Compito scritto dell'esame di Statistica e analisi dei dati (Crema) 23 giugno 2016	Prof. Giuseppe Boccignone	Corso di Laurea
Cognome:	Nome:	Matricola:

Istruzioni

- Il tempo riservato alla prova scritta è di 2 ore e 30 minuti. Durante la prova è possibile consultare libri e appunti.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede più passaggi di calcolo, non sarà preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o più frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sarà eventualmente rappresentato in quest'ultima forma).
- SOLO PER STUDENTI DEL CORSO ONLINE (VECCHIO ORDINAMENTO, Prof. Gianini): Indicare quale Modulo viene svolto al fine di completare una prova d'esame parziale precedentemente sostenuta con esito positivo
  - Modulo 1 ☐
  - Moduli 2 e 3 ☐

Modulo 1

Esercizio 1. (Probabilitá condizionata)
Un cappello contiene tre carte: una carta é nera su entrambi i lati; una carta é bianca su entrambi i lati; una carta é nera su un lato e bianca sull'altro. Le carte vengono mescolate nel cappello, poi una viene estratta a caso e appoggiata su di un tavolo.

(a) Se il lato visibile della carta é nero, qual é la probabilitá che l'altro lato sia bianco?

Soluzione: Uno studente brillante potrebbe risolvere in pochi secondi il problema ragionando come segue: la carta di interesse deve essere o la nero / nero o la nero / bianco: queste hanno pari probabilitá, dunque la probabilitá che, osservato il nero, l'altro lato sia bianco é $\frac{1}{2}$.
Lo studente, successivamente, per verificare la sua conclusione, effettua una simulazione del gioco delle carte (si veda lo script Matlab carte.m allegato) ma ottiene il risultato mostrato in Figura 1.

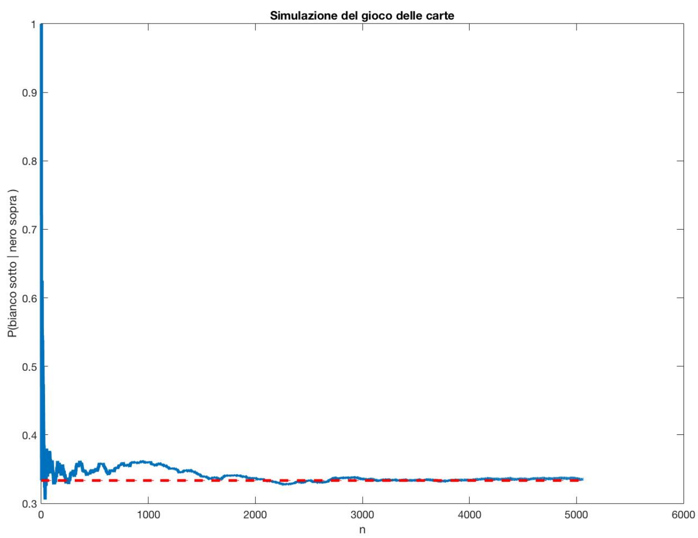

Figure 1: Simulazione del gioco delle tre carte: il valore di $P($"bianco sotto" $|$ "nero sopra"$)$ valutato in termini di frequenza relativa, per un numero di prove n crescente, tende al valore di $0.333\cdots \approx \frac{1}{3}$
.

Il valore dell'approssimazione frequentistica di $P(\text{“bianco sotto”} \mid \text{“nero sopra”})$ tende a $\frac{1}{3}$.
Per capire bene come si arriva a questo risultato, mettiamo da parte l'intuizione e definiamo con precisione lo spazio campionario e gli eventi di interesse.
Identifichiamo i lati delle carte

- N_1 e N_2 per la carta nero / nero
- B_1 e B_2 per la carta bianco / bianco
- N_3 e B_3 per la carta nero / bianco

In questo caso lo spazio campionario S é

$$S = \{N_1, N_2, N_3, B_1, B_2, B_3\}$$

L'evento $N =$ “nero sopra” é definibile come $N = \{N_1, N_2, N_3\}$.
L'evento $B =$ “bianco sotto” é definibile come $B = \{N_3, B_1, B_2\}$
Pertanto:

$$P(B \mid N) = \frac{P(B \cap N)}{P(N)}$$

Poiché $\#(\text{“bianco sotto”} \cap \text{“nero sopra”}) = 1$ e $\#(\text{“nero sopra”}) = 3$:

$$P(B \mid N) = \frac{1}{3}.$$

Esercizio 2. (Legge delle probabilitá totali)
Ci sono due squadre, 1 e 2, di rigoristi. La squadra i-sima ha $3i$ componenti. Per ogni rigore battuto, un rigorista della squadra i ha una probabilitá $\frac{1}{i+1}$ di fare goal, indipendentemente dai rigori battuti precedentemente.

(a) Un rigorista viene selezionato casualmente fra tutti i rigoristi disponibili. Sia G l'evento che venga segnato un goal. Calcolare la probabilitá $P(G)$ che un goal sia realizzato, quale che sia la squadra di provenienza del rigorista.

Soluzione: In prima battuta, la probabilitá dell'evento G che un rigorista segni un goal é condizionata dalla squadra a cui appartiene: $P(G \mid S_i) = \frac{1}{i+1}$. Pertanto:

$$P(G \mid S_1) = \frac{1}{2}$$
$$P(G \mid S_2) = \frac{1}{3}$$

Ciascuna squadra ha $3i$ componenti: dunque 3 per la squadra 1 e 6 della squadra 2. Indicando con S_i l'evento che sia stato selezionato in maniera casuale un rigorista della squadra i-esima, le probabilitá a priori sono

$$P(S_1) = \frac{3}{9} = \frac{1}{3}$$
$$P(S_2) = \frac{6}{9} = \frac{2}{3}$$

Usando la Legge delle Probabilitá totali:

$$P(G) = P(G \mid S_1)P(S_1) + P(G \mid S_2)P(S_2) = \frac{1}{2} \times \frac{1}{3} + \frac{1}{3} \times \frac{2}{3} = \frac{7}{18} = 0.388$$

(b) Vengono scelti a caso due rigoristi. Per $j = 1,2$, sia G_j l'evento che il rigorista j faccia goal. Trovare la probabilitá che il rigorista 1 e il rigorista 2 segnino entrambi un goal (*Suggerimento:* calcolare, mediante un chance tree, la probabilitá dell'evento S_{kl} che il primo rigorista sia stato selezionato dalla squadra k e il secondo rigorista dalla squadra l, con $k,l = 1,2$)

Soluzione: Per risolvere questa parte dl problema ricaviamo l'albero di probabilitá per due “estrazioni” successive e casuali di un rigorista dalle squadre 1 e 2 rappresentato in Figura 2,
Il diagramma ci dice che:

$$P(S_{11}) = \frac{1}{12}, P(S_{12}) = \frac{1}{4}, P(S_{21}) = \frac{1}{4}, P(S_{22}) = \frac{5}{12}$$

La probabilitá dell'evento congiunto $\{G_1, G_2\}$ é condizionata da quali squadre sono stati selezionati i rigoristi 1 e 2, ovvero dall'evento S_{kl} le cui probabilitá sono state derivate con il diagramma ad albero. Essendo l'esito di un rigore indipendente dall'altro: $P(G_1, G_2 \mid S_{kl}) = P(G_1 \mid S_{kl})P(G_2 \mid S_{kl})$. Pertanto:

$$P(G_1, G_2 \mid S_{11}) = \left(\frac{1}{2}\right)^2$$

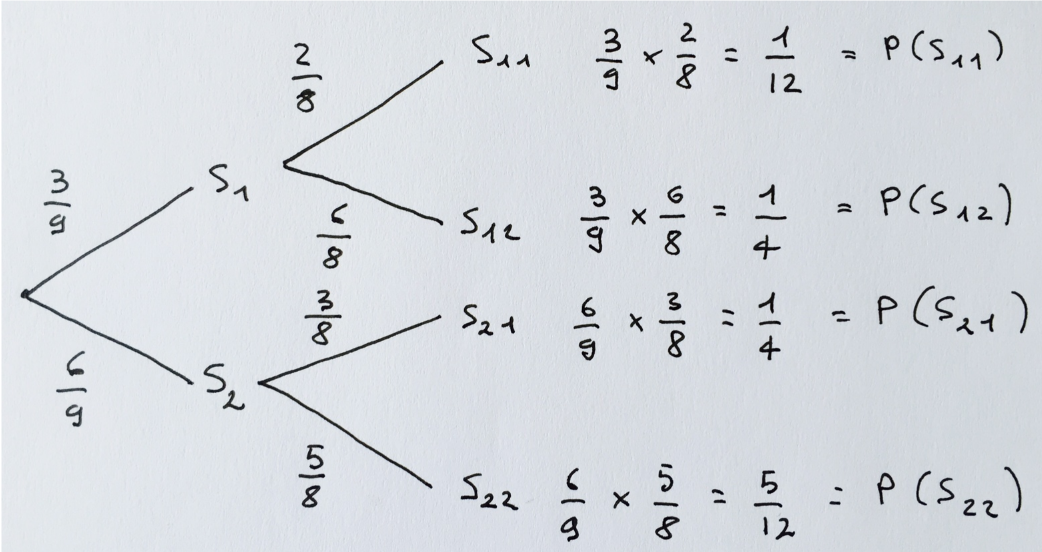

Figure 2: Chance tree per due "estrazioni" successive e casuali di un rigorista dalle squadre 1 e 2: S_k é l'evento che sia stato selezionato in maniera casuale un rigorista della squadra i-esima, S_{kl} che il primo rigorista sia stato selezionato dalla squadra k e il secondo rigorista dalla squadra l

$$P(G_1, G_2 \mid S_{12}) = \frac{1}{2} \times \frac{1}{3}$$
$$P(G_1, G_2 \mid S_{21}) = \frac{1}{3} \times \frac{1}{2}$$
$$P(G_1, G_2 \mid S_{22}) = \left(\frac{1}{3}\right)^2$$

Usando di nuovo la Legge delle Probabilitá totali:

$$P(G_1, G_2) = P(G_1, G_2 \mid S_{11})P(S_{11}) + P(G_1, G_2 \mid S_{12})P(S_{12}) + P(G_1, G_2 \mid S_{21})P(S_{21}) + P(G_1, G_2 \mid S_{22})P(S_{22}) = \frac{15}{96} = 0.1562$$

(c) Verificare se G_1 e G_2 siano eventi indipendenti. Commentare brevemente il risultato ottenuto.

Soluzione: Possiamo verificare la condizione di indipendenza stocastica:

$$P(G_1, G_2) = P(G_1)P(G_2)$$

E' evidente che $P(G_1) = P(G) = \frac{7}{18}$ e che per simmetria $P(G_1) = P(G_2)$, giacché lo scambio dei due rigoristi non muta nulla in termini di esito (se non si nvinti, usare ancora le probabilitá totali per calcolare $P(G_i) = P(G_i \mid S_{11})P(S_{11}) + P(G_i \mid S_{12})P(S_{12}) + P(G_i \mid S_{21})P(S_{21}) + P(G_i \mid S_{22})P(S_{22}) = \frac{7}{18}$).
Si ha allora che:

$$P(G_1, G_2) = 0.1562 \neq 0.1512 = (0.388)^2 = P(G_1)P(G_2)$$

dunque non sono indipendenti. Le due probabiliá ricavate sono vicine ma non esattamente uguali. G_1 e G_2 sono dipendenti perché se il primo rigore viene segnato, vi é maggior probabilit che abbia tirato un rigorista del gruppo 1. Questo rende piú probabile che il secondo giocatore venga selezionato dal gruppo 2, evento che riduce la probabilitá di segnare al secondo tiro.

Esercizio 3. (Teorema di Bayes)

Un robot autonomo controlla il livello di corrosione interna delle tubature dei sistemi di raffreddamento di una centrale nucleare. La corrosione non puó essere osservata direttamente, ma il robot effettua un test che puó fornisce indizi sulla specifica corrosione. Il test non é infallibile: nel 70% dei casi individua una corrosione quando é presente, ma puó anche sbagliare indicando una corrosione inesistente nel 20% dei casi. Risolvere il problema del robot a cui un operatore remoto, sotto l'ipotesi che il 10% delle tubature siano corrose, chiede di:

(a) Comunicargli la probabilitá che una parte delle tubature abbia delle corrosioni interne dopo che il test ha dato esito positivo (presenza di corrosioni). Determinare tale probabilitá.

Soluzione: Definiamo gli eventi $C =$ "la tubatura é corrosa" e $\overline{C} =$ "il test ha identificato la tubatura come corrosa". I dati del problema ci dicono che, a priori,

$$P(C) = 0.1$$

e dunque $P(\sim C) = 1 - 0.1 = 0.9$, dove $\sim C =$ "la tubatura non é corrosa". Inoltre

$$P(\overline{C} \mid C) = 0.7$$
$$P(\overline{C} \mid \sim C) = 0.2$$

Applicando la regola di Bayes:

$$P(C \mid \overline{C}) = \frac{P(\overline{C} \mid C)P(C)}{P(\overline{C})}$$

dove $P(\overline{C}) = P(\overline{C} \mid C)P(C) + P(\overline{C} \mid \sim C)P(\sim C) = 0.25$. Quindi

$$P(C \mid \overline{C}) = \frac{0.7 \times 0.1}{0.25} = 0.28$$

(b) Comunicargli la probabilitá che una parte delle tubature abbia delle corrosioni dopo che il test ha dato esito negativo (nessuna corrosione). Determinare tale probabilitá.

Soluzione: Si ripeta il procedimento precedente per il caso $P(C \mid \sim \overline{C})$

Moduli 2 e 3

Esercizio 1. (Funzione di densitá e di ripartizione)

Data la funzione: $\mathrm{f}_X(x) = \begin{cases} \frac{x}{2} - 1 & 2 \le x < k \\ 0 & \text{altrimenti} \end{cases}$.

(a) Si determini il valore di k che assicura che $f_X(x)$ rappresenta una funzione di densitá.

Soluzione: Imponiamo la condizione di normalizzazione della densitá $f_X(x)$:

$$\int_2^k \left(\frac{x}{2} - 1\right) dx = 1 \implies \left[\frac{x^2}{4}\right]_2^k - (k-2) = 1 \implies k^2 - 4k = 0$$

Risolvendo l'equazione, vi sono due possibili radici $k_1 = 0$ e $k_2 = 4$. Poiché $k_1 < 2 \le x < k$, é da scartare, dunque $k = k_2 = 4$. Pertanto la densitá correttamente normalizzata é:

$$f_X(x) = \frac{x}{2} - 1 \text{ con } 2 \le x < 4$$

(b) Si individui la corrispondente funzione di ripartizione $F_X(x)$

Soluzione: Per definizione $F_X(x) = \int_{-\infty}^x f_X(t)dt$. Nella fattispecie:

$$F_X(x) = \int_2^x \left(\frac{t}{2} - 1\right) dt$$

Integrando come fatto prima si ottiene:

$$F_X(x) = \frac{x^2}{4} - x + 1$$

per $2 \le x < 4$

(c) Si individui la mediana della variabile aleatoria X descritta dalla densitá $f_X(x)$

Soluzione: Per definizione la mediana x_{med} é il valore per cui

$$F_X(x) = \frac{1}{2} \implies \frac{x^2}{4} - x + 1 = \frac{1}{2}$$

Risolvendo l'equazione si trovano le soluzioni $x_1 = 2 + \sqrt{2}$ e $x_2 = 2 - \sqrt{2}\sqrt{2}$. Dunque $x_{med} = x_1$

Esercizio 2. (Distribuzioni di Poisson ed Esponenziale negativa)

Il numero di automobili che attraversano un particolare incrocio stradale in un'ora é mediamente pari a 30.

(a) Determinare la probabilitá che in un intervallo di tempo di cinque minuti nessuna automobile attraversi l'incrocio in questione.

Soluzione: Sia X la V.A. discreta che indica il numero di veicoli che passa lungo l'incrocio in 5 min. X segue una legge di Poisson, $X \sim Pois(\mu)$. Il numero medio di automobili che passa in 5 min. é $\frac{30}{12} = 2.5$. Dunque $X \sim Pois(2.5)$. Pertanto:

$$P(X = 0) = \frac{\mu^0}{0!}e^{-\mu} = e^{-2.5} = 0.082$$

(b) Qual é la probabilitá che in dieci minuti almeno due automobili passino lungo quel tratto di strada?

Soluzione: Se X denota il numero di veicoli che passa lungo l'incrocio in 5 min, indichiamo con $Y = 2X$ la V.A. discreta che indica il numero di veicoli che passa lungo l'incrocio in 10 min. Allora:

$$E[Y] = E[2X] = 2E[X] = 2\mu = 5$$

Dunque, $Y \sim Pois(5)$. La risposta alla domanda consiste nel calcolare

$$P(Y \geq 2) = 1 - P(Y < 2) = 1 - (P(Y = 0) + P(Y = 1)) = 1 - \left(e^{-5} + \frac{5^1}{1!}e^{-5}\right) = 1 - 6e^{-5} = 0.9595$$

(c) Assumendo che la variabile aleatoria T, rappresentante il tempo (sempre in minuti) trascorso tra il passaggio di un'auto e di quella successiva, si distribuisca con legge esponenziale negativa, qual é la probabilitá che tra il passaggio di un'auto e la successiva trascorra piú di un minuto?

Soluzione: Poiché $T \sim Exp(\lambda)$, i dati del problema ci dicono che il rate di passaggio é

$$\lambda = \frac{30}{60}\frac{\text{(auto)}}{\text{(min.)}}.$$

Dunque la V.A. T ha funzione di densitá $f_T(t) = 0.5e^{-0.5t}$, $t \geq 0$. La probabilitá che il tempo di attesa tra il passaggio di un un'auto e la successiva sia > 1 min. é $P(T > 1)$ che si puó ottenere dal calcolo diretto $P(T > 1) = \int_1^\infty f_T(t)dt$, oppure usando la funzione di sopravvivenza $S_T(t) = e^{-\lambda t}$:

$$P(T > 1) = 1 - F_T(1) = S_T(1) = e^{-0.5 \cdot 1} = 0.6065$$

Esercizio 3. (Distribuzione Normale, uso di tabelle)
Una linea di produzione fabbrica resistori da 1000 ohm (Ω). La fabbrica adotta una tolleranza del 10% (ovvero sono scartati i resistori con resistenza $R < 900\ \Omega$ e $R > 1100\ \Omega$).

(a) Sotto l'ipotesi che la resistenza R di un resistore sia una variabile aleatoria distribuita con legge normale di media 1000 Ω e varianza 2500, si calcoli la probabilitá che un resistore preso a caso venga scartato

Soluzione: Sia S l'evento in cui un resistore viene scartato. Allora:

$$S = \{R < 900\} \cup \{R > 1100\}$$

Poiché $\{R < 900\} \cap \{R > 1100\} = \emptyset$, si ha che

$$P(S) = P(\{R < 900\}) + P(\{R > 1100\}) = F_X(900) + (1 - F_X(1100)) = \Phi\left(\frac{900 - \mu}{\sigma}\right) + \left(1 - \Phi\left(\frac{1100 - \mu}{\sigma}\right)\right)$$

Poiché $\mu = 1000$, e $\sigma^2 = 2500 \implies \sigma = 50$, le variabili normalizzate sono

$$\frac{900 - 1000}{50} = -2$$

$$\frac{1100 - 1000}{50} = 2$$

.
Sfruttando la proprietá della cumulativa standardizzata $\Phi(-z) = 1 - \Phi(z)$, la precedente diventa:

$$P(S) = 2\,(1 - \Phi(2))$$

.
Accedendo alla tabella della Normale standard per $z = 2$, $\Phi(2) = 0.9772$ da cui $P(S) = 0.0455$

Compito scritto dell'esame di Statistica e analisi dei dati (Crema) 21 luglio 2016	Prof. Giuseppe Boccignone	Corso di Laurea
Cognome:	Nome:	Matricola:

Istruzioni

- Il tempo riservato alla prova scritta e di 2 ore e 30 minuti. Durante la prova e possibile consultare libri e appunti.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede piu passaggi di calcolo, non sara preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o piu frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sara eventualmente rappresentato in quest'ultima forma).
- SOLO PER STUDENTI DEL CORSO ONLINE (VECCHIO ORDINAMENTO, Prof. Gianini): Indicare quale Modulo viene svolto al fine di completare una prova d'esame parziale precedentemente sostenuta con esito positivo
  - Modulo 1 ☐
  - Moduli 2 e 3 ☐

Modulo 1

Esercizio 1. (Probabilitá condizionata)
Le probabilitá che tre uomini colpiscano un bersaglio sono, rispettivamente $\frac{1}{6}$ $\frac{1}{4}$ e $\frac{1}{3}$.

(a) Determinare la probabilitá che uno solo colpisca il bersaglio

Soluzione: Definiamo i seguenti eventi indipendenti:

- $A =$ `"il primo uomo colpisce il bersaglio"`
- $B =$ `"il secondo uomo colpisce il bersaglio"`
- $C =$ `"il terzo uomo colpisce il bersaglio"`

Il testo del problema ci dice che:

$$P(A) = \frac{1}{6}, P(B) = \frac{1}{4}, P(C) = \frac{1}{3}$$

Denotiamo con $\sim A, \sim B, \sim C$ gli eventi complementari. Ovviamente, $P(\sim A) = 1 - P(A)$, ecc.
L'evento di interesse é

$$T = \texttt{"uno solo colpisce il bersaglio"},$$

ed é definito mediante l'unione di tre eventi mutuamente esclusivi

$$T = \{A \cap \sim B \cap \sim C\} \cup \{\sim A \cap B \cap \sim C\} \cup \{\sim A \cap \sim B \cap C\}.$$

La sua probabilitá é data dunque da:

$$P(T) = P(\{A \cap \sim B \cap \sim C\}) + P(\{\sim A \cap B \cap \sim C\}) + P(\{\sim A \cap \sim B \cap C\})$$

Infine, usando l'indipendenza degli eventi:

$$P(T) = \frac{1}{6} \times (1 - \frac{1}{4}) \times (1 - \frac{1}{3}) + (1 - \frac{1}{6}) \times \frac{1}{4} \times (1 - \frac{1}{3}) + (1 - \frac{1}{6}) \times (1 - \frac{1}{4}) \times \frac{1}{3} = 0.43$$

(b) Determinare la probabilitá che uno solo colpisca il bersaglio e che questi sia il primo uomo.

Soluzione: La probabilitá che l'unico che ha colpito il bersaglio sia il primo uomo é la probabilitá condizionata

$$P(A \mid T) = \frac{P(A \cap T)}{P(T)} = \frac{P(\{A \cap \sim B \cap \sim C\})}{P(T)} = \frac{\frac{1}{6} \times (1 - \frac{1}{4}) \times (1 - \frac{1}{3})}{0.43} = 0.19$$

Esercizio 2. (Legge delle probabilitá totali)
In un primo turno elettorale il polo A ha avuto il 45% dei voti, e il polo B ha vinto con il 55% dei suffragi. Si ripetono le elezioni con i medesimi votanti, e dagli exit-poll risulta che:
- il 10% di coloro che avevano votato A hanno spostato il voto su B;
- il 20% dei vecchi elettori di B hanno votato A.

(a) Chi ha vinto, secondo gli exit-poll il secondo turno?

Soluzione: Definiamo i seguenti eventi:

- $A_1 =$ `"voto per A al primo turno"`
- $B_1 =$ `"voto per B al primo turno"`
- $A_2 =$ `"voto per A al secondo turno"`
- $B_2 =$ `"voto per B al secondo turno"`
- $E =$ `"voto cambiato"`

Dai dati del problema si ricava che:

- $P(A_1) = 0.45$
- $P(B_1) = 0.55$
- $P(E \mid A_1) = 0.10$
- $P(E \mid B_1) = 0.20$

Usando la legge delle probabilitá totali, la probabilitá che gli elettori abbiano votato A al secondo turno é:

$$P(A_2) = P(A_1)\,(1 - P(E \mid A1)) + P(B_1)P(E \mid B_1) = 0.45 \times 0.9 + 0.55 \times 0.20 = 0.515$$

Poiché gli eventi A_2 e B_2 sono mutuamente esclusivi, prendendo in considerazione gli exit-poll, vincerebbe A con il 51.5% contro B che raccoglierebbe il 48.5% dei consensi.

Esercizio 3. (Teorema di Bayes)
In un ufficio le pratiche relative ad una certa procedura amministrativa vengono affidate casualmente a tre impiegati che indicheremo con A, B e C. Statisticamente, la percentuale dei casi in cui la pratica viene completata entro una settimana da ciascun impiegato é riportata nella seguente tabella:

Impiegato	A	B	C
Perc.	40%	80%	30%

(a) Avendo ricevuto una pratica espletata entro una settimana, qual é secondo voi l'impiegato a cui era stata affidata e qual é la probabilitá della vostra conclusione?

Soluzione: Definiamo i seguenti eventi

- $S =$ `"pratica completata entro una settimana"`
- $A =$ `"pratica affidata all'impiegato A"`
- $B =$ `"pratica affidata all'impiegato B"`
- $C =$ `"pratica affidata all'impiegato C"`

L'affidamento pratiche é casuale dunque:

$$P(A) = P(B) = P(C) = \frac{1}{3}$$

Dalla tabella ricaviamo che:

$$P(S \mid A) = 0.4, P(S \mid B) = 0.8, P(S \mid C) = 0.3$$

Per rispondere al quesito si tratta di stabilire qual é la maggiore fra le probabilitá a posteriori

$$P(A \mid S) = \frac{P(S \mid A)P(A)}{P(S)}, P(B \mid S) = \frac{P(S \mid B)P(B)}{P(S)}, P(C \mid S) = \frac{P(S \mid C)P(C)}{P(S)}$$

Il denominatore, cioé la probabilitá che una pratica venga completata entro una settimana indipendentemente da quale impegato l'ha seguita é:

$$P(S) = P(S \mid A)P(A) + P(S \mid B)P(B) + P(S \mid C)P(C) = \frac{1}{3}(0.4 + 0.8 + 0.3) = \frac{1}{3} \times 1.5$$

Sostituendo tutti i dati:

$$P(A \mid S) = \frac{P(S \mid A)P(A)}{P(S)} = 0.267$$

$$P(B \mid S) = \frac{P(S \mid B)P(B)}{P(S)} = 0.533$$

$$P(C \mid S) = \frac{P(S \mid C)P(C)}{P(S)} = 0.2$$

Pertanto, sará stato l'impiegato B a seguire la pratica con probabilitá 0.533.

Moduli 2 e 3

Esercizio 1. (Distribuzione di Poisson, Funzioni generatrici)
Supponiamo di avere un panettone (di volume unitario $V = 1$) in cui ci siano uvette di due tipi, A e B, e che le uvette siano distribuite nel panettone secondo il modello di Poisson, ovvero la variabile aleatoria Poissoniana K_A conta le uvette di tipo A in una porzione di volume v e la variabile aleatoria Poissoniana K_B conta le uvette di tipo B, sempre in una porzione di volume v. Sappiamo che la densitá (numero medio / volume) delle uvette di tipo A é ϱ_A, e la densitá delle uvette di tipo B é ϱ_B.

(a) Calcolare la probabilitá che in una porzione di volume v del panettone ci siano esattamente n uvette.

Soluzione: Il problema si risolve calcolando la probabilitá

$$P(K_A + K_B = n)$$

Per determinare la distribuzione della somma di V.A. $K_A + K_B$ ricorriamo al metodo delle funzioni generatrici. Entrambe le V.A. hanno distribuzione poissoniana. Per una generica distribuzione di Poisson di parametro μ la funzione generatrice corrispondente é

$$G_K(u) = \sum_{k=0}^{\infty} \frac{\mu^k}{k!} e^{-\mu} u^k = e^{-\mu} \sum_{k=0}^{\infty} \frac{\mu^k}{k!} u^k = e^{-\mu} \sum_{k=0}^{\infty} \frac{(\mu u)^k}{k!} = e^{-\mu} e^{\mu u} = e^{\mu(u-1)},$$

dove si é fatto uso della serie esponenziale.
Poiché

$$K_A \sim Pois(k_A \mid \mu_A),$$

$$K_B \sim Pois(k_B \mid \mu_B),$$

la distribuzione della VA $K = K_A + K_B$ si puó ottenere mediante le funzioni generatrici:

$$G_K(u) = G_{K_A}(u) G_{K_B}(u) = e^{\mu_{K_A}(u-1)} e^{\mu_{K_B}(u-1)} = e^{(\mu_{K_A} + \mu_{K_B})(u-1)} = e^{\mu(u-1)}$$

dove si vede che la distribuzione $G_K(u)$ é ancora una distribuzione di Poisson, di media $\mu = \mu_{K_A} + \mu_{K_B}$.
(Il risultato é facilmente verificabile derivando k volte la funzione generatrice e ponendo $u = 0$ secondo la definizione

$$P_K(k) = \frac{1}{k!} \frac{d^k}{du^k} e^{\mu(u-1)}|_{u=0} = \frac{1}{k!} \mu^k e^{-\mu}$$

)
Dai dati del problema:

$$\mu_{K_A} = \varrho_A v, \mu_{K_B} = \varrho_B v.$$

Pertanto:

$$P(K_A + K_B = n) = \frac{((\varrho_A + \varrho_B)v)^n}{n!} e^{-(\varrho_A + \varrho_B)v}$$

Esercizio 2. (Distribuzione Normale, uso di tabelle)
La durata in giorni di una gravidanza stribuita come una normale di media 270 giorni e deviazione standard 10 giorni. Una signora un po' distratta cerca di capire chi sia il padre di suo figlio. Un suo partner é stato all'estero dal 290 al 240-esimo giorno antecedente la nascita del bambino.

(a) Se quest'uomo é veramente il padre del bambino qual é la probabilitá che la signora abbia avuto una gravidanza cosí corta o cosí lunga?

