri fattori principali o con l'interazione di due fattori principali. Conviene utilizzare il piano di risoluzione maggiore (IV) in quanto, a parità di numero di trattamenti, consente di stimare, senza confondimenti, gli effetti principali (ritenendo, come di prassi, trascurabili le interazioni di tre o più fattori).

**12.22** L'equazione simbolica del piano è  $I = DD = ABCD$  per cui il piano è di risoluzione IV. Pertanto nessun effetto principale è confuso con altri fattori principali o con l'interazione di due fattori ma, l'interazione di due fattori è confusa con quella di altri due ( $AB = CD$ ,  $AC = BD$ ,  $AD = BC$ ). Gli effetti principali e le interazioni a due a due risultano:

$$A = \frac{1}{4}[(211.33 + 63.87 + 148.73 + 70.80) - (57.27 + 45.93 + 41.50 + 17.50)] = +83.13;$$

$$B = \frac{1}{4}[(41.50 + 17.50 + 148.73 + 70.80) - (57.27 + 45.93 + 211.33 + 63.87)] = -24.97;$$

$$C = \frac{1}{4}[(45.93 + 17.50 + 63.87 + 70.80) - (57.27 + 41.50 + 211.33 + 148.73)] = -65.18;$$

$$D = \frac{1}{4}[(45.93 + 41.50 + 211.33 + 70.80) - (57.27 - 17.50 - 63.87 - 148.73)] = +20.6;$$

$$AB + CD = \frac{1}{4}[(57.27 + 45.93 + 148.73 + 70.80) - (41.50 + 17.50 + 211.30 + 63.87)] = -2.9;$$

**628** Soluzioni dei problemi proposti

$$AC + BD = \frac{1}{4}[(57.27 + 41.50 + 63.87 + 70.80) - (45.93 + 17.50 + 211.33 + 148.73)] = -47.5;$$

$$AD + BC = \frac{1}{4}[(57.27 + 17.50 + 211.33 + 70.80) - (45.93 + 41.50 + 63.87 + 148.73)] = +14.2.$$

**12.23** A partire dai dati riaggregati per l'ANOVA a 2 vie con interazione:

	$B_1 (-1)$	$B_2 (+1)$	medie
$A_1 (-1)$	65.30, 53.90, 52.60, 49.30, 44.30, 44.20 ( $\bar{x}_{1\bullet} = 51.60$ )	36.90, 41.60, 46.00, 18.40, 18.70, 15.40 ( $\bar{x}_{12\bullet} = 29.50$ )	40.55
$A_2 (+1)$	204.70, 213.10, 216.20, 59.20, 64.80, 67.60 ( $\bar{x}_{2\bullet} = 137.60$ )	158.60, 141.30, 146.30, 72.30, 71.20, 68.90 ( $\bar{x}_{22\bullet} = 109.77$ )	123.69
medie	94.60	69.64	82.12

si ottiene:

origine della varianza	devianza	gradi di libertà	varianza $S^2$	$z^*$	$p$ -value
$A$	41466.9	1	41466.9	19.19	0.0003
$B$	3740.0	1	3740.0	1.73	0.2032
$AB$	49.3	1	49.3	0.02	0.8815
errore	43221.4	20	2161.1		
totale	88477.6	23			

da cui si evince l'effetto significativo del fattore  $A$ . Non risultano invece significativi il fattore  $B$  e l'interazione  $AB$ . Ovviamente, tale analisi ha un senso se si ritengono trascurabili le interazioni  $BCD$  e  $ACD$  (confuse, rispettivamente, con i fattori  $A$  e  $B$ ) e l'interazione  $CD$  (confusa con  $AB$ ).

**12.24** Non è possibile in quanto il numero di gradi di libertà relativo agli effetti che dobbiamo stimare (5) è maggiore del numero di trattamenti disponibili (4).

**12.25** Non è possibile in quanto ad un piano ortogonale fattoriale  $2^3$  possono essere aggiunti al più 4 fattori, per un numero complessivo di 7 fattori.

**12.26** La risposta è affermativa. Il quadrato greco-latino applicato al piano fattoriale completo  $5^4$  seleziona un numero di trattamenti uguale a  $5^2$ , superiore al numero di gradi di libertà necessari per la stima degli effetti richiesti:  $1 + 4 \times (5 - 1) = 17$ .

Il risultato, per  $f \geq 4$  ed  $n \geq 3$ , vale in generale in quanto, per un generico piano fattoriale  $n^f$  il numero di trattamenti  $n^{f-2}$  previsto dal piano ridotto di tipo quadrato greco-latino risulta sempre maggiore del numero di gradi di libertà necessari per la stima degli effetti richiesti:  $1 + f \times (n - 1)$ . L'unico limite è legato all'esistenza, per un generico numero di livelli ( $n$ ), di quadrati latini ortogonali (cfr. Problema 12.30).

**12.27** Le  $4^{f-1}$  prove di un quadrato latino sono sufficienti a saturare tutti i gradi di libertà  $1+f(4-1)+\binom{f}{2}\times(4-1)\times(4-1)$  per  $f \geq 5$ .

**12.28** Il primo quadrato latino è ortogonale sia al secondo che al terzo; infatti, nelle celle corrispondenti, ogni coppia di livelli compare una ed una sola volta. Ne derivano i seguenti quadrati greco-latini:

$j$	1	2	3
$i$	$l_k$		
1	1 <sub>1</sub>	2 <sub>3</sub>	3 <sub>2</sub>
2	2 <sub>2</sub>	3 <sub>1</sub>	1 <sub>3</sub>
3	3 <sub>3</sub>	1 <sub>2</sub>	2 <sub>1</sub>

$j$	1	2	3
$i$	$l_k$		
1	1 <sub>1</sub>	3 <sub>3</sub>	2 <sub>2</sub>
2	3 <sub>2</sub>	2 <sub>1</sub>	1 <sub>3</sub>
3	2 <sub>3</sub>	1 <sub>2</sub>	3 <sub>1</sub>

o, data la simmetria, quelli realizzabili scambiando i livelli  $l$  con quelli  $k$ .

**12.29** Il piano fattoriale ridotto che risulta dall'applicazione del primo quadrato latino del problema precedente è il seguente:

condizioni sperimentali	A	B	C
$T_1$	1	1	1
$T_2$	1	2	3
$T_3$	1	3	2
$T_4$	2	1	2
$T_5$	2	2	1
$T_6$	2	3	3
$T_7$	3	1	3
$T_8$	3	2	2
$T_9$	3	3	1

Si verifica facilmente che le colonne di tale piano sperimentale sono linearmente indipendenti. Tale proprietà, di carattere del tutto generale, deriva dall'ortogonalità dei quadrati latini.

**12.30** Non è possibile in quanto per  $n = 6$  non esistono quadrati latini ortogonali (cfr. il noto problema di Eulero, detto anche "dei 36 ufficiali").

### Capitolo 13

**13.9** Alla statistica di Fisher corrisponde un  $p$ -value pari a  $1.04 \times 10^{-10}$ , quindi

**630** Soluzioni dei problemi proposti

l'ipotesi  $\mathcal{H}_0 = \{a = 0\}$  è rigettata.

**13.10** L'ipotesi  $\mathcal{H}_0 = \{a = 0\}$  è rigettata sia per  $\alpha = 0.05$  che per  $\alpha = 0.01$ , essendo positivi gli estremi dell'intervallo di confidenza in entrambi i casi. L'ipotesi  $\mathcal{H}_0 = \{b = 0\}$  non può essere rigettata né per  $\alpha = 0.05$  né per  $\alpha = 0.01$  essendo negativo l'estremo inferiore dell'intervallo di confidenza e positivo quello superiore in entrambi i casi.

**13.11** I 12 dati stimati della variabile  $Y$ , in corrispondenza dei 12 valori  $x_i$  del campione, sono:

$$\hat{y}_i = \{486, 506, 509, 512, 515, 523, 660, 694, 709, 695, 710, 713\}$$

ed in corrispondenza di  $x_i = 520$  è  $\hat{y}_i = 529$ .

**13.12** Le stime dei parametri e del coefficiente di correlazione sono:

$$\hat{a} = 1.57; \quad \hat{b} = 54.65; \quad \hat{\rho} = 0.88.$$

**13.13** Alla statistica di Fisher corrisponde un  $p$ -value pari a  $1.08 \times 10^{-8}$ , quindi l'ipotesi  $\mathcal{H}_0 = \{a = 0\}$  è rigettata.

**13.14** L'ipotesi  $\mathcal{H}_0 = \{a = 3\}$  è rigettata sia per  $\alpha = 0.05$  che per  $\alpha = 0.01$ , essendo il valore 3 esterno all'intervallo di confidenza in entrambi i casi. L'ipotesi  $\mathcal{H}_0 = \{b = 75\}$  è rigettata per  $\alpha = 0.05$  ma non per  $\alpha = 0.01$  essendo il valore 75 non incluso nell'intervallo di confidenza nel primo caso ed invece incluso nel secondo.

**13.15** L'intervallo di confidenza al livello  $1 - \alpha = 0.95$  della risposta corrispondente ad  $x_0 = 53$  è:

$$\begin{aligned} & (137.99 - 2.074 \times 8.878 \sqrt{1.0476}; \quad 137.99 + 2.074 \times 8.878 \sqrt{1.0476}) = \\ & = (119.14; \quad 156.83). \end{aligned}$$

**13.16** Le stime dei parametri e del coefficiente di correlazione sono:

$$\hat{a} = -0.102; \quad \hat{b} = -290.108; \quad \hat{\rho} = 0.51.$$

**13.17** Alla statistica di Fisher corrisponde un  $p$ -value pari a 0.136, quindi l'ipotesi  $\mathcal{H}_0 = \{a = 0\}$  non può essere rigettata, con un rischio di errore  $\alpha = 0.136$ .

**13.18** L'ipotesi  $\mathcal{H}_0 = \{a = 0.1\}$  è rigettata per  $\alpha = 0.05$  mentre non può essere rigettata per  $\alpha = 0.01$ . L'ipotesi  $\mathcal{H}_0 = \{b = -340\}$  non può essere rigettata né per  $\alpha = 0.05$  e né per  $\alpha = 0.01$ .

