Corso di Statistica per Ingegneria Fisica
Anno accademico 2020/2021

Esercitazione 5: Variabili Aleatorie Indipendenti e Teorema Centrale del Limite

Esercizio 1. Il signor Ugolino monitora la sua dieta considerando gli apporti calorici, misurati in *kcal*, dei suoi unici due pasti giornalieri, che si possono ritenere variabili aleatorie normali indipendenti.

(a) L'apporto calorico X del pranzo ha media 1000 e il signor Ugolino sa che la probabilità che superi 1100 è pari al 25%. Trovarne la deviazione standard. *[148.15]*

(b) Il signor Ugolino sa anche che la probabilità che l'apporto calorico Y della cena sia inferiore a 800 è pari a 0.3, e la probabilità che sia inferiore a 1300 è pari a 0.8. Trovare valore atteso e deviazione standard dell'apporto calorico della cena. *[$\mu = 992.31$; $\sigma = 366.30$]*

(c) Determinare la densità della variabile aleatoria $Z =$ apporto calorico totale di una giornata del signor Ugolino.

(d) Il fabbisogno calorico giornaliero del signor Ugolino è di 2200 *kcal*. Calcolare la probabilità che l'apporto calorico di una giornata sia inferiore a questo fabbisogno. *[70.0205%]*

Esercizio 2. Sia X_1, X_2, X_3, X_4, X_5 un campione di 5 misure, con media $\mathbb{E}[X_i] =: \mu$ e varianza $\mathrm{Var}(X_i) =: \sigma^2$ per ogni $i = 1, \ldots, 5$.

(a) Trovare un limite superiore per la probabilità che l'i-esima misura disti da μ per più di 2σ.
[$\mathbb{P}(|X_i - \mu| > 2\sigma) \leq 25\%$]

(b) Trovare un limite superiore per la probabilità che la media campionaria delle 5 misure disti da μ per più di 2σ. *[$\mathbb{P}\left(|\bar{X} - \mu| > 2\sigma\right) \leq 5\%$]*

Supponiamo ora di sapere che le 5 misure hanno densità gaussiana.

(c) Calcolare *esattamente* le probabilità dei due punti precedenti.
[$\mathbb{P}(|X_i - \mu| > 2\sigma) = 4.55\%$, $\mathbb{P}\left(|\bar{X} - \mu| > 2\sigma\right) = 2(1 - \Phi(2\sqrt{5})) \simeq 0$]

Esercizio 3. La specialità del cuoco Armando è la salsa di tonno. Nei vasetti da 10 g, la quantità di tonno ideale dovrebbe essere di 6 g. In realtà Armando è un po' impreciso e la quantità di tonno che utilizza è una variabile aleatoria X con densità:

$$f_X(x) = \begin{cases} \alpha x - 5 & \text{se } 5 \leq x < 6 \\ 7 - x & \text{se } 6 \leq x < 7 \\ 0 & \text{altrimenti} \end{cases}$$

(a) Trovare α tale per cui $f_X(x)$ sia una densità di probabilità e disegnarla. *[$\alpha = 1$]*

(b) Trovare la funzione di ripartizione $F_X(x)$ e disegnarla.

(c) Qual è la probabilità che in un vasetto di 10 g ci siano più di 8 g di tonno? E meno di 5.5? *[0;1/8]*

Armando viene contattato da un grosso negozio che ha deciso di puntare sulla sua salsa di tonno. Per questo vengono ordinati ad Armando 70 vasetti da 10 g.

(d) Calcolare la probabilità, eventualmente approssimata, che Armando debba utilizzare in tutto più di 425 g di tonno. *[0.072145]*

Esercizio 4. La percentuale di realizzazione nei tiri da due punti di un giocatore di pallacanestro è del 55%. Si calcolino:

(a) la probabilità che segni non più di 50 punti in 50 tiri; *[0.285]*

(b) il numero minimo di tiri che deve effettuare affinché la probabilità di segnare almeno 100 punti sia maggiore o uguale a 0.9. *[Con la correzione di continuità: $n \geq 102$. Senza correzione: $n \geq 103$.]*

Esercizio 5. Tre diversi tipi di allergia hanno colpito una popolazione. L'allergia di tipo A ha colpito il 20% della popolazione, l'allergia di tipo B ha colpito l'1% della popolazione, l'allergia di tipo C ha colpito il 95% della popolazione. Su un campione di 60 persone,

(a) con quale probabilità ci sono almeno 10 allergici di tipo A? *[Utilizzando il TCL si ottiene 79.013%.]*

(b) E almeno 3 allergici di tipo B? *[Qui non è possibile utilizzare il TCL. Bisogna calcolare la probabilità esatta con la binomiale (= 2.24202%) oppure utilizzare l'approssimazione tramite la Poisson (= 2.31153%)]*

(c) E almeno 58 allergici di tipo C? *[Anche in questo caso bisogna calcolare il risultato tramite la densità esatta (= 41.74358%) oppure utilizzando l'approssimazione di Poisson per un'opportuna variabile aleatoria (= 42.31901%). Quale?]*

Esercizio 6. Nel compilare il modello 730 il contribuente deve arrotondare ogni spesa all'intero più vicino. Supponiamo di modellizzare l'arrotondamento X, espresso in euro, come una variabile aleatoria uniforme sull'intervallo $(-0.5, 0.5)$.

(a) Si scriva la densità della variabile aleatoria X. *[$f_X(x) = \mathbf{1}_{(-1/2,1/2)}(x)$]*

(b) Si calcolino media e varianza della variabile aleatoria X. *[$\mathbb{E}[X] = 0$, $\mathrm{Var}(X) = 1/12$]*

(c) Si supponga ora di avere 50 voci di spesa nel modello 730 e di sommare i valori già arrotondati di ogni singola spesa. Calcolare, in modo approssimato, la probabilità che il numero ottenuto differisca dalla somma dei valori esatti per più di 3 euro. *[0.14156]*

Esercizio 7. Alla centrale del 118 di un piccolo paese, il numero di chiamate durante un'ora qualsiasi si può modellizzare con una variabile aleatoria di Poisson con media pari a 1.2 chiamate. Inoltre, i numeri di chiamate che arrivano in ore diverse sono tutti indipendenti tra loro.

(a) Qual è la probabilità che domani dalle 9:00 alle 10:00 non arrivi nessuna chiamata? *[30.119%]*

(b) Qual è invece la probabilità che domani dalle 9:00 alle 11:00 arrivino al massimo due chiamate? *[56.971%]*

(c) E qual è infine la probabilità che sempre domani, ma dalle 9:00 alle 24:00, arrivino più di 20 chiamate? *[27.760%]*

(d) Qual è il numero medio di chiamate che arrivano alla centrale nell'arco di 24 ore? *[28.8]*

Esercizio 8. Le Grand Bateau è un battello a due motori utilizzato per le crociere sulla Senna. I due motori lavorano indipendentemente e per ciascuno di loro il numero di guasti in una singola crociera è rappresentabile tramite una variabile aleatoria di Poisson, di media 0.1 per il primo motore e di media 0.2 per il secondo.

(a) Calcolare la probabilità che durante una crociera non si guasti il primo motore, la probabilità che non si guasti il secondo e la probabilità che non si guasti nessuno dei due motori. *[$\mathrm{e}^{-1/10}$; $\mathrm{e}^{-1/5}$; $\mathrm{e}^{-3/10}$]*

(b) Calcolare la probabilità che durante una crociera avvenga esattamente un guasto. *[$0.3\,\mathrm{e}^{-3/10}$]*

(c) Su 10 crociere del Gran Bateau, calcolare la probabilità che almeno 8 si concludano senza nessun guasto. *[0.49820]*

(d) Ancora su 10 crociere, calcolare la probabilità che si verifichino in totale almeno 2 guasti. (Attenzione: i due guasti possono anche verificarsi tutti nella stessa crociera!) *[$= 80.085\%$]*

(e) Su 100 crociere, calcolare la probabilità che si verifichino in totale almeno 20 guasti. (Stessa avvertenza del punto precedente.) *[$= 97.257\%$]*

Esercizio 9. Sia X la variabile aleatoria che rappresenta la durata (espressa in ore) di una lampadina Oscar. Supponiamo che la densità di X sia data dalla funzione:

$$f_X(x) = \begin{cases} 0 & \text{se } x < 0 \\ \frac{1}{100}\,\mathrm{e}^{-\frac{x}{100}} & \text{se } x \geq 0\,. \end{cases}$$

(a) Calcolare media, varianza e deviazione standard del tempo di vita di una lampadina Oscar *[$\mathbb{E}[X] = 100$, $\mathrm{Var}(X) = 100^2$, $\sqrt{\mathrm{Var}(X)} = 100$]*

Si supponga ora di impiegare n lampadine Oscar in sequenza, accendendone cioè una alla volta e sostituendo ogni lampadina appena si brucia.