Soluzione: Definiamo:

- $T =$ `durata della gravidanza`

Il concepimento prima dei 290 giorni oppure dopo il 240-esimo giorno dalla nascita puó essere definito come:

- $C = (T > 290) \cup (T < 240)$

La probabilitá di interesse é dunque:

$$P(C) = P((T > 290) \cup (T < 240)) = P(T > 290) + P(T < 240)$$

Standardizzando le variabili:

$$P(C) = P(\frac{T - 270}{10} > \frac{290 - 270}{10}) + P(\frac{T - 270}{10} < \frac{240 - 270}{10}) = P(\frac{T - 270}{10} > 2) + P(\frac{T - 270}{10} < -3),$$

ovvero, usando le proprietá della CDF normale standard:

$$P(C) = 1 - P(\frac{T - 270}{10} \leq 2) + P(\frac{T - 270}{10} < -3) = 1 - \Phi(2) + \Phi(-3) = 1 - \Phi(2) + 1 - \Phi(3).$$

Accedendo alla tabella della Normale standard per $z = 2$ e $z = 3$ e leggendo i valori di $\Phi(2)$ e $\Phi(3)$:

$$P(C) = 0.0241$$

Una probabilitá invero non molto elevata.

Esercizio 3. (Distribuzioni Esponenziale negativa e Gaussiana)

Due ipotesi sono in competizione per la spiegazione di un fenomeno sperimentale.
La prima assume che dato un input $T = n$ (n, numero intero, per semplicitá), l'output Y di un sistema dovrebbe essere una variabile aleatoria che segue una legge esponenziale negativa con parametro $\lambda = \frac{1}{n}$. L'ipotesi alternativa postula che a paritá di input, Y segue una legge Normale con parametri $\mu = n, \sigma^2 = n^2$.

(a) Come si differenziano le due ipotesi nel predire la probabilitá di un output tale che $Y > 2n$ per $T = n$?

Soluzione: Per un V.A. con legge esponenziale di parametro $\lambda = \frac{1}{n}$,

$$P(Y > 2n) = S(2n) = e^{-\lambda 2n} = e^{-\frac{1}{n}2n} = e^{-2} = 0.13534$$

Per una V.A. con legge Normale, riscrivendo $P(Y > 2n)$ in forma standard,

$$P\left(\frac{Y - n}{n} > \frac{2n - n}{n}\right) = P(Z > 1) = 1 - \Phi(1)$$

Usando la tabella della Normale standard si ha che

$$P(Y > 2n) = 0.15866$$

I due valori si differenziano ma non troppo.

(b) Viene effettuato un esperimento per $T = 10$ e si osserva a posteriori che $Y \in (9{,}12)$. Assumendo una probabilitá a priori uniforme sugli input $P(T = n) = \frac{1}{10}$, $n = 1{,}2{,}3 \cdots 10$, quale delle due ipotesi é piú probabile ovvero spiega meglio l'osservazione effettuata?

Soluzione: Si tratta in buona sostanza di calcolare la probabilitá a a posteriori $P(T = 10 \mid 9 < Y < 12)$ per entrambe le ipotesi usando la regola di Bayes:

$$P(T = 10 \mid 9 < Y < 12) = \frac{P(9 < Y < 12 \mid T = 10)P(T = 10)}{P(9 < Y < 12)}$$

Per l'ipotesi esponenziale la verosimiglianza vale:

$$P(9 < Y < 12 \mid T = n) = F_{T=n}(12) - F_{T=n}(9) = 1 - e^{-\lambda 12} - 1 + e^{-\lambda 9} = e^{-\frac{1}{n}9} - e^{-\frac{1}{n}12}$$

Il termine a denominatore (l'evidenza), ricordando che rappresenta la probabilitá totale

$$P(9 < Y < 12) = \sum_{n=1}^{10} P(9 < Y < 12 \mid T = n)P(T = n),$$

vale:

$$P(9 < Y < 12) = \left(F_T(12) - F_{T(9)}\right)\frac{1}{10} = \sum_{n=1}^{10} \frac{1}{10}\left(e^{-\frac{1}{n}9} - e^{-\frac{1}{n}12}\right).$$

Pertanto,

$$P(T = 10 \mid 9 < Y < 12) = \frac{\left(e^{-\frac{1}{n}9} - e^{-\frac{1}{n}12}\right)\frac{1}{10}}{\sum_{n=1}^{10}\frac{1}{10}\left(e^{-\frac{1}{n}9} - e^{-\frac{1}{n}12}\right)} \approx 0.15828.$$

Per l'ipotesi Normale si procede esattamente esattamente nello stesso modo, scrivendo la verosimiglianza forma standard:

$$P\left(9 < Y < 12 \mid T = n\right) = P\left(\frac{9-n}{n} < \frac{Y-n}{n} < \frac{12-n}{n} \mid T = n\right) = \Phi\left(\frac{12}{n} - 1\right) - \Phi\left(\frac{9}{n} - 1\right),$$

e dunque

$$P(9 < Y < 12) = \sum_{n=1}^{10} \frac{1}{10}\left(\Phi\left(\frac{12}{n} - 1\right) - \Phi\left(\frac{9}{n} - 1\right)\right).$$

Per l'ipotesi Gaussiana abbiamo infine:

$$P(T = 10 \mid 9 < Y < 12) = \frac{\Phi(0.2) - \Phi(-0.1)}{\sum_{n=1}^{10}\left(\Phi\left(\frac{12}{n} - 1\right) - \Phi\left(\frac{9}{n} - 1\right)\right)} \approx 0.1$$

Nel calcolare il denominatore, per velocizzare, notiamo che $z_n = \frac{12}{n} - 1$ da' origine alla sequenza

$$z_n = 11.0000, 5.0000, 3.0000, 2.0000, 1.4000, 1.0000, 0.7143, 0.5000, 0.3333, 0.2000,$$

mentre per il termine $z_n = \frac{9}{n} - 1$

$$z_n = 8.0000, 3.5000, 2.0000, 1.2500, 0.8000, 0.5000, 0.2857, 0.1250, 0, -0.1000.$$

Eliminiamo i termini uguali e utilizziamo l'approssimazione $P(Z < 3.5) \approx 1$. Per i termini restanti, leggiamo $\Phi(z_n)$ in tabella.
Ne concludiamo che l'ipotesi esponenziale é da preferirsi. Notiamo anche che l'ipotesi Gaussiana produce un valore a posteriori ($\approx 1_{\overline{10}}$) che non si differenzia dalla probabilitá a priori $P(T = n) = \frac{1}{10}$

Compito scritto dell'esame di Statistica e analisi dei dati (Crema) 6 settembre 2016	Prof. Giuseppe Boccignone	Corso di Laurea
Cognome:	Nome:	Matricola:

Istruzioni

- Il tempo riservato alla prova scritta e di 2 ore e 30 minuti. Durante la prova e possibile consultare libri e appunti.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede piu passaggi di calcolo, non sara preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o piu frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sara eventualmente rappresentato in quest'ultima forma).
- SOLO PER STUDENTI DEL CORSO ONLINE (VECCHIO ORDINAMENTO, Prof. Gianini): Indicare quale Modulo viene svolto al fine di completare una prova d'esame parziale precedentemente sostenuta con esito positivo
  – Modulo 1 ☐
  – Moduli 2 e 3 ☐

Modulo 1

Esercizio 1. (Principio di additivitá generalizzata)
Un sistema é formato da tre componenti disposti in parallelo. Il sistema funziona se almeno due dei tre componenti funzionano correttamente. Indicati con C_1, C_2, C_3 gli eventi "funzionamento corretto" del primo, secondo e terzo componente, rispettivamente, siano $P(C_1) = P(C_2) = P(C_3) = p$ con $p = 0.7$. Si assuma che gli eventi C_1, C_2, C_3 siano indipendenti tra di loro.

(a) Si mostri che la probabilitá che il sistema funzioni correttamente é maggiore della probabilitá di funzionamento dei singoli componenti.

Soluzione:
Definiamo l'evento $C =$ `"il sistema funziona correttamente"`. Il problema si risolve calcolando la probabilitá $P(C)$ e mostrando che $P(C) > p$.
Il sistema funziona se almeno due componenti funzionano correttamente. I casi possibili che si possono presentare sono 8, di questi, quelli in cui almeno due generatori sono attivi sono 4: il caso in cui funzionano tutti e tre, e i tre casi in cui si ha il malfunzionamente di uno solo. Dunque

$$C = \{C_1 \cap C_2 \cap C_3\} \cup \{\sim C_1 \cap C_2 \cap C_3\} \cup \{C_1 \cap \sim C_2 \cap C_3\} \cup \{C_1 \cap C_2 \cap \sim C_3\}.$$

Applicando il Principio di additivitá generalizzata, la probabilitá di corretto funzionamento del sistema é

$$P(C) = P(\{C_1 \cap C_2 \cap C_3\}) + P(\{\sim C_1 \cap C_2 \cap C_3\}) + P(\{C_1 \cap \sim C_2 \cap C_3\}) + P(\{C_1 \cap C_2 \cap \sim C_3\})$$

Infine, usando l'indipendenza degli eventi:

$$P(T) = p^3 + 3p^2(1-p) = 3p^2 - 2p^3 = 1.47 - 0.686 = 0.784 > p$$

Esercizio 2. (Probabilitá condizionata, regola del prodotto)
Un'urna contiene 1 pallina nera (N) e 2 palline bianche (B). Si estrae casualmente una pallina dall'urna e, dopo averne osservato il colore, la si rimette nell'urna aggiungendo altre 2 palline del colore estratto e 3 palline del colore non estratto.

i	Nere	Bianche
1	1	2
2	4	4
3	6	7
4	8	10

(a) Determinare la probabilitá che in 4 estrazioni successive, effettuate secondo la regola sopra stabilita, si ottenga la stringa (ordinata) $BNNB$

Soluzione: Indichiamo con B_i, N_i $(i = 1, \cdots, 4)$ gli eventi: `"si ha una pallina Bianca (Nera) alla i-esima estrazione"`. Dopo ogni estrazione cambia lo spazio campione, e se gli esiti delle prime tre estrazioni seguono la sequenza voluta $B_1 N_2 N_3$, il numero delle palline presenti nell'urna quando avviene la i-esima estrazione si modifica come mostrato nella tabella.
Allora si ha:

$$P(B_1) = \frac{2}{3}$$

$$P(N_2 \mid B_1) = \frac{4}{8} = \frac{1}{2}$$

$$P(N_3 \mid N_2 \cap B_1) = \frac{6}{13}$$

$$P(B_4 \mid N_3 \cap N_2 \cap B_1) = \frac{10}{18} = \frac{5}{9}$$

e di conseguenza la probabilitá che si verifichi la sequenza $BNNB$ vale, per la regola del prodotto:

$$P(B_1 \cap N_2 \cap N_3 \cap B_4) = \frac{2}{3} \times \frac{1}{2} \times \frac{6}{13} \times \frac{5}{9} = \frac{10}{117} \approx 0.08547$$

Esercizio 3. (Teorema di Bayes)
Quattro arcieri A, B, C, D scoccano la loro freccia contemporaneamente e hanno probabilitá, rispettivamente, $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$ e $\frac{1}{5}$ di colpire il bersaglio

(a) Che probabilitá c'é che dopo il tiro simultaneo risulti conficcata nel bersaglio esattamente una freccia?

Soluzione: Definiamo i seguenti eventi

- $C_A =$ `"A colpisce il bersaglio"`
- $C_B =$ `"B colpisce il bersaglio"`
- $C_D =$ `"C colpisce il bersaglio"`
- $C_C =$ `"D colpisce il bersaglio"`
- $C_{una} =$ `"Esattamente 1 freccia nel bersaglio"`

Dai dati del problema:

$$P(C_A) = \frac{1}{2}, P(C_B) = \frac{1}{3}, P(C_C) = \frac{1}{4}, P(C_D) = \frac{1}{5}$$

Indichiamo con $\sim C_A, \cdots$ gli eventi complementari, di probabilitá $P(\sim C_A) = 1 - P(C_A), \cdots$, e per semplicitá usiamo la notazione $P(\{X \cap Y\}) \equiv P(X,Y)$. Pertanto:

$$P(C_{una}) = P(C_A, \sim C_B, \sim C_C, \sim C_D) + P(\sim C_A, C_B, \sim C_C, \sim C_D) + P(\sim C_A, \sim C_B, C_C, \sim C_D) + P(\sim C_A, \sim C_B, \sim C_C, C_D) =$$

$$= \frac{1}{2} \times \frac{2}{3} \times \frac{3}{4} \times \frac{4}{5} + \frac{1}{2} \times \frac{1}{3} \times \frac{3}{4} \times \frac{4}{5} + \frac{1}{2} \times \frac{2}{3} \times \frac{1}{4} \times \frac{4}{5} + \frac{1}{2} \times \frac{2}{3} \times \frac{3}{4} \times \frac{1}{5} = \frac{50}{120}$$

(b) Se dopo il tiro simultaneo risulta conficcata nel bersaglio una e una sola freccia, che probabilitá c'é che si tratti di quella dell'arciere A?

Soluzione: Per rispondere al quesito si tratta di inferire usando la regola di Bayes:

$$P(C_A \mid U) = \frac{P(U \mid C_A) P(C_A)}{P(U)}$$

dove abbiamo definito l'evento $U =$ `"una e una sola freccia conficcata nel bersaglio"`.
Poiché

$$P(U \mid C_A) = P(\sim C_B, \sim C_C, \sim C_D)$$

e

$$P(U \mid C_A)P(C_A) = P(\sim C_B, \sim C_C, \sim C_D)P(C_A) = P(C_A, \sim C_B, \sim C_C, \sim C_D),$$

dove si é applicata la proprietá di indipendenza, ovvero il fatto che le prestazioni di B, C e D in un dato tiro non sono condizionate dall'esito di A (formalmente, $P(U \mid C_A) = P(\sim C_B, \sim C_C, \sim C_D \mid C_A) = P(\sim C_B, \sim C_C, \sim C_D)$).
Valendo la medesima proprietá nei casi $P(U \mid C_B), \cdots$, risulta che

$$P(U) = P(C_{una}) = \frac{50}{120}$$

calcolata precedentemente, come era da aspettarsi in termini logici (`"Esattamente 1 freccia nel bersaglio"` $\equiv$ `"una e una sola freccia conficcata nel bersaglio"`).
Sostituendo in Bayes:

$$P(C_A \mid U) = \frac{P(U \mid C_A)P(C_A)}{P(U)} = \frac{\frac{1}{2} \times \frac{2}{3} \times \frac{3}{4} \times \frac{4}{5}}{\frac{50}{120}} = \frac{\frac{24}{120}}{\frac{50}{120}} = \frac{24}{50} = 0.48$$

.

Moduli 2 e 3

Esercizio 1. (Distribuzione Binomiale)
Viene trasmesso un segnale binario, dove ciascun bit, in maniera indipendente, nel 10% dei casi viene ricevuto erroneamente.

(a) Se ciascun pacchetto é composto di 10 bit, qual é la distribuzione del numero di errori in ricezione ?

Soluzione: Per indipendenza, la distribuzione del numero di errori X segue una legge binomiale

$$P(X) = Bin(X; n, p) = Bin(X; 10, 0.1) = \binom{10}{x} 0.1^x 0.9^{(10-x)}$$

(b) Qual é la probabilitá che un pacchetto arrivi corrotto (cioé non tutti i bit sono ricevuti correttamente)?

Soluzione: La probabilitá che il pacchetto ricevuto sia corrotto é

$$1 - P(X = 0) = 1 - \left[\binom{10}{0} 0.1^0 0.9^{(10-0)}\right] = 1 - 0.9^{10} = 0.65132$$

(c) Si supponga che il decimo bit sia un bit di paritá: viene settato a 1 se gli altri 9 bit hanno un numero dispari di 1, a 0 se ne hanno un numero pari. Qual é la probabilitá che un pacchetto corrotto non sia rilevato come tale? (Diciamo che un pacchetto é non corrotto se nessun bit é stato corrotto e il bit di paritá é settato correttamente)

Soluzione:
E' semplice verificare che il bit di paritá indica erroneamente che il pacchetto é "buono" quando un numero pari di bit del pacchetto é stato invertito (includendo eventualmente il bit di paritá medesimo) durante la trasmissione. Questo puó accadere se hanno avuto luogo 2,4,6,8 o 10 errori. Pertanto la probabilitá di non rilevamento $P(\sim R)$ é, per additivitá di tali eventi disgiunti, $P(\sim R) = P(\{X = 2\} \cup \{X = 4\} \cup \{X = 6\} \cup \{X = 8\} \cup \{X = 10\}) = P(\{X = 2\}) + P(\{X = 4\}) + P(\{X = 6\}) + P(\{X = 8\}) + P(\{X = 10\})$. Poiché ciascun evento segue la legge binomiale stabilita precedentemente:

$$P(\sim R) = \binom{10}{2} 0.1^2 0.9^8 + \binom{10}{4} 0.1^4 0.9^6 + \binom{10}{6} 0.1^6 0.9^4 + \binom{10}{8} 0.1^8 0.9^2 + \binom{10}{10} 0.1^{10} 0.9^0 = 0.20501$$

Esercizio 2. (Distribuzione Normale)
Il contenuto reale di birra che un apposito macchinario mette in fusti da 5 litri puó essere considerato come una variabile aleatoria avente una distribuzione normale con una deviazione standard pari a 0.05 litri.

(a) Se solo il 2% dei fusti contiene meno di 5 litri, quale dovrebbe essere il contenuto medio dei fusti?

Soluzione:
Per determinare μ tale che $\Phi(z) = F_X\left(\frac{5-\mu}{\sigma}\right) = 0.02$, si cerca nella tabella della distribuzione normale standard il valore piú vicino a 0.02. Si trova la migliore approssimazione per 0.0202 che corrisponde a $z = -2.05$.
Qualora la tabella a disposizione consideri solo i valori per $z \geq 0$, si faccia uso della proprietá della cumulativa standardizzata $\Phi(-z) = 1-\Phi(z)$; in tal caso la migliore approssimazione si ha per $1-\Phi(2.05) = 1-0.9798 = 0.0202$.
Pertanto:

$$z = \frac{5-\mu}{\sigma} = \frac{5-\mu}{0.05} = -2.05.$$

Risolvendo l'equazione in μ, si trova che $\mu = 5.1$ litri.

Esercizio 3. (Teorema del limite centrale)
Un barattolo di pittura da 1 litro copre in media una superficie di circa 12 metri quadri con una deviazione standard di 0.7 metri quadri.

(a) Qual é la probabilitá che la media campionaria dell'area coperta da un campione di 40 barattoli da 1 litro assuma un valore compreso tra 11.78 e 12.10 metri quadri?

Soluzione:
In breve, il Teorema del limite centrale ci dice che per un numero n grande ($n \to \infty$) di V.A. indipendenti e identicamente distribuite vale

$$\begin{aligned} &\lim_{n\to\infty} P(a \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b) = \\ =&\lim_{n\to\infty} P(a \leq \frac{\sqrt{n}}{\sigma} \cdot \left(\frac{S_n}{n} - \mu\right) \leq b) = \\ =&\Phi(b) - \Phi(a), \end{aligned} \tag{1}$$

dove $S_n = \sum_{i=1}^{n} X_i$ e $\frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$ é la media campionaria. In buona sostanza, se $\overline{X}$ é la media del campione di taglia n estratto da una popolazione di media μ e varianza σ^2, il teorema garantisce che la variabile ridotta $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ sia una variabile aleatoria la cui funzione di distribuzione tende alla distribuzione normale standard Φ.
Nel caso in esame $\mu = 12$ metri quadri, $\sigma = 0.7$ metri quadri, $n = 40$.
Pertanto, la probabilitá che la media campionaria $\overline{X}$ assuma un valore compreso tra $\overline{X}_a = 11.78$ e $\overline{X}_b = 112.10$ é:

$$P(\overline{X}_a \leq \overline{X} \leq \overline{X}_b) = \Phi\left(\frac{\overline{X}_b - \mu}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{\overline{X}_a - \mu}{\sigma/\sqrt{n}}\right)$$

dove:

$$a = \left(\frac{\overline{X}_a - \mu}{\sigma/\sqrt{n}}\right) = \left(\frac{11.78 - 12}{0.7/\sqrt{40}}\right) \approx -1.987$$

$$b = \left(\frac{\overline{X}_b - \mu}{\sigma/\sqrt{n}}\right) = \left(\frac{12.10 - 12}{0.7/\sqrt{40}}\right) \approx 0.903$$

Accedendo alla tabella della distribuzione normale standard leggiamo $\Phi(-1.987) = 0.0239$ e $\Phi(0.903) = 0.8238$.
Si ricava dunque che

$$P(\overline{X}_a \leq \overline{X} \leq \overline{X}_b) = \Phi(0.903) - \Phi(-1.987) = 0.791$$

Compito scritto dell'esame di Statistica e analisi dei dati (Crema) 07 novembre 2016	Prof. Giuseppe Boccignone	Corso di Laurea
Cognome:	Nome:	Matricola:

Istruzioni

- Il tempo riservato alla prova scritta e di 2 ore e 30 minuti. Durante la prova e possibile consultare libri e appunti.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede piu passaggi di calcolo, non sara preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o piu frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sara eventualmente rappresentato in quest'ultima forma).
- SOLO PER STUDENTI DEL CORSO ONLINE (VECCHIO ORDINAMENTO, Prof. Gianini): Indicare quale Modulo viene svolto al fine di completare una prova d'esame parziale precedentemente sostenuta con esito positivo
  - Modulo 1 ☐
  - Moduli 2 e 3 ☐

Modulo 1

Esercizio 1.
Se gli eventi A e B sono incompatibili, allora $P(A) \leq P(\sim B)$.

(a) Vero o falso?

Soluzione: (Assiomi di probabilitá)
E' vero.
Perché se sono incompatibili allora:

$$A \subseteq \sim B,$$

dove $\sim B = S - B$.
Da cui si deduce, per gli assiomi della probabilitá che $P(A) \leq P(\sim B)$.

Esercizio 2.
Due urne contengono palline bianche e nere in proporzioni diverse. Siano p_1 e p_2 le probabilitá di estrarre una pallina bianca rispettivamente dall'urna 1 e dall'urna 2. Il giocatore vince se estraendo due palline almeno una é bianca. Egli puó scegliere tra due modalitá di estrazione:

A) Sceglie a caso una delle due urne, estrae una pallina, la rimette nell'urna da cui é stata estratta, quindi sceglie di nuovo a caso un'urna ed estrae la seconda pallina.

B) Sceglie a caso una delle due urne, estrae una pallina, la rimette nell'urna da cui é stata estratta, e, sempre dalla stessa urna, estrae una seconda pallina.

(a) Quale tra le due procedure é piú conveniente per la vittoria?

Soluzione: (Probabilitá totali)
Indichiamo con U_i, $i = 1{,}2$ la scelta di una delle due urne.
Definiamo inoltre gli eventi:

$$N_i = \text{“pallina nera alla i-esima estrazione”}$$

$$E = \text{“estrazione di almeno una pallina bianca”}$$

Si ha pertanto che:

$$P(U_i) = 0.5$$

$$P(E) = 1 - P(N_1 \cap N_2)$$

Seguendo la procedura A), le due estrazioni sono statisticamente indipendenti:

$$P(N_1 \cap N_2) = P(N_1)P(N_2) = P(N_1 \mid U_1)P(U_1) + P(N_1 \mid U_2)P(U_2) \times P(N_2 \mid U_1)P(U_1) + P(N_2 \mid U_2)P(U_2) =$$

$$= [(1-p_1)0.5 + (1-p_2)0.5]^2$$

da cui si ricava

$$P_A(E) = 1 - P(N_1 \cap N_2) = 1 - \left[\frac{(1-p_1)}{2} + \frac{(1-p_2)}{2}\right]^2 = p_1 + p_2 - \frac{(p_1+p_2)^2}{4}$$

Seguendo invece la procedura B), la probabilitá di estrarre due palline nere dalla medesima urna (con reimbussolamento della pallina estratta) vale:

$$P(N_1 \cap N_2 \mid U_i) = P(N_1 \mid U_i)P(U_i) = (1-p_i)^2$$

con $i = 1{,}2$. Pertanto, differentemente da A):

$$P(N_1 \cap N_2) = P(N_1 \cap N_2 \mid U_1)P(U_1) + P(N_1 \cap N_2 \mid U_2)P(U_2) = (1-p_1)^2 0.5 + (1-p_2)^2 0.5.$$

Dunque nel caso B)

$$P_B(E) = 1 - P(N_1 \cap N_2) = 1 - \left[\frac{(1-p_1)^2}{2} + \frac{(1-p_2)^2}{2}\right] = p_1 + p_2 - \frac{(p_1^2+p_2^2)}{2}.$$

La differenza fra le due probabilitá é

$$P_A(E) - P_B(E) = -\frac{(p_1+p_2)^2}{4} + \frac{(p_1^2+p_2^2)}{2} = \frac{(p_1-p_2)^2}{4} > 0$$

In conclusione, tra le due strategie é piú conveniente A) essendo $P_A(E) > P_B(E)$.

Esercizio 3.

Due ditte forniscono il medesimo prodotto. Se esso proviene dalla ditta A, la probabilitá che si guasti prima dell?istante t vale $1 - e^{-t}$; se invece proviene dalla ditta B questa probabilitá vale $1 - e^{-2t}$. Il prodotto puó essere acquistato con uguale probabilitá da A o da B, e non é nota la ditta fornitrice. Tuttavia, é stato osservato che il prodotto si guasta in un intervallo di tempo $1 \le t \le 2$.

(a) Determinare la probabilitá che esso sia stato acquistato dalla ditta A.

Soluzione: (Teorema di Bayes)
Definiamo l'evento
$G =$ "guasto in $1 \le t \le 2$"
I dati del problema ci dicono che, a priori, le probabilitá che il prodotto provenga da A o da B valgono

$$P(A) = P(B) = 0.5.$$

La probabilitá di guasto del prodotto A nell'intervallo di tempo $1 \le t \le 2$ vale

$$P(G \mid A) = 1 - e^{-2} - (1 - e^{-1}) = e^{-1} - e^{-2}.$$

Quella del prodotto B nello stesso intervallo é

$$P(G \mid B) = 1 - e^{-2\cdot 2} - (1 - e^{-2\cdot 1}) = e^{-2} - e^{-4}.$$

Applicando la regola di Bayes:

$$P(A \mid G) = \frac{P(G \mid A)P(A)}{P(G \mid A)P(A) + P(G \mid B)P(B)} = \frac{e^{-1} - e^{-2}}{e^{-1} - e^{-2} + e^{-2} - e^{-4}} = \frac{e^2(e-1)}{e^3-1} \approx 0.6652$$

Moduli 2 e 3

Esercizio 1.
Una moneta non truccata viene lanciata infinite volte.

(a) Calcolare la probabilitá di osservare 3 volte Testa esattamente e non prima del sesto lancio.

Soluzione: (Distribuzione binomiale) In generale, l'evento di interesse é

$$A = \text{“vengono osservate } k \text{ teste in } n \text{ lanci ma non prima di } n \text{ lanci ”}$$

Tale evento accade se e solo se i due eventi seguenti accadono congiuntamente:

$$B = \text{“vengono osservate } k-1 \text{ teste in } n-1 \text{ lanci ”}$$

$$C = \text{“viene osservata Testa al lancio } n\text{-esimo”}$$

Si ha pertanto che $P(A) = P(B,C)$ e poiché B e C sono indipendenti

$$P(A) = P(B)P(C)$$

Ogni singolo lancio é indipendente dai precedenti, dunque

$$P(C) = P(\text{Testa}) = p = \frac{1}{2}$$

essendo la moneta non truccata.
$P(B)$ ha distribuzione binomiale:

$$P(B) = \binom{n-1}{k-1} p^{k-1} q^{n-1-(k-1)}$$

Pertanto:

$$P(A) = P(B)P(C) = \binom{n-1}{k-1} p^{k-1} q^{n-1-(k-1)} p = \binom{n-1}{k-1} p^{k} q^{n-1-(k-1)} = 0.1562,$$

per $k = 3, n = 6, p = q = \frac{1}{2}$

Esercizio 2.
La variabile aleatoria (v.a.) $X(\omega)$ ha densitá di probabilitá: $\mathrm{f}_X(x) = \begin{cases} k(x-1)^2 & 0 \le x \le 2 \\ 0 & \text{altrimenti} \end{cases}$.

(a) Calcolare la probabilitá che $X(\omega)$ assuma valori in un intorno di raggio $\delta = 0.5$ del suo valor medio.

Soluzione: (Normalizzazone di Variabili aleatorie e media) Anzitutto, determiniamo il valore di k imponendo la condizione di normalizzazione della densitá $f_X(x)$:

$$k \int_0^2 (x-1)^2 dx = 1 \implies k = \frac{3}{2}.$$

Pertanto la densitá correttamente normalizzata é:
$\mathrm{f}_X(x) = \begin{cases} \frac{3}{2}(x-1)^2 & 0 \le x \le 2 \\ 0 & \text{altrimenti} \end{cases}$.
Il problema si risolve calcolando

$$P_X(|X - E[X]| < \delta) = \cdots \tag{1}$$

Il valore atteso di X é

$$E[X] = \frac{3}{2} \int_0^2 x(x-1)^2 dx = \frac{3}{2} \left[\frac{x^4}{4} - \frac{2x^3}{3} + \frac{x^2}{2} \right]_0^2 = 1.$$

la soluzione (1) si calcola pertanto integrando la densitá nell'intervallo $(1 - 0.5 \le x \le 1 + 0.5)$:

$$P_X(|X - E[X]| < \delta) = \frac{3}{2} \int_{\frac{1}{2}}^{\frac{3}{2}} (x-1)^2 dx = 3 \int_1^{\frac{3}{2}} (x-1)^2 dx = \frac{1}{8} \tag{2}$$

Esercizio 3.
Un vostro amico asserisce di aver ottenuto una media di 3.25 punti per lancio su 1000 lanci di un dado non truccato.