**13.19** L'intervallo di confidenza al livello  $1 - \alpha = 0.95$  della risposta corrispondente ad  $x_0 = 450$  è:

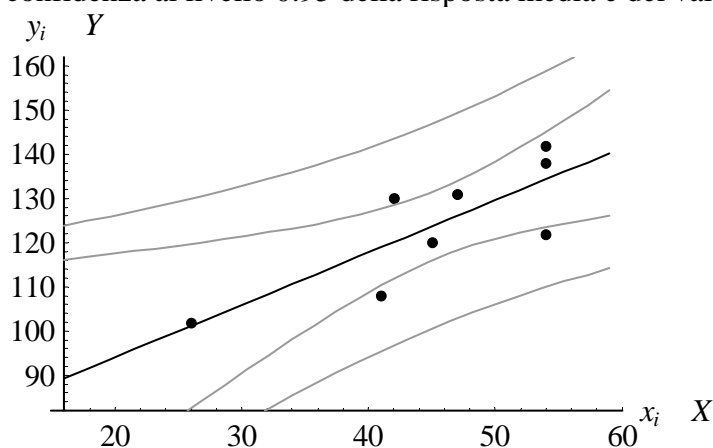


$$\begin{aligned} &(-336 - 2.306 \times 7.974 \sqrt{1.2547}; -336 + 2.306 \times 7.974 \sqrt{1.2547}) = \\ &= (-356.60; -315.4). \end{aligned}$$

**13.20** L'intervallo di confidenza al livello  $1 - \alpha = 0.95$  della risposta media corrispondente ad  $x_0 = 450$  è:

$$\begin{aligned} &(-336 - 2.306 \times 7.974 \sqrt{0.2547}; -336 + 2.306 \times 7.974 \sqrt{0.2547}) = \\ &= (-345.28; -326.72). \end{aligned}$$

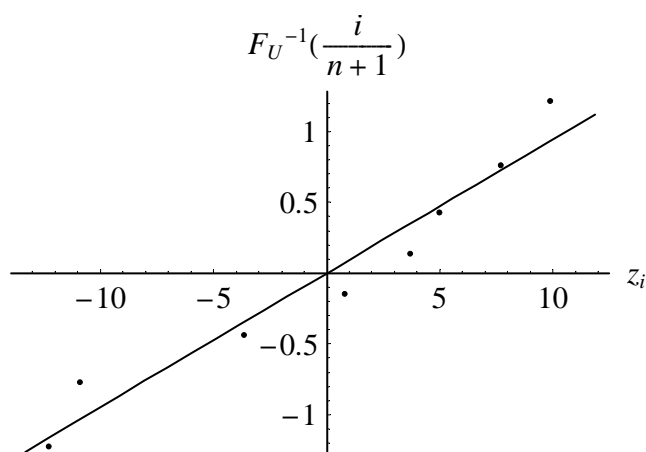
**13.21** Le stime dei parametri della retta di regressione e del coefficiente di correlazione sono  $\hat{a} = 1.18$ ;  $\hat{b} = 70.44$ ;  $\hat{\rho} = 0.81$ . Il grafico della retta di regressione, con relative bande di confidenza al livello 0.95 della risposta media e del valore singolo, è:



**13.22** I residui sono:

(0.799209, -10.9486, 9.86825, -3.68131, 4.95232, 7.67003, 3.67003, -12.33).

Essi hanno media  $-8.875 \times 10^{-6}$  ed il relativo grafico di probabilità con coefficiente di correlazione 0.97 è:

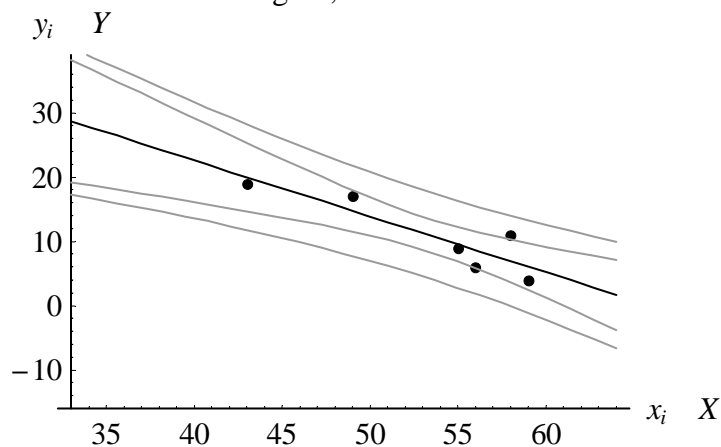


**13.23** Le stime dei parametri della retta di regressione e del coefficiente di correla-

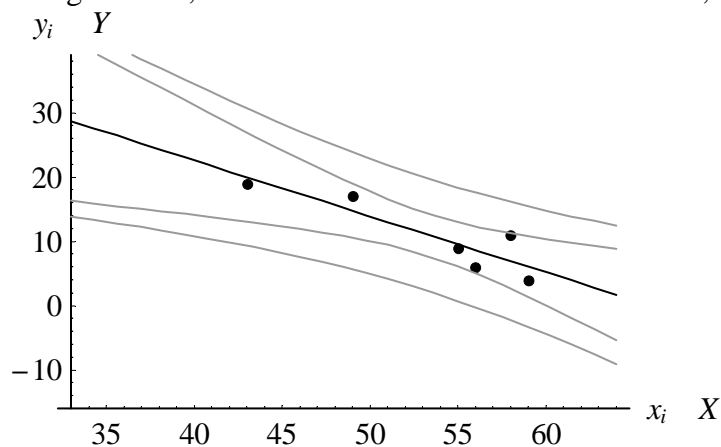
**632** Soluzioni dei problemi proposti

zione sono  $\hat{a} = -0.87$ ,  $\hat{b} = 57.48$ ,  $\hat{\rho} = -0.90$ .

Il grafico della retta di regressione, con relative bande di confidenza al livello 0.90 della risposta media e del valore singolo, è:



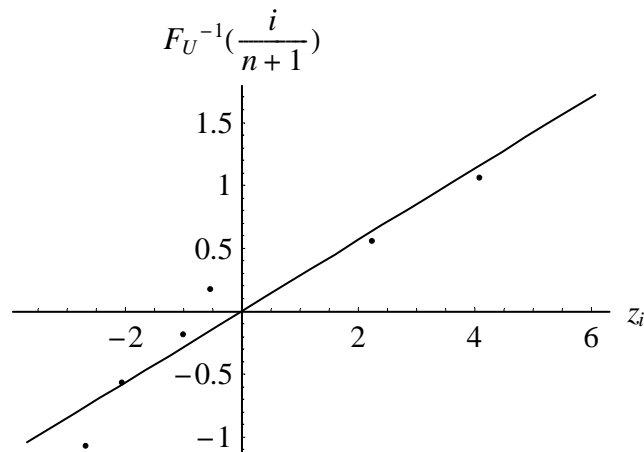
**13.24** La retta di regressione, con bande di confidenza al livello 0.95, è la seguente:



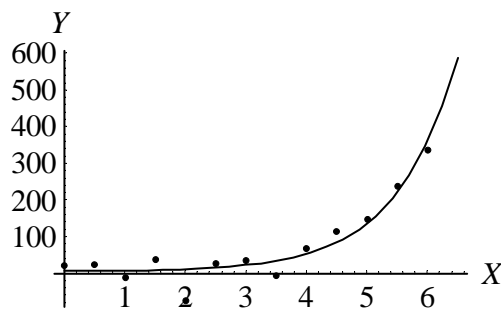
**13.25** I residui sono:

$(-1.00528, 2.22359, -0.547535, -2.67606, 4.0669, -2.06162)$ .

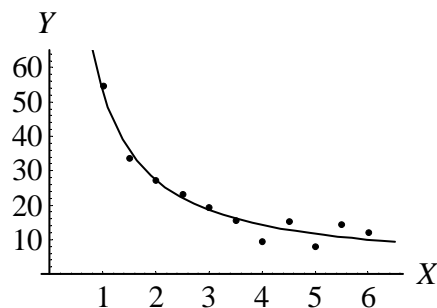
Essi hanno media  $-8.33 \times 10^{-7}$  ed il relativo grafico di probabilità con coefficiente di correlazione 0.96 è:



**13.26** I punti suggeriscono un andamento di tipo esponenziale, per cui la trasformazione  $\xi(X) = e^X$  ci permette di linearizzarli e di stimare la retta dei minimi quadrati  $Y = 4.75 + 0.878 \xi(X)$  passante per l'origine. Quindi, utilizzando le stime dei parametri di questa retta, possiamo tracciare la curva interpolatrice  $Y = 4.75 + 0.878 e^X$  cui compete una stima  $\hat{\rho}^2 = 0.913$ :

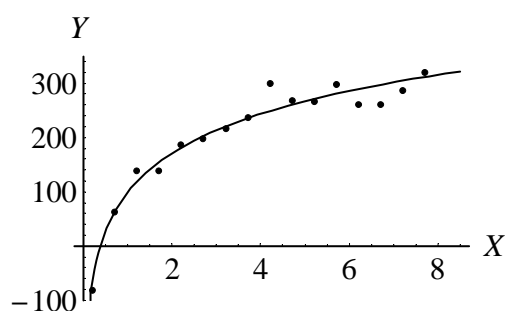


**13.27** I dati mostrano un andamento iperbolico, per cui la trasformazione  $\xi(X) = 1/X$  ci consente di linearizzarli. Utilizzando i parametri della regressione lineare stimiamo la curva interpolatrice  $Y = 1.39 + 51.8/X$ , cui compete una stima  $\hat{\rho}^2 = 0.962$ :