(b) Determinare la densità, anche approssimata, del numero di ore T di luce che si ottiene per n "molto grande". *[$T \approx N(100n, 100^2 n)$]*

(c) Usando l'approssimazione del punto precedente, calcolare la probabilità di ottenere luce per un anno intero (cioè 8˙760 ore) adoperando al massimo 80 lampadine. *[19.766%]*

(d) Usando ancora la stessa approssimazione, calcolare il numero minimo di lampadine necessarie per avere almeno 5˙000 ore di luce con probabilità maggiore uguale a 0.95. *[64]*

Esercizio 10. Il lavaggio completo di un'autovettura richiede due fasi successive, delle quali la prima ha una durata X di legge esponenziale di media 0.2 ore, mentre la seconda ha una durata Y, indipendente dalla precedente, di legge esponenziale di media 0.3 ore.

(a) Calcolare la media μ e la varianza σ^2 della durata T di un lavaggio completo. *[0.5; $13/100$]*

(b) Se l'officina deve eseguire il lavaggio completo di 30 autovetture, calcolare, in modo approssimato, la probabilità che l'intero lavoro venga completato in meno di 12 ore. *[0.064256]*

(c) Determinare il tempo minimo t nel quale l'officina riesce a completare il lavaggio delle 30 autovetture con probabilità maggiore o uguale a 0.95. *[$t \geq 15 + 1.645\sqrt{3.9} \simeq 18.25$]*

Esercizio 11. Un ingegnere civile costruisce un ponte che può sopportare un peso massimo di 100 tonnellate. Si supponga che il peso (espresso in tonn.) di un'automobile sia una variabile aleatoria di media 1 e deviazione standard 0.2. Quante auto devono transitare contemporaneamente sul ponte affinché ci sia una probabilità superiore al 5% che si ecceda il peso massimo sopportato? *[97]*

Esercizio 12. La distanza d di una stella misurata in parsec (pc) è calcolata come la media di una serie di misurazioni indipendenti e identicamente distribuite con media d e varianza 4. Quante osservazioni sono necessarie per essere sicuri al 95% che la media campionaria delle osservazioni approssimi d entro 0.5 pc? *[62]*

Soluzioni di alcuni esercizi

Soluzione 1.

Introduciamo le v.a.

$$X = \text{apporto calorico del pranzo} \qquad Y = \text{apporto calorico della cena}\,.$$

Le ipotesi sono

$$X \sim N(\mu_X, \sigma_X^2) \qquad Y \sim N(\mu_Y, \sigma_Y^2) \qquad X \text{ e } Y \text{ indipendenti}\,.$$

(a) Sappiamo che $\mu_X = 1000$ e $\mathbb{P}(X > 1100) = 0.25$. Quindi

$$0.25 = \mathbb{P}(X > 1100) = \mathbb{P}\left(\frac{X - \mu_X}{\sigma_X} > \frac{1100 - 1000}{\sigma_X}\right) = 1 - \Phi\left(\frac{1100 - 1000}{\sigma_X}\right) = 1 - \Phi\left(\frac{100}{\sigma_X}\right)$$
$$\Rightarrow \quad \Phi\left(\frac{100}{\sigma_X}\right) = 0.75 \quad \Rightarrow \quad \frac{100}{\sigma_X} = z_{0.75} \simeq 0.675$$
$$\Rightarrow \quad \sigma_X = 148.15\,.$$

Per calcolare $z_{0.75} \simeq 0.675$, abbiamo interpolato tra i due quantili più vicini al 75% sulle tavole della normale standard, e cioè $0.67 = z_{0.74857}$ e $0.68 = z_{0.75175}$.

(b) Sappiamo che $\mathbb{P}(Y < 800) = 0.3$ e $\mathbb{P}(Y < 1300) = 0.8$. Quindi

$$0.3 = \mathbb{P}(Y < 800) = \mathbb{P}\left(\frac{Y - \mu_Y}{\sigma_Y} < \frac{800 - \mu_Y}{\sigma_Y}\right) = \Phi\left(\frac{800 - \mu_Y}{\sigma_Y}\right)$$
$$\Rightarrow \quad \frac{800 - \mu_Y}{\sigma_Y} = z_{0.3} = -z_{1-0.3} = -z_{0.7} \simeq -0.525$$
$$0.8 = \mathbb{P}(Y < 1300) = \ldots$$
$$\Rightarrow \quad \frac{1300 - \mu_Y}{\sigma_Y} = z_{0.8} \simeq 0.84\,.$$

Si osservi che il quantile $z_{0.3}$ non è tabulato (le tavole della normale partono dall'ordine 50% in poi). Tuttavia, per la simmetria della normale standard, possiamo ricavarlo usando la relazione $z_\gamma = -z_{1-\gamma}$.

Ora mettiamo a sistema

$$\begin{cases} \frac{800 - \mu_Y}{\sigma_Y} = -0.525 \\ \frac{1300 - \mu_Y}{\sigma_Y} = 0.84 \end{cases} \qquad \Rightarrow \qquad \mu_Y = 992.31\,,\ \sigma_Y = 366.30\,.$$

(c) Abbiamo

$$Z = X + Y \sim N(\mu_Z, \sigma_Z^2) \qquad (\text{con } \mu_Z,\, \sigma_Z^2 \text{ parametri da determinare})\,,$$

in quanto X e Y sono *gaussiane e indipendenti*, e quindi è gaussiana anche qualunque loro combinazione lineare affine. Calcoliamo la media e la varianza di Z:

$$\mu_Z = \mathbb{E}[Z] = \mathbb{E}[X + Y] \underset{\substack{\text{linearità} \\ \text{di } \mathbb{E}}}{=} \mathbb{E}[X] + \mathbb{E}[Y] = 1000 + 992.31 = 1992.31$$
$$\sigma_Z^2 = \operatorname{Var}(Z) = \operatorname{Var}(X + Y) \underset{\substack{\text{quadraticità} \\ \text{di Var}}}{=} \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X, Y)$$
$$\underset{\substack{\text{indipendenza} \\ \text{di } X \text{ e } Y}}{=} \operatorname{Var}(X) + \operatorname{Var}(Y) = 148.15^2 + 366.30^2 = 156050.86\,.$$

In conclusione,

$$Z \sim N(1992.31,\, 156050.86)\,.$$

(d) La probabilità cercata è

$$\mathbb{P}(Z < 2200) = \mathbb{P}\left(\frac{Z-\mu_Z}{\sigma_Z} < \frac{2200-1992.31}{\sqrt{156050.86}}\right) = \Phi\left(\frac{2200-1992.31}{\sqrt{156050.86}}\right) = \Phi(0.526)$$
$$\simeq \frac{0.69847+0.70194}{2} = 0.700205 = 70.0205\%\,.$$

Soluzione 2.

1. Abbiamo l'uguaglianza di eventi

$$\text{“l'}i\text{-esima misura dista da } \mu \text{ per più di } 2\sigma\text{”} = \text{“}|X_i - \mu| > 2\sigma\text{”}$$

(la *distanza* è il modulo della differenza). La probabilità richiesta è pertanto $\mathbb{P}(|X_i-\mu|>2\sigma)$, e, se conoscessimo la densità delle X_i, potremmo calcolarla usando la solita formula

$$\mathbb{P}(|X_i-\mu|>2\sigma) = \begin{cases} \int_{-\infty}^{\mu-2\sigma} f_{X_i}(t)\mathrm{d}t + \int_{\mu+2\sigma}^{+\infty} f_{X_i}(t)\mathrm{d}t & \text{se } X_i \text{ è assolutamente continua} \\ \sum_{k<\mu-2\sigma} p_{X_i}(k) + \sum_{k>\mu+2\sigma} p_{X_i}(k) & \text{se } X_i \text{ è discreta}\,. \end{cases}$$

Ma non conosciamo la densità delle X_i: l'esercizio ci dà solo la loro media e varianza. Pertanto, non potremo trovare in nessun modo il valore esatto di $\mathbb{P}(|X_i-\mu|>2\sigma)$. Tuttavia, grazie alla disuguaglianza di Chebyshev, possiamo comunque dare un limite superiore per $\mathbb{P}(|X_i-\mu|>2\sigma)$. Infatti, Chebyshev dice che

$$\mathbb{P}(|X_i-\mathbb{E}[X_i]|>\varepsilon) \leq \frac{\mathrm{Var}(X_i)}{\varepsilon^2}\,,$$

e, poichè $\mathbb{E}[X_i]=\mu$ e $\mathrm{Var}(X_i)=\sigma^2$, sostituendo $\varepsilon = 2\sigma$ troviamo

$$\mathbb{P}(|X_i-\mu|>2\sigma) \leq \frac{\sigma^2}{(2\sigma)^2} = \frac{1}{4} = 25\%\,,$$

cioè $\mathbb{P}(|X_i-\mu|>2\sigma) \leq 25\%$.

2. Ora, non sapendo la densità delle X_i, a maggior ragione non sappiamo nemmeno quella della loro media campionaria $\bar{X}$. Infatti, solo $n=5$ misure sono troppo poche per applicare il TLC e concludere che $\bar{X}$ ha approssimativamente densità normale. Tuttavia, possiamo di nuovo applicare Chebishev, questa volta a $\bar{X}$, ricordando che

$$\mathbb{E}\left[\bar{X}\right] = \mathbb{E}[X_i] = \mu \qquad\qquad \mathrm{Var}\left(\bar{X}\right) = \frac{\mathrm{Var}(X_i)}{n} = \frac{\sigma^2}{5}\,.$$

La disuguaglianza di Chebyshev dà pertanto

$$\mathbb{P}\left(\left|\bar{X}-\mu\right|>2\sigma\right) = \mathbb{P}\left(\left|\bar{X}-\mathbb{E}\left[\bar{X}\right]\right|>2\sigma\right) \leq \frac{\mathrm{Var}\left(\bar{X}\right)}{(2\sigma)^2} = \frac{\frac{\sigma^2}{5}}{4\sigma^2} = \frac{1}{20} = 5\%,$$

cioè $\mathbb{P}\left(\left|\bar{X}-\mu\right|>2\sigma\right) \leq 5\%$.