(a) Qual é la probabilitá che stia mentendo?

Soluzione: (Teorema limite centrale)
Sia X_i la variabile aleatoria (v.a.) che denota l'esito dell'i-simo lancio di dado.
Sappiamo che le v.a. $X_1, X_2, \cdots, X_{1000}$ sono indipendenti e identicamente distribuite con distribuzione uniforme sull'insieme di interi $\{1,2,\cdots,6\}$ da cui é facile ricavare il valore atteso e la deviazione standard

$$E[X_i] = 3.5$$
$$\sigma_{X_i} = \sqrt{E[X_i^2] - E[X_i]^2} \approx 1.7078$$

In breve, il Teorema del limite centrale ci dice che per un numero n grande ($n \to \infty$) di V.A. indipendenti e identicamente distribuite di valore atteso μ e deviazione standard σ vale

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x) \tag{3}$$

dove $S_n = \sum_{i=1}^{n} X_i$. Nella fattispecie $n = 1000$ e S_n é il numero di punti totalizzati in 1000 lanci.
Sotto tali ipotesi, la probabilitá che il vostro amico affermi il vero si riduce a calcolare la probabilitá definita nel T.L.C riscritta come

$$P\left(\frac{\frac{S_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq x\right) = \Phi(x) \tag{4}$$

dove $\frac{S_n}{n} = 3.25 = \overline{X}$ é la media campionaria del campione di taglia $n = 1000$ dichiarata dal vostro amico. In buona sostanza il T.L.C. il teorema garantisce che la variabile ridotta $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ sia una variabile aleatoria la cui funzione di distribuzione tende alla distribuzione normale standard Φ. Pertanto,

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{3.25 - 3.5}{1.7078/\sqrt{1000}} = -4.629$$

La probabilitá che il vostro amico dichiari il vero é dunque pari a $\Phi(-4.629)$
Anche senza consultare le tabelle e tenendo presente che per la normale standard $\sigma = 1$ e dunque $z = z\sigma$, significa che stiamo valutando la probabilitá di un valore di media dichiarata che é di 4.629 deviazioni standard al di sotto della media $\mu = 3.5$.
Considerando la legge 3σ, é immediatamente evidente che la probabilitá $\Phi(-4.629)$ é molto bassa e pertanto la probabilitá che che il vostro amico dica il falso molto elevata $(1 - \Phi(-4.629))$.
Se poi, per scrupolo, si consultasse la tabella della normale ridotta, si leggerebbe che $\Phi(-4.629) = 1.84 \times 10^{-7}$

Compito scritto dell'esame di Statistica e analisi dei dati (Crema) 26 gennaio 2017	Prof. Giuseppe Boccignone	Corso di Laurea
Cognome:	Nome:	Matricola:

Istruzioni

- Il tempo riservato alla prova scritta e di 2 ore e 30 minuti. Durante la prova e possibile consultare libri e appunti.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede piu passaggi di calcolo, non sara preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o piu frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sara eventualmente rappresentato in quest'ultima forma).
- SOLO PER STUDENTI DEL CORSO ONLINE (VECCHIO ORDINAMENTO, Prof. Gianini): Indicare quale Modulo viene svolto al fine di completare una prova d'esame parziale precedentemente sostenuta con esito positivo
  - Modulo 1 ☐
  - Moduli 2 e 3 ☐

Modulo 1

Esercizio 1.

Dovete partecipare ad un torneo di scacchi in cui giocate contro altri tre giocatori (una sola partita con ciascuno). Conoscete le probabilitá di vincere con ognuno di loro e potete scegliere l'ordine in cui incontrarli. Si vince il torneo se si vincono due partite consecutive.

(a) Volendo massimizzare la probabilitá di vittoria, si mostri che l'ordinamento ottimale é quello che vi fa incontrare alla seconda partita il giocatore piú debole dei tre, mentre l'ordine in cui incontrate i restanti due giocatori non ha rilevanza.

Soluzione:

Denotiamo con

$$P(\texttt{"vittoria contro il giocatore incontrato alla partita i -esima"}) = P(i) = p_i$$

con $i = 1{,}2{,}3$.

La descrizione del problema ci dice che si vince la partita se si gioca la seconda partita con il giocatore piú debole - dunque con $p_2 \geq p_1, p_2 \geq p_3$ - e, **congiuntamente**, si vince con l'uno **oppure** l'altro giocatore incontrati alla partita 1 o 3. Quindi, formalmente, la probabilitá $P(V)$ di vincere il torneo é

$$\begin{aligned} P(V) = P(2 \cap (1 \cup 3)) = P(2)P(1 \cup 3) = P(2)\left[P(1) + P(3) - P(1) \cap P(3)\right] = \\ = p_2(p_1 + p_3 - p_1 p_3). \end{aligned}$$

L'ordinamento é ottimale se tale probabilitá non é inferiore alle probabilitá calcolate utilizzando i due ordinamenti alternativi $P(1 \cap (2 \cup 3)) = p_1(p_2 + p_3 - p_2 p_3)$ e $P(3 \cap (2 \cup 1)) = p_3(p_2 + p_1 - p_2 p_1)$, ovvero:

$$p_2(p_1 + p_3 - p_1 p_3) \geq p_1(p_2 + p_3 - p_2 p_3), \quad (1)$$
$$p_2(p_1 + p_3 - p_1 p_3) \geq p_3(p_2 + p_1 - p_2 p_1). \quad (2)$$

Infatti, la prima disuguaglianza é equivalente alla condizione

$$p_2 \geq p_1,$$

mentre la seconda disuguaglianza si riduce a

$$p_2 \geq p_3.$$

Le due condizioni (1) e (2) ci dicono che la probabilitá di vincere alla seconda partita deve essere massima, ovvero: per massimizzare la probabilitá di vincere il torneo il secondo giocatore da incontrare deve essere il piú debole dei tre.

Esercizio 2. Un lotto di cento componenti elettronici viene ispezionato da un compratore scegliendo quattro componenti a caso: se uno dei quattro componenti risulta essere difettoso, l'intero lotto viene scartato.

(a) Qual é la probabilitá che il lotto venga acquistato dal compratore se contiene cinque componenti difettosi?

Soluzione:
Definiamo gli eventi

$$A = \text{“il lotto viene acquistato”}$$

$$A_i = \text{“il componente i-simo non é difettoso”}$$

con $i = 1, 2, 3, 4$.
E' evidente che:

$$A = A_1 \cap A_2 \cap A_3 \cap A_4$$

Applicando banalmente la regola del prodotto

$$P(A) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_2, A_1)P(A_4 \mid A_3, A_2, A_1) = \frac{95}{100} \times \frac{94}{99} \times \frac{93}{98} \times \frac{92}{97} = 0.812$$

Esercizio 3. Marta cerca un documento che sa di avere archiviato in uno dei suoi raccoglitori. Presume di averlo lasciato nel raccoglitore j-esimo con probabilitá $p_j > 0$.
Purtroppo ciascun raccoglitore é cosí disordinato che anche se Marta decidesse correttamente di cercare il documento nel raccoglitore i-esimo che lo contiene, la probabilitá di trovarlo effettivamente nel raccoglitore sarebbe solo pari a r_i.
Alla fine Marta sceglie un particolare raccoglitore, diciamo i, ma il documento non viene trovato.

(a) Subordinatamente all'esito della ricerca, determinare la probabilitá che il documento fosse nel raccoglitore j nel caso $j \neq i$ e la probabilitá che il documento fosse in j nel caso $j = i$

Soluzione: (Teorema di Bayes)
Definiamo l'evento E che sintetizza l'esito della ricerca:

$$E = \text{“Il documento non viene trovato nel raccoglitore } i\text{”}$$

I dati del problema ci dicono che il documento é nel raccoglitore i con probabilitá p_i e che puó essere trovato con probabilitá r_i. Pertanto, la probabilitá dell'evento complementare

$$\sim E = \text{“Il documento viene trovato nel raccoglitore } i\text{”}$$

é pari a

$$P(\sim E) = p_i r_i,$$

ovvero:

$$P(E) = 1 - P(\sim E) = 1 - p_i r_i. \tag{3}$$

Denotiamo ora con B l'evento:

$$B = \text{“Il documento é nel raccoglitore } j\text{”}.$$

Vogliamo determinare nei due casi $j \neq i$ e $j = i$ la probabilitá condizionata

$$P(\text{ “Il documento é nel raccoglitore } j\text{”} \mid \text{ “Il documento non viene trovato nel raccoglitore } i\text{”}) = P(B \mid E),$$

ovvero vogliamo calcolare, usando Bayes, la probabilitá a posteriori

$$P(B \mid E) = \frac{P(E \mid B)P(B)}{P(E)} \tag{4}$$

1. $j \neq i$:
   In questo caso $E \cap B = B$. Se sappiamo che il documento é in j con $j \neq i$, la probabilitá che non venga trovato in i é $P(E \mid B) = 1$ (l'evento é certo). Sostituendo in (4) e usando la (3):

$$P(B \mid E) = \frac{1 \times P(B)}{P(E)} = \frac{p_j}{1 - p_i r_i}. \tag{5}$$

2. $j = i$:
   In questo caso la probabilitá

   $$P(E \mid B) = P(\text{ "Il documento non viene trovato nel raccoglitore } i\text{" } \mid \text{ "Il documento é nel raccoglitore } i\text{"}) = 1 - r_i$$

   Di nuovo, sostituendo in (4) e usando la (3):

   $$P(B \mid E) = \frac{(1 - r_i)P(B)}{P(E)} = \frac{(1 - r_i)p_j}{1 - p_i r_i}. \tag{6}$$

Moduli 2 e 3

Esercizio 1.
Un sistema di comunicazione consiste di un buffer che ospita i pacchetti provenienti dalla sorgente, e da una linea di comunicazione che recupera i pacchetti dal buffer e li trasmette a un ricevitore. Il sistema opera a coppie di *time-slot*. Nel primo slot il sistema mette nel buffer un certo numero di pacchetti generati dalla sorgente secondo una distribuzione di Poisson di parametro μ. La capacitá del buffer (numero massimo di pacchetti) é pari all'intero b_{max}; i pacchetti che arrivano a buffer pieno vengono scartati. Nel secondo slot temporale il sistema trasmette c pacchetti al ricevitore, dove c é un intero $0 < c < b_{max}$, oppure trasmette quelli effettivamente presenti se in numero inferiore a c.

(a) Assumendo che all'inizio del primo slot il buffer sia vuoto, trovare la PMF del numero di pacchetti bufferizzati alla fine del primo slot e la PMF alla fine del secondo slot.

Soluzione: Sia X la V.A. discreta che indica il numero di pacchetti bufferizzati alla fine del primo slot temporale Δt_1. La Figura 1 mostra il suo intervalli di variazione relativamente ai parametri del problema c, b_{max} Per determinare la PMF di X, P_X é sufficiente determinare: 1) la probabilitá di avere bufferizzato $k = 0,1,2,\cdots,b_{max}-1$ pacchetti; 2) la probabilitá di avere il buffer pieno che accade quando in Δt_1 vengono generati alla sorgente $k \geq b_{max}$ pacchetti.

1. $k = 0,1,2,\cdots,b_{max}-1$:
   La probabilitá di avere esattamente k pacchetti bufferizzati $P_X(X=k)$ é uguale alla probabilitá che k pacchetti siano stati generati alla sorgente durante Δt_1: ció implica, per quanto ci dice la specifica del problema, che X segua una legge di Poisson, $X \sim Pois(\mu)$ e dunque:
   $$P_X(X=k) = \frac{\mu^k}{k!}e^{-\mu} \tag{7}$$
   applicabile per $k = 0,1,2,\cdots,b_{max}-1$.
2. $k \geq b_{max}$:
   La probabilitá di avere il buffer pieno, ovvero che $X = b_{max}$ é la probabilitá che alla sorgente siano stati generati almeno b_{max} pacchetti, ovvero
   $$P_X(X=b_{max}) = \sum_{k=b_{max}}^{\infty} \frac{\mu^k}{k!}e^{-\mu} = 1 - \sum_{k=0}^{b_{max}-1} \frac{\mu^k}{k!}e^{-\mu}$$

Si noti come $P_X(X=0) + P_X(X=1) + \cdots + P_X(X=b_{max}-1) + P_X(X=b_{max}) = 1$, dunque P_X é una PMF correttamente normalizzata.
Chiamiamo ora Y la V.A. discreta che indica il numero di pacchetti bufferizzati alla fine del secondo slot temporale Δt_2. Il suo valore é determinabile come

$$Y = \text{ pacchetti bufferizzati durante } \Delta t_1 - \text{ pacchetti trasmessi durante } \Delta t_2 = X - min\{X,c\}.$$

Infatti se $X \leq c$, allora $min\{X,c\} = X$ e $Y = X - X = 0$. Se $X > c$, $min\{X,c\} = c$, quindi $Y = X - c$.
Per determinare la PMF di Y consideriamo i seguenti casi

1. $Y = 0$: la probabilitá che il buffer sia stato svuotato é uguale alla probabilitá che vi fossero $X \leq c$ pacchetti bufferizzati in Δt_1:
   $$P_Y(Y=0) = P_X(X \leq c) = \sum_{k=0}^{c} P_X(X=k) = \sum_{k=0}^{c} \frac{\mu^k}{k!}e^{-\mu}.$$

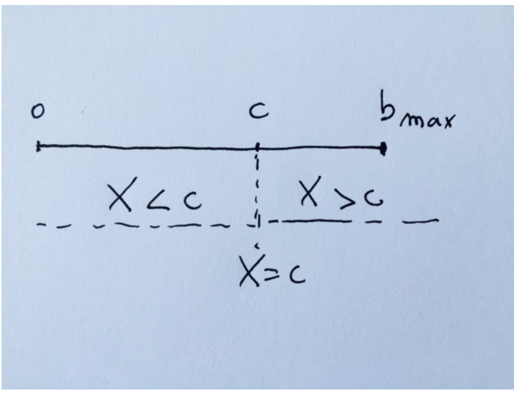

Figure 1: Valori che puó assumere la variabile aleatoria X nel problema del buffer

2. $0 < Y = k < b_{max} - c$:
La probabilitá che al termine di Δt_2 vi siano esattamente $Y = k$ pacchetti con $k = 1,2,\cdots,b_{max} - c - 1$ é uguale alla probabilitá che in Δt_1 siano stati bufferizzati $X = k + c$ pacchetti, con $k + c < b_{max}$

$$P_Y(Y = k) = P_X(X = k + c) = \frac{\mu^{k+c}}{(k + c)!}e^{-\mu}$$

3. $Y = b_{max} - c$:
E' il caso in cui in Δt_1 il buffer ștato riempito con $X = b_{max}$ pacchetti di cui c sono stati inviati in Δt_2:

$$P_Y(Y = b_{max} - c) = P_X(X = b_{max}) = 1 - \sum_{k=0}^{b_{max}-1} \frac{\mu^k}{k!}e^{-\mu}$$

(b) Qual é la probabilitá che alcuni pacchetti vengano scartati durante il primo slot?

Soluzione: La probabilitá che alcuni pacchetti vengano scartati durante il primo slot é uguale alla probabilitá che piú di b_{max} pacchetti siano stati generati alla sorgente in Δt_1 ovvero

$$P_X(X > b_{max}) = \sum_{k=b_{max}+1}^{\infty} \frac{\mu^k}{k!}e^{-\mu} = 1 - \sum_{k=0}^{b_{max}} \frac{\mu^k}{k!}e^{-\mu}$$

Esercizio 2.

Pippo si reca in banca per effettuare un versamento, e ha uguale probabilitá di trovare 0 o 1 clienti in coda prima di lui. Qualora vi sia effettivamente un cliente, il tempo di servizio é distribuito esponenzialmente con parametro λ.

(a) Qual é la funzione di distribuzione cumulativa (CDF) del tempo di attesa di Pippo?

Soluzione: (Distribuzione esponenziale) Sia X la VA che denota il tempo di attesa e Y il numero di clienti trovati. Chiaramente, per $x < 0$ $F_X(x) = 0$, mentre per $x \geq 0$,

$$F_X(x) = P(X \leq x) = \frac{1}{2}P(X \leq x \mid Y = 0) + \frac{1}{2}P(X \leq x \mid Y = 1). \tag{8}$$

Se non ci sono clienti

$$P(X \leq x \mid Y = 0) = 1,$$

e per specifica di problema, nel caso in cui vi sia una persona

$$P(X \leq x \mid Y = 1) = 1 - e^{-\lambda x},$$

Sostituendo nella (8), la CDF si scrive come

$$F_X(x) = \begin{cases} \frac{1}{2}\left(2 - e^{-\lambda x}\right), & \text{se } x \geq 0, \\ 0 & \text{altrimenti,} \end{cases} \tag{9}$$

ed ha un punto di discontinuitá in $x = 0$.

Esercizio 3.

Siano J e K due variabili aleatorie indipendenti le cui PMF sono definite come segue:

$$P_J(j) = \begin{cases} 0.2 & j = 1, \\ 0.6 & j = 2, \\ 0.2 & j = 3, \\ 0 & \text{altrimenti.} \end{cases} \tag{10}$$

$$P_K(k) = \begin{cases} 0.5 & k = -1, \\ 0.5 & k = 1, \\ 0 & \text{altrimenti.} \end{cases} \tag{11}$$

Si definisca una terza variabile aleatoria M come somma delle due

(a) Trovare il momento ordine tre di M.

Soluzione: (Funzioni generatrici dei momenti)
Per procedere, calcolo le funzioni generatrici dei momenti (FGM) delle due VA:

$$\varphi_J(u) = 0.2e^u + 0.6e^{2u} + 0.2e^{3u}$$

$$\varphi_K(u) = 0.5e^{-u} + 0.5e^u$$

Sia

$$M = J + K$$

la VA di interesse.
La FGM di M é allora:

$$\varphi_M(u) = \varphi_J(u)\varphi_K(u) = (0.2e^u + 0.6e^{2u} + 0.2e^{3u})(0.5e^{-u} + 0.5e^u) = 0.1 + 0.3e^u + 0.2e^{2u} + 0.3e^{3u} + 0.1e^{4u}$$

Per trovare il momento di ordine tre, differenziamo tre volte la FGM di M e ne calcoliamo il valore in $u = 0$:

$$E\left[M^3\right] = \left.\frac{d^3\varphi_M(u)}{du^3}\right|_{u=0} = 16.4$$

(b) Determinare la PMF di M, $P_M(m)$.

Soluzione:
Il valore della PMF $P_M(m)$ in ciascun m é esattamente il coefficiente dell'm-esimo termine e^{mu} in $\varphi_M(u)$:

$$P_M(m) = \begin{cases} 0.1 & m = 0{,}4, \\ 0.3 & m = 1{,}3, \\ 0.2 & m = 2, \\ 0 & \text{altrimenti.} \end{cases} \tag{12}$$

Compito scritto dell'esame di Statistica e analisi dei dati (Crema) 20 febbraio 2017	Prof. Giuseppe Boccignone	Corso di Laurea
Cognome:	Nome:	Matricola:

Istruzioni

- Il tempo riservato alla prova scritta e di 2 ore e 30 minuti. Durante la prova e possibile consultare libri e appunti.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede piu passaggi di calcolo, non sara preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o piu frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sara eventualmente rappresentato in quest'ultima forma).
- SOLO PER STUDENTI DEL CORSO ONLINE (VECCHIO ORDINAMENTO, Prof. Gianini): Indicare quale Modulo viene svolto al fine di completare una prova d'esame parziale precedentemente sostenuta con esito positivo
  - Modulo 1 ☐
  - Moduli 2 e 3 ☐

Modulo 1

Esercizio 1.
Un dado non truccato presenta sei facce: 4 verdi e 2 rosse. Si considerino le seguenti sequenze:

`R V R R R`

`R V R R R V`

`V R R R R R`

(a) Vincete 50 euro se, scelta una sequenza ed effettuato un numero di lanci pari alla sua lunghezza, l'esito dei lanci replica la sequenza scelta. Indicate la vostra scelta, motivandola, senza effettuare alcun calcolo probabilistico.

Soluzione:
Assumendo l'indipendenza di ciascun lancio dall'altro, le tre sequenze sono identiche e dunque equiprobabili, nei primi cinque lanci. La seconda e terza sequenza peró comportano un ulteriore lancio e, dunque, la moltiplicazione per un numero minore di uno (per definizione stessa di probabilitá). Dunque la prima sequenza é quella che garantisce maggiormente la vincita.
Una nota curiosa: in un celebre esperimento psicologico (A. Tversky e D. Kahneman, "Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment," *Psychological Review*, vol. 90, 1983, pp. 293–315), lo stesso quesito é stato posto a 260 studenti che non avevano nozioni di calcolo delle probabilitá. Il 63% di loro ha scelto la seconda sequenza, mostrando come l'uso della probabilitá che intuitivamente viene praticato nella vita reale non conduca a risultati accurati.

(b) Indicate ora la sequenza che ha maggior probabilitá di garantirvi la vincita derivando il risultato in modo formale (facendo i conti per esteso)

Soluzione:
I dati del problema ci dicono che:

$$P(R) = \frac{2}{6} = \frac{1}{3}$$

$$P(V) = \frac{2}{3}$$

La soluzione piú veloce é la seguente.
Per indipendenza dei lanci, la probabilitá dei primi cinque é identica: indichiamola con $P(H)$. Pertanto, per la prima sequenza vale

$$P(S_1) = P(H),$$

per la seconda

$$P(S_2) = P(H) \times P(V) = \frac{2}{3}P(H),$$

e per la terza

$$P(S_3) = P(H) \times P(R) = \frac{1}{3}P(H).$$

Da cui:

$$P(S_1) > P(S_2) > P(S_3).$$

Svolgendo, invece, nel dettaglio i conti, per la prima sequenza vale

$$P(S_1) = \frac{1}{3} \times \frac{2}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} = \frac{2}{243} = 0.0082,$$

per la seconda

$$P(S_2) = \frac{1}{3} \times \frac{2}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{2}{3} = \frac{4}{729} = 0.0055,$$

e per la terza

$$P(S_3) = \times\frac{2}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} = \frac{2}{729} = 0.0027.$$

Quindi $P(S_1)$ é la probabilitá massima.

Esercizio 2. Un esame consiste in un test a risposte multiple, con cinque possibili risposte per ogni domanda. Mentre sostenete l'esame, vi rendete conto di avere una probabilitá del 75% di formulare la risposta corretta per ogni domanda a cui sapete rispondere. Laddove non sapete rispondere, volete indicare una soluzione con probabilitá di essere corretta pari a $\frac{1}{5}$.

(a) Qual é la probabilitá di rispondere correttamente ad una domanda?

Soluzione:
Definiamo gli eventi

$$A = \text{“risposta corretta”}$$

$$B = \text{“so rispondere”}$$

Si ha che:

$$P(A) = P(A \cap B) + P(A \cap \sim B)$$

dove $\sim B$ é l'evento complementare di B.
Applicando la regola del prodotto e i dati forniti dal problema:

$$P(A \cap B) = P(A \mid B)P(B) = 1 \times 0.75 = 0.75,$$

$$P(A \cap \sim B) = P(A \mid \sim B)P(\sim B) = \frac{1}{5} \times 0.25 = 0.05.$$

Dunque

$$P(A) = P(A \cap B) + P(A \cap \sim B) = 0.75 + 0.05 = 0.8$$

Esercizio 3. Nel paese di Canicattí, il 30% degli elettori si dichiara di sinistra, il 50% di centro, e il 20% di destra. Alle ultime comunali hanno votato il 65% degli elettori di sinistra, l' 82% degli elettori di centro, e il 50% degli elettori di destra.

(a) Scegliendo a caso un elettore che ha votato, qual é la probabilitá che sia di centro?

Soluzione: (Teorema di Bayes)

Definiamo gli eventi:

$$V = \text{“La persona ha votato alle ultime elezioni”}$$
$$S = \text{“La persona e' di sinistra”}$$
$$C = \text{“La persona e' di centro”}$$
$$D = \text{“La persona e' di destra”}$$

La soluzione del problema consiste nel calcolare la probabilitá a posteriori $P(C \mid V)$, ovvero usando Bayes:

$$P(C \mid V) = \frac{P(V \mid C)P(C)}{P(V)}. \tag{1}$$

Sappiamo che:

$$P(S) = 0.3, P(C) = 0.5, P(D) = 0.2$$

$$P(V \mid S) = 0.65, P(V \mid C) = 0.82, P(V \mid D) = 0.5$$

Inoltre:

$$P(V) = P(V \mid S)P(S) + P(V \mid C)P(C) + P(V \mid D)P(D) = 0.705$$

Pertanto sostituendo in (1):

$$P(C \mid V) = \frac{0.82 \times 0.5}{0.705} = 0.5816 \tag{2}$$

Esercizio 1.
Una variabile aleatoria X ha densitá $f_X(x) = kx$ per $0 \leq x \leq 3$

(a) Trovare x_1 e x_2 tali che $P(X \leq x_1) = 0.1$ e $P(X \leq x_2) = 0.95$

Soluzione:
Per procedere, trovo la costante k imponendo la normalizzazione

$$\int_0^3 f_X(x)dx = \int_0^3 kxdx = 1$$

da cui ottengo $k = \frac{2}{9}$
Per qualunque $0 \leq x \leq 3$,

$$P(X \leq x) = \frac{2}{9}\int_0^x tdt = \frac{x^2}{9} = F_X(x)$$

.
Pertanto, $\frac{x_1^2}{9} = 0.1$ e $\frac{x_2^2}{9} = 0.95$. Da cui

$$x_1 = \sqrt{(}0.9) = 0.9487$$

$$x_2 = \sqrt{(}8.55) = 2.9240$$

(b) Determinare $P(|X - 1.8| < 0.6)$.

Soluzione: Usando la cumulativa $F_X(x)$ calcolata precedentemente:

$$P(|X - 1.8| < 0.6) = P(1.2 < X < 2.4) = \frac{1}{9}(2.4^2 - 1.2^2) = 0.48$$

Esercizio 2.
Si supponga che lo 0.3% di componenti elettronici prodotti da una azienda sia difettoso, dove il numero di componenti difettosi é una variabile aleatoria.

(a) Se i componenti vengono sistemati in scatole da 100, qual é la probabilitá che una scatola contenga un numero x di elementi difettosi?

Soluzione: (Distribuzione binomiale) Il numero di componenti difettosi segue una distribuzione binomiale. Ponendo il parametro $p = 0.003$ (probabilitá di essere difettoso) e $n = 100$

$$P(X = x) = Bin(100, 0.003)$$

(b) Qual é la probabilitá di non trovare elementi difettosi nella scatola?.

Soluzione: Nella fattispecie per i valori di N,p considerati conviene utilizzare un'approssimazione con distribuzione di Poisson, $Poiss(\mu)$, alla Binomiale. La distribuzione di Poisson avrá parametro

$$\mu = np = 100 \times 0.003 = 0.3,$$

ovvero

$$P(X = x) \approx \frac{e^{-0.3}(0.3)^x}{x!},$$

con $x = 0,1,2,$
Dunque,

$$P(X = 0) \approx 0.7408$$

(c) Si supponga di acquistare 8 scatole. Qual é la distribuzione del numero di scatole con nessun difetto?

Soluzione:
La distribuzione del numero di scatole, diciamo k, con nessun difetto segue di nuovo una distribuzione binomiale

$$P(K = k) = Bin(N, p_S)$$

dove nella fattispecie $N = 8$ e dove p_S indica la probabilitá di una scatola con nessun componente difettoso.
Nel quesito precedente abbiamo stabilito che la probabilitá di una scatola senza componenti difettosi é $P(X = 0) \approx 0.7408$, dunque $p_S \approx 0.7408$. Pertanto:

$$P(K = k) = Bin(8, 0.7408)$$

(d) Qual é il numero atteso di scatole che non contengono elementi difettosi?

Soluzione:

$$E[K] = np = 8 \times 0.7408$$

Esercizio 3. L'altezza di un uomo segue una distribuzione normale di media 178 cm con una deviazione dalla media di 8 cm. L'altezza di una donna ha invece una media di 165 cm con una deviazione pari a 7 cm. .

(a) Qual é, nella stessa popolazione, la proporzione di uomini la cui altezza é superiore a 185 cm?

Soluzione: Sia U la variabile aleatoria che denota l'altezza degli uomini.
Vogliamo determinare $P(U > 185)$, dove la distribuziione é Normale con $\mu = 178, \sigma = 8$. Standardizzando le variabili e usando le tabelle della distribuzione normale

$$P(U > 185) = P(\frac{U - 178}{8} > \frac{185 - 178}{8}) = P(Z < 0.875) \approx 0.19$$

(b) Qual é la proporzione di donne che sono piú alte della metá degli uomini?

Soluzione: Sia D la variabile aleatoria che denota l'altezza delle donne.
Sappiamo che media e mediana nella distribuzione normale coincidono dunque

$$P(U > u) = 0.5$$

quando $u = \mu = 178$
Pertanto, si tratta di determinare $P(D > 178)$:

$$P(D > 178) = P(Z > \frac{178 - 165}{7}) = P(Z > 1.857) \approx 0.032$$

.
Solo il 3.2% delle donne é piú alto di metá degli uomini.