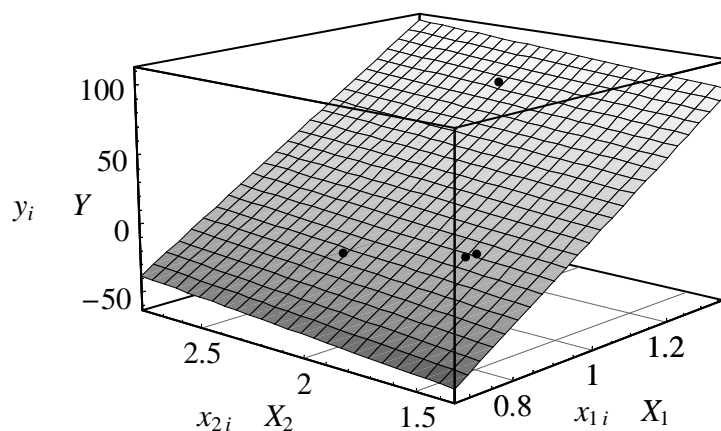


**634** Soluzioni dei problemi proposti

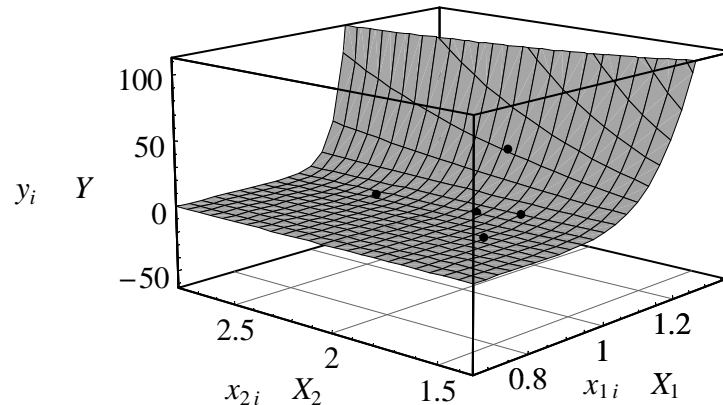
**13.28** I punti sperimentali seguono il tipico andamento logaritmico, da cui scaturisce il suggerimento di adottare la trasformazione  $\xi(X) = \ln(X)$  che rende lineare il loro grafico e ci permette di stimare la curva interpolatrice  $Y = 99.57 + 103.77 \ln(X)$  cui compete una stima  $\hat{\rho}^2 = 0.96$  :



**13.29** Il piano di regressione (stimato con la routine *Regress* di Mathematica 5.0) è  $Y = -207.504 + 208.277 X_1 + 11.105 X_2$  e la stima del relativo coefficiente di determinazione è  $\hat{\rho}^2 = 0.56$  .



**13.30** I parametri della superficie di risposta  $Y = aX_1^{bX_2/X_1}$  (stimati con la routine *FindMinimum* di Mathematica 5.0) sono  $a = 4.914$  e  $b = 12.987$  ed il corrispondente coefficiente di determinazione è  $\hat{\rho}^2 = 0.992$  .



### Capitolo 14

**14.9** La Cdf di  $t_3$  del PNOP, con  $z(t) = 2t^3$ , è:

$$F_3(t) = 1 - \sum_{i=0}^2 P_i(t) = 1 - \exp(-t^4/2) \sum_{i=0}^2 \frac{(t^4/2)^i}{i!}$$

da cui:

$$f_3(t) = \frac{dF_3(t)}{dt} = \exp(-t^4/2) \sum_{i=0}^2 \left[ t^{4i} (2t^3 - 4i/t) (1/2)^i / i! \right] = t^{11}/4 \exp(-t^4/2) .$$

**14.10** La soluzione del problema precedente ci fornisce la pdf del terzo tempo di guasto, da cui possiamo calcolare la richiesta probabilità:

$$\int_2^\infty f_3(t) dt = \int_2^\infty t^{11}/4 \exp(-t^4/2) dt = \left[ (-1 - t^4/2 - t^8/8) \exp(-t^4/2) \right]_2^\infty = 0.013754.$$

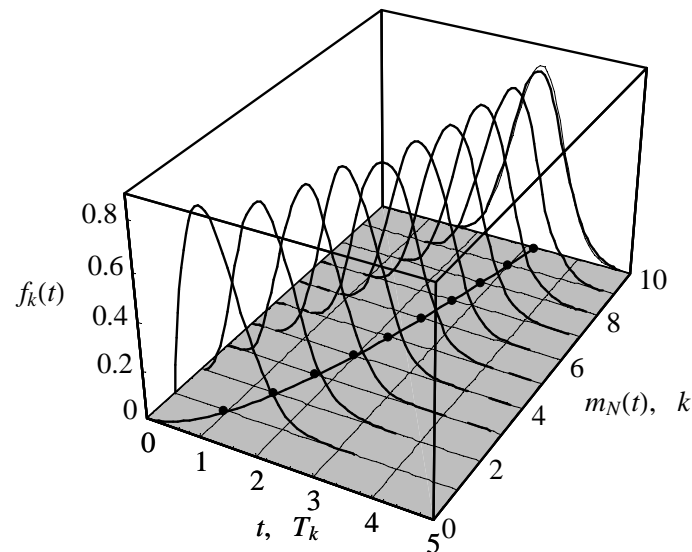
**14.11** Innanzitutto, possiamo ricavare la Cdf della v.a.  $X_k$  (intervallo tra l'arrivo  $T_{k-1}$  e  $T_k$ ), utilizzando la formulazione dell'affidabilità per intervallo, avendo posto  $t_0 = t_{k-1}$  e  $t_0 + \Delta t = t_{k-1} + x_k$ :

$$R(t_{k-1}, t_{k-1} + x_k) = 1 - F_{X_k}(x_k) = \exp \left[ - \int_{t_{k-1}}^{t_k} z(t) dt \right] = \exp \left[ - \left( \frac{t_k}{\alpha} \right)^\beta + \left( \frac{t_{k-1}}{\alpha} \right)^\beta \right]$$

da cui ricaviamo la funzione inversa  $F_{X_k}^{-1}(x_k)$ :

$$\ln \left[ \frac{1}{1 - F_{X_k}(x_k)} \right] = - \left( \frac{t_k}{\alpha} \right)^\beta + \left( \frac{t_{k-1}}{\alpha} \right)^\beta ; \quad t_{k_i} = \alpha \left[ - \ln \left( \frac{1}{1 - u_i} \right) + \left( \frac{t_{k-1}}{\alpha} \right)^\beta \right]^{1/\beta}$$

che ci consente di ottenere le determinazioni  $t_{k_i}$  della v.a.  $T_k$  dalle determinazioni  $u_i$  della v.a. Uniforme definita nell'intervallo (0, 1).

**636** Soluzioni dei problemi proposti


Ad esempio, sulla base di 5000 determinazioni pseudo-sperimentali di ciascuna v.a.  $T_k$  ( $k=1, \dots, 10$ ), con  $\alpha=1$  e  $\beta=2$ , ed utilizzando un'interpolazione cubica degli istogrammi delle densità di frequenza, si è ottenuto il grafico precedente che appare quasi indistinguibile da quello di Figura 13.11 realizzato utilizzando la formulazione analitica delle pdf. In quel grafico come in questo è riportata la funzione media  $m_N(t)$  e le  $f_k(t)$  dei primi dieci tempi di guasto  $T_k$ , le cui medie sono evidenziate con un punto in grassetto. Anche qui, grazie alla robustezza del teorema del limite centrale, possiamo notare come già per  $k=10$  le pdf risultano quasi indistinguibili da quella Normale di pari media e varianza, disegnata a tratto più sottile.

**14.12** La media della generica v.a.  $T_k$  è data da:

$$\begin{aligned} E\{T_k\} &= \int_0^{\infty} t f_k(t) dt = \int_0^{\infty} t dF_k(t) = -\int_0^{\infty} t dR_k(t) = \\ &= -[t R_k(t)]_0^{\infty} + \int_0^{\infty} R_k(t) dt = \int_0^{\infty} R_k(t) dt; \\ R_k(t) &= 1 - F_k(t) = \exp[-\lambda t^b] \sum_{i=0}^{k-1} \frac{[\lambda t^b]^i}{i!}; \quad \lim_{t \rightarrow \infty} [t R_k(t)] = 0 \end{aligned}$$

da cui abbiamo:

$$E\{T_k\} = \sum_{i=0}^{k-1} \int_0^{\infty} \exp[-\lambda t^b] \frac{[\lambda t^b]^i}{i!} dt = \frac{1}{\lambda^{1/b} b} \sum_{i=0}^{k-1} \left[ \frac{1}{i!} \Gamma(i+1/b) \right] = \frac{\Gamma(k+1/b)}{\lambda^{1/b} \Gamma(k)}.$$

Nel caso particolare in cui è  $b=1$  il processo è quello Omogeneo di Poisson ed essendo  $\Gamma(k+1) = k \Gamma(k)$  risulta  $E\{T_k\} = k/\lambda$  che corrisponde alla somma delle medie di  $k$  v.a. Esponenziali di parametro  $\lambda$ . In questo caso la funzione lineare discreta  $k = \lambda E\{T_k\}$  è sovrapposta alla retta  $m_N(t) = \lambda t$ . Infatti, rispetto al riferimento di ordinate ( $m_N(t) = k$ ) e di ascisse ( $t = E\{T_k\}$ ), le due funzioni hanno sempre la stessa

ordinata,  $\lambda E\{T_k\}$ , qualunque sia l'ascissa  $t = E\{T_k\}$ . In altre parole, il luogo (discreto) dei punti rappresentativi delle medie delle v.a.  $T_k$  è sovrapposto al luogo (continuo) dei punti rappresentativi delle medie delle v.a.  $N(t)$ .

Nel caso generale ( $b \neq 1$ ) le due funzioni differiscono leggermente ma tendono rapidamente a coincidere per  $b$  e/o  $k$  crescenti infinitamente.