3. Adesso l'esercizio ci dice qual è la densità delle X_i, cioè

$$X_i \sim N(\mu,\sigma^2)$$

(media e varianza di X_i sono le stesse μ e σ^2 dei punti precedenti). Di conseguanza, possiamo calcolare *esattamente* le probabilità $\mathbb{P}\left(|X_i - \mu| > 2\sigma\right)$ e $\mathbb{P}\left(\left|\bar{X} - \mu\right| > 2\sigma\right)$. Per la prima,

$$
\begin{aligned}
\mathbb{P}\left(|X_i - \mu| > 2\sigma\right) &= \mathbb{P}\left(\text{“}X_i - \mu < -2\sigma\text{”} \vee \text{“}X_i - \mu > +2\sigma\text{”}\right) \\
&= \mathbb{P}\left(X_i - \mu < -2\sigma\right) + \mathbb{P}\left(X_i - \mu > +2\sigma\right) && \begin{array}{l}3^\circ \text{ assioma della probabilità, perchè} \\ \text{“}X_i - \mu < -2\sigma\text{” e “}X_i - \mu > +2\sigma\text{”} \\ \text{sono eventi incompatibili}\end{array} \\
&= \mathbb{P}\left(\frac{X_i - \mathbb{E}\left[X_i\right]}{\sqrt{\operatorname{Var}\left(X_i\right)}} < \frac{-2\sigma}{\sqrt{\sigma^2}}\right) + \mathbb{P}\left(\frac{X_i - \mathbb{E}\left[X_i\right]}{\sqrt{\operatorname{Var}\left(X_i\right)}} > \frac{+2\sigma}{\sqrt{\sigma^2}}\right) && \text{standardizzazione di } X_i \\
&= \Phi\left(\frac{-2\sigma}{\sigma}\right) + \left[1 - \Phi\left(\frac{+2\sigma}{\sigma}\right)\right] = \Phi(-2) + 1 - \Phi(2) \\
&= 2\left(1 - \Phi(2)\right) = 2\left(1 - 0.97725\right) \\
&= 4.55\% \,.
\end{aligned}
$$

Per la seconda probabilità, osserviamo che per un campione gaussiano si ha *sempre* e *esattamente*

$$
\bar{X} \sim N\left(\mu \,, \frac{\sigma^2}{n}\right) \,,
$$

indipendentemente dal fatto che n sia grande o piccolo (è conseguenza del fatto che la somma di v.a. gaussiane indipendenti ha ancora densità gaussiana). Quindi,

$$
\begin{aligned}
\mathbb{P}\left(\left|\bar{X} - \mu\right| > 2\sigma\right) &= \mathbb{P}\left(\bar{X} - \mu < -2\sigma\right) + \mathbb{P}\left(\bar{X} - \mu > +2\sigma\right) && \text{come prima} \\
&= \mathbb{P}\left(\frac{\bar{X} - \mathbb{E}\left[\bar{X}\right]}{\sqrt{\operatorname{Var}\left(\bar{X}\right)}} < \frac{-2\sigma}{\sqrt{\frac{\sigma^2}{5}}}\right) + \mathbb{P}\left(\frac{\bar{X} - \mathbb{E}\left[\bar{X}\right]}{\sqrt{\operatorname{Var}\left(\bar{X}\right)}} > \frac{+2\sigma}{\sqrt{\frac{\sigma^2}{5}}}\right) && \text{standardizzazione di } \bar{X} \\
&= \Phi\left(\frac{-2\sigma}{\sigma/\sqrt{5}}\right) + \left[1 - \Phi\left(\frac{+2\sigma}{\sigma/\sqrt{5}}\right)\right] = \Phi(-2\sqrt{5}) + 1 - \Phi(2\sqrt{5}) \\
&= 2\left(1 - \Phi(2\sqrt{5})\right) = 2\left(1 - \Phi(4.47)\right) \simeq 2\left(1 - 1\right) \\
&= 0 \,.
\end{aligned}
$$

Soluzione 5.

(a) Introduciamo la v.a.

$$
X = \text{numero di allergici di tipo } A \text{ tra le 60 persone del campione} \,.
$$

Allora, X conta il numero di successi in una serie di $n = 60$ prove di Bernoulli (dove si ha successo all'i-esima prova se l'i-esima persona risulta allergica), ciascuna con probabilità di successo pari a $p = 0.20$. Pertanto, $X \sim B(60,\, 0.20)$. Tuttavia, poiché $n > 20$, $np = 12 > 5$ e $n(1 - p) = 48 > 5$, si può usare l'approssimazione normale della binomiale $B(n, p) \simeq N(np, np(1 - p)$, e concludere che

$$
X \sim B(60,\, 0.20) \simeq N(60 \cdot 0.20 \,, 60 \cdot 0.20 \cdot (1 - 0.20)) = N(12 \,, 9.6) \,.
$$

La probabilità cercata è pertanto

$$\begin{aligned}
\mathbb{P}(X \geq 10) &= \mathbb{P}(X \geq 9.5) && \text{correzione di continuità} \\
&= \mathbb{P}\left(\underbrace{\frac{X - \mathbb{E}[X]}{\sqrt{\operatorname{Var}(X)}}}_{\approx N(0,1)} \geq \frac{9.5 - 12}{\sqrt{9.6}}\right) && \text{TLC} \\
&= 1 - \Phi\left(\frac{9.5 - 12}{\sqrt{9.6}}\right) = 1 - \Phi(-0.807) = \Phi(0.807) = 0.79013 = 79.013\% \,.
\end{aligned}$$

(b) La v.a.

$$Y = \text{numero di allergici di tipo } B \text{ tra le 60 persone del campione}$$

ha di nuovo densità binomiale con $n = 60$, ma questa volta la probabilità di successo in ciascuna delle 60 prove è $p = 0.01$. In modo simile a prima, $Y \sim B(60,\, 0.01)$, n è ancora maggiore di 20, ma adesso $np = 0.6 < 5$, e quindi non possiamo usare l'approssimazione normale. Quindi, o facciamo il calcolo *esatto* con la binomiale, o usiamo l'approssimazione di Poisson $B(n, p) \simeq \mathcal{P}(np)$, in quanto n è grande e $np = 0.6 \simeq 1$. Nel primo caso,

$$\begin{aligned}
\mathbb{P}(Y \geq 3) &= 1 - \mathbb{P}(Y \leq 2) \\
&= 1 - \left[\underbrace{\binom{60}{0}}_{1} 0.01^0 (1 - 0.01)^{60-0} + \underbrace{\binom{60}{1}}_{60} 0.01^1 (1 - 0.01)^{60-1} + \underbrace{\binom{60}{2}}_{(60 \cdot 59)/(2 \cdot 1)} 0.01^2 (1 - 0.01)^{60-2}\right] \\
&= 0.0224202 = 2.24202\% \,.
\end{aligned}$$

Nel secondo, con $B(60,\, 0.01) \simeq \mathcal{P}(0.6)$,

$$\begin{aligned}
\mathbb{P}(Y \geq 3) = 1 - \mathbb{P}(Y \leq 2) &\simeq 1 - \mathrm{e}^{-0.6}\left(\underbrace{\frac{0.6^0}{0!}}_{1/1} + \underbrace{\frac{0.6^1}{1!}}_{0.6/1} + \underbrace{\frac{0.6^2}{2!}}_{0.6^2/2}\right) \\
&= 0.0231153 = 2.31153\% \,.
\end{aligned}$$

(c) Ora, la v.a.

$$Z = \text{numero di allergici di tipo } C \text{ tra le 60 persone del campione}$$

ha sempre densità di tipo binomiale, cioè $Z \sim B(60,\, 0.95)$. Tuttavia,

$$\begin{aligned}
np = 57 \gg 1 \quad &\Rightarrow \quad \text{non si può usare l'approssimazione di Poisson} \\
n(1 - p) = 3 < 5 \quad &\Rightarrow \quad \text{non si può usare l'approssimazione gaussiana} \,.
\end{aligned}$$

Non resta altro da fare che il calcolo esatto:

$$\begin{aligned}
\mathbb{P}(Z \geq 58) &= \underbrace{\binom{60}{58}}_{=\binom{60}{2}} 0.95^{58} (1 - 0.95)^{60-58} + \underbrace{\binom{60}{59}}_{=\binom{60}{1}} 0.95^{59} (1 - 0.95)^{60-59} + \underbrace{\binom{60}{60}}_{=\binom{60}{0}} 0.95^{60} (1 - 0.95)^{60-60} \\
&= 0.4174358 = 41.74358\% \,.
\end{aligned}$$

Alternativamente, se vogliamo semplificare i calcoli usando qualche approssimazione, possiamo osservare che la nuova v.a.

$$T = \text{numero di } \textit{non} \text{ allergici di tipo } C \text{ tra i } 60 = 60 - Z$$

ha sempre densità binomiale con $n = 60$, ma la probabilità di successo a ogni prova è ora $q = 1 - p = 0.05$. Poiché $nq = 3 \simeq 1$, per T vale l'approssimazione di Poisson $T \sim B(60,\, 0.05) \simeq \mathcal{P}(60 \cdot 0.05) = \mathcal{P}(3)$. La probabilità cercata si può quindi trovare anche come

$$
\begin{aligned}
\mathbb{P}\,(Z \geq 58) &= \mathbb{P}\,(-Z \leq -58) = \mathbb{P}\,(60 - Z \leq 60 - 58) = \mathbb{P}\,(T \leq 2) \\
&\simeq \mathrm{e}^{-3}\left(1 + 3 + \frac{3^2}{2}\right) = 0.4231901 = 42.31901\%\,.
\end{aligned}
$$

Soluzione 6.