Compito scritto dell'esame di Statistica e analisi dei dati (Crema) 12 giugno 2017	Prof. Giuseppe Boccignone	Corso di Laurea
Cognome:	Nome:	Matricola:

Istruzioni

- Il tempo riservato alla prova scritta e di 2 ore e 30 minuti. Durante la prova e possibile consultare libri e appunti.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede piu passaggi di calcolo, non sara preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o piu frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sara eventualmente rappresentato in quest'ultima forma).
- SOLO PER STUDENTI DEL CORSO ONLINE (VECCHIO ORDINAMENTO, Prof. Gianini): Indicare quale Modulo viene svolto al fine di completare una prova d'esame parziale precedentemente sostenuta con esito positivo
  - Modulo 1 ☐
  - Moduli 2 e 3 ☐

Modulo 1

Esercizio 1.
Da un mazzo di 52 carte (ossia tredici per ognuno dei quattro semi) se ne sceglie una a caso.

(a) Quanto vale la probabilitá di estrarre una figura e un fiori?

Soluzione:
L' evento "figura" non influisce sulla probabilitá dell'evento "fiori", per cui essi sono statisticamente indipendenti. Ne segue:

$$P(\{\mathsf{figura}\} \cap \{\mathsf{fiori}\}) = P(\{\mathsf{figura}\})P(\{\mathsf{fiori}\}) = \frac{12}{52} \cdot \frac{13}{52} = \frac{3}{52}$$

(b) Quanto vale la probabilitá di estrarre una figura o una carta di fiori?

Soluzione:
Usando il risultato precedente:

$$P(\{\mathsf{figura}\} \cup \{\mathsf{fiori}\}) = P(\{\mathsf{figura}\}) + P(\{\mathsf{fiori}\}) - P(\{\mathsf{figura}\} \cap \{\mathsf{fiori}\}) =$$

$$\frac{12}{52} + \frac{13}{52} - \frac{3}{52} = \frac{11}{26}$$

Esercizio 2. Nel lancio ripetuto di due dadi non truccati, sappiamo che la somma dei risultati é un numero pari..

(a) Quanto vale la probabilitá di aver totalizzato 8?

Soluzione:
La probabilitá di aver totalizzato 8 é:

$$P(\{8\}) = P(\{6+2\} \cup \{5+3\} \cup \{4+4\} \cup \{3+5\} \cup \{2+6\}) = \frac{5}{36}$$

Sapendo che é uscito un numero pari, poiché $\{8\} \subset \{\mathsf{pari}\}$ si ha:

$$P(\{8\} \mid \{\mathsf{pari}\}) = \frac{P(\{8\} \cap \{\mathsf{pari}\})}{P(\{\mathsf{pari}\})} = \frac{P(\{8\})}{0.5} = \frac{5}{18}$$

Esercizio 3. Si utilizza un prodotto fornito in percentuali uguali da due ditte A e B. E' stato calcolato che, scelto a caso un esemplare difettoso, la probabilitá che esso sia stato fornito dalla ditta A vale 0.25.

(a) Se la produzione del prodotto da parte della ditta A ha un difetto di qualitá del 5%, qual é il difetto di qualitá nella produzione della ditta B?

Soluzione: (Teorema di Bayes)
I dati del problema ci dicono che

$$P(A \mid \mathsf{difettoso}) = 0.25$$

$$P(\mathsf{difettoso} \mid A) = 0.05$$

Vogliamo calcolare $P(\mathsf{difettoso} \mid B)$.
Usando Bayes:

$$P(A \mid \mathsf{difettoso}) = \frac{P(\mathsf{difettoso} \mid A)P(A)}{P(\mathsf{difettoso} \mid A)P(A) + P(\mathsf{difettoso} \mid B)P(B)} = 0.25. \tag{1}$$

Risolvendo l'equazione (1) rispetto alla probabilitá richiesta:

$$P(\mathsf{difettoso} \mid B)P(B) = \frac{0.05}{0.25} - 0.05 = 0.15$$

Il difetto di qualitá nella produzione della ditta B ṕari al 15%

Moduli 2 e 3

Esercizio 1. Una variabile aleatoria X ha densitá

$$f_X(x) = \begin{cases} x - \frac{1}{4}x^3, & per \quad 0 \le x \le 2 \\ 0 & \quad altrove \end{cases}$$

(a) Determinare valore atteso, varianza e mediana di X.

Soluzione:

$$E[X] = \int_0^2 x\left(x - \frac{1}{4}x^3\right) dx = \left[\frac{x^3}{3} - \frac{x^5}{20}\right]_0^2 = \frac{16}{15}$$

La varianza é $\sigma_X^2 = E[X^2] - (E[X])^2 = E[X^2] - \left(\frac{16}{15}\right)^2$, dove

$$E[X^2] = \int_0^2 x^2\left(x - \frac{1}{4}x^3\right) dx = \left[\frac{x^4}{4} - \frac{x^6}{24}\right]_0^2 = \frac{4}{3}$$

Pertanto,

$$\sigma_X^2 = \frac{4}{3} - \left(\frac{16}{15}\right)^2 = 16\left(\frac{1}{12} - \frac{16}{225}\right) \approx 0.195$$

Per calcolare la mediana x_{med}, si impone $F_X(x_{med}) = \frac{1}{2}$, dunque:

$$F_X(x_{med}) = \int_0^{x_{med}} \left(x - \frac{1}{4}x^3\right) dx = \left[\frac{x^2}{2} - \frac{x^4}{16}\right]_0^{x_{med}} = \frac{1}{2}\left(x_{med}^2 - \frac{x_{med}^4}{16}\right) = \frac{1}{2}$$

Si deve pertanto risolvere l'equazione

$$x_{med}^4 - 8x_{med}^2 + 8 = 0$$

cercando la radice che appartiene all'intervallo $0 \le x \le 2$.
Ponendo $t = x_{med}^2$

$$t = \begin{cases} 4 + 2\sqrt{2} \\ 4 - 2\sqrt{2} \end{cases}$$

da cui

$$\begin{cases} x_{med}^{(1,2)} = \pm\sqrt{4 + 2\sqrt{2}} \approx \pm 2.613 \\ x_{med}^{(3,4)} = \pm\sqrt{4 - 2\sqrt{2}} \approx \pm 1.0924 \end{cases}$$

Scartando le soluzioni $x_{med}^{(1,2)}$ giacché fuori dell'intervallo $0 \le x \le 2$, l'unica soluzione ammissibile é

$$x_{med}^{(3)} = \sqrt{4 - 2\sqrt{2}} \approx 1.0924$$

Esercizio 2.
Il giocatore A lancia un dado non truccato per 4 volte, e vince se esce almeno una volta il 6. Il giocatore B lo lancia 8 volte, e vince se il 6 esce almeno due volte.

(a) Chi ha maggiore probabilitá di vincere e perché?

Soluzione: (Distribuzione binomiale) In ogni lancio la probabilitá che esca il 6 vale $p = 1/6$ (equiprobabilitá di 6 eventi). La probabilitá di avere $X = 0$ successi in $n = 4$ prove indipendenti vale, si determina in base alla distribuzione binomiale:

$$Bin\left(X = 0 \mid n = 4, p = \frac{1}{6}\right) = \binom{4}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^4 = \left(\frac{5}{6}\right)^4 \approx 0.48226$$

Dunque, la probabilitá di vittoria per A é

$$P(A) = 1 - 0.48226 \approx 0.51774$$

Per il giocatore B, la probabilitá di avere non piú di $X = 1$ successi in $n = 8$ prove (perdendo cosí la scommessa) é

$$P_X(0 \leq X \leq 1) = Bin\left(X = 0 \mid n = 8, p = \frac{1}{6}\right) + Bin\left(X = 1 \mid n = 8, p = \frac{1}{6}\right)$$

ovvero

$$P_X(0 \leq X \leq 1) = \binom{8}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^8 + \binom{8}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^7 \approx 0.6046$$

Pertanto, la probabilitá di vittoria per B é

$$P(B) = 1 - P_X(0 \leq X \leq 1) \approx 0.3936$$

Esercizio 3. Si dispone di un campione di 100 misure di una variabile temporale X di una popolazione di cui non conosciamo la distribuzione, ma la cui deviazione standard é nota e vale $\sigma_X = 120$ secondi. .

(a) Qual é la probabilitá che la media campionaria differisca per piú di 3 secondi dal valore atteso teorico (incognito) dei tempi misurati?

Soluzione:
Siano:

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

la media campionaria;

$$\mu_X = E[X]$$

il valore atteso teorico.
Il problema ci chiede di determinare la probabilitá

$$P_X\left(\left|\frac{S_n}{n} - \mu\right| > 3\right)$$

Non conosciamo la legge di probabilitá secondo cui si distribuisce la popolazione, ma il Teorema Centrale Limite (TCL) ci assicura che per n sufficientemente grande la distribuzione degli scarti fra media campionaria $\frac{S_n}{n}$ e valore atteso teorico μ_X segue una legge Gaussiana:

$$P\left(a \leq \frac{\sqrt{n}}{\sigma_X} \cdot \left(\frac{S_n}{n} - \mu\right) \leq b\right) \approx \Phi(b) - \Phi(a).$$

Nel nostro caso,

$$\frac{\sigma_X}{\sqrt{n}} = \frac{120}{\sqrt{100}} = 12$$

é la deviazione standard del campione.
Usando il TCL con $a = -3, b = 3$ e passando alla variabile standardizzata $Z_n \sim N(0,1)$

$$Z_n = \frac{\frac{S_n}{n} - \mu_X}{\frac{\sigma_X}{\sqrt{n}}},$$

il problema iniziale si riduce a calcolare la seguente:

$$P_X\left(\left|\frac{S_n}{n} - \mu_X\right| > 3\right) = P_Z\left(\frac{\sigma_X}{\sqrt{n}}|Z_n| > 3\right) = P_Z(|Z_n| > 0.25).$$

Nel caso Gaussiano, $P_Z(|Z_n| > z) = 2(1 - \Phi(z))$, dunque :

$$P_Z(|Z_n| > 0.25) = 2(1 - \Phi(0.25)).$$

Usando le tabelle della CDF standard Φ leggiamo che

$$\Phi(0.25) = 0.5987.$$

Pertanto,

$$P_X\left(\left|\frac{S_n}{n} - \mu\right| > 3\right) = 2(1 - 0.5987) = 0.8026.$$

Compito scritto dell'esame di Statistica e analisi dei dati (Crema) 27 luglio 2017	Prof. Giuseppe Boccignone	Corso di Laurea
Cognome:	Nome:	Matricola:

Istruzioni

- Il tempo riservato alla prova scritta e di 2 ore e 30 minuti. Durante la prova e possibile consultare libri e appunti.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede piu passaggi di calcolo, non sara preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o piu frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sara eventualmente rappresentato in quest'ultima forma).
- SOLO PER STUDENTI DEL CORSO ONLINE (VECCHIO ORDINAMENTO, Prof. Gianini): Indicare quale Modulo viene svolto al fine di completare una prova d'esame parziale precedentemente sostenuta con esito positivo
  - Modulo 1 ☐
  - Moduli 2 e 3 ☐

Modulo 1

Esercizio 1.

Si supponga di avere un mazzo di 40 carte di cui 30 blu e 10 rosse. Si estrae una carta: se esce carta blu si lancia una moneta altrimenti un dado regolare.

(a) Con quale probabilitá esce testa?

Soluzione:

Definiamo i seguenti eventi

- $B =$ "esce carta blu"
- $T =$ "esce testa nel lancio della moneta"
- $D_i =$ "esce faccia i nel lancio del dado" , $i = 1, 2, \cdots, 6$
- $E =$ "esce testa nel gioco"

Osserviamo che

$$E = B \cap T$$

Allora:

$$P(E) = P(B \cap T) = P(T \mid B)P(B) = \frac{1}{2} \cdot \frac{30}{40} = \frac{3}{8}$$

(b) Con quale probabilitá esce il numero 6?

Soluzione:

Definiamo l'evento $F =$ "esce numero 6 nel gioco"

In modo analogo al caso precedente, $F = D_6 \cap \sim B$, dunque:

$$P(F) = P(D_6 \mid \sim B)P(\sim B) = \frac{1}{6} \cdot \frac{10}{40} = \frac{1}{24}$$

Esercizio 2. In un canale di trasmissione binario asimmetrico (unidirezionale, $T \to R$) tutti gli errori sono sempre $1 \to 0$ o $0 \to 1$ ed entrambi possono accadere; esempi sono costituiti da fibre ottiche, dischi ottici, circuiti e memorie VLSI e ROM. Si supponga che un segnale binario T, emesso come 1 (ovvero $T = 1$) con probabilitá 0.75, sia inviato su un canale di trasmissione asimmetrico nel quale la probabilitá di errore nella trasmissione di 1 vale 0.08. In generale, il segnale trasmesso é ricevuto nella forma $R = 1$ con probabilitá 0.70. Calcolare:

(a) La probabilitá che il segnale 0 sia ricevuto come 1

Soluzione: Si tratta di calcolare $P(R = 1 \mid T = 0)$.
Sappiamo che $P(T = 1) = 0.75$, $P(R = 1) = 0.7$, $P(R = 0 \mid T = 1) = 0.08$ Da cui

$$P(T = 0) = 1 - P(T = 1) = 0.25.$$

$$P(R = 1 \mid T = 1) = 1 - P(R = 0 \mid T = 1) = 1 - 0.08 = 0.92$$

Per le probabilitá totali

$$0.7 = P(R = 1) = P(R = 1 \mid T = 1)P(T = 1) + P(R = 1 \mid T = 0)P(T = 0) = 0.92 \times 0.75 + P(R = 1 \mid T = 0) \times 0.25$$

Risolvendo:

$$P(R = 1 \mid T = 0) = 0.04$$

(b) La probabilitá totale di errore nella ricezione del segnale.

Soluzione:

$$P(\{errore\}) = P(\{R = 1,T = 0\} \cup \{R = 0,T = 1\}) = P(\{R = 1,T = 0\}) + P(\{R = 0,T = 1\}) =$$

$$= P(R = 1 \mid T = 0)P(T = 0) + P(R = 0 \mid T = 1)P(T = 1) = 0.07$$

Esercizio 3. Un gruppo di bagnanti é costituito per il 65% da persone di carnagione scura (S) e le altre sono di carnagione chiara (C). L'uso inappropriato di creme solari fa sí che si abbia una percentuale di persone ustionate dal sole (U) del 10% se di carnagione scura e del 60% se di carnagione chiara

(a) Sapendo che un bagnante a caso si é ustionato prendendo il sole, con che probabilitá ha carnagione chiara?

Soluzione: (Teorema di Bayes)
I dati del problema ci dicono che

$$P(S) = 0.65$$

$$P(C) = 1 - P(S) = 0.35$$

$$P(U \mid S) = 0.1$$

$$P(U \mid C) = 0.6$$

Vogliamo calcolare $P(C \mid U)$.
Usando Bayes:

$$P(C \mid U) = \frac{P(U \mid C)P(C)}{P(U \mid C)P(C) + P(U \mid S)P(S)} = 0.76. \tag{1}$$

Moduli 2 e 3

Esercizio 1. Una variabile aleatoria X, con valori nell'intervallo $1 \leq x \leq 2$, ha densitá

$$f_X(x) = \frac{k}{x^2}$$

.

(a) Calcolare valore atteso $E[X]$ e deviazione standard σ_X

Soluzione: Si calcola la costante k utilizzando la proprietá di normalizzazione:

$$1 = \int_1^2 f_X(x)dx = \int_1^2 \frac{k}{x^2}dx = k\left[-\frac{1}{x}\right]_1^2 = k(-\frac{1}{2}+1) = \frac{k}{2},$$

da cui $k = 2$

$$E[X] = \int_1^2 x\frac{2}{x^2}dx = 2\log 2 \approx 1.386$$

$$\sigma_X^2 = E\left[X^2\right] - E[X]^2 = E\left[X^2\right] - 4\log^2 2 = 2\int_1^2 dx - 4\log^2 2 = 2(1-2\log^2 2) \approx 0.078$$

$$\sigma_X = \sqrt{0.078} \approx 0.28.$$

Esercizio 2.

Si effettuano 600 lanci di un dado non truccato.

(a) Calcolare un valore approssimato della probabilitá che il "5" esca un numero di volte compreso tra 94 e 106

Soluzione: (Distribuzione binomiale in approssimazione normale) In ogni lancio la probabilitá che esca il 5 vale $p = 1/6$ (equiprobabilitá di 6 eventi). Poiché $n = 600$ é sufficientemente grande, approssimiamo con una Gaussiana $\mathcal{N}(x \mid \mu = np, \sigma^2 = npq)$ dove

$$\mu = np = 600 \cdot \frac{1}{6} = 100$$

$$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{100 \cdot \frac{1}{6} \cdot \frac{5}{6}} \approx 9.1287$$

Procediamo alla standardizzazione

$$z_1 = \frac{94-100}{9.1287} = -0.657$$

$$z_2 = \frac{106-100}{9.1287} = 0.657$$

Quindi usando le tabelle della normale ridotta

$$P(94 \leq X \leq 100) = F_Z(0.657) - F_Z(-0.657) = 2F_Z(0.657) - 1 \approx 0.49$$

Esercizio 3. Il responsabile dell'illuminazione di un palazzo ha a disposizione 3 lampadine di riserva. Ciascuna delle lampadine utilizzate dura in media 200 ore e il responsabile deve aspettare 24 ore perché gli vengano consegnate nuove lampadine di riserva.

(a) Qual é la probabilitá che il responsabile non sia in grado di sostituire le lampadine fulminate?

Soluzione: Il problema sorge se si fulminano piú di tre lampadine entro le 24 ore: in tal caso il responsabile non sará in grado di sostituirle. Dobbiamo dunque determinare la probabilitá congiunta

$$P(\{X_1 \leq 0.24\}, \{X_2 \leq 0.24\}, \{X_3 \leq 0.24\}).$$

dove X_1,X_2,X_3 sono i tempi di attesa della rottura della prima, seconda e terza lampadina.

Si assuma una distribuzione esponenziale negativa dei tempi di attesa e si ragioni, per semplicitá di calcolo in termini di centinaia di ore.

Il tempo medio é pari a $E[X] = 2$ (in centinaia di ore). Il parametro della distribuzione esponenziale é dunque

$$\lambda = \frac{1}{2} = 0.5$$

La probabilitá del tempo di rottura di una singola lampadina entro le 24 ore é

$$P(X \leq 0.24) = 1 - e^{-\lambda \cdot x} = 1 - e^{-0.5 \cdot 0.24} = 1 - 0.887 = 0.113$$

Poiché la rottura di ciascuna lampadina é indipendente dalle altre:

$$P(\{X_1 \leq 0.24\}, \{X_2 \leq 0.24\}, \{X_3 \leq 0.24\}) = P(\{X_1 \leq 0.24\}) \cdot P(\{X_2 \leq 0.24\}) \cdot P(\{X_3 \leq 0.24\}) = (0.113)^3 = 0.0014$$

Compito scritto dell'esame di Statistica e analisi dei dati (Crema) 01 settembre 2017	Prof. Giuseppe Boccignone	Corso di Laurea
Cognome:	Nome:	Matricola:

Istruzioni

- Il tempo riservato alla prova scritta e di 2 ore e 30 minuti. Durante la prova e possibile consultare libri e appunti.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede piu passaggi di calcolo, non sara preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o piu frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sara eventualmente rappresentato in quest'ultima forma).
- SOLO PER STUDENTI DEL CORSO ONLINE (VECCHIO ORDINAMENTO, Prof. Gianini): Indicare quale Modulo viene svolto al fine di completare una prova d'esame parziale precedentemente sostenuta con esito positivo
  - Modulo 1 ☐
  - Moduli 2 e 3 ☐

Modulo 1

Esercizio 1. Una moneta da 2 euro (il suo diametro é 25,75 mm) viene lanciata su un pavimento ricoperto con 100 mattonelle quadrate di lato $10cm$.

(a) Qual é la probabilitá che la moneta vada a finire internamente ad una mattonella? (cioé non tagli i lati dei quadrati)

Soluzione: Immaginiamo di lanciare la moneta su una mattonella quadrata di lato $10cm$. Fissiamo l'attenzione sul centro C della moneta e osserviamo che la probabilitá che esso cada in una regione R di area $a\ cm^2$, contenuta nel quadrato, é proporzionale all'area, ossia

$$P(C \in R) = \frac{a}{10^2}$$

La moneta cade all'interno della mattonella quando R é un quadrato concentrico alla mattonella e con lato $(10 - 2{,}575)cm$. Pertanto:

$$P(C \in R) = \frac{(10 - 2{,}575)^2}{10^2} \approx 0.551$$

Il risultato precedente continua a valere nel caso di un pavimento formato da n mattonelle quadrate, infatti

$$P(C \in R) = \frac{(10 - 2{,}575)^2 \times n}{10^2 \times n} \approx 0.551$$

Esercizio 2. Una scatola contiene: due palline bianche numerate “0”; otto palline bianche numerate “1”; quattro palline rosse numerate “0”; sei palline rosse numerate “1”. Si estrae una pallina casualmente dalla scatola.

(a) Gli eventi “esce una pallina bianca” e “esce una pallina numerata con 0” sono indipendenti?

Soluzione: Il numero di palline é uguale a 20. Definiamo gli eventi seguenti:

$$A = \text{“esce una pallina bianca”};$$

$$B = \text{“esce una pallina numerata con 0”}.$$

Siccome ci sono 10 palline bianche e 6 palline numerate con 0,

$$P(A) = \frac{1}{2}, P(B) = \frac{6}{20}.$$

Si definisca l'evento $\{A \cap B\} =$ “esce una pallina bianca numerata con 0”. Allora:

$$P(A \cap B) = \frac{2}{20} \neq \frac{3}{20} = P(A)P(B)$$

Dunque A e B non sono indipendenti.

(b) Spiegare brevemente il perché del risultato ottenuto in (a).

Soluzione: A e B non sono indipendenti in quanto la proporzione di palline numerate con 0 é diversa tra le palline bianche e le palline nere.

Esercizio 3. Consideriamo un test diagnostico per evidenziare la presenza di una patologia nel sangue. Il test ha probabilitá 0.8 di individuare correttamente la presenza della patologia, ma sappiamo che dá risultato positivo (ossia, indica la presenza della patologia) anche nel 20% degli esami compiuti su soggetti sani. Supponiamo che la presenza della patologia nella popolazione sia pari al 5%

(a) Determinare la probabilitá che un soggetto sia malato dopo che il test ha dato risultato positivo

Soluzione: (Teorema di Bayes)
Indichiamo con M l'evento "soggetto malato" e con T l'evento "test positivo".
I dati del problema ci dicono che $P(M) = 0.05, P(T \mid M) = 0.8, P(T \mid\sim M) = 0.2$
Usando Bayes:

$$P(M \mid T) = \frac{P(T \mid M)P(M)}{P(T \mid M)P(M) + P(T \mid\sim M)P(\sim M)} = \frac{4}{23} \approx 0.174. \tag{1}$$

(b) Determinare la probabilitá che un soggetto sia sano dopo che il test ha dato risultato negativo

Soluzione:
Usando sempre Bayes:

$$P(\sim M \mid\sim T) = \frac{P(\sim T \mid\sim M)P(\sim M)}{P(\sim T \mid\sim M)P(\sim M) + P(\sim T \mid M)P(M)} = \frac{76}{77} \approx 0.987. \tag{2}$$

Moduli 2 e 3

Esercizio 1. In media, si impiegano 10 minuti per guidare da casa all'universitá.

(a) Supponiamo per semplicitá che il tempo necessario, T , segua una legge esponenziale. Se parto da casa alle 8 : 45, qual é la probabilitá di arrivare in universitá per le 9?

Soluzione: T segue una legge esponenziale di parametro

$$\lambda = \frac{1}{10};$$

Pertanto

$$P(T < 15) = 1 - e^{-\frac{15}{10}} \approx 77.7\%$$

(b) A che ora devo partire per avere una probabilitá del 99% di arrivare in tempo?

Soluzione: Per arrivare in tempo con una probabilitá del 99%, devo trovare qual é quel tempo t per cui $P(T \leq t) = 0.99$, ovvero

$$(1 - e^{-\lambda t}) = 0.99$$

$$-\lambda t = \log(0.01)$$

$$t = 10\log(100) = 46.0517min.$$

da cui partendo verso le 8 : 10 ho (statisticamente)una ragionevole certezza di puntualitá.

Esercizio 2.

Al servizio di soccorso stradale di una certa cittá, aperto 24 ore su 24, arrivano in media 48 chiamate al giorno, due in media all'ora

(a) Calcolare la probabilitá che nella prima ora arrivino almeno due chiamate

Soluzione: (Distribuzione di Poisson) Possiamo assumere che le chiamate si distribuiscano con legge di Poisson. La variabile aleatoria X che descrive il numero di chiamate per ora ha media

$$\mu = 2,$$

pertanto la distribuzione di X si scrive

$$p_X(x) = \frac{2^x e^{-2}}{x!}.$$

Le probabilitá che non arrivino chiamate o che ne arrivi solo una sono date rispettivamente da

$$P(X = 0) = p_X(0) = \frac{2^0 e^{-2}}{0!} = e^{-2},$$

$$P(X = 1) = p_X(1) = \frac{2^1 e^{-2}}{1!} = 2e^{-2}.$$

Usando la probabilitá complementare si ha

$$P(X > 1) = 1 - P(0 \leq X \leq 1) = 1 - 3e^{-2}.$$

(b) Calcolare la probabilitá che il tempo di attesa fino alla prima chiamata di un certo giorno sia di almeno un'ora

Soluzione: Il quesito si risolve immediatamente notando che

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

Esercizio 3. Una macchina produce barre di acciaio a sezione circolare la cui lunghezza ottimale dovrebbe essere di 5 metri ed il diametro della sezione di 4 centimetri. Le barre effettivamente prodotte, che si suppongono tra loro indipendenti, hanno una lunghezza aleatoria con distribuzione normale di media $\mu_1 = 5\ m$ e scarto standard $\sigma_1 = 4\ cm$. Il diametro della sezione é una variabile aleatoria, indipendente dalla precedente, e con distribuzione normale di media $\mu_2 = 4\ cm$ e scarto standard $\sigma_2 = 0{,}8\ cm$. Una generica barra prodotta puó essere direttamente venduta senza modifiche se la sua lunghezza é compresa tra $4{,}95m$ e $5{,}05m$ e la sua sezione tra $2{,}8\ cm$ e $5{,}2\ cm$.

(a) Si verifichi che la probabilitá p di poter mettere in vendita senza modifiche una generica barra prodotta é 0.68

Soluzione: La variabile aleatoria l che descrive la lunghezza di una barra ha densitá di probabilitá normale. Dunque, standardizzando l con

$$z_l = \frac{l - \mu_1}{\sigma_1}$$

e facendo uso delle tavole della normale standard si ha che

$$P(495cm \leq l \leq 505cm) = \Phi(\frac{5}{4}) - \Phi(-\frac{5}{4}) \approx 0.788$$

Analogamente, per il diametro d della sezione:

$$P(2{,}8cm \leq d \leq 5{,}2cm) = \Phi(\frac{3}{2}) - \Phi(-\frac{3}{2}) \approx 0.866$$

Per indipendenza di l,d:

$$P(\{495cm \leq l \leq 505cm\} \cap \{2{,}8cm \leq d \leq 5{,}2cm\}) = P(495cm \leq l \leq 505cm) \times P(2{,}8cm \leq d \leq 5{,}2cm) \approx 0.788 \times 0.866 = 0.68$$

(b) Indicata con f_n la frequenza relativa delle barre direttamente vendibili su n barre prodotte, si esprima, in funzione di p, una stima della numerositá n necessaria affinché la probabilitá che f_n disti da p piú di 0.05 sia non superiore a 0.05

Soluzione: Sia X la variabile aleatoria che descrive il numero di barre vendibili, tale che $f_n = \frac{X}{n}$.
Tale variabile aleatoria ha distribuzione binomiale (numero barre vendibili / non vendibili) con media

$$\mu_X = np$$

e varianza

$$\sigma_X^2 = np(1-p)$$

dove $p \approx 0.68$ é la probabilitá calcolata in (a).
Il quesito richiede di determinare la probabilitá

$$P(|\frac{X}{n} - p| \geq 0.05) \leq 0.05 \tag{3}$$

Possiamo applicare il Teorema di Chebychev nella forma

$$P_X(|X - \mu_X| \geq r) \leq \frac{\sigma_X^2}{r^2},$$

ovvero

$$P_X(|X - np| \geq r) \leq \frac{np(1-p)}{r^2},$$

dove per confronto con eq. (3) si vede che $\frac{r}{n} = 0.05$. Il problema impone pertanto che il termine a destra della diseguaglianza di Chebychev sia ≤ 0.05. Dunque si ha che

$$\frac{np(1-p)}{n^2(0.05)^2} \leq 0.05$$

da cui, sostituendo i valori numerici noti si ottiene che

$$n \geq 1741$$

Compito scritto dell'esame di Statistica e analisi dei dati (Crema) 01 settembre 2017	Prof. Giuseppe Boccignone	Corso di Laurea
Cognome:	Nome:	Matricola:

Istruzioni

- Il tempo riservato alla prova scritta e di 2 ore e 30 minuti. Durante la prova e possibile consultare libri e appunti.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede piu passaggi di calcolo, non sara preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o piu frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sara eventualmente rappresentato in quest'ultima forma).
- SOLO PER STUDENTI DEL CORSO ONLINE (VECCHIO ORDINAMENTO, Prof. Gianini): Indicare quale Modulo viene svolto al fine di completare una prova d'esame parziale precedentemente sostenuta con esito positivo
  - Modulo 1 ☐
  - Moduli 2 e 3 ☐

Modulo 1

Esercizio 1. Supponiamo che in una partita di calcio la probabilitá che vinca la squadra di casa sia 0.5 e la probabilitá che vinca la squadra ospite sia 0.2.

(a) Qual é la probabilitá di pareggio?