**14.13** In presenza di rimessa, la vincita,  $Y_{10}$ , maturata dopo dieci estrazioni si distribuisce come una v.a. Binomiale di parametri  $n = 10$  e  $p = 2/5$ :

$$\Pr\{Y_{10} = y\} = \binom{10}{y} \left(\frac{2}{5}\right)^y \left(1 - \frac{2}{5}\right)^{10-y}; \quad y = 0, 1, 2, \dots, 10.$$

**14.14** Indicata con  $Y_i$  la vincita maturata dopo  $i$  estrazioni e con  $Y_{8,10}$  l'incremento di vincita maturato con le estrazioni 8-10, si può scrivere:

$$\begin{aligned} \Pr\{(Y_7 = x) \cap (Y_{10} = y)\} &= \Pr\{(Y_7 = x)\} \Pr\{(Y_{10} = y) | (Y_7 = x)\} = \\ &= \Pr\{(Y_7 = x)\} \Pr\{(Y_{8,10} = y - x) | (Y_7 = x)\} = \Pr\{(Y_7 = x)\} \Pr\{(Y_3 = y - x)\} = \\ &= \binom{7}{x} \left(\frac{2}{5}\right)^x \left(1 - \frac{2}{5}\right)^{7-x} \binom{3}{y-x} \left(\frac{2}{5}\right)^{y-x} \left(1 - \frac{2}{5}\right)^{3-(y-x)}; \\ & \quad x = 0, 1, 2, \dots, 7; \quad y = 0, 1, 2, \dots, 10; \quad y \geq x. \end{aligned}$$

In virtù della validità della uguaglianza:

$$\Pr\{(Y_3 = y - x)\} = \Pr\{(Y_{8,10} = y - x) | (Y_7 = x)\}$$

il processo è anche detto senza memoria (o anche ad incrementi indipendenti e stazionari).

**14.15** In caso di assenza di rimessa la vincita,  $Y_{10}$ , maturata dopo dieci estrazioni si distribuisce come una v.a. Ipergeometrica:

$$\Pr\{Y_{10} = y\} = \frac{\binom{20}{y} \binom{50-20}{10-y}}{\binom{50}{10}}; \quad y = 0, 1, 2, \dots, 10.$$

**14.16** Utilizzando la stessa notazione del problema 14.14 si può scrivere:

**638** Soluzioni dei problemi proposti

$$\begin{aligned}
 \Pr\{(Y_7 = x) \cap (Y_{10} = y)\} &= \Pr\{Y_7 = x\} \Pr\{Y_{10} = y | (Y_7 = x)\} = \\
 &= \Pr\{Y_7 = x\} \Pr\{Y_{8,10} = y - x | (Y_7 = x)\} = \\
 &= \frac{\binom{20}{x} \binom{30}{7-x} \binom{20-x}{y-x} \binom{30-7+x}{3-y+x}}{\binom{50}{7}};
 \end{aligned}$$

$$x = 0, 1, 2, \dots, 7; \quad y = 0, 1, 2, \dots, 10; \quad y \geq x.$$

Come si vede, in assenza di rimessa il processo ha memoria (la distribuzione dell'incremento  $Y_{8,10}$  dipende da  $Y_7$ , numero di biglie nere estratte in precedenza).

**14.17** Indicata con  $Y_i$  la vincita maturata dopo  $i$  estrazioni e con  $Y_{i,i}$  l'incremento di vincita maturato con la  $i$ -esima estrazione, abbiamo che:

a) nel caso di estrazione con rimessa, per ogni  $i$  la v.a.  $Y_{i,i}$  è una v.a. di Bernoulli di parametro  $p = 2/5$ :

$$\Pr\{Y_{i,i} = y\} = \left(\frac{2}{5}\right)^y \left(1 - \frac{2}{5}\right)^{1-y}; \quad y = 0, 1;$$

b) nel caso di estrazione senza rimessa, per ogni  $i \leq 50$  si ha:

$$\begin{aligned}
 \Pr\{(Y_{i,i} = 1)\} &= \\
 &= \sum_{x=0}^{i-1} \Pr\{(Y_{i,i} = 1) | (Y_{i-1} = x)\} \Pr\{Y_{i-1} = x\} = \\
 &= \sum_{x=0}^{i-1} \frac{20-x}{50-(i-1)} \left[ \frac{\binom{20}{x} \binom{30}{(i-1)-x}}{\binom{50}{i-1}} \right] = \\
 &= \sum_{x=0}^{i-1} \frac{20-x}{50-(i-1)} \frac{20!}{(20-x)!x!} \frac{30!}{(30-(i-1)+x)!((i-1)-x)!} \frac{(50-(i-1))!(i-1)!}{50!} = \\
 &= \frac{20}{50} \sum_{x=0}^{i-1} \frac{(20-1)!}{(20-x-1)!x!} \frac{30!}{(30-(i-1)+x)!((i-1)-x)!} \frac{(50-(i-1)-1)!(i-1)!}{(50-1)!} = \\
 &= \frac{20}{50} \sum_{x=0}^{i-1} \left[ \frac{\binom{20-1}{x} \binom{30}{(i-1)-x}}{\binom{50-1}{i-1}} \right] = \frac{2}{5}
 \end{aligned}$$

da cui ricaviamo:

$$\Pr\{(Y_{i,i} = 0)\} = 1 - \Pr\{(Y_{i,i} = 1)\} = \frac{3}{5}.$$

Pertanto, sia nel caso a) che nel caso b)  $Y_{i,i}$  è una v.a. di Bernoulli di parametro  $p = 2/5$ . La differenza tra i due processi risiede nel fatto che nel caso a) le v.a. di

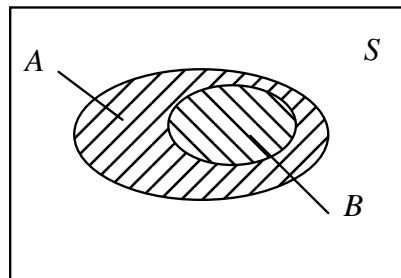


Bernoulli che si ottengono al variare di  $i$  sono  $s$ -indipendenti mentre nel caso b) sono  $s$ -dipendenti.

**14.18** Introdotte le seguenti notazioni:

$$A = \{T_i \leq t\}; \quad B = \{T_{i+1} \leq t\}$$

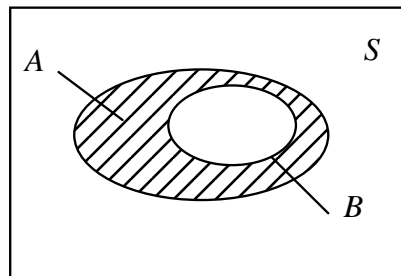
e considerando che  $T_i \leq T_{i+1}$ , è facile dedurre che  $B$  implica  $A$ , infatti se è vero  $B$  è sicuramente vero anche  $A$  mentre se è vero  $A$  non è detto che sia vero  $B$ . Sul diagramma di Venn i due eventi devono pertanto essere rappresentati come segue:



**14.19** Utilizzando le stesse notazioni dell'esercizio precedente possiamo scrivere:

$$\Pr\{N(t) = i\} = \Pr\{(T_i \leq t) \cap (T_{i+1} > t)\} = \Pr\{A \cap \bar{B}\}.$$

Osservando il seguente diagramma di Venn, in cui l'evento tratteggiato è  $A \cap \bar{B}$ :



deduciamo che:

$$\Pr\{A \cap \bar{B}\} = \Pr\{A\} - \Pr\{B\}.$$

Pertanto, ricordando che:

$$\Pr\{A\} = \Pr\{T_i \leq t\} = F_i(t) \quad \text{e} \quad \Pr\{B\} = \Pr\{T_{i+1} \leq t\} = F_{i+1}(t)$$

abbiamo:

$$\begin{aligned} \Pr\{N(t) = i\} &= \Pr\{(T_i \leq t) \cap (T_{i+1} > t)\} = \Pr\{A \cap \bar{B}\} = \Pr\{A\} - \Pr\{B\} = \\ &= F_i(t) - F_{i+1}(t). \end{aligned}$$

**14.20** Essendo  $r(t)dt$  il differenziale della funzione media, possiamo scrivere:

**640** Soluzioni dei problemi proposti

$$\begin{aligned}
 r(t)dt &= dm_N(t) = m_N(t+dt) - m_N(t) = \\
 &= E[N(t+dt)] - E[N(t)] = E[N(t, t+dt)] = \\
 &= \sum_{k=0}^{\infty} k \cdot \Pr\{N(t, t+dt) = k\}
 \end{aligned}$$

dove risulta:

$$E[N(t+dt)] - E[N(t)] = E[N(t, t+dt)]$$

per la linearità dell'operatore speranza matematica.

**14.21** La probabilità richiesta può essere formulata come:

$$\begin{aligned}
 \Pr\{[N(t) = i] \cap [N(t+\Delta t) = i+j]\} &= \Pr\{[N(t) = i] \cap [N(t, t+\Delta t) = j]\} = \\
 &= \Pr\{N(t) = i\} \Pr\{N(t, t+\Delta t) = j\} = \\
 &= \frac{\left(\int_0^t \lambda dt\right)^i}{i!} \exp\left(-\int_0^t \lambda dt\right) \frac{\left(\int_t^{t+\Delta t} \lambda dt\right)^j}{j!} \exp\left(-\int_t^{t+\Delta t} \lambda dt\right) = \\
 &= \frac{(\lambda t)^i}{i!} \exp(-\lambda t) \frac{(\lambda \Delta t)^j}{j!} \exp(-\lambda \Delta t)
 \end{aligned}$$

con la uguaglianza:

$$\Pr\{[N(t) = i] \cap [N(t, t+\Delta t) = j]\} = \Pr\{N(t) = i\} \Pr\{N(t, t+\Delta t) = j\}$$

garantita dal fatto che in un processo di Poisson le variabili che contano gli eventi in intervalli disgiunti sono  $s$ -indipendenti. In particolare, se la funzione intensità è costante risulta anche:

$$\Pr\{N(t, t+\Delta t) = j\} = \Pr\{N(\Delta t) = j\}.$$

In virtù di questa proprietà il processo Omogeneo di Poisson è detto senza memoria.

**14.22** Nelle condizioni ipotizzate, per  $t \leq 10$ , abbiamo:

$$\begin{aligned}
 \Pr\{T_1 \leq t \mid N(10) = 1\} &= \frac{\Pr\{(N(t) = 1) \cap (N(t, 10) = 0)\}}{\Pr\{N(10) = 1\}} = \\
 &= \frac{\Pr\{(N(t) = 1)\} \Pr\{N(10-t) = 0\}}{\Pr\{N(10) = 1\}} = \frac{\lambda t \exp(-\lambda t) \exp[-\lambda(10-t)]}{\lambda 10 \exp(-\lambda 10)} = \frac{t}{10}.
 \end{aligned}$$

La v.a. tempo al primo guasto,  $T_1$ , dato  $N(10) = 1$ , ha Cdf Uniforme nell'intervallo  $(0, 10)$  (che per  $t > 10$  vale 1).