(a) La densità uniforme continua sull'intervallo (a, b) è $f(x) = \frac{1}{b-a}\mathbb{1}_{[a,b]}(x)$, dunque con $a = -0.5$, $b = 0.5$ abbiamo

$$
f_X(x) = \frac{1}{0.5 - (-0.5)}\mathbb{1}_{[-0.5,0.5]}(x) = \mathbb{1}_{[-0.5,0.5]}(x) = \begin{cases} 1 & \text{se } x \in [-0.5, 0.5] \\ 0 & \text{altrimenti}\,. \end{cases}
$$

(b) Calcoliamo

$$
\begin{aligned}
\mathbb{E}\,[X] &= \int_{-\infty}^{+\infty} x\, f_X(x)\mathrm{d}x = 0 \qquad \text{perché } f_X(-x) = f_X(x) \text{ (simmetria rispetto a } x = 0) \\
\mathbb{E}\left[X^2\right] &= \int_{-\infty}^{+\infty} x^2\, f_X(x)\mathrm{d}x = \int_{-0.5}^{+0.5} x^2 \mathrm{d}x = \left[\frac{x^3}{3}\right]_{x=-0.5}^{x=+0.5} = \frac{1}{12} \\
\mathrm{Var}\,(X) &= \mathbb{E}\left[X^2\right] - \mathbb{E}\,[X]^2 = \frac{1}{12} - 0^2 = \frac{1}{12}\,.
\end{aligned}
$$

Avremmo anche potuto copiare direttamente dal formulario:

$$
\mathbb{E}\,[X] = \frac{a+b}{2} = \frac{-0.5 + 0.5}{2} = 0 \qquad\qquad \mathrm{Var}\,(X) = \frac{(b-a)^2}{12} = \frac{[0.5 - (-0.5)]^2}{12} = \frac{1}{12}\,.
$$

(c) Ora, abbiamo 50 voci di spesa, ognuna delle quali differisce dal corrispondente valore esatto c_i in misura del corrispondente arrotondamento X_i. In altre parole, nell'i-esima voce di spesa non abbiamo il valore esatto c_i, bensì quello approssimato $C_i = c_i + X_i$. L'esercizio ci dice che possiamo assumere le v.a. $X_1, X_2, \ldots, X_{50}$ tutte *identicamente distibuite* con la densità trovata al punto (a). L'ipotesi (ragionevole!) che possiamo fare in più noi è che gli arrotondamenti in voci di spesa diverse non si influenzino tra loro; in altre parole, possiamo supporre che le $X_1, X_2, \ldots, X_{50}$ siano anche *indipendenti*. In questo caso, per il TLC la v.a.

$$
S_{50} = \text{somma dei 50 arrotondamenti} = X_1 + X_2 + \ldots + X_{50}
$$

ha approssimativamente densità normale

$$
S_{50} \approx N(50 \cdot \mathbb{E}\,[X_1]\;,\; 50 \cdot \mathrm{Var}\,(X_1)) = N\left(50 \cdot 0\,,\; 50 \cdot \frac{1}{12}\right) = N\,(0\,,\; 4.1667)\;.
$$

Quindi,

$$
\begin{aligned}
&\mathbb{P}\,(\text{“la somma dei valori arrotondati differisce dalla somma dei valori esatti per più di } 3\,€\text{”})\\
&= \mathbb{P}\left(\left|\sum_{i=1}^{50} C_i - \sum_{i=1}^{50} c_i\right| > 3\right) = \mathbb{P}\left(\left|\sum_{i=1}^{50} X_i\right| > 3\right) = \mathbb{P}\,(|S_{50}| > 3)\\
&= 1 - \mathbb{P}\,(|S_{50}| \leq 3) = 1 - \mathbb{P}\,(-3 \leq S_{50} \leq +3)\\
&= 1 - \mathbb{P}\left(\frac{-3-0}{\sqrt{4.1667}} \leq \underbrace{\frac{S_{50} - \mathbb{E}\,[S_{50}]}{\sqrt{\operatorname{Var}(S_{50})}}}_{\approx N(0,1)} \leq \frac{+3-0}{\sqrt{4.1667}}\right) \qquad \text{TLC}\\
&= 1 - \left[\Phi\left(\frac{+3-0}{\sqrt{4.1667}}\right) - \Phi\left(\frac{-3-0}{\sqrt{4.1667}}\right)\right] = 1 - [\Phi(1.470) - \Phi(-1.470)]\\
&= 1 - \{\Phi(1.470) - [1 - \Phi(1.470)]\} = 2[1 - \Phi(1.470)] = 2(1 - 0.92922)\\
&= 0.14156 = 14.56\%\,.
\end{aligned}
$$

Soluzione 7.

(a) Intoduciamo la v.a.

$$X_1 = \text{numero di chiamate che arriveranno domani tra le 9:00 e le 10:00}\,.$$

Sappiamo dal testo che

$$X_1 \sim \mathcal{P}(1.2)\,.$$

Di conseguenza,

$$
\begin{aligned}
&\mathbb{P}\,(\text{“domani non arriva nessuna chiamata dalle 9:00 alle 10:00”}) = \mathbb{P}\,(X_1 = 0)\\
&\qquad = \mathrm{e}^{-1.2}\,\frac{1.2^0}{0!} = \mathrm{e}^{-1.2} = 0.30119 = 30.119\%\,.
\end{aligned}
$$

(b) Ora introduciamo la nuova v.a.

$$X_2 = \text{numero di chiamate che arriveranno domani tra le 10:00 e le 11:00}\,.$$

Il testo dell'esercizio ci dice che anche

$$X_2 \sim \mathcal{P}(1.2)\,.$$

Inoltre, possiamo fare noi l'ipotesi aggiuntiva che le due v.a. X_1 e X_2 siano indipendenti (cioè supporre che il numero di chiamate che arrivano in due ore *diverse e non sovrapposte* non si influenzino tra loro). Di conseguenza, la v.a.

$$S_2 = X_1 + X_2 = \text{numero di chiamate che arriveranno domani tra le 9:00 e le 11:00}$$

ha ancora densità di Poisson, essendo la somma di due v.a. poissoniane e indipendenti. La sua media è

$$\mathbb{E}\,[S_2] = \mathbb{E}\,[X_1 + X_2] = \mathbb{E}\,[X_1] + \mathbb{E}\,[X_2] = 1.2 + 1.2 = 2.4\,,$$

e per una poissoniana sappiamo che coincide col (l'unico) parametro della densità. Di conseguenza,

$$S_2 \sim \mathcal{P}(2.4)\,,$$

e la probabilità cercata è

$$
\begin{aligned}
\mathbb{P}\left(\text{“domani dalle 9:00 alle 11:00 arrivano al massimo 2 chiamate”}\right) &= \mathbb{P}\left(S_2 \leq 2\right) \\
&= \mathrm{e}^{-2.4}\left(1 + 2.4 + \frac{2.4^2}{2!}\right) = 0.56971 = 56.971\% \,.
\end{aligned}
$$

(c) Ora, introduciamo ancora le nuove v.a.

$$X_i = \text{numero di chiamate che arriveranno domani tra le } [9 + (i-1)]\text{:00 e le } [10 + (i-1)]\text{:00} \,.$$

per $i = 3, 4, \ldots, 15$ (dalle 9:00 alle 24:00 passano 15 ore piene; le v.a. X_1 e X_2 sono quelle che abbiamo già visto ai punti precedenti). Come prima, le $X_1, X_2, \ldots, X_{15}$ hanno tutte la stessa densità

$$X_i \sim \mathcal{P}(1.2) \qquad \forall i = 1, \ldots, 15 \,,$$

e inoltre si possono considerare indipendenti. In altre parole, la successione $X_1, X_2, \ldots, X_{15}$ costituisce un *campione aleatorio*. Per lo stesso motivo di prima, la v.a.