Soluzione: Indichiamo con X l'evento "pareggio", con VC l'evento "vince la squadra di casa" e con VO l'evento "vince la squadra ospite" Allora:

$$P(\{X\}) = P(\sim (\{VC\} \cup \{VO\}) = 1 - P(\{VC\} \cup \{VO\}) = 1 - 0.5 - 0.2 = 0.3$$

Esercizio 2. Una recente indagine ha rivelato che il 14% delle segretarie ha dolore al polso. Inoltre, il 6% delle segretarie intervistate ha dolore al polso e al tempo stesso assume regolarmente un farmaco antinfiammatorio.

(a) Qual é la probabilitá che una segretaria che ha dolore al polso assuma regolarmente un farmaco antinfiammatorio?

Soluzione: Definiamo gli eventi seguenti:

$$A = \text{"la segretaria ha il dolore al polso"};$$

$$B = \text{"la segretaria assume il farmaco"}.$$

$$P(A) = 0.14, P(A,B) = 0.06.$$

La probabilitá richiesta é la condizionata:

$$P(B \mid A) = \frac{P(B,A)}{P(A)} = 0.428$$

Esercizio 3. In una popolazione ci sono il 50% di maschi e il 50% di femmine. Supponiamo che il 5% degli uomini e il 10% delle donne siano daltonici. Si sceglie a caso una persona daltonica.

(a) Determinare la probabilitá che sia un maschio

Soluzione: (Teorema di Bayes)
Indichiamo con M l'evento "maschio", con F l'evento "femmina" e con D l'evento "daltonico".
I dati del problema ci dicono che $P(M) = 0.5 = P(F), P(D \mid M) = 0.05, P(D \mid F) = 0.10$
Usando Bayes:

$$P(M \mid D) = \frac{P(D \mid M)P(M)}{P(D \mid M)P(M) + P(D \mid F)P(\sim F)} = \frac{0.025}{0.075} = \frac{1}{3}. \quad (1)$$

Moduli 2 e 3

Esercizio 1.
Un gioco consiste nel lanciare una moneta e nel generare un numero casuale Y tra zero e uno. La vincita é $2Y$ se compare testa e 1 se compare croce.

(a) Calcolare la vincita attesa.

Soluzione: (Valore atteso) Sia T l'evento: "compare testa" definito su un classico spazio di probabilitá $(S, \mathcal{F}, P)$, $S = \{T,C\}$ e sia V la vincita
Usiamo la classica variabile aleatoria indicatrice

$$X_1(\omega) = I_T(\omega) = \begin{cases} 1 & \text{se } \omega \in T \\ 0 & \text{se } \omega \notin T(\omega \in \sim T) \end{cases}. \tag{2}$$

e la VA indicatrice speculare

$$X_2(\omega) = I_{\sim T}(\omega) = \begin{cases} 0 & \text{se } \omega \in T \\ 1 & \text{se } \omega \notin T(\omega \in \sim T) \end{cases}. \tag{3}$$

Allora:

$$V = 2YI_T - I_{\sim T}$$

Poiché X e I_T sono v.a. indipendenti

$$E[V] = E[2YI_T - I_{\sim T}] = E[2YI_T] - E[I_{\sim T}] = 2E[Y]E[I_T] - E[I_{\sim T}]$$

Il valore atteso della VA indicatrice X_1 si calcola facilmente:

$$E[X_1] = 0 \cdot P_X(X_1 = 0) + 1 \cdot P_X(X_1 = 1) = 0 \cdot (1 - P(T)) + 1 \cdot P(T) = P(T) = \frac{1}{2},$$

ovvero, il valor medio della funzione indicatrice é la probabilitá di successo. Analogamente, per $I_{\sim A}$.
Sappiamo che il valor atteso di una VA distribuita uniformemente, $Y \sim Unif(a,b)$ é

$$E[Y] = \frac{b+a}{2}$$

dunque se $Y \sim Unif(0,1)$, $E[Y] = \frac{1}{2}$
Quindi:

$$E[V] = 2 \times \frac{1}{2} \times \frac{1}{2} - \frac{1}{2} = 0$$

Esercizio 2. Il tempo atteso di vita di un certo componente elettronico, misurata in ore, é 120. Quando il componente si danneggia, viene immediatamente sostituito.

(a) Calcolare il tempo medio di attesa per la terza sostituzione.

Soluzione: Definiamo:

1. X_1 la durata in vita di un componente elettronico messo in uso al tempo zero;
2. X_2 la durata in vita di un componente elettronico messo in uso dopo la prima sostituzione;
3. X_3 la durata in vita di un componente elettronico messo in uso dopo la seconda sostituzione.

Il tempo di attesa per la terza sostituzione X é

$$X = X_1 + X_2 + X_3.$$

Poiché

$$E[X_1] = E[X_2] = E[X_3] = 120h,$$

allora:

$$E[X] = E[X_1 + X_2 + X_3] = E[X_1] + E[X_2] + E[X_3] = 360h.$$

Esercizio 3. Una variabile aleatoria X ha una distribuzione ignota con media 250 e deviazione standard 20.

(a) Determinare le probabilitá :

- $P(210 < X < 290)$
- $P(220 < X < 280)$

Soluzione: Se la distribuzione di X é ignota si usa la disuguaglianza di Chebychev nella forma

$$P_X(|X-\mu|< k\sigma) \geq 1-\frac{1}{k^2}.$$

con $\mu = 250, \sigma = 20$.
Entrambi gli intervalli hanno come punto centrale la media della distribuzione. Per esempio, nel primo caso il punto centrale vale $(210+290)/2 = 250$
Si trova il raggio dell'intorno di 250 che é $(290-210)/2 = 40$. Quindi l'intervallo é 250 ± 40
Si esprime il raggio come un multiplo della deviazione standard:

$$40 = k \times 20$$

da cui $k = 2$ (il raggio 40 doppio della deviazione standard).
Usando Chebychev:

$$P(210 < X < 290) \geq 1-\frac{1}{k^2} = 1-\frac{1}{4} = 0.75$$

Analogamente:

$$P(220 < X < 280) \geq 1-\frac{1}{k^2} = 1-\frac{1}{1.5^2} = 0.555$$

(b) Determinare le stesse probabilitá sapendo che X segue una legge normale $X \sim \mathcal{N}(\mu, \sigma)$

Soluzione: Usando le tavole della Gaussiana standard:

$$P(210 < X < 290) = P(-2 < Z < 2) = 2(0.9772) - 1 = 0.9544$$

$$P(220 < X < 280) = P(-1.5 < Z < 1.5) = 2(0.9332) - 1 = 0.8664$$

Compito scritto dell'esame di Statistica e analisi dei dati (Crema) 22 gennaio 2018	Prof. Giuseppe Boccignone	Corso di Laurea
Cognome:	Nome:	Matricola:

Istruzioni

- Il tempo riservato alla prova scritta e di 2 ore e 30 minuti. Durante la prova e possibile consultare libri e appunti.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede piu passaggi di calcolo, non sara preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o piu frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sara eventualmente rappresentato in quest'ultima forma).
- SOLO PER STUDENTI DEL CORSO ONLINE (VECCHIO ORDINAMENTO, Prof. Gianini): Indicare quale Modulo viene svolto al fine di completare una prova d'esame parziale precedentemente sostenuta con esito positivo
  - Modulo 1 ☐
  - Moduli 2 e 3 ☐

Modulo 1

Esercizio 1. Abbiamo sul tavolo 9 carte coperte: due di esse sono di cuori, tre di fiori e quattro di picche.

(a) Calcolare la probabilitá che, scelte simultaneamente due carte a caso, siano di seme diverso.

Soluzione: Indichiamo con QQ l'evento "estrazione di due cuori", con FF l'evento "estrazione di due fiori" e con PP l'evento "estrazione di due picche"
Lo spazio campione S é costituito da $\binom{9}{2} = 36$ eventi possibili ovvero il numero di combinazioni di 9 elementi a 2 a 2.
Inoltre si hanno
- $\binom{2}{2} = 1$ evento QQ
- $\binom{3}{2} = 3$ eventi FF
- $\binom{4}{2} = 6$ eventi PP

La probabilitá di estrarre due carte dello stesso seme vale:

$$P(\text{stesso seme}) = P(\{QQ\} \cup \{FF\} \cup \{PP\}) = P(\{QQ\}) + P(\{FF\}) + P(\{PP\}) = \frac{1}{36} + \frac{3}{36} + \frac{6}{36} = \frac{5}{18}$$

Pertanto:

$$P(\text{seme diverso}) = 1 - P(\text{stesso seme}) = \frac{13}{18}$$

Esercizio 2. Un costruttore viene fornito per gli stessi tipi di pezzi per l'80% dalla ditta A e per il restante 20% dalla ditta B. Tali pezzi vengono poi depositati assieme nello stesso magazzino. Per il passato é stato notato che i prodotti di A erano per il 5% difettosi, mentre quelli della ditta B lo erano nella misura del 9%

(a) Avendo scelto un pezzo a caso dal magazzino ed avendo riscontrato che é difettoso, qual é la probabilitá che sia stato fornito da B?

Soluzione: (Teorema di Bayes)
Indichiamo con D l'evento "pezzo difettoso", con A l'evento "fornito da A" e con B l'evento "fornito da B".
I dati del problema ci dicono che $P(D \mid A) = 0.05, P(D \mid B) = 0.09, P(A) = 0.80, P(B) = 0.20$
Usando Bayes:

$$P(B \mid D) = \frac{P(D \mid B)P(B)}{P(D \mid A)P(A) + P(D \mid B)P(B)} = \frac{0.018}{0.0518} = 0.3475. \tag{1}$$

Esercizio 3. Si lanciano due dadi non truccati.

(a) Qual é la probabilitá che sia uscito almeno un 6 sapendo che sono uscite due facce diverse?

Soluzione: (Prob. condiz.)

Definiamo gli eventi seguenti:

$$E = \text{“i dadi hanno valore diverso”};$$

$$F = \text{“uscito almeno un 6”}.$$

E' facile calcolare che

$$P(E) = \frac{5}{6}, P(E,F) = \frac{10}{36}.$$

La probabilitá richiesta é la condizionata:

$$P(F \mid E) = \frac{P(E,F)}{P(E)} = \frac{1}{3}$$

(b) Qual é la probabilitá che il primo dado abbia valore 6 sapendo che la somma dei numeri usciti é m, con $m \in \{2,...,12\}$?

Soluzione: (Teorema di Bayes)

Definiamo gli eventi seguenti:

$$A = \text{“il primo dado ha valore 6”};$$

$$B = \text{“la somma dei numeri usciti é m ”}, m \in \{2,...,12\}$$

Usando Bayes,

$$P(A \mid B_m) = \frac{P(B_m \mid A)P(A)}{P(B_m)}. \tag{2}$$

1. Se $m \leq 6$, allora $P(B_m \mid A) = 0 \implies P(A \mid B_m) = 0$
2. Se $m > 6$, allora $P(B_m \mid A) = \frac{1}{6}$ e $P(B_m) = \frac{12-m+1}{36}$.
   Poiché $P(A) = \frac{1}{6}$, $\implies P(A \mid B_m) = \frac{1}{12-m+1}$

Moduli 2 e 3

Esercizio 1. La durata Xdi vita (in mesi) di una batteria segue una distribuzione esponenziale $Exp(\lambda = 7)$

(a) Calcolare la probabilitá che la batteria abbia una durata di 3 mesi dato che é giá durata un mese.

Soluzione: Ricordiamo che per la distribuzione esponenziale vale la legge della mancanza della memoria:

$$P(X > 3 \mid X > 1) = P(X > 2) = 1 - P(X \leq 2) = e^{-14}$$

(b) Qual é la probabilitá che, su 30 batterie, meno di 5 durino piú di $\frac{1}{3}$ di mese?

Soluzione:

Sia Y la v.a. che conta il numero di batterie che durano piú di $\frac{1}{3}$ di mese.

Questa é discreta e si distribuisce secondo legge binomiale $Y \sim Bin(n,p)$, con parametri:

- $n = 30$
- $p = P(X > \frac{1}{3}) = 1 - P(X \leq \frac{1}{3}) = e^{-\frac{7}{3}}$

Pertanto:

$$P(Y \leq 5) = \sum_{i=0}^{4} \binom{30}{i} \cdot (e^{-\frac{7}{3}})^i \cdot (1 - e^{-\frac{7}{3}})^{30-i} = 0.8397$$

Esercizio 2. In un lungo manoscritto, si é scoperto che solo il 13.5% delle pagine contengono errori tipografici.

(a) Trovate la percentuale di pagine che hanno esattamente 1 errore.

Soluzione: Sia X la V.A. discreta che indica il numero di errori per pagina. Poiché la percentuale di errori é bassa, possiamo assumere che X segue una legge di Poisson, $X \sim Pois(\mu)$, ovvero

$$Pois(\mu) = P_X(X = k) = \frac{\mu^k}{k!} e^{-\mu} \tag{3}$$

Possiamo calcolare il parametro μ, usando il fatto che, per $k = 0$ (nessun errore)

$$P(X = 0) = 1 - P(X > 0) = 1 - 0.135 = 0.865 = e^{-\mu}$$

da cui $\mu = 0.145$. Dunque:

$$P_X(X = 1) = \frac{0.145^1}{1!} e^{-0.145} = 0.125 \tag{4}$$

cioé il 12.5% delle pagine ha esattamente un errore.

Esercizio 3.

Il voto dell'esame di Statistica e Analisi dei dati segue una distribuzione normale di media $\mu = 20$ ed é inoltre noto che il 70% degli studenti supera l?esame.

(a) Stabilire quale voto viene superato dal 10% degli studenti.

Soluzione:

Riportiamoci ad una normale standard $\mathcal{N}(0,1)$ in modo da poter consultare le tavole numeriche. Sia $Y \sim \mathcal{N}(0,1)$.

I dati del problema ci dicono che

$$P(X < 18) = P(Y < -\frac{2}{\sigma}) = 0.3$$

Dalla tabella leggiamo che $z_{0.3} = -z_{0.7} = 0.52 = -\frac{2}{\sigma}$, da cui $\sigma = 3.84615$.

Sia v il voto con cui viene superato dal 10% degli studenti. Quindi

$$P(X > v) = 1 - P(X \leq v) = 1 - P(Y \leq \frac{v - 20}{3.84615}) = 0.1$$

ovvero

$$P(Y \leq \frac{v - 20}{3.84615}) = 0.9$$

Dalla tabella $z_{0.9} = 1.28 = \frac{v-20}{3.84615}$, da cui $v = 24.92301$

Compito scritto dell'esame di Statistica e analisi dei dati (Crema) 23 giugno 2016	Prof. Giuseppe Boccignone	Corso di Laurea
Cognome:	Nome:	Matricola:

Istruzioni

- Il tempo riservato alla prova scritta è di 2 ore e 30 minuti. Durante la prova è possibile consultare libri e appunti.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede più passaggi di calcolo, non sarà preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o più frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sarà eventualmente rappresentato in quest'ultima forma).
- SOLO PER STUDENTI DEL CORSO ONLINE (VECCHIO ORDINAMENTO, Prof. Gianini): Indicare quale Modulo viene svolto al fine di completare una prova d'esame parziale precedentemente sostenuta con esito positivo
  - Modulo 1 ☐
  - Moduli 2 e 3 ☐

Modulo 1

Esercizio 1. (Probabilitá condizionata)
Un cappello contiene tre carte: una carta é nera su entrambi i lati; una carta é bianca su entrambi i lati; una carta é nera su un lato e bianca sull'altro. Le carte vengono mescolate nel cappello, poi una viene estratta a caso e appoggiata su di un tavolo.

(a) Se il lato visibile della carta é nero, qual é la probabilitá che l'altro lato sia bianco?

Soluzione: Uno studente brillante potrebbe risolvere in pochi secondi il problema ragionando come segue: la carta di interesse deve essere o la nero / nero o la nero / bianco: queste hanno pari probabilitá, dunque la probabilitá che, osservato il nero, l'altro lato sia bianco é $\frac{1}{2}$.
Lo studente, successivamente, per verificare la sua conclusione, effettua una simulazione del gioco delle carte (si veda lo script Matlab carte.m allegato) ma ottiene il risultato mostrato in Figura 1.

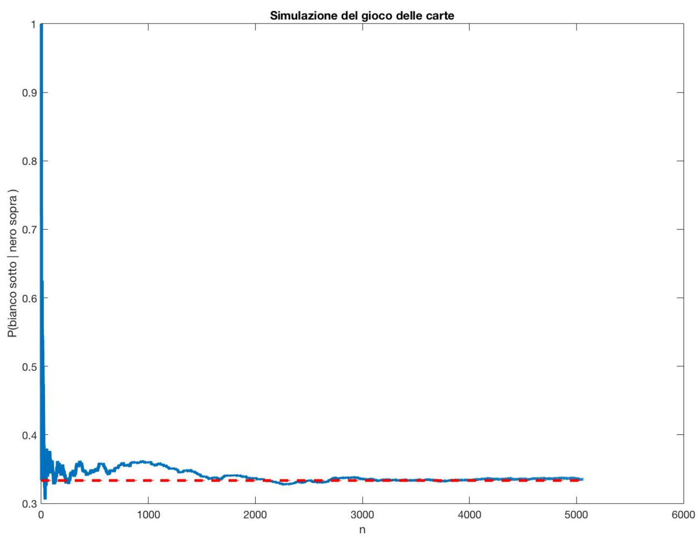

Figure 1: Simulazione del gioco delle tre carte: il valore di P("bianco sotto" | "nero sopra") valutato in termini di frequenza relativa, per un numero di prove n crescente, tende al valore di $0.333\cdots \approx \frac{1}{3}$

.

Il valore dell'approssimazione frequentistica di $P(\text{"bianco sotto"} \mid \text{"nero sopra"})$ tende a $\frac{1}{3}$.
Per capire bene come si arriva a questo risultato, mettiamo da parte l'intuizione e definiamo con precisione lo spazio campionario e gli eventi di interesse.
Identifichiamo i lati delle carte

- N_1 e N_2 per la carta nero / nero
- B_1 e B_2 per la carta bianco / bianco
- N_3 e B_3 per la carta nero / bianco

In questo caso lo spazio campionario S é

$$S = \{N_1, N_2, N_3, B_1, B_2, B_3\}$$

L'evento $N =$ "nero sopra" é definibile come $N = \{N_1, N_2, N_3\}$.
L'evento $B =$ "bianco sotto" é definibile come $B = \{N_3, B_1, B_2\}$
Pertanto:

$$P(B \mid N) = \frac{P(B \cap N)}{P(N)}$$

Poiché $\#(\text{"bianco sotto"} \cap \text{"nero sopra"}) = 1$ e $\#(\text{"nero sopra"}) = 3$:

$$P(B \mid N) = \frac{1}{3}.$$

Esercizio 2. (Legge delle probabilitá totali)
Ci sono due squadre, 1 e 2, di rigoristi. La squadra i-sima ha $3i$ componenti. Per ogni rigore battuto, un rigorista della squadra i ha una probabilitá $\frac{1}{i+1}$ di fare goal, indipendentemente dai rigori battuti precedentemente.

(a) Un rigorista viene selezionato casualmente fra tutti i rigoristi disponibili. Sia G l'evento che venga segnato un goal. Calcolare la probabilitá $P(G)$ che un goal sia realizzato, quale che sia la squadra di provenienza del rigorista.

Soluzione: In prima battuta, la probabilitá dell'evento G che un rigorista segni un goal é condizionata dalla squadra a cui appartiene: $P(G \mid S_i) = \frac{1}{i+1}$. Pertanto:

$$P(G \mid S_1) = \frac{1}{2}$$
$$P(G \mid S_2) = \frac{1}{3}$$

Ciascuna squadra ha $3i$ componenti: dunque 3 per la squadra 1 e 6 della squadra 2. Indicando con S_i l'evento che sia stato selezionato in maniera casuale un rigorista della squadra i-esima, le probabilitá a priori sono

$$P(S_1) = \frac{3}{9} = \frac{1}{3}$$
$$P(S_2) = \frac{6}{9} = \frac{2}{3}$$

Usando la Legge delle Probabilitá totali:

$$P(G) = P(G \mid S_1)P(S_1) + P(G \mid S_2)P(S_2) = \frac{1}{2} \times \frac{1}{3} + \frac{1}{3} \times \frac{2}{3} = \frac{7}{18} = 0.388$$

(b) Vengono scelti a caso due rigoristi. Per $j = 1{,}2$, sia G_j l'evento che il rigorista j faccia goal. Trovare la probabilitá che il rigorista 1 e il rigorista 2 segnino entrambi un goal (*Suggerimento:* calcolare, mediante un chance tree, la probabilitá dell'evento S_{kl} che il primo rigorista sia stato selezionato dalla squadra k e il secondo rigorista dalla squadra l, con $k{,}l = 1{,}2$)

Soluzione: Per risolvere questa parte dl problema ricaviamo l'albero di probabilitá per due "estrazioni" successive e casuali di un rigorista dalle squadre 1 e 2 rappresentato in Figura 2,
Il diagramma ci dice che:

$$P(S_{11}) = \frac{1}{12}, P(S_{12}) = \frac{1}{4}, P(S_{21}) = \frac{1}{4}, P(S_{22}) = \frac{5}{12}$$

La probabilitá dell'evento congiunto $\{G_1, G_2\}$ é condizionata da quali squadre sono stati selezionati i rigoristi 1 e 2, ovvero dall'evento S_{kl} le cui probabilitá sono state derivate con il diagramma ad albero. Essendo l'esito di un rigore indipendente dall'altro: $P(G_1, G_2 \mid S_{kl}) = P(G_1 \mid S_{kl})P(G_2 \mid S_{kl})$. Pertanto:

$$P(G_1, G_2 \mid S_{11}) = \left(\frac{1}{2}\right)^2$$

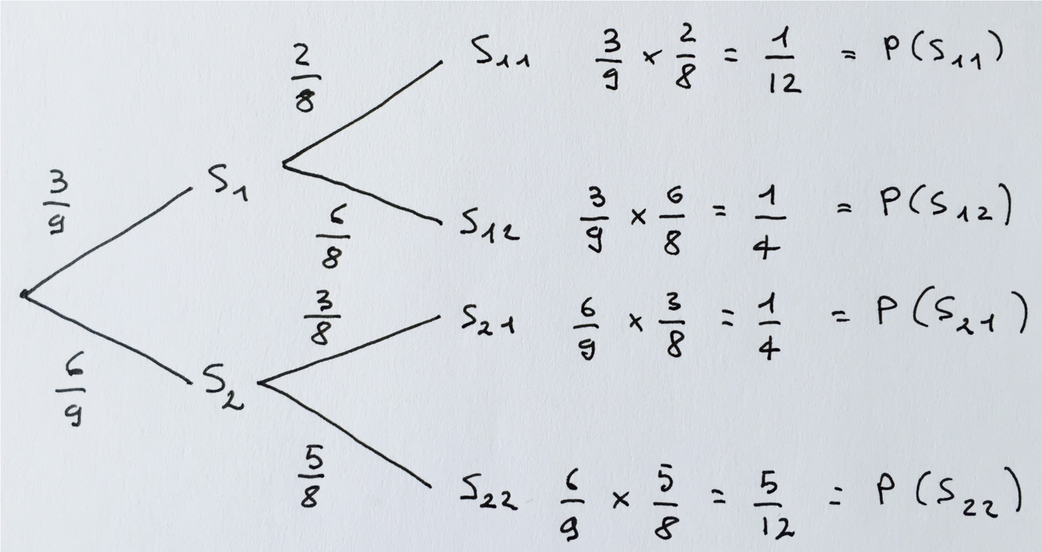

Figure 2: Chance tree per due "estrazioni" successive e casuali di un rigorista dalle squadre 1 e 2: S_k é l'evento che sia stato selezionato in maniera casuale un rigorista della squadra i-esima, S_{kl} che il primo rigorista sia stato selezionato dalla squadra k e il secondo rigorista dalla squadra l

$$P(G_1, G_2 \mid S_{12}) = \frac{1}{2} \times \frac{1}{3}$$
$$P(G_1, G_2 \mid S_{21}) = \frac{1}{3} \times \frac{1}{2}$$
$$P(G_1, G_2 \mid S_{22}) = \left(\frac{1}{3}\right)^2$$

Usando di nuovo la Legge delle Probabilitá totali:

$$P(G_1, G_2) = P(G_1, G_2 \mid S_{11})P(S_{11}) + P(G_1, G_2 \mid S_{12})P(S_{12}) + P(G_1, G_2 \mid S_{21})P(S_{21}) + P(G_1, G_2 \mid S_{22})P(S_{22}) = \frac{15}{96} = 0.1562$$

(c) Verificare se G_1 e G_2 siano eventi indipendenti. Commentare brevemente il risultato ottenuto.

Soluzione: Possiamo verificare la condizione di indipendenza stocastica:

$$P(G_1, G_2) = P(G_1)P(G_2)$$

E' evidente che $P(G_1) = P(G) = \frac{7}{18}$ e che per simmetria $P(G_1) = P(G_2)$, giacché lo scambio dei due rigoristi non muta nulla in termini di esito (se non si nvinti, usare ancora le probabilitá totali per calcolare $P(G_i) = P(G_i \mid S_{11})P(S_{11}) + P(G_i \mid S_{12})P(S_{12}) + P(G_i \mid S_{21})P(S_{21}) + P(G_i \mid S_{22})P(S_{22}) = \frac{7}{18}$).
Si ha allora che:

$$P(G_1, G_2) = 0.1562 \neq 0.1512 = (0.388)^2 = P(G_1)P(G_2)$$

dunque non sono indipendenti. Le due probabiliá ricavate sono vicine ma non esattamente uguali. G_1 e G_2 sono dipendenti perché se il primo rigore viene segnato, vi é maggior probabilit che abbia tirato un rigorista del gruppo 1. Questo rende piú probabile che il secondo giocatore venga selezionato dal gruppo 2, evento che riduce la probabilitá di segnare al secondo tiro.

Esercizio 3. (Teorema di Bayes)
Un robot autonomo controlla il livello di corrosione interna delle tubature dei sistemi di raffreddamento di una centrale nucleare. La corrosione non puó essere osservata direttamente, ma il robot effettua un test che puó fornisce indizi sulla specifica corrosione. Il test non é infallibile: nel 70% dei casi individua una corrosione quando é presente, ma puó anche sbagliare indicando una corrosione inesistente nel 20% dei casi. Risolvere il problema del robot a cui un operatore remoto, sotto l'ipotesi che il 10% delle tubature siano corrose, chiede di:

(a) Comunicargli la probabilitá che una parte delle tubature abbia delle corrosioni interne dopo che il test ha dato esito positivo (presenza di corrosioni). Determinare tale probabilitá.

Soluzione: Definiamo gli eventi $C =$ "la tubatura é corrosa" e $\overline{C} =$ "il test ha identificato la tubatura come corrosa".
I dati del problema ci dicono che, a priori,

$$P(C) = 0.1$$

e dunque $P(\sim C) = 1 - 0.1 = 0.9$, dove $\sim C =$ "la tubatura non é corrosa". Inoltre

$$P(\overline{C} \mid C) = 0.7$$
$$P(\overline{C} \mid\sim C) = 0.2$$

Applicando la regola di Bayes:

$$P(C \mid \overline{C}) = \frac{P(\overline{C} \mid C)P(C)}{P(\overline{C})}$$

dove $P(\overline{C}) = P(\overline{C} \mid C)P(C) + P(\overline{C} \mid\sim C)P(\sim C) = 0.25$. Quindi

$$P(C \mid \overline{C}) = \frac{0.7 \times 0.1}{0.25} = 0.28$$

(b) Comunicargli la probabilitá che una parte delle tubature abbia delle corrosioni dopo che il test ha dato esito negativo (nessuna corrosione). Determinare tale probabilitá.

Soluzione: Si ripeta il procedimento precedente per il caso $P(C \mid\sim \overline{C})$

Moduli 2 e 3

Esercizio 1. (Funzione di densitá e di ripartizione)

Data la funzione: $\mathrm{f}_X(x) = \begin{cases} \frac{x}{2} - 1 & 2 \leq x < k \\ 0 & \text{altrimenti} \end{cases}$.

(a) Si determini il valore di k che assicura che $f_X(x)$ rappresenta una funzione di densitá.

Soluzione: Imponiamo la condizione di normalizzazione della densitá $f_X(x)$:

$$\int_2^k \left(\frac{x}{2} - 1\right) dx = 1 \implies \left[\frac{x^2}{4}\right]_2^k - (k-2) = 1 \implies k^2 - 4k = 0$$

Risolvendo l'equazione, vi sono due possibili radici $k_1 = 0$ e $k_2 = 4$. Poiché $k_1 < 2 \leq x < k$, é da scartare, dunque $k = k_2 = 4$. Pertanto la densitá correttamente normalizzata é:

$$f_X(x) = \frac{x}{2} - 1 \text{ con } 2 \leq x < 4$$

(b) Si individui la corrispondente funzione di ripartizione $F_X(x)$

Soluzione: Per definizione $F_X(x) = \int_{-\infty}^x f_X(t)dt$. Nella fattispecie:

$$F_X(x) = \int_2^x \left(\frac{t}{2} - 1\right) dt$$

Integrando come fatto prima si ottiene:

$$F_X(x) = \frac{x^2}{4} - x + 1$$

per $2 \leq x < 4$

(c) Si individui la mediana della variabile aleatoria X descritta dalla densitá $f_X(x)$

Soluzione: Per definizione la mediana x_{med} é il valore per cui

$$F_X(x) = \frac{1}{2} \implies \frac{x^2}{4} - x + 1 = \frac{1}{2}$$

Risolvendo l'equazione si trovano le soluzioni $x_1 = 2 + \sqrt{2}$ e $x_2 = 2 - \sqrt{2}\sqrt{2}$. Dunque $x_{med} = x_1$

Esercizio 2. (Distribuzioni di Poisson ed Esponenziale negativa)

Il numero di automobili che attraversano un particolare incrocio stradale in un'ora é mediamente pari a 30.