**14.23** In generale, per un processo Omogeneo di Poisson (cfr. Problema 14.21) vale la seguente:

$$\Pr\{(N(t) = i) \cap (N(t+\Delta t) = i+j)\} = \Pr\{N(t) = i\} \Pr\{N(\Delta t) = j\}$$

da cui ricaviamo:

$$\begin{aligned} \Pr\{(N(t) = i) | (N(t + \Delta t) = i + j)\} &= \frac{\Pr\{(N(t) = i) \cap (N(t + \Delta t) = i + j)\}}{\Pr\{N(t + \Delta t) = i + j\}} = \\ &= \frac{\Pr\{N(t) = i\} \Pr\{N(\Delta t) = j\}}{\Pr\{N(t + \Delta t) = i + j\}} = \frac{\frac{(\lambda t)^i}{i!} \exp(-\lambda t) \frac{(\lambda \Delta t)^j}{j!} \exp(-\lambda \Delta t)}{\frac{[\lambda(t + \Delta t)]^{i+j}}{(i+j)!} \exp[-\lambda(t + \Delta t)]} = \\ &= \frac{(i+j)!}{i! j!} \left(\frac{t}{t + \Delta t}\right)^i \left(\frac{\Delta t}{t + \Delta t}\right)^j = \binom{i+j}{i} \left(\frac{t}{t + \Delta t}\right)^i \left(1 - \frac{t}{t + \Delta t}\right)^{(i+j)-i}. \end{aligned}$$

Quindi, la richiesta probabilità è:

$$\Pr\{(N(4) = 2) | (N(10) = 5)\} = \binom{5}{2} \left(\frac{4}{10}\right)^2 \left(1 - \frac{4}{10}\right)^{5-2}.$$

**14.24** La Cdf del  $k$ -imo tempo di guasto di un generico processo di conteggio è definita dalla (14.39). Nel caso di un processo Omogeneo di Poisson di intensità  $\lambda$ , abbiamo:

$$P_i(t) = \frac{(\lambda t)^i}{i!} \exp(-\lambda t).$$

La Cdf del tempo al quinto guasto è pertanto:

$$F_5(t) = \sum_{i=5}^{\infty} \frac{(\lambda t)^i}{i!} \exp(-\lambda t) = 1 - \sum_{i=0}^4 \frac{(\lambda t)^i}{i!} \exp(-\lambda t).$$

**14.25** La Cdf del tempo al  $k$ -esimo guasto per un generico processo è definita dalla (14.39). Nel caso di un processo Non-Omogeneo di Poisson di intensità  $z(t)$ , abbiamo:

$$P_i(t) = \frac{\left(\int_0^t z(x) dx\right)^i}{i!} \exp\left(-\int_0^t z(x) dx\right).$$

Possiamo pertanto scrivere:

$$F_k(t) = \exp\left(-\int_0^t z(x) dx\right) \cdot \sum_{i=k}^{\infty} \frac{\left(\int_0^t z(x) dx\right)^i}{i!} = 1 - \exp\left(-\int_0^t z(x) dx\right) \cdot \sum_{i=0}^{k-1} \frac{\left(\int_0^t z(x) dx\right)^i}{i!}$$

da cui, essendo:

$$z(t) = \left(\frac{b}{a}\right) \left(\frac{x}{a}\right)^{b-1}$$

**642** Soluzioni dei problemi proposti

abbiamo:

$$F_k(t) = \exp\left(-\left(\frac{t}{a}\right)^b\right) \sum_{i=k}^{\infty} \frac{\left[\left(\frac{t}{a}\right)^b\right]^i}{i!} = 1 - \exp\left(-\left(\frac{t}{a}\right)^b\right) \sum_{i=0}^{k-1} \frac{\left[\left(\frac{t}{a}\right)^b\right]^i}{i!}$$

che, posto  $k = 5$ , fornisce la Cdf del tempo al quinto guasto.

**14.26** Indicata con  $z(t)$  l'intensità del processo di Poisson, possiamo scrivere:

$$f_k(t) = \frac{dF_k(t)}{dt} = \frac{d}{dt} \left[ \sum_{i=k}^{\infty} \Pr\{N(t) = i\} \right] = \frac{d}{dt} \left[ \exp\left(-\int_0^t z(x) dx\right) \sum_{i=k}^{\infty} \frac{1}{i!} \left(\int_0^t z(x) dx\right)^i \right].$$

Tenuto conto del fatto che:

$$\frac{d}{dt} \left( \int_0^t z(x) dx \right) = z(t)$$

abbiamo:

$$\begin{aligned} f_k(t) &= z(t) \exp\left(-\int_0^t z(x) dx\right) \sum_{i=k}^{\infty} \left[ \frac{1}{(i-1)!} \left(\int_0^t z(x) dx\right)^{i-1} - \frac{1}{i!} \left(\int_0^t z(x) dx\right)^i \right] = \\ &= z(t) \exp\left(-\int_0^t z(x) dx\right) \left[ \sum_{i=k}^{\infty} \frac{1}{(i-1)!} \left(\int_0^t z(x) dx\right)^{i-1} - \sum_{i=k}^{\infty} \frac{1}{i!} \left(\int_0^t z(x) dx\right)^i \right] = \\ &= z(t) \exp\left(-\int_0^t z(x) dx\right) \frac{1}{(k-1)!} \left(\int_0^t z(x) dx\right)^{k-1}. \end{aligned}$$

Infatti risulta:

$$\begin{aligned} \sum_{i=k}^{\infty} \frac{1}{(i-1)!} \left(\int_0^t z(x) dx\right)^{i-1} - \sum_{i=k}^{\infty} \frac{1}{i!} \left(\int_0^t z(x) dx\right)^i &= \sum_{j=k-1}^{\infty} \frac{1}{j!} \left(\int_0^t z(x) dx\right)^j - \sum_{i=k}^{\infty} \frac{1}{i!} \left(\int_0^t z(x) dx\right)^i = \\ &= \frac{1}{(k-1)!} \left(\int_0^t z(x) dx\right)^{k-1} + \sum_{j=k}^{\infty} \frac{1}{j!} \left(\int_0^t z(x) dx\right)^j - \sum_{i=k}^{\infty} \frac{1}{i!} \left(\int_0^t z(x) dx\right)^i = \frac{1}{(k-1)!} \left(\int_0^t z(x) dx\right)^{k-1}. \end{aligned}$$

Nel caso del processo Omogeneo di Poisson, l'intensità è costante e pari a  $\lambda$ , per cui abbiamo:

$$f_k(t) = \lambda \exp(-\lambda t) \frac{(\lambda t)^{k-1}}{(k-1)!} = \lambda \exp(-\lambda t) \frac{(\lambda t)^{k-1}}{\Gamma(k)}$$

con  $\Gamma(k) = (k-1)!$ . La v.a.  $T_k$  di un processo Omogeneo di Poisson è una Gamma (Paragrafo 5.2) di parametri  $k$  e  $\lambda$ . Per ottenere la pdf  $f_5(t)$  del tempo al quinto

guasto, basta porre  $k = 5$ .

**14.27** Nel caso dell'assegnato processo Non-Omogeneo di Poisson, abbiamo:

$$z(t) = \left(\frac{b}{a}\right) \left(\frac{t}{a}\right)^{b-1}$$

$$\int_0^t z(x) dx = m_N(t) = \left(\frac{t}{a}\right)^b$$

che sostituita nell'espressione generale (cfr. Problema 14.26) della pdf del tempo al  $k$ -esimo guasto di un processo Non-Omogeneo di Poisson, fornisce:

$$f_k(t) = \frac{b}{a} \left(\frac{t}{a}\right)^{b-1} \frac{\left[\left(\frac{t}{a}\right)^b\right]^{k-1}}{(k-1)!} \exp\left[-\left(\frac{t}{a}\right)^b\right] = \frac{b}{a} \frac{\left(\frac{t}{a}\right)^{k \cdot b - 1}}{\Gamma(k)} \exp\left[-\left(\frac{t}{a}\right)^b\right].$$

**14.28** Le ipotesi assunte suggeriscono l'impiego del processo Omogeneo di Poisson (14.61). Per questo processo, indicata con  $\lambda$  la funzione intensità, l'affidabilità  $R(t, t + \Delta t) = \Pr\{N(t, t + \Delta t) = 0\}$ , può essere espressa come:

$$R(t, t + \Delta t) = \Pr\{N(t, t + \Delta t) = 0\} =$$

$$= \frac{\left(\int_t^{t+\Delta t} \lambda dt\right)^0}{0!} \exp\left(-\int_t^{t+\Delta t} \lambda dt\right) = \exp\left\{-[\lambda(t + \Delta t) - \lambda t]\right\} =$$

$$= \exp(-\lambda \Delta t) = \exp[-m_N(\Delta t)].$$

**14.29** Per un processo Non-Omogeneo di Poisson si ha (Paragrafo 14.5):

$$R(t, t + \Delta t) = \Pr\{N(t, t + \Delta t) = 0\} = \frac{\left(\int_t^{t+\Delta t} z(t) dt\right)^0}{0!} \exp\left(-\int_t^{t+\Delta t} z(t) dt\right) =$$

$$= \exp\left\{-\left[\int_0^{t+\Delta t} z(t) dt - \int_0^t z(t) dt\right]\right\} = \exp\left\{-[m_N(t + \Delta t) - m_N(t)]\right\}.$$

Nel caso in esame, abbiamo:

$$R(t, t + \Delta t) = \exp\left\{-\left[\left(\frac{t + \Delta t}{a}\right)^b - \left(\frac{t}{a}\right)^b\right]\right\}.$$