$$S_{15} = X_1 + X_2 + \ldots + X_{15} = \text{numero di chiamate che arriveranno domani tra le 9:00 e le 24:00}$$

ha densità poissoniana

$$S_{15} \sim \mathcal{P}(\underbrace{1.2 + 1.2 + \ldots + 1.2}_{\text{15 volte}}) = \mathcal{P}(18) \,,$$

e la probabilità richiesta è

$$\mathbb{P}\left(S_{15} > 20\right) = \sum_{k=21}^{+\infty} \mathrm{e}^{-18}\,\frac{18^k}{k!} = 1 - \sum_{k=0}^{20} \mathrm{e}^{-18}\,\frac{18^k}{k!} \,. \tag{*}$$

Tuttavia, calcolare esattamente la somma di 21 termini dell'equazione precedente è molto lungo e laborioso (può farlo solo una macchina). È estremamente più comodo usare l'approssimazione normale della poissoniana

$$\mathcal{P}(\lambda) \simeq N(\lambda, \lambda) \qquad \text{se } \lambda > 5$$

(i parametri della densità normale sono la stessa media e varianza della Poisson, che ricordiamo coincidono entrambe con λ). Con questa approssimazione,

$$S_{15} \sim \mathcal{P}(18) \simeq N(18, 18)$$

e la probabilità richiesta è

$$
\begin{aligned}
\mathbb{P}\left(S_{15} > 20\right) &= \mathbb{P}\left(S_{15} \geq 20.5\right) && \text{correzione di continuità, perché } S_{15} \text{ è discreta} \\
&= \mathbb{P}\Bigg(\underbrace{\frac{S_{15} - \mathbb{E}\left[S_{15}\right]}{\sqrt{\mathrm{Var}\left(S_{15}\right)}}}_{\approx N(0,1)} \geq \frac{20.5 - 18}{\sqrt{18}}\Bigg) && \text{approssimazione normale della Poisson} \\
&= 1 - \Phi\left(\frac{20.5 - 18}{\sqrt{18}}\right) = 1 - \Phi(0.589) = 1 - 0.72240 \\
&= 0.27760 = 27.760\% \,.
\end{aligned}
$$

(Col calcolo esatto, l'espressione (*) avrebbe dato 26.928%.)

(d) Il numero di chiamate che arrivano nella centrale in 24 ore è la v.a.

$$S_{24} = Y_1 + Y_2 + \ldots + Y_{24}\,,$$

dove

$$Y_i = \text{numero di chiamate che arrivano tra le } (i-1)\text{:00 e le } i\text{:00}\,.$$

Come prima, le $Y_1, Y_2, \ldots, Y_{24}$ sono i.i.d. (cioè costituiscono un campione aleatorio), con $Y_i \sim \mathcal{P}(1.2)$ per ogni i. Il valor medio richiesto è

$$\mathbb{E}\left[S_{24}\right] = \mathbb{E}\left[Y_1 + Y_2 + \ldots + Y_{24}\right] = \mathbb{E}\left[Y_1\right] + \mathbb{E}\left[Y_2\right] + \ldots + \mathbb{E}\left[Y_{24}\right] = \underbrace{1.2 + 1.2 + \ldots + 1.2}_{24 \text{ volte}} = 28.8\,.$$

Notare che per il calcolo della media precedente si usa solo la linearità di $\mathbb{E}$, che vale anche per v.a. *non* indipendenti. In altre parole, per risolvere questo punto dell'esercizio non sarebbe neanche necessario assumere che le $Y_1, Y_2, \ldots, Y_{24}$ siano indipendenti.

Soluzione 8.

(a) Introdiciamo le v.a.

$$\begin{aligned} X &= \text{numero di guasti del primo motore in una crociera} \\ Y &= \text{numero di guasti del secondo motore in una crociera} \\ Z = X + Y &= \text{numero di guasti totali in una crociera} \end{aligned}$$

I dati del testo sono

$$X \sim \mathcal{P}(0.1) \qquad Y \sim \mathcal{P}(0.2) \qquad X \text{ e } Y \text{ sono indipendendenti}\,.$$

Le prime due probabilità richieste si calcolano direttamente:

$$\begin{aligned} \mathbb{P}\left(\text{“in una crociera non si guasta il primo motore”}\right) &= \mathbb{P}\left(X = 0\right) = \mathrm{e}^{-0.1}\,\frac{0.1^0}{0!} = \mathrm{e}^{-0.1} \\ \mathbb{P}\left(\text{“in una crociera non si guasta il secondo motore”}\right) &= \mathbb{P}\left(Y = 0\right) = \mathrm{e}^{-0.2}\,\frac{0.2^0}{0!} = \mathrm{e}^{-0.2}\,. \end{aligned}$$

Inoltre, poiché Z è la somma di due v.a. poissoniane indipendenti,

$$Z \sim \mathcal{P}(0.1 + 0.2) = \mathcal{P}(0.3)\,.$$

Quindi,

$$\mathbb{P}\left(\text{“in una crociera non si guasta nessuno dei due motori”}\right) = \mathbb{P}\left(Z = 0\right) = \mathrm{e}^{-0.3}\,\frac{0.3^0}{0!} = \mathrm{e}^{-0.3}\,.$$

Per calcolare quest'ultima probabilità, si poteva anche usare direttamente l'indipendenza di X e Y nel modo seguente:

$$\begin{aligned} &\mathbb{P}\left(\text{“in una crociera non si guasta nessuno dei due motori”}\right) = \\ &\quad = \mathbb{P}\left(\text{“in una crociera non si guasta il } 1^\circ \text{ motore”} \wedge \text{“in una crociera non si guasta il } 2^\circ \text{ motore”}\right) \\ &\quad = \mathbb{P}\left(\text{“}X = 0\text{”} \wedge \text{“}Y = 0\text{”}\right) = \mathbb{P}\left(X = 0\right)\mathbb{P}\left(Y = 0\right) \qquad \text{indipendenza di } X \text{ e } Y \\ &\quad = \mathrm{e}^{-0.1} \cdot \mathrm{e}^{-0.2} = \mathrm{e}^{-0.3}\,. \end{aligned}$$

(b) Il modo più semplice per calcolare la probabilità richiesta è

$$\mathbb{P}\left(\text{“in una crociera avviene esattamente un guasto”}\right) = \mathbb{P}\left(Z = 1\right) = \mathrm{e}^{-0.3}\,\frac{0.3^1}{1!} = 0.3\,\mathrm{e}^{-0.3}\,.$$

Tuttavia, come al punto precendente si può anche usare direttamente l'indipendenza di X e Y, osservando prima di tutto che si ha l'uguaglianza di eventi

$$\text{“}\,Z = 1\,\text{”} \;=\; \left(\text{“}\,X = 1\,\text{”} \wedge \text{“}\,Y = 0\,\text{”}\right) \vee \left(\text{“}\,X = 0\,\text{”} \wedge \text{“}\,Y = 1\,\text{”}\right)$$

(attenzione che l'ordine delle parentesi a secondo membro è essenziale!). Dopodiché, calcoliamo

$$\begin{aligned}
\mathbb{P}\left(Z = 1\right) &= \mathbb{P}\left(\left(\text{“}\,X = 1\,\text{”} \wedge \text{“}\,Y = 0\,\text{”}\right) \vee \left(\text{“}\,X = 0\,\text{”} \wedge \text{“}\,Y = 1\,\text{”}\right)\right) \\
&= \mathbb{P}\left(\text{“}\,X = 1\,\text{”} \wedge \text{“}\,Y = 0\,\text{”}\right) + \mathbb{P}\left(\text{“}\,X = 0\,\text{”} \wedge \text{“}\,Y = 1\,\text{”}\right) \qquad \text{3° assioma della probabilità} \\
&= \mathbb{P}\left(X = 1\right)\mathbb{P}\left(Y = 0\right) + \mathbb{P}\left(X = 0\right)\mathbb{P}\left(Y = 1\right) \qquad \text{indipendenza di } X \text{ e } Y \\
&= p_X(1)p_Y(0) + p_X(0)p_Y(1) \\
&= \left(\mathrm{e}^{-0.1}\,\frac{0.1^1}{1!}\right)\cdot\left(\mathrm{e}^{-0.2}\right) + \left(\mathrm{e}^{-0.1}\right)\cdot\left(\mathrm{e}^{-0.2}\,\frac{0.2^1}{1!}\right) \\
&= 0.3\,\mathrm{e}^{-0.3}
\end{aligned}$$

(attenzione che il 3° assioma si può usare solo perché gli eventi “$X = 1$” $\wedge$ “$Y = 0$” e “$X = 0$” $\wedge$ “$Y = 1$” sono *incompatibili*). Riotteniamo quindi (naturalmente!) lo stesso risultato, ma questa volta in modo molto più complicato.