(a) Determinare la probabilitá che in un intervallo di tempo di cinque minuti nessuna automobile attraversi l'incrocio in questione.

Soluzione: Sia X la V.A. discreta che indica il numero di veicoli che passa lungo l'incrocio in 5 min. X segue una legge di Poisson, $X \sim Pois(\mu)$. Il numero medio di automobili che passa in 5 min. é $\frac{30}{12} = 2.5$. Dunque $X \sim Pois(2.5)$. Pertanto:

$$P(X = 0) = \frac{\mu^0}{0!}e^{-\mu} = e^{-2.5} = 0.082$$

(b) Qual é la probabilitá che in dieci minuti almeno due automobili passino lungo quel tratto di strada?

Soluzione: Se X denota il numero di veicoli che passa lungo l'incrocio in 5 min, indichiamo con $Y = 2X$ la V.A. discreta che indica il numero di veicoli che passa lungo l'incrocio in 10 min. Allora:

$$E[Y] = E[2X] = 2E[X] = 2\mu = 5$$

Dunque, $Y \sim Pois(5)$. La risposta alla domanda consiste nel calcolare

$$P(Y \geq 2) = 1 - P(Y < 2) = 1 - (P(Y = 0) + P(Y = 1)) = 1 - \left(e^{-5} + \frac{5^1}{1!}e^{-5}\right) = 1 - 6e^{-5} = 0.9595$$

(c) Assumendo che la variabile aleatoria T, rappresentante il tempo (sempre in minuti) trascorso tra il passaggio di un'auto e di quella successiva, si distribuisca con legge esponenziale negativa, qual é la probabilitá che tra il passaggio di un'auto e la successiva trascorra piú di un minuto?

Soluzione: Poiché $T \sim Exp(\lambda)$, i dati del problema ci dicono che il rate di passaggio é

$$\lambda = \frac{30}{60}\frac{\text{(auto)}}{\text{(min.)}}.$$

Dunque la V.A. T ha funzione di densitá $f_T(t) = 0.5e^{-0.5t}$, $t \geq 0$. La probabilitá che il tempo di attesa tra il passaggio di un un'auto e la successiva sia > 1 min. é $P(T > 1)$ che si puó ottenere dal calcolo diretto $P(T > 1) = \int_1^{\infty} f_T(t)dt$, oppure usando la funzione di sopravvivenza $S_T(t) = e^{-\lambda t}$:

$$P(T > 1) = 1 - F_T(1) = S_T(1) = e^{-0.5 \cdot 1} = 0.6065$$

Esercizio 3. (Distribuzione Normale, uso di tabelle)
Una linea di produzione fabbrica resistori da 1000 ohm (Ω). La fabbrica adotta una tolleranza del 10% (ovvero sono scartati i resistori con resistenza $R < 900\ \Omega$ e $R > 1100\ \Omega$).

(a) Sotto l'ipotesi che la resistenza R di un resistore sia una variabile aleatoria distribuita con legge normale di media $1000\ \Omega$ e varianza 2500, si calcoli la probabilitá che un resistore preso a caso venga scartato

Soluzione: Sia S l'evento in cui un resistore viene scartato. Allora:

$$S = \{R < 900\} \cup \{R > 1100\}$$

Poiché $\{R < 900\} \cap \{R > 1100\} = \emptyset$, si ha che

$$P(S) = P(\{R < 900\}) \ + P(\{R > 1100\}) = F_X(900) + (1 - F_X(1100)) = \Phi\left(\frac{900 - \mu}{\sigma}\right) + \left(1 - \Phi\left(\frac{1100 - \mu}{\sigma}\right)\right)$$

Poiché $\mu = 1000$, e $\sigma^2 = 2500 \implies \sigma = 50$, le variabili normalizzate sono

$$\frac{900 - 1000}{50} = -2$$

$$\frac{1100 - 1000}{50} = 2$$

.
Sfruttando la proprietá della cumulativa standardizzata $\Phi(-z) = 1 - \Phi(z)$, la precedente diventa:

$$P(S) = 2\,(1 - \Phi(2))$$

.
Accedendo alla tabella della Normale standard per $z = 2$, $\Phi(2) = 0.9772$ da cui $P(S) = 0.0455$

Compito scritto dell'esame di Statistica e analisi dei dati **15 febbraio 2023**	**Prof. Giuseppe Boccignone**	**Corso di Laurea**
Cognome:	**Nome:**	**Matricola:**

Istruzioni

- Il tempo riservato alla prova scritta e' di 2 ore. Durante la prova e' possibile consultare il formulario ed utilizzare la calcolatrice. Non e' possibile consultare libri e appunti.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede piu' passaggi di calcolo, non sara' preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o piu' frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sara' eventualmente rappresentato in quest'ultima forma).

Problemi

Esercizio 1. Il giocatore A lancia un dado non truccato per 4 volte, e vince se esce almeno una volta il 6. Il giocatore B lo lancia 8 volte, e vince se il 6 esce almeno due volte.

(a) Chi ha maggiore probabilitá di vincere e perché?

Soluzione: (Distribuzione binomiale) In ogni lancio la probabilitá che esca il 6 vale $p = 1/6$ (equiprobabilitá di 6 eventi). La probabilitá di avere $X = 0$ successi in $n = 4$ prove indipendenti vale, si determina in base alla distribuzione binomiale:

$$Bin\left(X = 0 \mid n = 4, p = \frac{1}{6}\right) = \binom{4}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^4 = \left(\frac{5}{6}\right)^4 \approx 0.48226$$

Dunque, la probabilitá di vittoria per A é

$$P(A) = 1 - 0.48226 \approx 0.51774$$

Per il giocatore B, la probabilitá di avere non piú di $X = 1$ successi in $n = 8$ prove (perdendo cosí la scommessa) é

$$P_X(0 \leq X \leq 1) = Bin\left(X = 0 \mid n = 8, p = \frac{1}{6}\right) + Bin\left(X = 1 \mid n = 8, p = \frac{1}{6}\right)$$

ovvero

$$P_X(0 \leq X \leq 1) = \binom{8}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^8 + \binom{8}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^7 \approx 0.6046$$

Pertanto, la probabilitá di vittoria per B é

$$P(B) = 1 - P_X(0 \leq X \leq 1) \approx 0.3936$$

Esercizio 2. Una variabile aleatoria X ha densitá

$$f_X(x) = \begin{cases} x - \frac{1}{4}x^3, & per \quad 0 \leq x \leq 2 \\ 0 & altrove \end{cases}$$

(a) Determinare valore atteso, varianza e mediana di X.

Soluzione:

$$E[X] = \int_0^2 x\left(x - \frac{1}{4}x^3\right)dx = \left[\frac{x^3}{3} - \frac{x^5}{20}\right]_0^2 = \frac{16}{15}$$

La varianza é $\sigma_X^2 = E[X^2] - (E[X])^2 = E[X^2] - \left(\frac{16}{15}\right)^2$, dove

$$E[X^2] = \int_0^2 x^2\left(x - \frac{1}{4}x^3\right)dx = \left[\frac{x^4}{4} - \frac{x^6}{24}\right]_0^2 = \frac{4}{3}$$

Pertanto,

$$\sigma_X^2 = \frac{4}{3} - \left(\frac{16}{15}\right)^2 = 16\left(\frac{1}{12} - \frac{16}{225}\right) \approx 0.195$$

Per calcolare la mediana x_{med}, si impone $F_X(x_{med}) = \frac{1}{2}$, dunque:

$$F_X(x_{med}) = \int_0^{x_{med}} \left(x - \frac{1}{4}x^3\right) dx = \left[\frac{x^2}{2} - \frac{x^4}{16}\right]_0^{x_{med}} = \frac{1}{2}\left(x_{med}^2 - \frac{x_{med}^4}{16}\right) = \frac{1}{2}$$

Si deve pertanto risolvere l'equazione

$$x_{med}^4 - 8x_{med}^2 + 8 = 0$$

cercando la radice che appartiene all'intervallo $0 \le x \le 2$.
Ponendo $t = x_{med}^2$

$$t = \begin{cases} 4 + 2\sqrt{2} \\ 4 - 2\sqrt{2} \end{cases}$$

da cui

$$\begin{cases} x_{med}^{(1,2)} = \pm\sqrt{4 + 2\sqrt{2}} \approx \pm 2.613 \\ x_{med}^{(3,4)} = \pm\sqrt{4 - 2\sqrt{2}} \approx \pm 1.0924 \end{cases}$$

Scartando le soluzioni $x_{med}^{(1,2)}$ giacché fuori dell'intervallo $0 \le x \le 2$, l'unica soluzione ammissibile é

$$x_{med}^{(3)} = \sqrt{4 - 2\sqrt{2}} \approx 1.0924$$

Esercizio 3. Una linea di produzione fabbrica resistori da 1000 ohm (Ω). La fabbrica adotta una tolleranza del 10% (ovvero sono scartati i resistori con resistenza $R < 900\ \Omega$ e $R > 1100\ \Omega$).

(a) Sotto l'ipotesi che la resistenza R di un resistore sia una variabile aleatoria distribuita con legge normale di media $1000\ \Omega$ e varianza 2500, si calcoli la probabilitá che un resistore preso a caso venga scartato

Soluzione: Sia S l'evento in cui un resistore viene scartato. Allora:

$$S = \{R < 900\} \cup \{R > 1100\}$$

Poiché $\{R < 900\} \cap \{R > 1100\} = \emptyset$, si ha che

$$P(S) = P(\{R < 900\}) + P(\{R > 1100\}) = F_X(900) + (1 - F_X(1100)) = \Phi\left(\frac{900 - \mu}{\sigma}\right) + \left(1 - \Phi\left(\frac{1100 - \mu}{\sigma}\right)\right)$$

Poiché $\mu = 1000$, e $\sigma^2 = 2500 \implies \sigma = 50$, le variabili normalizzate sono

$$\frac{900 - 1000}{50} = -2$$

$$\frac{1100 - 1000}{50} = 2$$

.
Sfruttando la proprietá della cumulativa standardizzata $\Phi(-z) = 1 - \Phi(z)$, la precedente diventa:

$$P(S) = 2\left(1 - \Phi(2)\right)$$

.
Accedendo alla tabella della Normale standard per $z = 2$, $\Phi(2) = 0.9772$ da cui $P(S) = 0.0455$

Esercizio 4. Il numero di automobili che attraversano un particolare incrocio stradale in un'ora é mediamente pari a 30.

(a) Determinare la probabilitá che in un intervallo di tempo di cinque minuti nessuna automobile attraversi l'incrocio in questione.

Soluzione: Sia X la V.A. discreta che indica il numero di veicoli che passa lungo l'incrocio in 5 min. X segue una legge di Poisson, $X \sim Pois(\mu)$. Il numero medio di automobili che passa in 5 min. é $\frac{30}{12} = 2.5$. Dunque $X \sim Pois(2.5)$. Pertanto:

$$P(X=0) = \frac{\mu^0}{0!}e^{-\mu} = e^{-2.5} = 0.082$$

(b) Qual é la probabilitá che in dieci minuti almeno due automobili passino lungo quel tratto di strada?

Soluzione: Se X denota il numero di veicoli che passa lungo l'incrocio in 5 min, indichiamo con $Y = 2X$ la V.A. discreta che indica il numero di veicoli che passa lungo l'incrocio in 10 min. Allora:

$$E[Y] = E[2X] = 2E[X] = 2\mu = 5$$

Dunque, $Y \sim Pois(5)$. La risposta alla domanda consiste nel calcolare

$$P(Y \geq 2) = 1 - P(Y < 2) = 1 - (P(Y=0) + P(Y=1)) = 1 - \left(e^{-5} + \frac{5^1}{1!}e^{-5}\right) = 1 - 6e^{-5} = 0.9595$$

(c) Assumendo che la variabile aleatoria T, rappresentante il tempo (sempre in minuti) trascorso tra il passaggio di un'auto e di quella successiva, si distribuisca con legge esponenziale negativa, qual é la probabilitá che tra il passaggio di un'auto e la successiva trascorra piú di un minuto?

Soluzione: Poiché $T \sim Exp(\lambda)$, i dati del problema ci dicono che il rate di passaggio é

$$\lambda = \frac{30}{60}\frac{(\text{auto})}{(\text{min.})}.$$

Dunque la V.A. T ha funzione di densitá $f_T(t) = 0.5e^{-0.5t}$, $t \geq 0$. La probabilitá che il tempo di attesa tra il passaggio di un un'auto e la successiva sia > 1 min. é $P(T > 1)$ che si puó ottenere dal calcolo diretto $P(T > 1) = \int_1^\infty f_T(t)dt$, oppure usando la funzione di sopravvivenza $S_T(t) = e^{-\lambda t}$:

$$P(T > 1) = 1 - F_T(1) = S_T(1) = e^{-0.5 \cdot 1} = 0.6065$$

Esercizio 5. Una macchina viene utilizzata su una linea di produzione inglese per riempire scatole. La deviazione standard del peso é di 0.3 once. Il problema da risolvere è mantenere bassa la variabilitá del peso di prodotto nella scatola. Dopo aver implementato un miglioramento della linea, si estrae un campione aleatorio di 20 scatole, misurando una varianza campionaria di 0.045 once

(a) Assumendo che i pesi si distribuiscano con legge normale, si tragga una conclusione con un livello di confidenza al 95% sull'efficacia del miglioramento produttivo.

Soluzione: Abbiamo che

$$n = 20, s^2 = 0.045, \sigma = 0.3$$

Utilizziamo l'intervallo di confidenza

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

Determiniamo $\alpha/2$. Da $95\% = 100(1-\alpha)\%$, si ha:

$$\begin{aligned}(100)(1-\alpha) &= 95\\ 1-\alpha &= 0.95\\ \alpha &= 0.05\\ \alpha/2 &= 0.025\end{aligned}$$

I g.d.l. sono $\nu = n - 1 = 19$. Usando la tabella della distribuzione Chi-quadro, il valore critico per 0.025 con 19 g.d.l. é $\chi^2_{0.025} = 32.825$, e $\chi^2_{1-0.025} = \chi^2_{0.975} = 8.907$.
L'intervallo di interesse é

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$
$$\frac{(19)\cdot(0.045)}{\chi^2_{0.025}} < \sigma^2 < \frac{(19)\cdot(0.045)}{\chi^2_{0.975}}$$
$$\frac{(19)\cdot(0.045)}{32.825} < \sigma^2 < \frac{(19)\cdot(0.045)}{8.907}$$
$$0.012 < \sigma^2 < 0.045$$

Pertanto l'intervallo di confidenza al 95% per σ é approssimabile come

$$0.110 < \sigma < 0.212$$

Poiché 0.3 cade al di fuori di tale intervallo, si puó concludere che da un punto di vista statistico, il nuovo processo abbia significativamente ridotto e dunque migliorata la variabilità

Esercizio 6. Bob ha calcolato un intervallo di confidenza per la media di una popolazione normale, a partire da un campione di 8 osservazioni, dopo averne stimato la varianza con la varianza campionaria $s^2 = 4.2$. Il risultato é $(16.08, 18.82)$.

(a) Con quale confidenza é stato ottenuto l'intervallo?

Soluzione:

La taglia del campione $n = 8$ é piccola e la varianza é ignota essendo stata stimata mediante la varianza campionaria $s^2 = 4.2$. Inoltre, la popolazione da cui proviene il campione é normale. Pertanto l'IC per la media deve essere stato stimato utilizzando la distribuzione t di Student con $\nu = n - 1 = 7$ gradi di libertá:

$$\mu = \bar{x} \pm \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2}}(\nu = 7) = \bar{x} \pm \frac{\sqrt{4.2}}{\sqrt{8}} t_{\frac{\alpha}{2}}(\nu = 7)$$

Gli estremi dell'IC indicato corrispondono a

$$\bar{x} - \frac{\sqrt{4.2}}{\sqrt{8}} t_{\frac{\alpha}{2}}(\nu = 7) = 16.08$$
$$\bar{x} + \frac{\sqrt{4.2}}{\sqrt{8}} t_{\frac{\alpha}{2}}(\nu = 7) = 18.82$$

Risolvendo il sistema a due incognite si ottiene

$$\bar{x} = 17.45$$
$$t_{\frac{\alpha}{2}}(\nu = 7) = 1.895$$

Dalla tabella della t di Student con $\nu = 7$ leggiamo che $t_{\frac{\alpha}{2}} = 1.895$ per $\frac{\alpha}{2} = 0.05$, cioé $\alpha = 0.10$
Dunque l'intervallo é stato ottenuto con una confidenza pari al $(100)(1-\alpha) = (100)(1-0.10) = 90\%$

Compito scritto dell'esame di Statistica e analisi dei dati **17 gennaio 2023**	**Prof. Giuliano Grossi**	**Corso di Laurea**
Cognome:	**Nome:**	**Matricola:**

Istruzioni

- Il tempo riservato alla prova scritta e' di 2 ore. Durante la prova e' possibile consultare il formulario ed utilizzare la calcolatrice. Non e' possibile consultare libri e appunti.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede piu' passaggi di calcolo, non sara' preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o piu' frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sara' eventualmente rappresentato in quest'ultima forma).

Problemi

Esercizio 1.

Dovete partecipare ad un torneo di scacchi in cui giocate contro altri tre giocatori (una sola partita con ciascuno). Conoscete le probabilitá di vincere con ognuno di loro e potete scegliere l'ordine in cui incontrarli. Si vince il torneo se si vincono due partite consecutive.

(a) Volendo massimizzare la probabilitá di vittoria, si mostri che l'ordinamento ottimale é quello che vi fa incontrare alla seconda partita il giocatore piú debole dei tre, mentre l'ordine in cui incontrate i restanti due giocatori non ha rilevanza.

Soluzione:
Denotiamo con

$$P(\texttt{"vittoria contro il giocatore incontrato alla partita i -esima"}) = P(i) = p_i$$

con $i = 1,2,3$.
La descrizione del problema ci dice che si vince la partita se si gioca la seconda partita con il giocatore piú debole - dunque con $p_2 \geq p_1, p_2 \geq p_3$ - e, **congiuntamente**, si vince con l'uno **oppure** l'altro giocatore incontrati alla partita 1 o 3. Quindi, formalmente, la probabilitá $P(V)$ di vincere il torneo é

$$P(V) = P(2 \cap (1 \cup 3)) = P(2)P(1 \cup 3) = P(2)\left[P(1) + P(3) - P(1) \cap P(3)\right] =$$
$$= p_2(p_1 + p_3 - p_1 p_3).$$

L'ordinamento é ottimale se tale probabilitá non é inferiore alle probabilitá calcolate utilizzando i due ordinamenti alternativi $P(1 \cap (2 \cup 3)) = p_1(p_2 + p_3 - p_2 p_3)$ e $P(3 \cap (2 \cup 1)) = p_3(p_2 + p_1 - p_2 p_1)$, ovvero:

$$p_2(p_1 + p_3 - p_1 p_3) \geq p_1(p_2 + p_3 - p_2 p_3), \quad (1)$$
$$p_2(p_1 + p_3 - p_1 p_3) \geq p_3(p_2 + p_1 - p_2 p_1). \quad (2)$$

Infatti, la prima disuguaglianza é equivalente alla condizione

$$p_2 \geq p_1,$$

mentre la seconda disuguaglianza si riduce a

$$p_2 \geq p_3.$$

Le due condizioni (**??**) e (**??**) ci dicono che la probabilitá di vincere alla seconda partita deve essere massima, ovvero: per massimizzare la probabilitá di vincere il torneo il secondo giocatore da incontrare deve essere il piú debole dei tre.

Esercizio 2.

Un robot autonomo controlla il livello di corrosione interna delle tubature dei sistemi di raffreddamento di una centrale nucleare. La corrosione non puó essere osservata direttamente, ma il robot effettua un test che puó fornisce indizi sulla specifica corrosione. Il test non é infallibile: nel 70% dei casi individua una corrosione quando é presente, ma puó anche sbagliare indicando una corrosione inesistente nel 20% dei casi. Risolvere il problema del robot a cui un operatore remoto, sotto l'ipotesi che il 10% delle tubature siano corrose, chiede di:

(a) Comunicargli la probabilitá che una parte delle tubature abbia delle corrosioni interne dopo che il test ha dato esito positivo (presenza di corrosioni). Determinare tale probabilitá.

Soluzione: Definiamo gli eventi $C =$ "la tubatura é corrosa" e $\overline{C} =$ "il test ha identificato la tubatura come corrosa". I dati del problema ci dicono che, a priori,

$$P(C) = 0.1$$

e dunque $P(\sim C) = 1 - 0.1 = 0.9$, dove $\sim C =$ "la tubatura non é corrosa". Inoltre

$$P(\overline{C} \mid C) = 0.7$$
$$P(\overline{C} \mid\sim C) = 0.2$$

Applicando la regola di Bayes:

$$P(C \mid \overline{C}) = \frac{P(\overline{C} \mid C)P(C)}{P(\overline{C})}$$

dove $P(\overline{C}) = P(\overline{C} \mid C)P(C) + P(\overline{C} \mid\sim C)P(\sim C) = 0.25$. Quindi

$$P(C \mid \overline{C}) = \frac{0.7 \times 0.1}{0.25} = 0.28$$

(b) Comunicargli la probabilitá che una parte delle tubature abbia delle corrosioni dopo che il test ha dato esito negativo (nessuna corrosione). Determinare tale probabilitá.

Soluzione: Si ripeta il procedimento precedente per il caso $P(C \mid\sim \overline{C})$

Esercizio 3. Si dispone di un campione di 100 misure di una variabile temporale X di una popolazione di cui non conosciamo la distribuzione, ma la cui deviazione standard é nota e vale $\sigma_X = 120$ secondi. .

(a) Qual é la probabilitá che la media campionaria differisca per piú di 3 secondi dal valore atteso teorico (incognito) dei tempi misurati?

Soluzione:
Siano:

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

la media campionaria;

$$\mu_X = E[X]$$

il valore atteso teorico.
Il problema ci chiede di determinare la probabilitá

$$P_X\left(\left|\frac{S_n}{n} - \mu\right| > 3\right)$$

Non conosciamo la legge di probabilitá secondo cui si distribuisce la popolazione, ma il Teorema Centrale Limite (TCL) ci assicura che per n sufficientemente grande la distribuzione degli scarti fra media campionaria $\frac{S_n}{n}$ e valore atteso teorico μ_X segue una legge Gaussiana:

$$P\left(a \leq \frac{\sqrt{n}}{\sigma_X}\cdot\left(\frac{S_n}{n} - \mu\right) \leq b\right) \approx \Phi(b) - \Phi(a).$$

Nel nostro caso,

$$\frac{\sigma_X}{\sqrt{n}} = \frac{120}{\sqrt{100}} = 12$$

é la deviazione standard del campione.
Usando il TCL con $a = -3, b = 3$ e passando alla variabile standardizzata $Z_n \sim N(0,1)$

$$Z_n = \frac{\frac{S_n}{n} - \mu_X}{\frac{\sigma_X}{\sqrt{n}}},$$

il problema iniziale si riduce a calcolare la seguente:

$$P_X\left(\left|\frac{S_n}{n} - \mu_X\right| > 3\right) = P_Z\left(\frac{\sigma_X}{\sqrt{n}}|Z_n| > 3\right) = P_Z(|Z_n| > 0.25).$$

Nel caso Gaussiano, $P_Z(|Z_n| > z) = 2(1 - \Phi(z))$, dunque :

$$P_Z(|Z_n| > 0.25) = 2(1 - \Phi(0.25)).$$

Usando le tabelle della CDF standard Φ leggiamo che

$$\Phi(0.25) = 0.5987.$$

Pertanto,

$$P_X\left(\left|\frac{S_n}{n} - \mu\right| > 3\right) = 2(1 - 0.5987) = 0.8026.$$

Esercizio 4.
Vengono sottoposti a confronto i consumi delle autovetture A e B alla velocitá costante di $120 Km/h$. Si ritiene che i consumi dei due tipi di autovetture possa essere descritto da variabili aleatorie con distribuzione normale con la stessa varianza. L'auto di tipo A in 20 prove consuma mediamente $6.5 litri/100Km$, quella di tipo B in 22 prove consuma mediamente $6.6 litri/100Km$. Le relative varianze campionarie sono rispettivamente di 0.30 e 0.28.

(a) Possiamo ritenere che le due autovetture abbiano lo stesso consumo medio al livello di significativitá del 5%?

Soluzione:
Abbiamo che:

- $\bar{x}_A = 6.5, s_A^2 = 0.30, n_A = 20$
- $\bar{x}_B = 6.6, s_B^2 = 0.28, n_B = 22$

Siamo nel caso di varianze ignote ma eguali, e la differenza fra medie é valutabile mediante una statistica t-Student, dove i gradi di libertá ν sono dati da $\nu = n_A + n_B - 2 = 40$, $\alpha = 0.05$, per cui, $t_{\nu,\frac{\alpha}{2}} = t_{40,0.025} \approx 2.021$
La deviazione standard pooled vale:

$$S_{pool} = \sqrt{\frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2}} = 0.5381$$

L'intervallo di confidenza per la differenza fra le medie $\mu_A - \mu_B$ a livello $(100)(1 - \alpha) = (100)(1 - 0.05) = 95\%$ vale pertanto

$$\bar{x}_A - \bar{x}_B \pm t_{40,0.025} S_{pool}\sqrt{\frac{1}{n_A} + \frac{1}{n_B}} = -0.1 \pm 0.33586$$

ovvero l'IC ha estremi $[-0.43596; 0.23596]$ per cui si puó ragionevolmente ritenere che le due autovetture abbiano differenza in consumo medio trascurabile.
Equivalentemente, si osserva che la statistica della differenza fra medie normalizzata

$$\left|\frac{\bar{x}_A - \bar{x}_B}{S_{pool}\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}\right| = 0.612 < t_{40,0.025} = 2.021$$

Esercizio 5. Un produttore di batterie per auto garantisce che il suo prodotto dura in media 3 anni con una deviazione standard di 1 anno. Si campionano casualmente 5 batterie e ne risulta che abbiano una durata di $1.9, 2.4, 3.0, 3.5, 4.2$ anni.

(a) Assumendo che la vita della popolazione di batterie sia distribuita con legge normale, si tragga una conclusione con un livello di confidenza al 95% sull'asserzione del produttore che $\sigma = 1$.

Soluzione:
Utilizziamo l'intervallo di confidenza

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

con $\sigma^2 = \sigma = 1$
Determiniamo la varianza del campione s^2 con

$$s^2 = \frac{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}{n(n-1)}$$

Per $n = 5$:

$$\sum_{i=1}^{5} x_i = 1.9 + 2.4 + 3.0 + 3.5 + 4.2 = 15$$

e

$$\sum_{i=1}^{5} x_i^2 = 1.9^2 + 2.4^2 + 3.0^2 + 3.5^2 + 4.2^2 = 48.26$$

Pertanto

$$\begin{aligned} s^2 &= \frac{n\sum_{i=1}^{5} x_i^2 - (\sum_{i=1}^{5} x_i)^2}{n(n-1)} \\ &= \frac{5 \cdot 48.26 - (15)^2}{5(5-1)} \\ &= \frac{16.3}{20} \\ &= 0.815 \end{aligned}$$

Determiniamo $\alpha/2$. Da $95\% = 100(1-\alpha)\%$, si ha:

$$\begin{aligned} (100)(1-\alpha) &= 95 \\ 1-\alpha &= 0.95 \\ \alpha &= 0.05 \\ \alpha/2 &= 0.025 \end{aligned}$$

I g.d.l. sono $\nu = n - 1 = 4$. Usando la tabella della distribuzione Chi-quadro, il valore critico per 0.025 con 4 g.d.l. é $\chi^2_{0.025} = 11.143$, e $\chi^2_{1-0.025} = \chi^2_{0.975} = 0.484$.
L'intervallo di interesse é

$$\begin{aligned} \frac{(n-1)s^2}{\chi^2_{\alpha/2}} &< \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}} \\ \frac{(5-1)\cdot 0.815}{\chi^2_{0.025}} &< \sigma^2 < \frac{(5-1)\cdot 0.815}{\chi^2_{0.975}} \\ \frac{4\cdot 0.815}{11.143} &< \sigma^2 < \frac{4\cdot 0.815}{0.484} \\ 0.29 &< \sigma^2 < 6.74 \end{aligned}$$

Poiché 1 cade nell'intervallo di confidenza al 95% si puó concludere che quanto dichiarato dal produttore é plausibile statisticamente.

Esercizio 6. Un costruttore sta considerando l'acquisto di speciali barre metalliche da due diversi fornitori. Un campione di 12 barre di lunghezza dichiarata pari a $127mm$ viene acquistato da ciascuno dei due fornitori e poi misurato. La deviazione standard della lunghezza delle barre del primo fornitore risulta essere $s_1 = 0.13mm$, mentre quella delle barre del secondo fornitore é di $s_2 = 0.17mm$.