**14.30** La Cdf richiesta è:

**644** Soluzioni dei problemi proposti

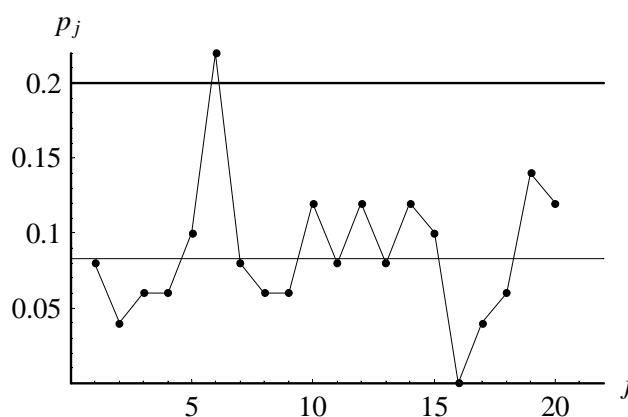
$$\begin{aligned}
 \Pr\{T_2 \leq t | T_1 = t_1\} &= \Pr\{N(t_1, t) \geq 1 | T_1 = t_1\} = \\
 &= 1 - \Pr\{N(t_1, t) = 0 | T_1 = t_1\} = 1 - \Pr\{N(t_1, t) = 0\} = \\
 &= 1 - \exp\left\{-\int_{t_1}^t z(x) dx\right\} = 1 - \exp\left\{-\left[\int_0^t z(x) dx - \int_0^{t_1} z(x) dx\right]\right\} = \\
 &= 1 - \exp\left\{-[m_N(t) - m_N(t_1)]\right\}; \quad t \geq t_1.
 \end{aligned}$$

**Capitolo 15**

**15.9** Dai dati calcoliamo i seguenti valori  $p_j = d_j/n$ :

0.08, 0.04, 0.06, 0.06, 0.10, 0.20, 0.08, 0.06, 0.06, 0.12,  
 0.08, 0.12, 0.08, 0.12, 0.10, 0.0, 0.04, 0.06, 0.10, 0.12

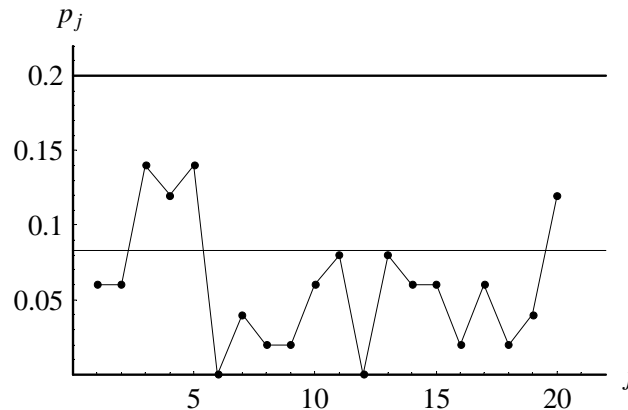
che riportati sulla carta già costruita evidenziano che si è verificato un episodio di fuori controllo in concomitanza del sesto prelievo.



**15.10** Dai 20 campioni prelevati successivamente alla modifica del processo otteniamo i seguenti valori  $p_j = d_j/n$ :

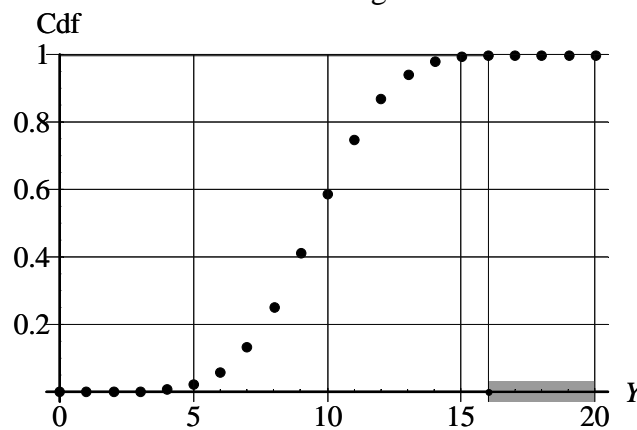
0.06, 0.06, 0.14, 0.12, 0.14, 0.0, 0.04, 0.02, 0.02, 0.06,  
 0.08, 0.0, 0.08, 0.06, 0.06, 0.02, 0.06, 0.02, 0.04, 0.12

a cui corrispondono sulla carta di controllo già calcolata i seguenti 20 punti:



La deriva “imposta” al processo sembra effettivamente essersi realizzata come era nelle aspettative dei tecnici.

**15.11** Se non ci fosse deriva, i punti dovrebbero capitare al disopra o al disotto della linea centrale seguendo una legge di probabilità Bernoulliana di parametro 0.5. Quindi, dal diagramma della Cdf Binomiale di parametri 0.5 e 20 deduciamo che l’evento osservato (16 punti al disotto della linea centrale su 20) appartiene ad una coda di probabilità  $\alpha = 0.001$ . Pertanto la deriva risulta significativa al livello  $1 - \alpha = 0.999$ .



**15.12** Dai dati deduciamo le seguenti stime per la costruzione della carta  $c$ :

$$\bar{c} = \frac{1}{m} \sum_{j=1}^m c_j = 0.50; \quad \hat{\sigma}_{c_j} = \sqrt{\bar{c}} = 0.707;$$

$$\bar{c} - 3\hat{\sigma}_{c_j} = -1.62; \quad \bar{c} + 3\hat{\sigma}_{c_j} = 2.62.$$

Possiamo quindi tracciare la carta  $c$  (con limite inferiore nullo essendo risultata negativa la differenza  $\bar{c} - 3\hat{\sigma}_{c_j}$ ) che mostra una produzione verosimilmente in controllo.

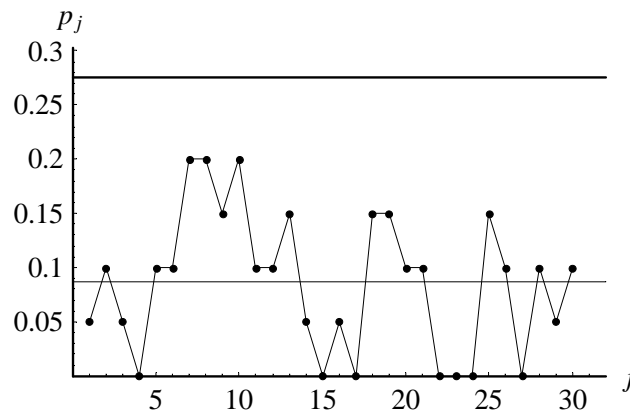
**15.13** Dai dati raccolti sui 30 campioni di 20 autovetture otteniamo le stime:

**646** Soluzioni dei problemi proposti

$$\bar{p} = \frac{1}{nm} \sum_{j=1}^m d_j = 0.087; \quad \hat{\sigma}_{p_j} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0.0629;$$

$$\bar{p} - 3\hat{\sigma}_{p_j} = -0.102; \quad \bar{p} + 3\hat{\sigma}_{p_j} = 0.275$$

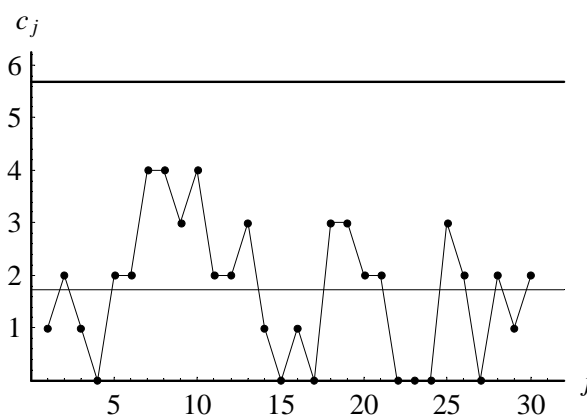
con cui costruiamo la seguente carta  $p$  (con limite inferiore nullo essendo risultata negativa la differenza  $\bar{p} - 3\hat{\sigma}_{p_j}$ ):



**15.14** Assumendo come unità produttiva il “lotto” di 20 autovetture, deduciamo le seguenti stime per la costruzione della carta  $c$ :

$$\bar{c} = \frac{1}{m} \sum_{j=1}^m c_j = 1.73; \quad \hat{\sigma}_{c_j} = \sqrt{\bar{c}} = 1.317; \quad \bar{c} - 3\hat{\sigma}_{c_j} = -2.22; \quad \bar{c} + 3\hat{\sigma}_{c_j} = 5.68$$

tracciamo quindi la carta  $c$  (con limite inferiore nullo essendo risultata negativa la differenza  $\bar{c} - 3\hat{\sigma}_{c_j}$ ):



Notiamo che questa carta appare del tutto simile alla precedente carta  $p$ , come del resto dovevamo aspettarci essendo molto bassa la frazione media di esemplari difettosi (0.087).

**15.15** La nuova carta è nota come “carta  $np$ ” ed ha la seguente linea centrale e limiti

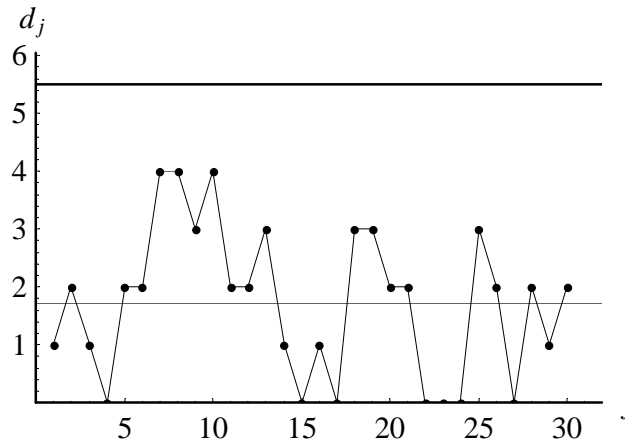


di controllo:

$$n\bar{p} = \frac{1}{m} \sum_{j=1}^m d_j = 1.74; \quad \hat{\sigma}_{d_j} = \sqrt{n\bar{p}(1-\bar{p})} = 1.258;$$

$$n\bar{p} - 3\hat{\sigma}_{d_j} = -2.03; \quad n\bar{p} + 3\hat{\sigma}_{d_j} = 5.51.$$

Confrontando questo grafico con quello della carta  $c$  del problema precedente, otteniamo un'ulteriore conferma della buona approssimazione ottenibile adottando la carta  $c$  per il controllo di regimi a bassa difettosità.