(c) Sia ora T la v.a.

$$T = \text{numero di crociere che si concludono senza nessun guasto fra le 10 crociere totali}\,.$$

Allora, considerando ciascuna crociera come una prova, in cui si ha successo se la crociera si conclude senza nessun guasto, abbiamo una successione di $n = 10$ prove di Bernoulli (cioè indipendenti e ciascuna con la stessa probabilità di successo di qualunque altra); la probabilità di successo in una singola prova è la $\mathbb{P}\left(Z = 0\right) = \mathrm{e}^{-0.3}$ trovata al punto (a). La v.a. T conta il numero di successi nelle 10 prove; di conseguenza,

$$T \sim B(10,\,\mathrm{e}^{-0.3})\,,$$

e la probabilità cercata è

$$\begin{aligned}
\mathbb{P}\left(T \geq 8\right) &= \sum_{k=8}^{10} p_T(k) \\
&= \binom{10}{8}(\mathrm{e}^{-0.3})^8(1-\mathrm{e}^{-0.3})^{10-8} + \binom{10}{9}(\mathrm{e}^{-0.3})^9(1-\mathrm{e}^{-0.3})^{10-9} + \binom{10}{10}(\mathrm{e}^{-0.3})^{10}(1-\mathrm{e}^{-0.3})^{10-10}\,.
\end{aligned}$$

Calcoliamo a parte

$$\binom{10}{8} = \binom{10}{10-8} = \binom{10}{2} = \frac{10\cdot 9}{2!} = 45 \qquad \binom{10}{9} = \binom{10}{1} = 10 \qquad \binom{10}{10} = 1$$

e quindi

$$\begin{aligned}
\mathbb{P}\left(T \geq 8\right) &= 45\cdot(\mathrm{e}^{-0.3})^8(1-\mathrm{e}^{-0.3})^2 + 10\cdot(\mathrm{e}^{-0.3})^9(1-\mathrm{e}^{-0.3}) + (\mathrm{e}^{-0.3})^{10} \\
&= 0.49820 = 49.820\%\,.
\end{aligned}$$

(d) Per rispondere a questa domanda, dobbiamo introdurre la nuova v.a.

$$S = \text{numero complessivo di guasti che accadono nelle 10 crociere}\,,$$

che non c'entra nulla con la T del punto precedente. Infatti, se per esempio succede che nella prima crociera capitano 8 guasti, nella seconda 20 e poi non ne succedono più fino all'ultima crociera, allora si è verificato che $T = 2$, mentre $S = 28 \Rightarrow S \neq T$. Quello che ora chiede l'esercizio è

$$\mathbb{P}\left(\text{“si verificano almeno 2 guasti nelle 10 crociere”}\right) = \mathbb{P}\left(S \geq 2\right)\,.$$

Per calcolarla, osserviamo che S è la somma dei guasti nelle singole crociere, cioè

$$S = Z_1 + Z_2 + \ldots + Z_{10} \qquad \text{dove} \qquad Z_i \sim \mathcal{P}(0.3)\ \forall i \quad \text{e} \quad Z_1, Z_2, \ldots, Z_{10} \quad \text{sono indipendenti}\,.$$

In altre parole, $Z_1, Z_2, \ldots, Z_{10}$ costituisce un *campione aleatorio* poissoniano. Abbiamo già visto al punto (a) che la somma di poissoniane indipendenti è ancora poissoniana, dunque

$$S \sim \mathcal{P}(10 \cdot 0.3) = \mathcal{P}(3)\,.$$

La probabilità cercata è quindi

$$\mathbb{P}\left(S \geq 2\right) = 1 - \mathbb{P}\left(S \leq 1\right) = 1 - \mathrm{e}^{-3}(1+3) = 0.80085 = 80.085\%\,.$$

Osservazione: Sia la v.a. T del punto (c) sia la S di questo punto sono funzioni delle $Z_1, Z_2, \ldots, Z_{10}$; tuttavia sono funzioni *diverse*. Infatti,

$$\begin{aligned} T &= \mathbb{1}_{\{0\}}(Z_1) + \mathbb{1}_{\{0\}}(Z_2) + \ldots + \mathbb{1}_{\{0\}}(Z_{10}) \\ S &= Z_1 + Z_2 + \ldots + Z_{10}\,, \end{aligned}$$

dove

$$\mathbb{1}_{\{0\}}(Z_i) = \begin{cases} 1 & \text{se } Z_i = 0 \\ 0 & \text{se } Z_i \neq 0\,. \end{cases}$$

(e) Se ora introduciamo anche le v.a. $Z_{11}, Z_{12}, \ldots, Z_{100}$ che contano i numeri di guasti nell'11ma, 12ma …100ma crociera, rispettivamente, il numero di guasti totali in 100 crociere è la somma

$$S_{100} = Z_1 + Z_2 + \ldots + Z_{100}\,,$$

in cui di nuovo le v.a. $Z_1, Z_2, \ldots, Z_{100}$ costituiscono un campione aleatorio con $Z_i \sim \mathcal{P}(0.3)$ per ogni i. Similmente a prima,

$$S_{100} \sim \mathcal{P}(100 \cdot 0.3) = \mathcal{P}(30)$$

e la probabilità cercata è

$$\mathbb{P}\left(S_{100} \geq 20\right) = 1 - \mathbb{P}\left(S_{100} \leq 19\right) = \sum_{k=0}^{19} \mathrm{e}^{-3}\,\frac{3^k}{k!}\,.$$

Tuttavia, questa volta il calcolo esatto è molto lungo e laborioso. È molto più semplice usare l'approssimazione normale della densità di Poisson, che ci dice che $\mathcal{P}(\lambda) \simeq N(\lambda, \lambda)$ se $\lambda > 5$. In questo caso,

$$S_{100} \sim \mathcal{P}(30) \simeq N(30,\, 30)\,,$$

e pertanto

$$
\begin{aligned}
\mathbb{P}(S_{100} \geq 20) &= \mathbb{P}(S_{100} \geq 19.5) \qquad \text{approx. di continuità, perché } S_{100} \text{ è discreta} \\
&= \mathbb{P}\Bigg(\underbrace{\frac{S_{100} - \mathbb{E}[S_{100}]}{\sqrt{\mathrm{Var}(S_{100})}}}_{\approx N(0,1)} \geq \frac{19.5 - 30}{\sqrt{30}} \Bigg) \qquad \text{TLC} \\
&= 1 - \Phi\left(\frac{19.5 - 30}{\sqrt{30}} \right) = 1 - \Phi(-1.917) = \Phi(1.917) \\
&= 0.97257 = 97.257\% \,.
\end{aligned}
$$

Soluzione 9.

(a) Riconosciamo in f_X la densità esponenziale di parametro $\lambda = 1/100$, in quanto con questa scelta di λ si ha

$$
f_X(x) = \begin{cases} 0 & \text{se } x < 0 \\ \lambda \mathrm{e}^{-\lambda x} & \text{se } x \geq 0 \,. \end{cases}
$$

Il valore atteso, la varianza e la deviazione standard di tale densità si trovano sul formulario, e sono rispettivamente

$$
\mathbb{E}[X] = \frac{1}{\lambda} = \frac{1}{1/100} = 100 \,, \qquad \mathrm{Var}(X) = \frac{1}{\lambda^2} = \frac{1}{(1/100)^2} = 10\,000 \,, \qquad \mathrm{Sd}(X) = \sqrt{\mathrm{Var}(X)} = 100 \,.
$$

(b) Se X_i è la durata dell'i-esima lampadina adoperata in sequenza, il numero totale di ore di luce ottenute con n lampadine è la somma

$$
T_n = X_1 + X_2 + \ldots + X_n \,,
$$

dove ciascuna X_i ha la stessa densità del punto precedente, e tutte le v.a. $X_1, X_2, \ldots, X_n$ sono tra loro indipendenti. Se n è grande, possiamo dunque applicare il TLC e approssimare la densità di T_n con una normale:

$$
T_n \underset{\text{TLC}}{\approx} N(\mathbb{E}[T_n] \,,\, \mathrm{Var}(T_n)) = N(n\mathbb{E}[X_1] \,,\, n\mathrm{Var}(X_1)) = N(100\,n \,,\, 10\,000\,n) \,.
$$

(c) Se usiamo un numero $n \geq 30$ di lampadine, si ha

$$
\begin{aligned}
&\mathbb{P}(\text{“si ottiene luce per un anno intero”}) = \\
&\qquad = \mathbb{P}(\text{“la durata totale delle } n \text{ lampadine è sufficiente a coprire un anno”}) \\
&\qquad = \mathbb{P}(T_n > 8\,760) = \mathbb{P}\Bigg(\underbrace{\frac{T_n - \mathbb{E}[T_n]}{\sqrt{\mathrm{Var}(T_n)}}}_{\approx N(0,1)} > \frac{8\,760 - 100\,n}{\sqrt{10\,000\,n}} \Bigg) \simeq 1 - \Phi\left(\frac{8\,760 - 100\,n}{100\sqrt{n}} \right) \,.
\end{aligned}
$$

Quindi con $n = 80$ lampadine la probabilità richiesta è

$$
\begin{aligned}
\mathbb{P}(\text{“si ottiene luce per un anno intero”}) &\simeq 1 - \Phi\left(\frac{8\,760 - 100 \cdot 80}{100 \cdot \sqrt{80}} \right) = 1 - \Phi(0.8497) = 1 - 0.80234 \\
&= 19.766\% \,.
\end{aligned}
$$

Si noti che la funzione $n \mapsto \mathbb{P}(T_n > 8\,760) = 1 - \Phi\left(\frac{8\,760 - 100\,n}{100\sqrt{n}}\right)$ è crescente, in quanto $n \mapsto \frac{8\,760 - 100\,n}{100\sqrt{n}}$ è decrescente (infatti, lo è sia $n \mapsto \frac{8\,760}{100\sqrt{n}}$ sia $n \mapsto -\frac{100\,n}{100\sqrt{n}} = -\sqrt{n}$), $1 - \Phi$ è decrescente (perché Φ è crescente), e dunque la composizione $n \mapsto \mathbb{P}(T_n > 8\,760) =$ (decrescente $\circ$ decrescente) è crescente. Ciò significa che più aumenta il numero di lampadine usate in sequenza, più aumenta la probabilità di ottenere luce per un anno intero, come del resto ci si poteva aspettare.