(a) Questi dati indicano che la lunghezza di una barra del primo fornitore é soggetta a maggior variabilitá rispetto a quella del secondo fornitore? (Si assuma normalitá e un livello di significativitá 0.05)

Soluzione:
I dati del problema ci dicono che: $n_1 = 12$, $n_2 = 12$, $s_1^2 = (0.13)^2$, $s_2^2 = (0.17)^2$.
I gdl sono $\nu_1 = \nu_2 = 11$
Considerando la statistica $F = \frac{s_1^2}{s_2^2} = 0.585$ (nell'ipotesi $\frac{\sigma_1^2}{\sigma_2^2} = 1$, notiamo immediatamente che é largamente inferiore al valore critico $f_{0.025}(11, 11) = 3.53$ valore oltre il quale la probabilitá che i campioni siano stati generati da distribuzioni con $\sigma_1 = \sigma_2$ (ipotesi nulla) sarebbe molto bassa.
Verifichiamo con il calcolo dell'intervallo di confidenza (IC)
L'IC per il rapporto delle due varianze, con $\alpha = 0.05$ é:

$$\frac{s_1^2}{s_2^2}\frac{1}{f_{0.025}(\nu_1,\nu_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} f_{0.025}(\nu_2,\nu_1)$$

dove $f_{0.025}(\nu_2, \nu_1) = f_{0.025}(\nu_1, \nu_2) = f_{0.025}(11, 11) = 3.53$
Sostituendo:

$$0.16565 < \frac{\sigma_1^2}{\sigma_2^2} < 2.0642$$

L'intervallo ammette la possibilitá che $\frac{\sigma_1}{\sigma_2} = 1$, dunque i dati indicano che statisticamente non si puó rigettare l'ipotesi che le lunghezze delle barre dei due fornitori abbiano la stessa variabilitá.

Compito scritto dell'esame di Statistica e analisi dei dati 01 settembre 2022	Prof. Giuseppe Boccignone	Corso di Laurea
Cognome:	**Nome:**	**Matricola:**

Istruzioni

- Il tempo riservato alla prova scritta e' di 2 ore. Durante la prova e' possibile consultare il formulario ed utilizzare la calcolatrice. Non e' possibile consultare libri e appunti.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede piu' passaggi di calcolo, non sara' preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o piu' frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sara' eventualmente rappresentato in quest'ultima forma).

Problemi

Esercizio 1.

Un cappello contiene tre carte: una carta é nera su entrambi i lati; una carta é bianca su entrambi i lati; una carta é nera su un lato e bianca sull'altro. Le carte vengono mescolate nel cappello, poi una viene estratta a caso e appoggiata su di un tavolo.

(a) Se il lato visibile della carta é nero, qual é la probabilitá che l'altro lato sia bianco?

Soluzione: Uno studente potrebbe essere tentato di risolvere in pochi secondi il problema ragionando come segue: la carta di interesse deve essere o la nero / nero o la nero / bianco: queste hanno pari probabilitá, dunque la probabilitá che, osservato il nero, l'altro lato sia bianco é $\frac{1}{2}$.
Lo studente, successivamente, per verificare la sua conclusione, effettua una simulazione del gioco delle carte, ma ottiene il risultato mostrato in Figura 1.

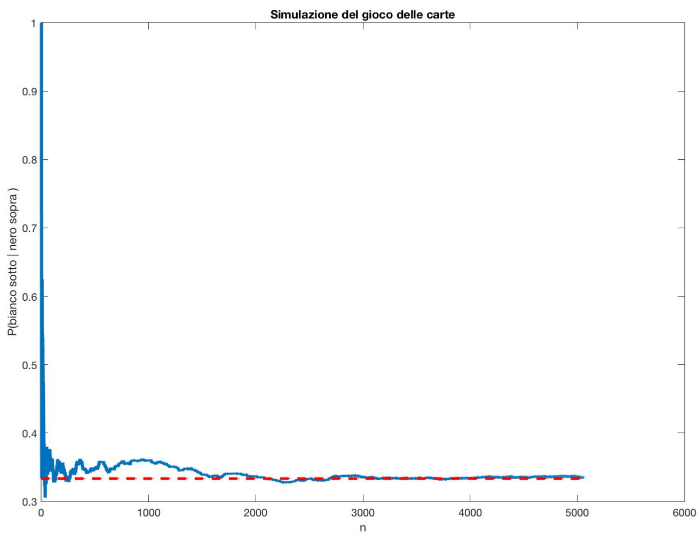

Figure 1: Simulazione del gioco delle tre carte: il valore di P(“bianco sotto” | “nero sopra”) valutato in termini di frequenza relativa, per un numero di prove n crescente, tende al valore di $0.333 \cdots \approx \frac{1}{3}$

.

Il valore dell'approssimazione frequentistica di P(“bianco sotto” | “nero sopra”) tende a $\frac{1}{3}$.
Per capire bene come si arriva a questo risultato, mettiamo da parte l'intuizione e definiamo con precisione lo spazio campionario e gli eventi di interesse.
Identifichiamo i lati delle carte

- N_1 e N_2 per la carta nero / nero

- B_1 e B_2 per la carta bianco / bianco
- N_3 e B_3 per la carta nero / bianco

In questo caso lo spazio campionario S é

$$S = \{N_1, N_2, N_3, B_1, B_2, B_3\}$$

L'evento $N =$ "nero sopra" é definibile come $N = \{N_1, N_2, N_3\}$.
L'evento $B =$ "bianco sotto" é definibile come $B = \{N_3, B_1, B_2\}$
Pertanto:

$$P(B \mid N) = \frac{P(B \cap N)}{P(N)}$$

Poiché $\#(\text{"bianco sotto"} \cap \text{"nero sopra"}) = 1$ e $\#(\text{"nero sopra"}) = 3$:

$$P(B \mid N) = \frac{1}{3}.$$

Esercizio 2.

Un robot autonomo controlla il livello di corrosione interna delle tubature dei sistemi di raffreddamento di una centrale nucleare. La corrosione non puó essere osservata direttamente, ma il robot effettua un test che puó fornisce indizi sulla specifica corrosione. Il test non é infallibile: nel 70% dei casi individua una corrosione quando é presente, ma puó anche sbagliare indicando una corrosione inesistente nel 20% dei casi. Risolvere il problema del robot a cui un operatore remoto, sotto l'ipotesi che il 10% delle tubature siano corrose, chiede di:

(a) Comunicargli la probabilitá che una parte delle tubature abbia delle corrosioni interne dopo che il test ha dato esito positivo (presenza di corrosioni). Determinare tale probabilitá.

Soluzione: Definiamo gli eventi $C =$ "la tubatura é corrosa" e $\overline{C} =$ "il test ha identificato la tubatura come corrosa". I dati del problema ci dicono che, a priori,

$$P(C) = 0.1$$

e dunque $P(\sim C) = 1 - 0.1 = 0.9$, dove $\sim C =$ "la tubatura non é corrosa". Inoltre

$$P(\overline{C} \mid C) = 0.7$$
$$P(\overline{C} \mid\sim C) = 0.2$$

Applicando la regola di Bayes:

$$P(C \mid \overline{C}) = \frac{P(\overline{C} \mid C)P(C)}{P(\overline{C})}$$

dove $P(\overline{C}) = P(\overline{C} \mid C)P(C) + P(\overline{C} \mid\sim C)P(\sim C) = 0.25$. Quindi

$$P(C \mid \overline{C}) = \frac{0.7 \times 0.1}{0.25} = 0.28$$

(b) Comunicargli la probabilitá che una parte delle tubature abbia delle corrosioni dopo che il test ha dato esito negativo (nessuna corrosione). Determinare tale probabilitá.

Soluzione: Si ripeta il procedimento precedente per il caso $P(C \mid\sim \overline{C})$

Esercizio 3.

Un sistema di comunicazione consiste di un buffer che ospita i pacchetti provenienti dalla sorgente, e da una linea di comunicazione che recupera i pacchetti dal buffer e li trasmette a un ricevitore. Il sistema opera a coppie di *time-slot*. Nel primo slot il sistema depone nel buffer un certo numero di pacchetti generati dalla sorgente secondo una distribuzione di Poisson di parametro μ. La capacitá del buffer (numero massimo di pacchetti) é pari b_{max}; i pacchetti che arrivano a buffer pieno vengono scartati. Nel secondo slot temporale il sistema trasmette c pacchetti al ricevitore, dove c é un intero $0 < c < b_{max}$, oppure trasmette quelli effettivamente presenti se in numero inferiore a c.

(a) Assumendo che all'inizio del primo slot il buffer sia vuoto, trovare la PMF del numero di pacchetti bufferizzati alla fine del primo slot e la PMF alla fine del secondo slot.

Soluzione: Sia X la V.A. discreta che indica il numero di pacchetti bufferizzati alla fine del primo slot temporale Δt_1. La Figura 2 mostra il suo intervalli di variazione relativamente ai parametri del problema c, b_{max} Per determinare la PMF di X, P_X é sufficiente determinare: 1) la probabilitá di avere bufferizzato $k = 0,1,2,\cdots,b_{max}-1$ pacchetti; 2) la probabilitá di avere il buffer pieno che accade quando in Δt_1 vengono generati alla sorgente $k \geq b_{max}$ pacchetti.

1. $k = 0,1,2,\cdots, b_{max}-1$:
   La probabilitá di avere esattamente k pacchetti bufferizzati $P_X(X=k)$ é uguale alla probabilitá che k pacchetti siano stati generati alla sorgente durante Δt_1: ció implica, per quanto ci dice la specifica del problema, che X segua una legge di Poisson, $X \sim Pois(\mu)$ e dunque:
   $$P_X(X=k) = \frac{\mu^k}{k!}e^{-\mu} \tag{1}$$
   applicabile per $k = 0,1,2,\cdots, b_{max}-1$.
2. $k \geq b_{max}$:
   La probabilitá di avere il buffer pieno, ovvero che $X = b_{max}$ é la probabilitá che alla sorgente siano stati generati almeno b_{max} pacchetti, ovvero
   $$P_X(X=b_{max}) = \sum_{k=b_{max}}^{\infty} \frac{\mu^k}{k!}e^{-\mu} = 1 - \sum_{k=0}^{b_{max}-1} \frac{\mu^k}{k!}e^{-\mu}$$

Si noti come $P_X(X=0) + P_X(X=1) + \cdots + P_X(X=b_{max}-1) + P_X(X=b_{max}) = 1$, dunque P_X é una PMF correttamente normalizzata.
Chiamiamo ora Y la V.A. discreta che indica il numero di pacchetti bufferizzati alla fine del secondo slot temporale Δt_2. Il suo valore é determinabile come

$$Y = \text{ pacchetti bufferizzati durante } \Delta t_1 - \text{ pacchetti trasmessi durante } \Delta t_2 = X - min\{X,c\}.$$

Infatti se $X \leq c$, allora $min\{X,c\} = X$ e $Y = X - X = 0$. Se $X > c$, $min\{X,c\} = c$, quindi $Y = X - c$.
Per determinare la PMF di Y consideriamo i seguenti casi

1. $Y = 0$: la probabilitá che il buffer sia stato svuotato é uguale alla probabilitá che vi fossero $X \leq c$ pacchetti bufferizzati in Δt_1:
   $$P_Y(Y=0) = P_X(X \leq c) = \sum_{k=0}^{c} P_X(X=k) = \sum_{k=0}^{c} \frac{\mu^k}{k!}e^{-\mu}.$$
2. $0 < Y = k < b_{max} - c$:
   La probabilitá che al termine di Δt_2 vi siano esattamente $Y = k$ pacchetti con $k = 1,2,\cdots, b_{max} - c - 1$ é uguale alla probabilitá che in Δt_1 siano stati bufferizzati $X = k + c$ pacchetti, con $k + c < b_{max}$
   $$P_Y(Y=k) = P_X(X=k+c) = \frac{\mu^{k+c}}{(k+c)!}e^{-\mu}$$
3. $Y = b_{max} - c$:
   E' il caso in cui in Δt_1 il buffer śtato riempito con $X = b_{max}$ pacchetti di cui c sono stati inviati in Δt_2:
   $$P_Y(Y=b_{max}-c) = P_X(X=b_{max}) = 1 - \sum_{k=0}^{b_{max}-1} \frac{\mu^k}{k!}e^{-\mu}$$

(b) Qual é la probabilitá che alcuni pacchetti vengano scartati durante il primo slot?
Soluzione: La probabilitá che alcuni pacchetti vengano scartati durante il primo slot é uguale alla probabilitá che piú di b_{max} pacchetti siano stati generati alla sorgente in Δt_1 ovvero

$$P_X(X > b_{max}) = \sum_{k=b_{max}+1}^{\infty} \frac{\mu^k}{k!}e^{-\mu} = 1 - \sum_{k=0}^{b_{max}} \frac{\mu^k}{k!}e^{-\mu}$$

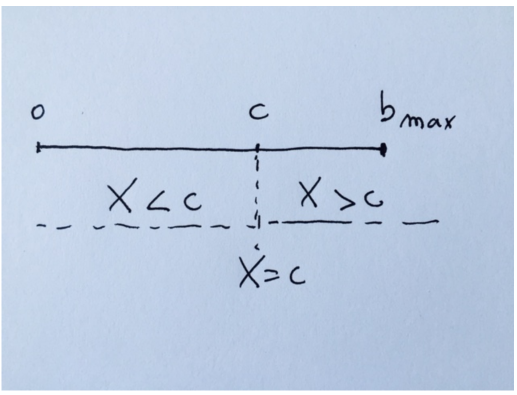

Figure 2: Valori che puó assumere la variabile aleatoria X nel problema del buffer

Esercizio 4.
Una linea di produzione fabbrica resistori da 1000 ohm (Ω). La fabbrica adotta una tolleranza del 10% (ovvero sono scartati i resistori con resistenza $R < 900\ \Omega$ e $R > 1100\ \Omega$).

(a) Sotto l'ipotesi che la resistenza R di un resistore sia una variabile aleatoria distribuita con legge normale di media $1000\ \Omega$ e varianza 2500, si calcoli la probabilitá che un resistore preso a caso venga scartato

Soluzione: Sia S l'evento in cui un resistore viene scartato. Allora:

$$S = \{R < 900\} \cup \{R > 1100\}$$

Poiché $\{R < 900\} \cap \{R > 1100\} = \emptyset$, si ha che

$$P(S) = P(\{R < 900\}) \ + P(\{R > 1100\}) = F_X(900) + (1 - F_X(1100)) = \Phi\left(\frac{900-\mu}{\sigma}\right) + \left(1 - \Phi\left(\frac{1100-\mu}{\sigma}\right)\right)$$

Poiché $\mu = 1000$, e $\sigma^2 = 2500 \implies \sigma = 50$, le variabili normalizzate sono

$$\frac{900-1000}{50} = -2$$

$$\frac{1100-1000}{50} = 2$$

.
Sfruttando la proprietá della cumulativa standardizzata $\Phi(-z) = 1 - \Phi(z)$, la precedente diventa:

$$P(S) = 2\,(1 - \Phi(2))$$

.
Accedendo alla tabella della Normale standard per $z = 2$, $\Phi(2) = 0.9772$ da cui $P(S) = 0.0455$

Esercizio 5.
Vengono sottoposti a confronto i consumi delle autovetture A e B alla velocitá costante di $120 Km/h$. Si ritiene che i consumi dei due tipi di autovetture possa essere descritto da variabili aleatorie con distribuzione normale con la stessa varianza. L'auto di tipo A in 20 prove consuma mediamente $6.5 litri/100Km$, quella di tipo B in 22 prove consuma mediamente $6.6 litri/100Km$. Le relative varianze campionarie sono rispettivamente di 0.30 e 0.28.

(a) Possiamo ritenere che le due autovetture abbiano lo stesso consumo medio al livello di significativitá del 5%?

Soluzione:
Abbiamo che:

- $\bar{x}_A = 6.5, s_A^2 = 0.30, n_A = 20$
- $\bar{x}_B = 6.6, s_B^2 = 0.28, n_B = 22$

Siamo nel caso di varianze ignote ma eguali, e la differenza fra medie é valutabile mediante una statistica t-Student, dove i gradi di libertá ν sono dati da $\nu = n_A + n_B - 2 = 40$, $\alpha = 0.05$, per cui, $t_{\nu,\frac{\alpha}{2}} = t_{40,0.025} \approx 2.021$
La deviazione standard pooled vale:

$$S_{pool} = \sqrt{\frac{(n_A-1)s_A^2 + (n_B-1)s_B^2}{n_A + n_B - 2}} = 0.5381$$

L'intervallo di confidenza per la differenza fra le medie $\mu_A - \mu_B$ a livello $(100)(1-\alpha) = (100)(1-0.05) = 95\%$ vale pertanto

$$\bar{x}_A - \bar{x}_B \pm t_{40,0.025} S_{pool} \sqrt{\frac{1}{n_A} + \frac{1}{n_B}} = -0.1 \pm 0.33586$$

ovvero l'IC ha estremi $[-0.43596; 0.23596]$ per cui si puó ragionevolmente ritenere che le due autovetture abbiano differenza in consumo medio trascurabile.
Equivalentemente, si osserva che la statistica della differenza fra medie normalizzata

$$\left| \frac{\bar{x}_A - \bar{x}_B}{S_{pool}\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \right| = 0.612 < t_{40,0.025} = 2.021$$

Esercizio 6. Un costruttore sta considerando l'acquisto di speciali barre metalliche da due diversi fornitori. Un campione di 12 barre di lunghezza dichiarata pari a $127 mm$ viene acquistato da ciascuno dei due fornitori e poi misurato. La deviazione standard della lunghezza delle barre del primo fornitore risulta essere $s_1 = 0.13 mm$, mentre quella delle barre del secondo fornitore é di $s_2 = 0.17 mm$.

(a) Questi dati indicano che la lunghezza di una barra del primo fornitore é soggetta a maggior variabilitá rispetto a quella del secondo fornitore? (Si assuma normalitá e un livello di significativitá 0.05)

Soluzione:
I dati del problema ci dicono che: $n_1 = 12$, $n_2 = 12$, $s_1^2 = (0.13)^2$, $s_2^2 = (0.17)^2$.
I gdl sono $\nu_1 = \nu_2 = 11$
Considerando la statistica $F = \frac{s_1^2}{s_2^2} = 0.585$ (nell'ipotesi $\frac{\sigma_1^2}{\sigma_2^2} = 1$, notiamo immediatamente che é largamente inferiore al valore critico $f_{0.025}(11, 11) = 3.53$ valore oltre il quale la probabilitá che i campioni siano stati generati da distribuzioni con $\sigma_1 = \sigma_2$ (ipotesi nulla) sarebbe molto bassa.
Verifichiamo con il calcolo dell'intervallo di confidenza (IC)
L'IC per il rapporto delle due varianze, con $\alpha = 0.05$ é:

$$\frac{s_1^2}{s_2^2} \frac{1}{f_{0.025}(\nu_1, \nu_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} f_{0.025}(\nu_2, \nu_1)$$

dove $f_{0.025}(\nu_2, \nu_1) = f_{0.025}(\nu_1, \nu_2) = f_{0.025}(11, 11) = 3.53$
Sostituendo:

$$0.16565 < \frac{\sigma_1^2}{\sigma_2^2} < 2.0642$$

L'intervallo ammette la possibilitá che $\frac{\sigma_1}{\sigma_2} = 1$, dunque i dati indicano che statisticamente non si puó rigettare l'ipotesi che le lunghezze delle barre dei due fornitori abbiano la stessa variabilitá.

Compito scritto dell'esame di Statistica e analisi dei dati 12 luglio 2022	Prof. Giuseppe Boccignone	Corso di Laurea
Cognome:	Nome:	Matricola:

Istruzioni

- Il tempo riservato alla prova scritta e' di 2 ore. Durante la prova e' possibile consultare il formulario ed utilizzare la calcolatrice. Non e' possibile consultare libri e appunti.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede piu' passaggi di calcolo, non sara' preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o piu' frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sara' eventualmente rappresentato in quest'ultima forma).

Problemi

Esercizio 1.

Un sistema é formato da tre componenti disposti in parallelo. Il sistema funziona se almeno due dei tre componenti funzionano correttamente. Indicati con C_1, C_2, C_3 gli eventi "funzionamento corretto" del primo, secondo e terzo componente, rispettivamente, siano $P(C_1) = P(C_2) = P(C_3) = p$ con $p = 0.7$. Si assuma che gli eventi C_1, C_2, C_3 siano indipendenti tra di loro.

(a) Si mostri che la probabilitá che il sistema funzioni correttamente é maggiore della probabilitá di funzionamento dei singoli componenti.

Soluzione:
Definiamo l'evento $C =$ `"il sistema funziona correttamente"`. Il problema si risolve calcolando la probabilità $P(C)$ e mostrando che $P(C) > p$.
Il sistema funziona se almeno due componenti funzionano correttamente. I casi possibili che si possono presentare sono 8, di questi, quelli in cui almeno due generatori sono attivi sono 4: il caso in cui funzionano tutti e tre, e i tre casi in cui si ha il malfunzionamente di uno solo. Dunque

$$C = \{C_1 \cap C_2 \cap C_3\} \cup \{\sim C_1 \cap C_2 \cap C_3\} \cup \{C_1 \cap \sim C_2 \cap C_3\} \cup \{C_1 \cap C_2 \cap \sim C_3\}.$$

Applicando il Principio di additivitá generalizzata, la probabilitá di corretto funzionamento del sistema é

$$P(C) = P(\{C_1 \cap C_2 \cap C_3\}) + P(\{\sim C_1 \cap C_2 \cap C_3\}) + P(\{C_1 \cap \sim C_2 \cap C_3\}) + P(\{C_1 \cap C_2 \cap \sim C_3\})$$

Infine, usando l'indipendenza degli eventi:

$$P(T) = p^3 + 3p^2(1-p) = 3p^2 - 2p^3 = 1.47 - 0.686 = 0.784 > p$$

Esercizio 2.

In un ufficio le pratiche relative ad una certa procedura amministrativa vengono affidate casualmente a tre impiegati che indicheremo con A, B e C. Statisticamente, la percentuale dei casi in cui la pratica viene completata entro una settimana da ciascun impiegato é riportata nella seguente tabella:

Impiegato	A	B	C
Perc.	40%	80%	30%

(a) Avendo ricevuto una pratica espletata entro una settimana, qual é secondo voi l'impiegato a cui era stata affidata e qual é la probabilitá della vostra conclusione?

Soluzione: Definiamo i seguenti eventi

- $S =$ `"pratica completata entro una settimana"`
- $A =$ `"pratica affidata all'impiegato A"`
- $B =$ `"pratica affidata all'impiegato B"`
- $C =$ `"pratica affidata all'impiegato C"`

L'affidamento pratiche é casuale dunque:

$$P(A) = P(B) = P(C) = \frac{1}{3}$$

Dalla tabella ricaviamo che:

$$P(S \mid A) = 0.4, P(S \mid B) = 0.8, P(S \mid C) = 0.3$$

Per rispondere al quesito si tratta di stabilire qual é la maggiore fra le probabilitá a posteriori

$$P(A \mid S) = \frac{P(S \mid A)P(A)}{P(S)}, P(B \mid S) = \frac{P(S \mid B)P(B)}{P(S)}, P(C \mid S) = \frac{P(S \mid C)P(C)}{P(S)}$$

Il denominatore, cioé la probabilitá che una pratica venga completata entro una settimana indipendentemente da quale impegato l'ha seguita é:

$$P(S) = P(S \mid A)P(A) + P(S \mid B)P(B) + P(S \mid C)P(C) = \frac{1}{3}(0.4 + 0.8 + 0.3) = \frac{1}{3} \times 1.5$$

Sostituendo tutti i dati:

$$P(A \mid S) = \frac{P(S \mid A)P(A)}{P(S)} = 0.267$$

$$P(B \mid S) = \frac{P(S \mid B)P(B)}{P(S)} = 0.533$$

$$P(C \mid S) = \frac{P(S \mid C)P(C)}{P(S)} = 0.2$$

Pertanto, sará stato l'impiegato B a seguire la pratica con probabilitá 0.533.

Esercizio 3. L'altezza di un uomo segue una distribuzione normale di media 178 cm con una deviazione dalla media di 8 cm. L'altezza di una donna ha invece una media di 165 cm con una deviazione pari a 7 cm. .

(a) Qual é, nella stessa popolazione, la proporzione di uomini la cui altezza é superiore a 185 cm?

Soluzione: Sia U la variabile aleatoria che denota l'altezza degli uomini.
Vogliamo determinare $P(U > 185)$, dove la distribuziione é Normale con $\mu = 178, \sigma = 8$. Standardizzando le variabili e usando le tabelle della distribuzione normale

$$P(U > 185) = P(\frac{U - 178}{8} > \frac{185 - 178}{8}) = P(Z < 0.875) \approx 0.19$$

(b) Qual é la proporzione di donne che sono piú alte della metá degli uomini?

Soluzione: Sia D la variabile aleatoria che denota l'altezza delle donne.
Sappiamo che media e mediana nella distribuzione normale coincidono dunque

$$P(U > u) = 0.5$$

quando $u = \mu = 178$
Pertanto, si tratta di determinare $P(D > 178)$:

$$P(D > 178) = P(Z > \frac{178 - 165}{7}) = P(Z > 1.857) \approx 0.032$$

.
Solo il 3.2% delle donne é piú alto di metá degli uomini.

Esercizio 4.
Un vostro amico asserisce di aver ottenuto una media di 3.25 punti per lancio su 1000 lanci di un dado non truccato.

(a) Qual é la probabilitá che stia mentendo?

Soluzione: (Teorema limite centrale)
Sia X_i la variabile aleatoria (v.a.) che denota l'esito dell'i-simo lancio di dado.
Sappiamo che le v.a. $X_1,X_2,\cdots,X_{1000}$ sono indipendenti e identicamente distribuite con distribuzione uniforme sull'insieme di interi $\{1,2,\cdots,6\}$ da cui é facile ricavare il valore atteso e la deviazione standard

$$E[X_i] = 3.5$$
$$\sigma_{X_i} = \sqrt{E[X_i^2] - E[X_i]^2} \approx 1.7078$$

In breve, il Teorema del limite centrale ci dice che per un numero n grande ($n \to \infty$) di V.A. indipendenti e identicamente distribuite di valore atteso μ e deviazione standard σ vale

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x\right) = \Phi(x) \tag{1}$$

dove $S_n = \sum_{i=1}^{n} X_i$. Nella fattispecie $n = 1000$ e S_n é il numero di punti totalizzati in 1000 lanci.
Sotto tali ipotesi, la probabilitá che il vostro amico affermi il vero si riduce a calcolare la probabilitá definita nel T.L.C riscritta come

$$P\left(\frac{\frac{S_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} \le x\right) = \Phi(x) \tag{2}$$

dove $\frac{S_n}{n} = 3.25 = \overline{X}$ é la media campionaria del campione di taglia $n = 1000$ dichiarata dal vostro amico. In buona sostanza il T.L.C. il teorema garantisce che la variabile ridotta $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ sia una variabile aleatoria la cui funzione di distribuzione tende alla distribuzione normale standard Φ. Pertanto,

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{3.25 - 3.5}{1.7078/\sqrt{1000}} = -4.629$$

La probabilitá che il vostro amico dichiari il vero é dunque pari a $\Phi(-4.629)$
Anche senza consultare le tabelle e tenendo presente che per la normale standard $\sigma = 1$ e dunque $z = z\sigma$, significa che stiamo valutando la probabilitá di un valore di media dichiarata che é di 4.629 deviazioni standard al di sotto della media $\mu = 3.5$.
Considerando la legge 3σ, é immediatamente evidente che la probabilitá $\Phi(-4.629)$ é molto bassa e pertanto la probabilitá che che il vostro amico dica il falso molto elevata $(1 - \Phi(-4.629))$.
Se poi, per scrupolo, si consultasse la tabella della normale ridotta, si leggerebbe che $\Phi(-4.629) = 1.84 \times 10^{-7}$

Esercizio 5.
La quantitá di stoffa per produrre poltrone segue una legge normale. Su un campione casuale di 15 poltrone, si é riscontrato che l'ammontare medio del materiale é $912cm^2$, con una deviazione standard di $64cm^2$.

(a) Qual é l'intervallo di confidenza al 99% per la media della quantitá di materiale?

Soluzione:
Sappiamo che $n = 15, \bar{x} = 912cm^2, s = 64cm^2$. Il campione di taglia $n = 15$ é piccolo e l'intervallo é basato sulla t di Student con $\nu = n - 1 = 14$ gradi di libertá.
Con $\alpha = 0.01$, si ricava dalla tabella della t che il quantile é $t_\alpha = t_{.005} = 2.977$.
L'errore standard é $SE = \frac{\bar{x}}{\sqrt{n}} = \frac{912}{\sqrt{15}} = 16.52473$
L' intervallo di confidenza al 99% per la media é dunque

$$\bar{x} - 2.977 \cdot SE < \mu < \bar{x} + 2.977 \cdot SE,$$

ovvero

$$\mu = 912 \pm 49.2cm^2$$

Esercizio 6. Il termostato di un condizionatore d'aria é supposto essere tarato sul valore di soglia $\mu = 25°C$ (cioé l'apparecchio entra in funzione quando la temperatura dell'ambiente supera i $25°C$). Un tecnico misura in 8 occasioni diverse la temperatura X alla quale il termostato scatta, ottenendo i valori seguenti ($°C$):

24.6 24.8 25.2 25.4 25.5 24.0 24.7 25.3

(a) Formulando un adeguato modello per X, stabilire se, al 95% di confidenza, il valore di μ sia plausibile statisticamente

Soluzione:
Assumendo una distribuzione normale della temperatura, $X \sim \mathcal{N}(\mu, \sigma^2)$, poiché la varianza σ^2 é incognita, la stima intervallare di μ al 95% di confidenza si ottiene considerando

$$P\left(\bar{x} - t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right) = 1 - \alpha = 0.95$$

Poiché al 95% di confidenza $\alpha = 0.05$, allora $t_{\frac{\alpha}{2}} = t_{0.025}$, da determinarsi dalla distribuzione di Student con $\nu = n - 1 = 7$ gdl:

$$t_{0.025} = 2.365$$

Dai dati conosciuti si ottengono media e deviazione standard campionaria:

$$\bar{x} \approx 24.93$$

$$s \approx 0.83$$

Pertanto

$$24.93 - 2.365 \times \frac{0.83}{2.64} < \mu < 24.93 + 2.365 \times \frac{0.83}{2.64}$$

ovvero $24.2 < \mu < 25.66$ e dunque il valore $\mu = 25°C$ puó essere ritenuto plausibile.