Confrontando questo grafico con quello della carta  $c$  del problema precedente, otteniamo un'ulteriore conferma della buona approssimazione ottenibile adottando la carta  $c$  per il controllo di regimi a bassa difettosità.

**15.16** È preferibile il primo piano perché ha probabilità di accettazione minore nei confronti di un lotto con il 20% di pezzi difettosi:

$$P_{a_1} = \sum_{k=0}^4 \frac{\binom{200}{k} \binom{1000-200}{50-k}}{\binom{1000}{50}} = 0.016;$$

$$P_{a_2} = \sum_{k=0}^2 \frac{\binom{200}{k} \binom{1000-200}{25-k}}{\binom{1000}{25}} = 0.095.$$

**15.17** Dai dati otteniamo:

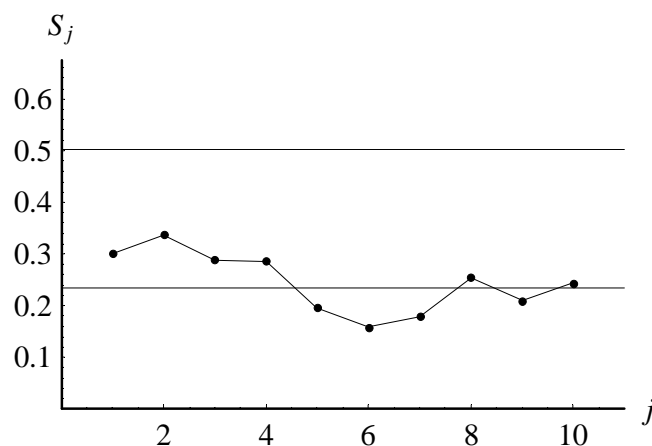
$$\hat{\sigma}_0 = \sqrt{\frac{1}{m} \sum_{j=1}^m s_j^2} = 0.252; \quad \bar{s} = \hat{\sigma}_0 \sqrt{(2n-3)/(2n-2)} = 0.236;$$

$$\hat{\sigma}_{S_j} = \hat{\sigma}_0 / \sqrt{2(n-1)} = 0.089; \quad \bar{s} - 3\hat{\sigma}_{S_j} = -0.031; \quad \bar{s} + 3\hat{\sigma}_{S_j} = 0.503.$$

La carta  $S$  (con limite inferiore nullo essendo risultata negativa la differenza

**648** Soluzioni dei problemi proposti

$\bar{S} - 3\hat{\sigma}_{S_j}$ ) evidenzia una sostanziale stabilità dello scarto tipo dell'errore dimensionale.

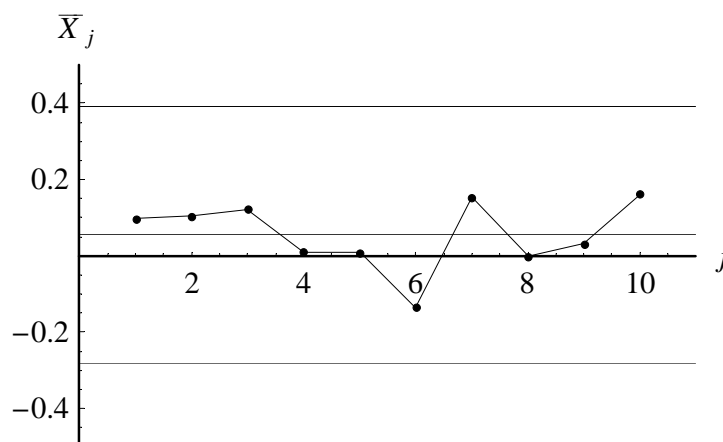


**15.18** Dai dati otteniamo:

$$\bar{x} = \sum_{j=1}^m \bar{x}_j / m = 0.055; \quad \hat{\sigma}_{\bar{X}_j} = \hat{\sigma}_0 / \sqrt{n} = 0.113;$$

$$\bar{x} - 3 \hat{\sigma}_{\bar{X}_j} = -0.284; \quad \bar{x} + 3 \hat{\sigma}_{\bar{X}_j} = 0.394.$$

La carta  $\bar{X}$  evidenzia una sostanziale stabilità del processo di foratura.



**15.19** Con i dati disponibili calcoliamo i limiti di controllo della carta  $\bar{X}$  :

$$\bar{x} = \sum_{j=1}^{50} \sum_{i=1}^5 x_{ij} / 250 = 502.5; \quad \hat{\sigma}_0 = \sqrt{\sum_{j=1}^{50} S_j^2 / 50} = 2.5; \quad \hat{\sigma}_{\bar{X}_j} = \hat{\sigma}_0 / \sqrt{5} = 1.1;$$

$$\bar{x} - 3 \hat{\sigma}_{\bar{X}_j} = 499.2; \quad \bar{x} + 3 \hat{\sigma}_{\bar{X}_j} = 505.8.$$

L'efficacia della carta è valutata mediante la probabilità:

$$1 - \beta = 1 - \Pr \left\{ \frac{502.5 - 3 \times 2.5 / \sqrt{5} - (502.5 + 5)}{2.5 / \sqrt{5}} \leq U \leq \frac{502.5 + 3 \times 2.5 / \sqrt{5} - (502.5 + 5)}{2.5 / \sqrt{5}} \right\} =$$

$$= 1 - \Pr \{-7.47 \leq U \leq -1.47\} = 1 - 0.0708 = 0.9292.$$

**15.20** La carta di controllo  $p$  consente di monitorare la frazione di registrazioni incomplete. Poiché la numerosità campionaria non è costante è necessario adottare una carta  $p$  a limiti variabili in cui:

$$\hat{p}_j = d_j / n_j = \{0.771, 0.766, 0.788, 0.554, 0.546, 0.514, \\ 0.544, 0.400, 0.455, 0.297, 0.322, 0.500\};$$

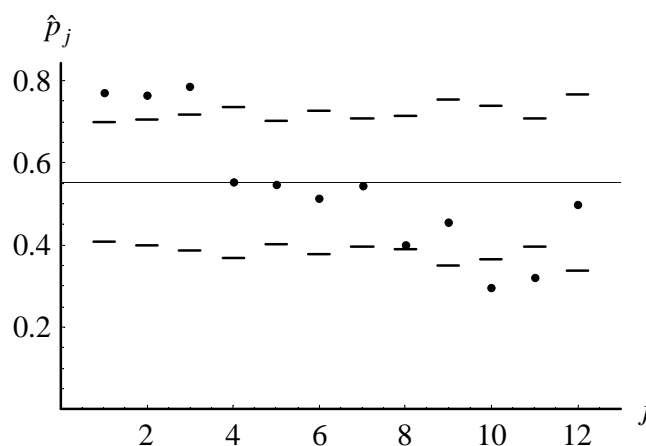
$$\hat{\sigma}_{\hat{p}_j} = \sqrt{\bar{p}(1-\bar{p})/n_j} = \{0.049, 0.051, 0.056, 0.062, 0.050, 0.059, \\ 0.052, 0.054, 0.067, 0.062, 0.052, 0.072\};$$

$$\bar{p} = \frac{\sum_{j=1}^m d_j}{\sum_{j=1}^m n_j} = 0.552;$$

$$\bar{p} + 3 \hat{\sigma}_{\hat{p}_j} = \{0.698, 0.706, 0.719, 0.737, 0.704, 0.728, \\ 0.710, 0.714, 0.754, 0.739, 0.710, 0.768\};$$

$$\bar{p} - 3 \hat{\sigma}_{\hat{p}_j} = \{0.407, 0.399, 0.386, 0.367, 0.401, 0.377, \\ 0.395, 0.391, 0.351, 0.366, 0.395, 0.337\}.$$

La carta di controllo mostra un'evidente instabilità del processo.



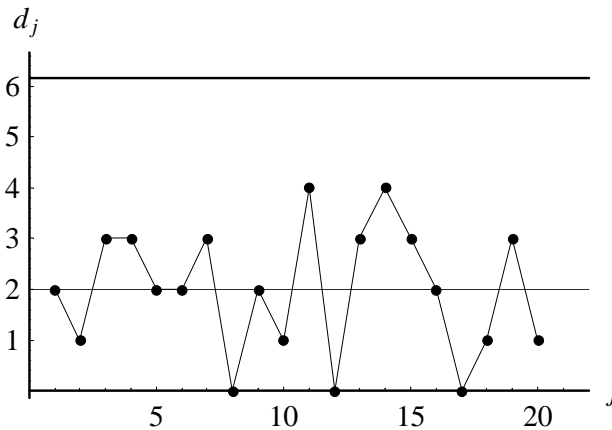
**15.21** La caratteristica di qualità da monitorare è il numero di componenti scartati, a tal fine può essere utilizzata una carta  $np$  (o in maniera equivalente si può monitorare la frazione di pezzi scartati attraverso una carta  $p$ ):

**650** Soluzioni dei problemi proposti

$$\bar{p} = \sum_{j=1}^m d_j / nm = 0.033; \quad \hat{\sigma}_{d_j} = \sqrt{n\bar{p}(1-\bar{p})} = 1.384;$$

$$n\bar{p} = 1.980; \quad n\bar{p} - 3\hat{\sigma}_{d_j} = -2.172; \quad n\bar{p} + 3\hat{\sigma}_{d_j} = 6.132.$$

La carta  $np$  (con limite inferiore nullo essendo risultata negativa la differenza  $n\bar{p} - 3\hat{\sigma}_{d_j}$ ) evidenzia che il processo è in controllo.