(d) Ora il problema è trovare il numero minimo n di lampadine tale che

$$\mathbb{P}(T_n > 5\,000) \geq 0.95\,. \tag{$*$}$$

Il quesito suggerisce di risolvere anche questa disequazione usando l'approssimazione normale del TLC. In linea di principio, però, non è detto che n debba essere per forza > 30 per risolvere $(*)$. In altre parole, potrebbe succedere che il minimo n che risolve $(*)$ sia piccolo (p.es., $n = 5$), e dunque non sia lecito usare il TLC per ricavarlo. Per escludere questa possibilità, dimostriamo preliminarmente che, per risolvere $(*)$, è necessario che n sia > 30. Per farlo, usiamo la disuguaglianza di Chebishev per limitare dall'alto la probabilità

$$\mathbb{P}(T_n > 5\,000) = \mathbb{P}(T_n - \mathbb{E}[T_n] > 5\,000 - 100\,n)\ .$$

Se $n < 50$, la differenza $5\,000 - 100\,n$ a secondo membro dell'evento "$T_n - \mathbb{E}[T_n] > 5000 - 100\,n$" è positiva. In tal caso, abbiamo dunque l'implicazione di eventi

$$\text{“}T_n - \mathbb{E}[T_n] > 5\,000 - 100\,n\text{”} \quad \Rightarrow \quad \text{“}\,|T_n - \mathbb{E}[T_n]| > 5\,000 - 100\,n\text{”} \tag{$**$}$$

(infatti, se $\varepsilon > 0$, allora $x > \varepsilon$ implica $|x| > \varepsilon$). Se un evento ne implica un altro, sappiamo che la probabilità dell'evento implicante è più piccola di quella dell'evento implicato. Di conseguenza,

$$\mathbb{P}(T_n > 5\,000) = \mathbb{P}(T_n - \mathbb{E}[T_n] > 5\,000 - 100\,n) \leq \mathbb{P}(|T_n - \mathbb{E}[T_n]| > 5\,000 - 100\,n)\ .$$

A questo punto, usiamo la disuguaglianza di Chebyshev $\mathbb{P}(|Z - \mathbb{E}[Z]| > \varepsilon) < \frac{\mathrm{Var}(Z)}{\varepsilon^2}$ per ottenere la limitazione

$$\mathbb{P}(|T_n - \mathbb{E}[T_n]| > 5\,000 - 100\,n) < \frac{\mathrm{Var}(T_n)}{(5\,000 - 100\,n)^2} = \frac{10\,000\,n}{(5\,000 - 100\,n)^2} = \frac{n}{(50 - n)^2}\,.$$

Se combiniamo questa disuguaglianza con quella scritta tre righe sopra e imponiamo in più che sia $\mathbb{P}(T_n > 5\,000) \geq 0.95$ come richiesto dall'esercizio, troviamo la catena di disequazioni

$$\begin{aligned} 0.95 \leq \mathbb{P}(T_n > 5\,000) &= \mathbb{P}(T_n - \mathbb{E}[T_n] > 5\,000 - 100\,n) \leq \mathbb{P}(|T_n - \mathbb{E}[T_n]| > 5\,000 - 100\,n) \\ &< \frac{\mathrm{Var}(T_n)}{(5\,000 - 100\,n)^2} = \frac{n}{(50 - n)^2} \end{aligned}$$

e quindi

$$0.95 < \frac{n}{(50 - n)^2}\,.$$

Questa disuguaglianza si può riscrivere

$$0.95\,n^2 - 96\,n + 2\,375 < 0\,.$$

Le soluzioni dell'equazione di secondo grado associata sono

$$n_\pm = \frac{96 \pm \sqrt{96^2 - 4 \cdot 0.95 \cdot 2\,375}}{2 \cdot 0.95} = \begin{cases} 57.80014 \\ 43.25249 \end{cases}$$

e dunque dobbiamo prendere i valori (interi!) di n compresi nell'intervallo $(43.25249\,,\,57.80014)$, e cioè $44 \leq n \leq 57$. A questo punto, ricordiamoci che tutti i calcoli precedenti sono stati fatti supponendo $n < 50$, perché altrimenti l'implicazione $(**)$ non sarebbe stata vera. Quindi, ricapitolando: se $n < 50$ risolve la disequazione $(*)$ di partenza, allora deve essere $n \geq 44$, che è abbastanza grande per applicare il TLC; altrimenti, se $(*)$ è risolta da $n \geq 50$, tutto quello che abbiamo visto finora non è più valido, ma questo non ha alcuna importanza perché $n \geq 50$ è già grande per conto proprio. In conclusione, se n risolve $(*)$, allora è necessariamente ≥ 44 e dunque possiamo usare il TLC per trovarlo. Facciamolo:

$$0.95 \leq \mathbb{P}\left(T_n > 5\,000\right) = \mathbb{P}\Bigg(\underbrace{\frac{T_n - \mathbb{E}\left[T_n\right]}{\sqrt{\mathrm{Var}\left(T_n\right)}}}_{\approx N(0,1)} > \frac{5\,000 - 100\,n}{\sqrt{10\,000\,n}}\Bigg) = 1 - \Phi\left(\frac{50-n}{\sqrt{n}}\right)$$

$$\begin{aligned}
&\Leftrightarrow \quad \Phi\left(\frac{50-n}{\sqrt{n}}\right) \leq 1 - 0.95 = 0.05 \\
&\Leftrightarrow \quad \frac{50-n}{\sqrt{n}} \leq z_{0.05} = -z_{1-0.05} = -z_{0.95} = -1.645 \\
&\Leftrightarrow \quad 1.645\sqrt{n} \leq n - 50\,. \qquad\qquad (***)
\end{aligned}$$

Per risolvere l'ultima disequazione, dobbiamo innanzitutto imporre che il membro di destra sia positivo, cioè

$$n - 50 \geq 0 \quad \Leftrightarrow \quad n \geq 50\,,$$

e poi, una volta assunta questa condizione, possiamo quadrare ambo i membri, ottenendo

$$n^2 - 102.70603\,n + 2\,500 \geq 0\,.$$

Le soluzioni di quest'ultima disequazione sono tutti gli n esterni ai due valori

$$n_\pm = \frac{102.70603 \pm \sqrt{102.70603^2 - 4 \cdot 1 \cdot 2\,500}}{2 \cdot 1} = \begin{cases} 63.06335 \\ 39.64268 \end{cases}$$

cioè tutti gli interi nell'unione dei due intervalli $[0, 39] \cup [64, +\infty)$. D'altra parte, ricordandoci che per $n < 50$ la disequazione $(***)$ non ha soluzioni, scartiamo il primo intervallo e troviamo che $n \geq 64$. Notare che questo valore è grande e consistente con la condizione (necessaria ma non sufficiente!) $n \geq 44$ che già avevamo trovato con Chebyshev. Quindi, concludiamo che il numero minimo di lampadine richiesto dall'esercizio è 64.

Soluzione 11.

Supponiamo che sul ponte passino n automobili, e siano $X_1, X_2, \ldots, X_n$ i pesi di ciascuna di esse. Allora, possiamo supporre che le v.a. $X_1, X_2, \ldots, X_n$ siano indipendenti (i pesi di macchine diverse naturalmente non si influenzano tra di loro) e identicamente distribuite (cioè le macchine provengono tutte da una popolazione *omogenea*). In altre parole, possiamo supporre che le $X_1, X_2, \ldots, X_n$ costituiscano un *campione aleatorio*. Il testo dell'esercizio non ci dice qual è la densità delle X_i, ma si limita darci la loro media e deviazione standard:

$$\mathbb{E}\left[X_i\right] = 1 \qquad \sqrt{\mathrm{Var}\left(X_i\right)} = 0.2 \qquad \text{per ogni } i = 1, \ldots, n\,.$$

Introdotta la v.a.

$$S_n = \text{peso totale delle } n \text{ macchine} = X_1 + X_2 + \ldots + X_n\,,$$

abbiamo l'uguaglianza di eventi

$$\text{“il peso totale delle } n \text{ macchine supera la portata massima del ponte”} = \text{“} S_n > 100 \text{”}\,.$$

L'esercizio ci chiede dunque di determinare n in modo tale che

$$\mathbb{P}\,(S_n > 100) > 5\%\,.$$

Per farlo, dovremmo conoscere la densità di S_n. Se n è piccolo, non sapendo la densità delle X_i non c'è nessuna speranza di conoscere quella di S_n. Ma se n è grande, allora per il TLC si ha

$$S_n \approx N(n\mathbb{E}\,[X_1]\ ,\, n\mathrm{Var}\,(X_1)) = N(n\,,\,(0.2)^2\,n) \tag{$*$}$$

qualunque sia la densità delle X_i. E in effetti è proprio questa la situazione dell'esercizio: se infatti n fosse troppo piccolo (diciamo $n < 30$), allora si avrebbe $\mathbb{E}\,[S_n] = n < 30$, $\mathrm{Var}\,(S_n) = (0.2)^2 n < (0.2)^2 \cdot 30 = 1.2$, e quindi, per la disuguaglianza di Chebishev,

$$\begin{aligned}
&\mathbb{P}\left(|S_n - n| > 5\cdot\sqrt{1.2}\right) = \mathbb{P}\left(|S_n - \mathbb{E}\,[S_n]\,| > 5\cdot\sqrt{1.2}\right) < \frac{\mathrm{Var}\,(S_n)}{5^2\cdot 1.2} < \frac{1}{5^2} = 4\% \\
&\quad\Rightarrow\quad \mathbb{P}\,(S_n > 100) \ll 4\% \quad \text{perché “}S_n > 100\text{” implica “}|S_n - n| > 5\cdot\sqrt{1.2}\text{” se } n < 30\,.
\end{aligned}$$

Ma noi non vogliamo $\mathbb{P}\,(S_n > 100) \ll 4\%$. Di conseguenza, possiamo tranquillamente assumere $n > 30$, in modo tale che valga $(*)$. Notare che si tratta di un discorso del tutto simile a quello fatto nella soluzione dell'Esercizio 9 per giustificare l'ipotesi che n sia grande. Ora, imponiamo

$$\begin{aligned}
5\% < \mathbb{P}\,(S_n > 100) &= \mathbb{P}\Bigg(\underbrace{\frac{S_n - \mathbb{E}\,[S_n]}{\sqrt{\mathrm{Var}\,(S_n)}}}_{\approx N(0,1)} > \frac{100-n}{0.2\sqrt{n}}\Bigg) = 1 - \Phi\left(\frac{100-n}{0.2\sqrt{n}}\right) \\
\Rightarrow\quad & \Phi\left(\frac{100-n}{0.2\sqrt{n}}\right) < 0.95 \\
\Rightarrow\quad & \frac{100-n}{0.2\sqrt{n}} < z_{0.95} = 1.645 \\
\Rightarrow\quad & 100 - n < 0.329\sqrt{n}\,.
\end{aligned}$$

Osserviamo che se $n > 100$ l'ultima disequazione è banalmente verificata (a sinistra si ha un numero negativo, a destra uno positivo). Consideriamo invece il caso $n \leq 100$. Allora possiamo elevare a quadrato ambo i membri:

$$(100-n)^2 < (0.329\sqrt{n})^2 \quad\Rightarrow\quad n^2 - 200.108241\,n + 10000 < 0\,.$$

Le soluzioni sono tutti gli n tali che

$$\begin{aligned}
96.764 = \frac{200.108241 - \sqrt{200.108241^2 - 4\cdot 10000}}{2} &< n \\
&< \frac{200.108241 + \sqrt{200.108241^2 - 4\cdot 10000}}{2} = 103.345\,.
\end{aligned}$$

In realtà, ora stiamo considerando solo il caso $n \leq 100$, mentre abbiamo già visto che tutti gli $n > 100$ vanno bene. Gli n per cui $\mathbb{P}\,(S_n > 100) > 5\%$ sono pertanto tutti quelli tali che

$$n > 96.764\,.$$

Poiché n è intero, il più piccolo valore ammissibile è $n > 97$.

Soluzione 12.

Sia X_i il risultato dell'i-esima misura di distanza dalla stella, e sia d la sua vera distanza incognita. Se facciamo n misure, abbiamo un campione aleatorio $X_1, \ldots, X_n$. L'esercizio chiede quante misure dobbiamo fare se vogliamo che la media campionaria $\overline{X}_n$ disti da d per meno di $0.5\,\mathrm{pc}$ *con probabilità (almeno) del* 95%. In altre parole, dobbiamo risolvere rispetto a n la disequazione

$$\mathbb{P}\left(\left|\overline{X}_n - d\right| < 0.5\,\mathrm{pc}\right) \geq 95\%\,, \tag{$*$}$$

in cui $\left|\overline{X}_n - d\right|$ è appunto l'errore che si commette approssimando il valore vero d con la media campionaria $\overline{X}_n$ delle nostre n misure.

Il testo dell'esercizio ci dice che $\mathbb{E}\left[X_i\right] = d$ (cioè le misure non sono affette da errore sistematico) e che $\mathrm{Var}\left(X_i\right) = 4\,\mathrm{pc}^2$. Per le ben note proprietà, la media campionaria avrà dunque valore atteso

$$\mathbb{E}\left[\overline{X}_n\right] = \mathbb{E}\left[X_i\right] = d$$

e varianza

$$\mathrm{Var}\left(\overline{X}_n\right) = \frac{\mathrm{Var}\left(X_i\right)}{n} = \frac{4}{n}\,.$$

La disuguaglianza di Chebyshev ci permette di affermare che, prendendo un numero di misure sufficientemente elevato, possiamo rendere la probabilità $\mathbb{P}\left(\left|\overline{X}_n - d\right| < 0.5\right)$ grande a piacere. Infatti,

$$\begin{aligned}\mathbb{P}\left(\left|\overline{X}_n - d\right| < 0.5\right) &= 1 - \mathbb{P}\left(\left|\overline{X}_n - d\right| \geq 0.5\right) = 1 - \mathbb{P}\left(\left|\overline{X}_n - \mathbb{E}\left[\overline{X}_n\right]\right| \geq 0.5\right) \\ &\underset{\text{Chebyshev}}{\geq} 1 - \frac{\mathrm{Var}\left(\overline{X}_n\right)}{0.5^2} = 1 - \frac{4/n}{0.25} = 1 - \frac{16}{n}\,.\end{aligned}$$

Se vogliamo essere sicuri che $\mathbb{P}\left(\left|\overline{X}_n - d\right| < 0.5\right) \geq 95\%$, allora basterà che ($=$ condizione *sufficiente*)

$$1 - \frac{16}{n} \geq 95\% \quad \Rightarrow \quad n \geq 320\,.$$

Almeno 320 misure sono dunque certamente sufficienti a risolvere l'equazione $(*)$.

Notiamo che la disuguaglianza di Chebyshev non fa nessuna ipotesi sulla densità delle X_i, che del resto noi non conosciamo, né su quella della loro media campionaria $\overline{X}_n$. La disuguaglianza che abbiamo ottenuto per n è dunque del tutto universale. L'unico svantaggio in questo, però, è che Chebyshev in generale fornisce un limite inferiore molto sub-ottimale. Ciò va bene se vogliamo trovare una minorazione per $\mathbb{P}\left(\left|Z - \mathbb{E}\left[Z\right]\right| < \varepsilon\right)$ e non abbiamo nessuna informazione sulla densità di Z. Ma se, come in questo caso, la nostra Z è in realtà la media campionaria $\overline{X}_n$, allora il TLC ci permette di approssimare $\mathbb{P}\left(\left|\overline{X}_n - d\right| < 0.5\right)$ in modo molto più preciso di quanto faccia Chebyshev, e dunque ci lascia trovare per n un valore minimo molto più piccolo.

Supponendo $n \geq 30$, il TLC ci dice infatti che

$$\overline{X}_n \approx N\left(\mathbb{E}\left[\overline{X}_n\right], \mathrm{Var}\left(\overline{X}_n\right)\right) = N\left(d, \frac{16}{n}\right)\,,$$

e dunque

$$\begin{aligned}\mathbb{P}\left(\left|\overline{X}_n - d\right| < 0.5\right) &= \mathbb{P}\left(-0.5 < \overline{X}_n - d < 0.5\right) = \mathbb{P}\left(\frac{-0.5}{\sqrt{\frac{4}{n}}} < \underbrace{\frac{\overline{X}_n - \mathbb{E}\left[\overline{X}_n\right]}{\sqrt{\mathrm{Var}\left(\overline{X}_n\right)}}}_{\approx N(0,1)} < \frac{0.5}{\sqrt{\frac{4}{n}}}\right) \\ &\simeq \Phi\left(\frac{0.5}{\sqrt{\frac{4}{n}}}\right) - \Phi\left(-\frac{0.5}{\sqrt{\frac{4}{n}}}\right) = \Phi\left(\frac{\sqrt{n}}{4}\right) - \Phi\left(-\frac{\sqrt{n}}{4}\right) = 2\,\Phi\left(\frac{\sqrt{n}}{4}\right) - 1\,.\end{aligned}$$

Per risolvere $(*)$, dobbiamo quindi imporre

$$2\,\Phi\left(\frac{\sqrt{n}}{4}\right) - 1 \geq 95\% \quad \Rightarrow \quad \Phi\left(\frac{\sqrt{n}}{4}\right) \geq 0.975 \quad \Rightarrow \quad \frac{\sqrt{n}}{4} \geq z_{0.975} = 1.96 \quad \Rightarrow \quad n \geq 61.4656\,.$$

Bastano pertanto $n = 62$ misure. Da notare che questo risultato non è (ovviamente!) in contraddizione con quello trovato con Chebyshev, e soprattutto è un valore ≥ 30, in accordo con l'ipotesi che abbiamo fatto per applicare il TLC.