**15.22** Utilizzando la CdF della variabile aleatoria Binomiale di parametri  $n = 60$  e  $p = 0.033 \times (1 + 0.05) = 0.035$ , l'efficacia della carta  $np$  può essere valutata attraverso la probabilità:

$$1 - \beta = 1 - \sum_{d=0}^6 \binom{60}{d} 0.035^d (1 - 0.035)^{60-d} = 0.0049.$$

**15.23** La caratteristica di qualità da monitorare è il numero di difetti. I componenti esaminati hanno superfici differenti pertanto occorre valutare il numero di difetti rispetto ad una stessa unità di riferimento. Per ciascuna parte esaminata calcoliamo il numero medio di imperfezioni per  $cm^2$ ,  $u_j = c_j / n_j$ :

$$\{0.070, 0.060, 0.056, 0.058, 0.058, 0.084, 0.065, 0.064, \\ 0.062, 0.065, 0.065, 0.067, 0.074, 0.075, 0.072, 0.062\}$$

Per monitorare il numero di imperfezioni per  $cm^2$  può essere utilizzata una carta  $u$  con linea centrale e limiti di controllo variabili così definiti:

$$\bar{u} = \sum_{j=1}^m c_j / \sum_{j=1}^m n_j = 0.065;$$

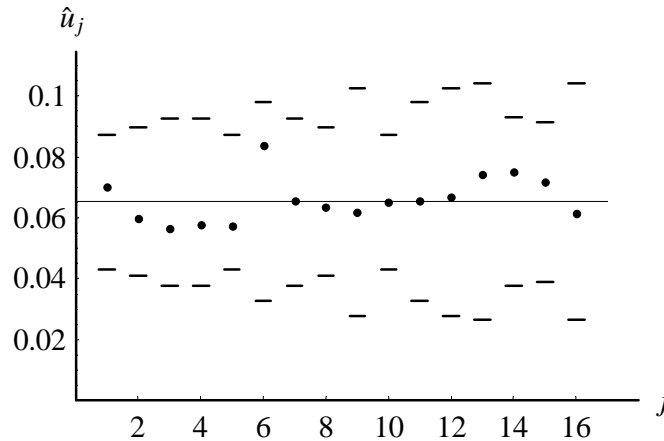
$$\bar{u} - 3\hat{\sigma}_{u_j} = \{0.043, 0.041, 0.038, 0.038, 0.043, 0.033, 0.038, 0.041, \\ 0.028, 0.043, 0.033, 0.028, 0.026, 0.037, 0.039, 0.026\};$$

$$\bar{u} + 3\hat{\sigma}_{u_j} = \{0.087, 0.090, 0.093, 0.093, 0.087, 0.098, 0.093, 0.090, \\ 0.103, 0.087, 0.098, 0.103, 0.104, 0.093, 0.092, 0.104\}$$

essendo:

$$\hat{\sigma}_{u_j} = \sqrt{\bar{u}/n_j} = \{0.007, 0.008, 0.009, 0.009, 0.007, 0.011, 0.009, 0.008, 0.012, 0.007, 0.011, 0.012, 0.013, 0.009, 0.009, 0.013\}.$$

La carta di controllo  $u$  evidenzia che il processo è in controllo.

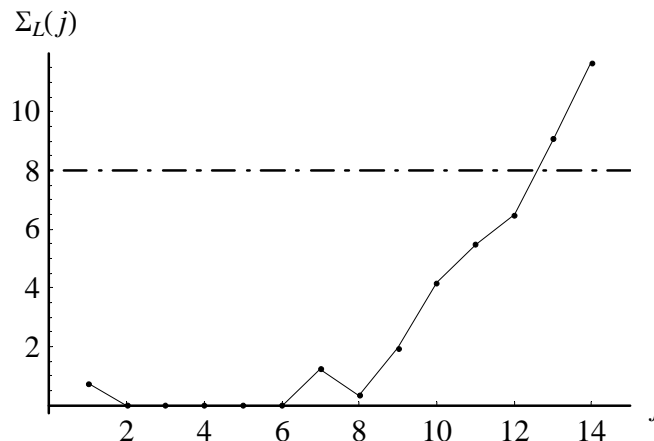


**15.24** Dai dati calcoliamo i valori della somma superiore,  $\sum_H(j)$ , e di quella inferiore,  $\sum_L(j)$ :

$$\sum_H(j) = \{0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0\};$$

$$\sum_L(j) = \{0.73, 0, 0, 0, 0, 0, 1.24, 0.34, 1.94, 4.16, 5.48, 6.48, 9.08, 11.68\}.$$

I valori  $\sum_L(j)$  riportati sulla carta CUSUM unilaterale *inferiore* insieme al limite decisionale,  $L = 4|\Delta\mu| = 8^\circ\text{C}$ , evidenziano una significativa riduzione della temperatura.



**15.25** In condizioni di controllo statistico, la linea centrale e i limiti di controllo della carta  $p$  assumono i seguenti valori:

**652** Soluzioni dei problemi proposti

$$p_0 = 0.22; \quad \sigma_p = \sqrt{p_0(1-p_0)/n} = 0.059;$$

$$p_0 - 3\sigma_p = 0.043; \quad p_0 + 3\sigma_p = 0.397$$

Se dovesse verificarsi un incremento della difettosità del 25%,  $p = 0.22 \times (1 + 0.25) = 0.275$ , per poter individuare tale deriva, sarà necessario, in media, esaminare un numero di campioni pari al periodo di ritorno di un fuori controllo:

$$T_{1-\beta} = \frac{1}{1-\beta} \cong 53$$

essendo

$$\beta = \Pr \left\{ 0.043 \leq \frac{D}{n} \leq 0.397 \mid p = 0.275 \right\} = \Pr \{ 2 \leq D \leq 20 \mid p = 0.275 \} =$$

$$= \sum_{d=2}^{20} \binom{50}{d} 0.275^d (1-0.275)^{50-d} = 0.981.$$

Quindi, poiché la frequenza di campionamento è 30 minuti, l'incremento di difettosità sarà segnalato dalla carta dopo circa 26 ore e 30 minuti dal suo verificarsi.

**15.26** In condizioni di controllo statistico, la linea centrale ed i limiti di controllo della carta  $c$  assumono i seguenti valori:

$$c_0 = 10; \quad \sigma_c = \sqrt{c_0} = 3.16;$$

$$c_0 - 3\sigma_c = 0.51; \quad c_0 + 3\sigma_c = 19.49.$$

Se dovesse verificarsi un incremento della difettosità del 40%,  $c = 10(1 + 0.40) = 14$ , per poter individuare tale deriva, sarà necessario, in media, esaminare un numero di campioni pari al periodo di ritorno di un fuori controllo:

$$T_{1-\beta} = \frac{1}{1-\beta} \cong 21;$$

essendo

$$\beta = \Pr \{ 0 \leq D \leq 20 \mid c = 14 \} = \sum_{d=0}^{20} \frac{c^d e^{-c}}{d!} = 0.952.$$

**15.27** Dalle stime campionarie ricaviamo le stime della media  $\mu_0$  e dello scarto tipo  $\sigma_0$  del processo:

$$\hat{\mu}_0 = 3050.0/25 = 122.0; \quad \hat{\sigma}_0 = \sqrt{28/25} = 1.1.$$

Nell'ipotesi che la CdF della v.a.  $X$  sia ben interpretata da un modello Normale, la stima della capacità del processo è data da:

$$C_{pk} = \min \left\{ \frac{LS - \hat{\mu}_0}{3\hat{\sigma}_0}, \frac{\hat{\mu}_0 - LI}{3\hat{\sigma}_0} \right\} = \min \left\{ \frac{120 \times (1 + 0.03) - 122}{3 \times 1.1}, \frac{122 - 120 \times (1 - 0.03)}{3 \times 1.1} \right\} =$$

$$= \min \{ 0.48, 1.70 \} = 0.48.$$

L'ipotesi di normalità consente di stimare la frazione di fogli non conformi come segue:

$$\begin{aligned}\hat{p} &= 1 - \Pr\{LI \leq X \leq LS\} = 1 - \left[ F_U\left(\frac{LS - \hat{\mu}_0}{\hat{\sigma}_0}\right) - F_U\left(\frac{LI - \hat{\mu}_0}{\hat{\sigma}_0}\right) \right] = \\ &= 1 - [F_U(1.45) - F_U(-5.09)] = 0.0735.\end{aligned}$$

Il valore della frazione di non conformi è elevato, coerentemente alla scarsa capacità del processo.

**15.28** La probabilità di accettare il lotto è calcolata utilizzando la CdF del modello Ipergeometrico ( $N = 100$ ,  $D = 0.05 \times 100 = 5$ ,  $n = 10$ ):

$$P_a = F_K(1) = \sum_{k=0}^1 \frac{\binom{5}{k} \binom{100-5}{10-k}}{\binom{100}{10}} = 0.923.$$

**15.29** Essendo la frazione di campionamento pari a  $n/N = 5/150 = 0.033$  (ossia inferiore al 10%), è possibile approssimare la CdF del modello Ipergeometrico con quella del modello Binomiale. In base alla regola di collaudo, la probabilità di accettazione del lotto può essere espressa come:

$$P_a \cong \left(1 - \frac{D}{N}\right)^5.$$

In corrispondenza di un rischio del cliente pari a  $\beta = 0.10$  il livello di qualità rifiutabile ( $RQL$ ) è approssimativamente pari a:

$$RQL \cong 1 - 0.10^{1/5} = 0.37.$$

**15.30** In base alla regola di collaudo fissata, in corrispondenza di una frazione di difettosi  $D/N = 0.05$ , la probabilità di accettare il lotto è calcolata utilizzando la CdF del modello Ipergeometrico ( $N = 40$ ,  $D = 0.05 \times 40 = 2$ ,  $n = 3$ ):

$$P_a = F_K(0) = \frac{\binom{2}{0} \binom{38}{3}}{\binom{40}{3}} = 0.854.$$

Il rischio del fornitore coincide con la probabilità che il lotto sia rifiutato ed è pari a:

$$\beta = 1 - P_a = 0.146.$$

**15.31** Risolvendo in termini di  $Y$  e  $n$  le due equazioni:

**654** Soluzioni dei problemi proposti

---

$$y \ln\left(\frac{0.01}{0.06}\right) + (n-y) \ln\left(\frac{1-0.01}{1-0.06}\right) = \ln\left(\frac{1-0.05}{0.10}\right);$$

$$y \ln\left(\frac{0.01}{0.06}\right) + (n-y) \ln\left(\frac{1-0.01}{1-0.06}\right) = \ln\left(\frac{0.05}{1-0.10}\right)$$

si ottiene il piano di analisi sequenziale richiesto:

$$y = 1.57 + 0.028 n$$

$$y = -1.22 + 0.028 n.$$

**15.32** In base al piano di analisi sequenziale definito nel problema precedente è necessario continuare a campionare per poter raggiungere un'evidenza sperimentale più elevata.