(b) Stabilire quanto vale al massimo la precisione p (ovvero il reciproco della deviazione standard) del termostato al 95% di confidenza

Soluzione:
Assumendo $\mu = 25°C$ per l'analisi precedente, $X \sim \mathcal{N}(25, \sigma^2)$.
Sappiamo che $p = \frac{1}{\sigma}$, dunque il valore di precisione massima é quello di deviazione standard minima, $p_{max} = \frac{1}{\sigma_{min}}$.
Il problema richiede di determinare

$$P(p < p_{max}) = 1 - \alpha = 0.95$$

Dalla definizione di p questo equivale alla stima unilaterale sul limite inferiore della varianza (σ^2_{min}) :

$$P(p < p_{max}) = P\left(\sqrt{\frac{(n-1)s^2}{\chi^2_\alpha}} < \sigma\right) = 1 - \alpha$$

Ne deriva che al 95% di confidenza il limite superiore di precisione vale $p_{max} = \sqrt{\frac{\chi^2_{0.05}}{(n-1)s^2}}$, dunque, leggendo dalla tabella della distribuzione Chi-quadro, per $\nu = 7$, $\chi^2_{0.05}(7) = 14.067$ e usando s^2 calcolato precedentemente.

$$p < \sqrt{\frac{\chi^2_{0.05}}{(n-1)s^2}} \approx 1.69°C$$

Compito scritto dell'esame di Statistica e analisi dei dati **14 giugno 2022**	**Prof. Giuseppe Boccignone**	**Corso di Laurea**
Cognome:	**Nome:**	**Matricola:**

Istruzioni

- Il tempo riservato alla prova scritta e' di 2 ore. Durante la prova e' possibile consultare il formulario ed utilizzare la calcolatrice. Non e' possibile consultare libri e appunti.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede piu' passaggi di calcolo, non sara' preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o piu' frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sara' eventualmente rappresentato in quest'ultima forma).

Problemi

Esercizio 1. Di due eventi A e B si sa che $P(A \cup B) = \alpha$ e che $P(A \mid B) = P(B \mid A) = p > 0$.

(a) Calcolare $P(A), P(B), P(A \cap B)$

Soluzione: Si ha che

$$P(A \cap B) = P(B \mid A)P(A) = P(A \mid B)P(B) \implies pP(A) = pP(B) \implies P(A) = P(B)$$

Dunque

$$\alpha = P(A \cup B) = P(A) + P(B) - P(A \cap B) = 2P(A) - P(A \cap B) = 2P(A) - P(B \mid A)P(A) = (2-p)P(A)$$

$$\implies P(A) = \frac{\alpha}{2-p} = P(B)$$

Infine

$$P(A \cap B) = P(B \mid A)P(A) = \frac{\alpha p}{2-p}$$

Esercizio 2. Una variabile aleatoria discreta X puó assumere solo i valori $x = 1, 2, 3$. Inoltre, si ha che $E[X] = 1.5, var(X) = 0.5$.

(a) Calcolare la distribuzione di probabilitá di X.

Soluzione:
La la distribuzione di probabilitá discreta é

$$P(X = 1) = p_1, P(X = 2) = p_2, P(X = 3) = p_3$$

con condizione di normalizzazione

$$p_1 + p_2 + p_3 = 1 \tag{1}$$

Sappiamo che:

$$E[X] = \sum_{x=1}^{3} xP(X = x) = 1p_1 + 2p_2 + 3p_3 = 1.5 \tag{2}$$

Inoltre $var(X) = E\left[X^2\right] - (E[X])^2 = E\left[X^2\right] - (1.5)^2 = 0.5$, con $E\left[X^2\right] = \sum_{x=1}^{3} x^2 P(X = x) = 1p_1 + 4p_2 + 9p_3$, dunque

$$1p_1 + 4p_2 + 9p_3 - (1.5)^2 = 0.5 \tag{3}$$

Risolvendo il sistema lineare di equazioni 1, 2, 3, si ottiene

$$p_1 = 0.625, p_2 = 0.250, p_3 = 0.125$$

Esercizio 3.
Sparando a un bersaglio ho il 20% di probabilitá di fare centro.

(a) Se sparo due volte, che probabilitá ho di fare almeno un centro?

Soluzione: (Distribuzione binomiale) Il numero X_n di centri fatti con n colpi segue una legge binomiale, $X_n \sim Bin(X_n; n, 0.2)$.
Quindi

$$P(X_2 \geq 1) = 1 - P(X_2 = 0) = 1 - Bin(0; 2, 0.2) = 1 - (0.8)^2 = 0.36$$

(b) Che numero minimo di colpi devo sparare per aver piú del 50% di probabilitá di fare almeno un centro?

Soluzione: Si deve determinare

$$P(X_n \geq 1) > 0.5$$

Procedendo come prima

$$P(X_n \geq 1) = 1 - P(X_n = 0) = 1 - Bin(0; n, 0.2) = 1 - (0.8)^n$$

dunque

$$1 - (0.8)^n > 5 \implies (0.8)^n < 0.5 \implies n > \frac{\ln 0.5}{\ln 0.8} \approx 3.1$$

Esercizio 4. Un transistor ha una vita media nominale di 1000 giorni.

(a) Se in un test di durata il transistore ha giá funzionato ininterrottamente per 500 giorni, qual é la probabilit'a che si guasti nei successivi 200 giorni?

Soluzione: Assumendo che i guasti seguano un processo poissoniano, la durata T (in giorni) segue una legge esponenziale $Exp(T; \lambda)$ di parametro

$$\lambda = \frac{1}{10^3};$$

Per l'assenza di memoria che caratterizza tale distribuzione, la probabilitá che un transistor vecchio di 500 giorni duri ancora al massimo 200 giorni é uguale alla probabilitá che un transistor nuovo duri al massimo 200 giorni, Pertanto

$$P(T \leq 200) = 1 - e^{-\frac{200}{10^3}} \approx 18.13\%$$

(b) Se 5 transistor sono sottoposti al test di durata, qual é la probabilitá che almeno uno di essi funzioni piú di 2000 giorni? (Si assuma che i transistor si guastino indipendentemente.)

Soluzione: Sia p la probabilit'a che un transistore funzioni al massimo 2000 giorni

$$p = P(T \leq 200) = 1 - e^{-\frac{2000}{10^3}} = 1 - e^{-2}$$

Per indipendenza, la probabilitá che tutti i 5 transistor funzionino al massimo 2000 giorni é uguale a p^5, e quindi la probabilitá dell'evento contrario (almeno un transistor funziona piú di 2000 giorni), vale

$$1 - p^5 \approx 51.67\%$$

Esercizio 5. Un costruttore sta considerando l'acquisto di speciali barre metalliche da due diversi fornitori. Un campione di 12 barre di lunghezza dichiarata pari a $127mm$ viene acquistato da ciascuno dei due fornitori e poi misurato. La deviazione standard della lunghezza delle barre del primo fornitore risulta essere $s_1 = 0.13mm$, mentre quella delle barre del secondo fornitore é di $s_2 = 0.17mm$.

(a) Questi dati indicano che la lunghezza di una barra del primo fornitore é soggetta a maggior variabilitá rispetto a quella del secondo fornitore? (Si assuma normalitá e un livello di significativitá 0.05)

Soluzione:
I dati del problema ci dicono che: $n_1 = 12$, $n_2 = 12$, $s_1^2 = (0.13)^2$, $s_2^2 = (0.17)^2$.
I gdl sono $\nu_1 = \nu_2 = 11$
Considerando la statistica $F = \frac{s_1^2}{s_2^2} = 0.585$ (nell'ipotesi $\frac{\sigma_1^2}{\sigma_2^2} = 1$, notiamo immediatamente che é largamente inferiore al valore critico $f_{0.025}(11, 11) = 3.53$ valore oltre il quale la probabilitá che i campioni siano stati generati da distribuzioni con $\sigma_1 = \sigma_2$ (ipotesi nulla) sarebbe molto bassa.
Verifichiamo con il calcolo dell'intervallo di confidenza (IC)
L'IC per il rapporto delle due varianze, con $\alpha = 0.05$ é:

$$\frac{s_1^2}{s_2^2}\frac{1}{f_{0.025}(\nu_1, \nu_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} f_{0.025}(\nu_2, \nu_1)$$

dove $f_{0.025}(\nu_2, \nu_1) = f_{0.025}(\nu_1, \nu_2) = f_{0.025}(11, 11) = 3.53$
Sostituendo:

$$0.16565 < \frac{\sigma_1^2}{\sigma_2^2} < 2.0642$$

L'intervallo ammette la possibilitá che $\frac{\sigma_1}{\sigma_2} = 1$, dunque i dati indicano che statisticamente non si puó rigettare l'ipotesi che le lunghezze delle barre dei due fornitori abbiano la stessa variabilitá.

Esercizio 6.

Vengono sottoposti a confronto i consumi delle autovetture A e B alla velocitá costante di $120 Km/h$. Si ritiene che i consumi dei due tipi di autovetture possa essere descritto da variabili aleatorie con distribuzione normale con la stessa varianza. L'auto di tipo A in 20 prove consuma mediamente $6.5 litri/100Km$, quella di tipo B in 22 prove consuma mediamente $6.6 litri/100Km$. Le relative varianze campionarie sono rispettivamente di 0.30 e 0.28.

(a) Possiamo ritenere che le due autovetture abbiano lo stesso consumo medio al livello di significatività del 5%?

Soluzione:
Abbiamo che:

- $\bar{x}_A = 6.5, s_A^2 = 0.30, n_A = 20$
- $\bar{x}_B = 6.6, s_B^2 = 0.28, n_B = 22$

Siamo nel caso di varianze ignote ma eguali, e la differenza fra medie é valutabile mediante una statistica t-Student, dove i gradi di libertá ν sono dati da $\nu = n_A + n_B - 2 = 40$, $\alpha = 0.05$, per cui, $t_{\nu, \frac{\alpha}{2}} = t_{40, 0.025} \approx 2.021$
La deviazione standard pooled vale:

$$S_{pool} = \sqrt{\frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2}} = 0.5381$$

L'intervallo di confidenza per la differenza fra le medie $\mu_A - \mu_B$ a livello $(100)(1 - \alpha) = (100)(1 - 0.05) = 95\%$ vale pertanto

$$\bar{x}_A - \bar{x}_B \pm t_{40, 0.025} S_{pool} \sqrt{\frac{1}{n_A} + \frac{1}{n_B}} = -0.1 \pm 0.33586$$

ovvero l'IC ha estremi $[-0.43596; 0.23596]$ per cui si puó ragionevolmente ritenere che le due autovetture abbiano differenza in consumo medio trascurabile.
Equivalentemente, si osserva che la statistica della differenza fra medie normalizzata

$$\left| \frac{\bar{x}_A - \bar{x}_B}{S_{pool} \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \right| = 0.612 < t_{40, 0.025} = 2.021$$

Compito scritto dell'esame di Statistica e analisi dei dati 15 febbraio 2022	Prof. Giuseppe Boccignone	Corso di Laurea
Cognome:	Nome:	Matricola:

Istruzioni

- Il tempo riservato alla prova scritta e' di 2 ore. Durante la prova e' possibile consultare il formulario ed utilizzare la calcolatrice. Non e' possibile consultare libri e appunti.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede piu' passaggi di calcolo, non sara' preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o piu' frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sara' eventualmente rappresentato in quest'ultima forma).

Problemi

Esercizio 1. Un dado é truccato in modo che la probabilitá di ogni faccia sia proporzionale al suo punteggio. .

(a) Determinare se é maggiore la probabilitá che in un lancio esca un numero primo o un numero non primo

Soluzione: Si ha che

$$P(x) = kx, \quad x = 1, \cdots, 6.$$

Determiniamo k con la condizione di normalizzazione $\sum_{x=1}^{6} P(x) = 1$:

$$\sum_{x=1}^{6} kx = 1 \implies 21k = 1 \implies k = \frac{1}{21}$$

Definiamo l'evento

- $E =$ "esce numero primo" $= \{1,2,3,5\}$

Allora

$$P(E) = P(1) + P(2) + P(3) + P(5) = \frac{1+2+3+5}{21} = \frac{11}{21}$$

$$P(\sim E) = 1 - P(E) = \frac{10}{21} \implies P(\sim E) < P(E)$$

(b) Sapendo che é uscito un numero primo, calcolare la probabilitá che sia uscito 2.

Soluzione:

$$P(2 \mid E) = \frac{P(2 \cap E)}{P(E)} = \frac{P(2)}{P(E)} = \frac{2/21}{11/21} = \frac{2}{11}$$

Esercizio 2. La ditta di Bob produce e vende un componente elettronico la cui vita media é di 2.5 anni. La ditta si impegna, per garanzia, a sostituirlo se esso cessa di funzionare entro due anni.

(a) Qual é la probabilitá che la ditta debba intervenire per un singolo pezzo?

Soluzione: Supponiamo un tempo di vita T del prodotto che segue una legge esponenziale $T \sim Exp(\lambda) = \lambda e^{-\lambda t}, \quad t > 0$

Sappiamo che $E[T] = 2.5$ anni. Da questo ricaviamo λ

$$E[T] = \frac{1}{\lambda} = 2.5 \implies \lambda = 0.4$$

Quindi la probabilitá di intervenire entro i due anni di garanzia per sostituire il pezzo é

$$P(T \leq 2) = \int_0^2 0.4 e^{-0.4t} dt = 1 - e^{-0.8} = 0.55$$

(b) Che tempo di garanzia dovrebbe stabilire la ditta in modo da dover intervenire in non piú del 10% dei casi?

Soluzione: Si deve determinare il tempo t per cui $P(T \leq t) \leq 0.1$, ovvero, usando la cumulativa dell'esponenziale:

$$1 - e^{-\lambda t} = 1 - e^{-0.4t} \leq 0.1$$

Risolvendo la disequazione, si ricava:

$$t < 2.5 \ln \frac{10}{9} = 0.263$$

Esercizio 3. Sia data una variabile aleatoria X.

(a) Calcolare quanto vale al massimo la probabilitá che lo scarto di X dalla propria media valga almeno 2 deviazioni standard.

Soluzione: Lo scarto di X dalla sua media μ é $|X - \mu|$.
Si chiede di valutare la probabilitá $P_X(|X - \mu| \geq 2\sigma)$.
La distribuzione di X é ignota dunque si usa la disuguaglianza di Chebychev nella forma

$$P_X(|X - \mu| \geq 2\sigma) \leq \frac{\sigma^2}{(2\sigma)^2} = \frac{1}{4} = 25\%$$

Esercizio 4.
In un videogioco Alice ha a disposizione 6 tiri, ognuno dei quali colpisce a caso uno di due bersagli (A e B). Si vince se il bersaglio A é colpito almeno 4 volte su 6.

(a) Qual é la probabilitá che Alice vinca?

Soluzione: (Distribuzione binomiale)
Ogni tiro (indipendente) ha successo (colpisce il bersaglio) con probabilitá che vale $p = 1/2$.
Usando il modello binomiale $X \sim Bin\left(n = 6, p = \frac{1}{2}\right)$ la probabilitá di vittoria é semplicemente

$$P(X \geq 4) = \sum_{x=4}^{6} \binom{6}{x} \left(\frac{1}{2}\right)^x \left(1 - \frac{1}{2}\right)^{6-x} = \sum_{x=4}^{6} \binom{6}{x} \left(\frac{1}{2}\right)^6 = \frac{11}{32} \approx 34.37\%$$

(b) Se la partita costa 1 euro e in caso di vittoria si ricevono 3 euro, quanto puó vincere in media Alice?

Soluzione:
Sia Y la v.a. che misura il guadagno di Alice. Allora Y si distribuisce come segue:

- in caso di vittoria: $P(Y = 3 - 1) = P(Y = 2) = \frac{11}{32}$, come determinato sopra;
- in caso di sconfitta: $P(Y = -1) = 1 - \frac{11}{32} = \frac{21}{32}$.

Il guadagno atteso di Alice é

$$E[Y] = 2 \times \frac{11}{32} - 1 \times \frac{21}{32} = \frac{1}{32}$$

Esercizio 5. Uno strumento produce mine per le matite di una certa marca con una deviazione standard della lunghezza delle mine pari a $\sigma = 0.1$ cm. Misurando dieci mine, si sono ottenuti i seguenti valori (in cm)

$$12.21, 12.33, 12.84, 12.97, 13.22, 12.93, 13.07, 13.52, 13.23, 13.01$$

(a) Determinare un intervallo di confidenza al 95% per la lunghezza media delle mine.

Soluzione:
Assumiamo il modello gaussiano per la lunghezza X delle mine: $X \sim \mathcal{N}(\mu, (0.1)^2)$ (σ nota).
Sappiamo che $n = 10$ e la media campionaria delle misure vale $\bar{x} = \frac{1}{10} \sum_{i=1}^{10} x_i = 12.933$.
L' IC al 95% si ottiene per

$$\begin{aligned} (100)(1 - \alpha) &= 95 \\ 1 - \alpha &= 0.95 \\ \alpha &= 0.05 \\ \alpha/2 &= 0.025 \end{aligned}$$

Il valore critico $z_{0.025}$ é il valore che lascia un'area a destra pari a 0.025 (a sinistra pari a $1 - 0.025 = 0.975$).

Usando la tabella della Normale standard lo z-value che lascia a destra 0.025 é $z_{0.025} = 1.96$.
L'IC cercato é:

$$12.933 - 1.96\frac{0.1}{\sqrt{10}} < \mu < 12.933 + 1.96\frac{0.1}{\sqrt{10}}$$

$$12.933 - 0.062 < \mu < 12.933 + 0.062$$

$$12.871 < \mu < 12.995$$

(b) Mantenendo il medesimo livello di confidenza, quante mine occorrerebbe misurare per avere un intervallo non piú ampio di 0.1 cm?

Soluzione:
La condizione che l'ampiezza dell'intervallo di confidenza al 95% non superi 0.1 cm, richiede che

$$2 \times 1.96\frac{0.1}{\sqrt{n}} \leq \frac{1}{10} \implies n \geq (2 \times 1.96)^2 \approx 15.4$$

Esercizio 6.
Bob ha calcolato un intervallo di confidenza per la media di una popolazione normale, a partire da un campione di 8 osservazioni, dopo averne stimato la varianza con la varianza campionaria $s^2 = 4.2$. Il risultato é $(16.08, 18.82)$.

(a) Con quale confidenza é stato ottenuto l'intervallo?

Soluzione:
La taglia del campione $n = 8$ é piccola e la varianza é ignota essendo stata stimata mediante la varianza campionaria $s^2 = 4.2$. Inoltre, la popolazione da cui proviene il campione é normale. Pertanto l'IC per la media deve essere stato stimato utilizzando la distribuzione t di Student con $\nu = n - 1 = 7$ gradi di libertá:

$$\mu = \bar{x} \pm \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2}}(\nu = 7) = \bar{x} \pm \frac{\sqrt{4.2}}{\sqrt{8}} t_{\frac{\alpha}{2}}(\nu = 7)$$

Gli estremi dell'IC indicato corrispondono a

$$\bar{x} - \frac{\sqrt{4.2}}{\sqrt{8}} t_{\frac{\alpha}{2}}(\nu = 7) = 16.08$$

$$\bar{x} + \frac{\sqrt{4.2}}{\sqrt{8}} t_{\frac{\alpha}{2}}(\nu = 7) = 18.82$$

Risolvendo il sistema a due incognite si ottiene

$$\bar{x} = 17.45$$

$$t_{\frac{\alpha}{2}}(\nu = 7) = 1.895$$

Dalla tabella della t di Student con $\nu = 7$ leggiamo che $t_{\frac{\alpha}{2}} = 1.895$ per $\frac{\alpha}{2} = 0.05$, cioé $\alpha = 0.10$
Dunque l'intervallo é stato ottenuto con una confidenza pari al $(100)(1 - \alpha) = (100)(1 - 0.10) = 90\%$

Compito scritto dell'esame di Statistica e analisi dei dati **18 gennaio 2022**	**Prof. Giuseppe Boccignone**	**Corso di Laurea**
Cognome:	**Nome:**	**Matricola:**

Istruzioni

- Il tempo riservato alla prova scritta e' di 2 ore. Durante la prova e' possibile consultare il formulario ed utilizzare la calcolatrice. Non e' possibile consultare libri e appunti.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede piu' passaggi di calcolo, non sara' preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o piu' frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sara' eventualmente rappresentato in quest'ultima forma).

Problemi

Esercizio 1. In un recente sondaggio sul ministro della salute il 58% dei rispondenti ha dichiarato di aver fiducia nel suo operato. Le donne costituiscono il 58% del campione, e tra queste il 46% ha riposto positivamente (fiducia).

(a) Si seleziona a caso una persona tra quelle intervistate. Qual é la probabilitá che la persona selezionata sia maschio?

Soluzione: Gli unici eventi che ci interessano sono

- $M =$ "rispondente maschio"
- $D =$ "rispondente donna"

Sappiamo che $P(D) = 0.58$. Dunque:

$$P(M) = 1 - P(D) = 1 - 0.58 = 0.42$$

Esercizio 2.

Due ditte forniscono il medesimo prodotto. Se esso proviene dalla ditta A, la probabilitá che si guasti prima dell'istante t vale $1 - e^{-t}$; se invece proviene dalla ditta B questa probabilitá vale $1 - e^{-2t}$. Il prodotto puó essere acquistato con uguale probabilitá da A o da B, e non é nota la ditta fornitrice. Tuttavia, é stato osservato che il prodotto si guasta in un intervallo di tempo $1 \leq t \leq 2$.

(a) Determinare la probabilitá che esso sia stato acquistato dalla ditta A.

Soluzione: (Teorema di Bayes)
I dati del problema ci dicono che, a priori, le probabilitá che il prodotto provenga da A o da B valgono

$$P(A) = P(B) = 0.5.$$

Possiamo ragionare in due modi:
1. Approccio puramente probabilistico
Definiamo l'evento
$G =$ "guasto in $1 \leq t \leq 2$"
Nel caso di A $1 - e^{-t} = P($ "guasto prima di$t \mid A)$. L'evento G é dato da

$$G = \{ \text{"guasto prima di}t = 2\} - \{ \text{"guasto prima di}t = 1\}$$

La probabilitá di guasto del prodotto A nell'intervallo di tempo $1 \leq t \leq 2$ vale allora

$$P(G \mid A) = 1 - e^{-2} - (1 - e^{-1}) = e^{-1} - e^{-2}.$$

2. Approccio basato su modello di probabilitá e introduzione di una variabile aleatoria X
Dalla natura del problema e dalla forma $P(X < t \mid A) = 1 - e^{-t}$, si riconosce che tale probabilitá definisce la classica **funzione di ripartizione** $F_X(t) = 1 - e^{-t}$ di una v.a. $X \sim Exp(x; \lambda)$ con **densitá** esponenziale negativa e^{-t} con $\lambda = 1$.
Pertanto,

$$P(1 \leq X \leq 2 \mid A) = F_X(2) - F_X(1) = 1 - e^{-2} - (1 - e^{-1}) = e^{-1} - e^{-2}$$

.
Oppure, in modo equivalente, integrando direttamente la **densità**:

$$P(1 \leq X \leq 2 \mid A) = \int_1^2 e^{-t} dt = e^{-1} - e^{-2}$$

,
Stesso ragionamento vale per determinare $P(1 \leq X \leq 2 \mid B)$ (in tal caso $\lambda = 2$).
Dunque, per il prodotto B nello stesso intervallo é

$$P(G \mid B) = 1 - e^{-2 \cdot 2} - (1 - e^{-2 \cdot 1}) = e^{-2} - e^{-4}.$$

Applicando la regola di Bayes:

$$P(A \mid G) = \frac{P(G \mid A)P(A)}{P(G \mid A)P(A) + P(G \mid B)P(B)} = \frac{e^{-1} - e^{-2}}{e^{-1} - e^{-2} + e^{-2} - e^{-4}} = \frac{e^2(e-1)}{e^3 - 1} \approx 0.6652$$

Esercizio 3. Sia X la variabile che conta il numero di volte in cui si ottiene una doppia testa in 100 lanci di una coppia di monete regolari.

(a) Calcolare la probabilitá di ottenere una doppia testa almeno 20 volte.

Soluzione: Una doppia Testa (= “successo”) nel lancio di due monete regolari si ottiene con probabilitá 1/4.
La risposta richiederebbe l'uso della distribuzione binomiale $Bin(n,p)$, con parametri $n = 100$ e $p = 1/4 = 0.25$.
Poiché n è grande posso approssimare con una normale di parametri $\mu = np = 25, \sigma^2 = npq = 18.75$:

$$Bin(x; 0.25, 100 \approx \mathcal{N}(x; 25, 18.75)$$

Pertanto, usando normale standard e tabelle, con correzione di continuità $19.5 < 20$:

$$P(X \geq 20) = 1 - P(X < 20) \approx 1 - \Phi(\frac{19.5 - 25}{\sqrt{18.75}}) \approx 0.90$$

(con la Binomiale si otterrebbe 0.9000469)

Esercizio 4. Sia $X \sim \mathcal{N}(x; 1{,}2^2)$ e si definisca la v.a. $Y = X^2 - 3X$.

(a) Calcolare il valore atteso $E[Y]$ e la probabilità $P(Y > 0)$.

Soluzione: Sappiamo che per $\mathcal{N}(x; 1{,}2^2)$, $E[X] = 1$, $var(X) = 2^2$
Per il calcolo del valore atteso si ha:

$$E[Y] = E\left[X^2 - 3X\right] = E\left[X^2\right] - 3E[X]$$

Poichè $var(X) = E\left[X^2\right] - E[X]^2$,

$$E\left[X^2\right] - 3E[X] = var(X) + E[X]^2 - 3E[X] = 4 + 1 - 3 = 2$$

.
Per calcolare $P(Y > 0)$ si ha che

$$Y > 0 \implies X^2 - 3X = X(X-3) > 0,$$

e la disuguaglianza è valida per i casi (disgiunti) $X < 0$ e $X > 3$ Pertanto:

$$P(Y > 0) = P(X < 0) + P(X > 3)$$

La distribuzione di X é normale, dunque standardizzando e usando le tabelle:

$$P(Y > 0) = \Phi(\frac{0-1}{2}) + 1 - \Phi(\frac{3-1}{2}) \approx 0.47$$

Esercizio 5. Una popolazione assume un farmaco per mantenere il valor medio del battito cardiaco (battito/min, bpm) tra 80 e 90 bpm. Il bpm medio rilevato su un campione di 60 soggetti é pari a 91 bpm. Da dati storici, si sa che la variazione del bpm della popolazione segue una legge normale con deviazione standard $\sigma = 5$.

(a) Si verifichi se, relativamente al bpm medio, il farmaco possa ritenersi efficace con livello di confidenza al 90% e 95%

Soluzione:
Sappiamo che $n = 60$ e $\bar{x} = 91$.
L' IC al 90% si ottiene per

$$\begin{aligned}(100)(1-\alpha) &= 90 \\ 1-\alpha &= 0.9 \\ \alpha &= 0.1 \\ \alpha/2 &= 0.05\end{aligned}$$

Il valore critico $z_{0.05}$ é il valore che lascia un'area a destra pari a 0.05 (a sinistra pari a $1 - 0.05 = 0.95$).
Usando la tabella della Normale standard lo z-value che lascia a destra 0.05 é $z_{0.05} = 1.64$.
L'IC cercato é:

$$91 - 1.64\frac{5}{\sqrt{60}} < \mu < 91 + 1.64\frac{5}{\sqrt{60}}$$
$$89.94 < \mu < 92.06$$

Analogamente, per l'IC al 95%, $\alpha = 0.05$ e $z_{\alpha/2} = 1.96$.
Pertanto,

$$91 - 1.96\frac{5}{\sqrt{60}} < \mu < 91 + 1.96\frac{5}{\sqrt{60}}$$
$$89.74 < \mu < 92.26$$

Sulla base di tali risultati i farmaci con IC al 90% e al 95% sono al limite dell'efficacia. Considerando che in entrambi i casi si ha un intervallo di incertezza molto ampio, complessivamente, l'ipotesi di efficacia puó essere rifiutata.

Esercizio 6. Un produttore di robot per le pulizie domestiche garantisce che il suo prodotto dura in media 3 anni con una deviazione standard di 1 anno. Si campionano casualmente 5 robottini e ne risulta che abbiano una durata di $1.9, 2.4, 3.0, 3.5, 4.2$ anni.

(a) Assumendo che la vita della popolazione di robot sia distribuita con legge normale, si tragga una conclusione con un livello di confidenza al 95% sull'asserzione del produttore che $\sigma = 1$.

Soluzione:
Utilizziamo l'intervallo di confidenza

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

con $\sigma^2 = \sigma = 1$
Determiniamo la varianza del campione s^2 con

$$s^2 = \frac{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}{n(n-1)}$$

Per $n = 5$:

$$\sum_{i=1}^5 x_i = 1.9 + 2.4 + 3.0 + 3.5 + 4.2 = 15$$

e

$$\sum_{i=1}^5 x_i^2 = 1.9^2 + 2.4^2 + 3.0^2 + 3.5^2 + 4.2^2 = 48.26$$

Pertanto

$$\begin{aligned}s^2 &= \frac{n\sum_{i=1}^5 x_i^2 - (\sum_{i=1}^5 x_i)^2}{n(n-1)} \\ &= \frac{5 \cdot 48.26 - (15)^2}{5(5-1)} \\ &= \frac{16.3}{20} \\ &= 0.815\end{aligned}$$

Determiniamo $\alpha/2$. Da $95\% = 100(1-\alpha)\%$, si ha:

$$\begin{aligned}(100)(1-\alpha) &= 95\\ 1-\alpha &= 0.95\\ \alpha &= 0.05\\ \alpha/2 &= 0.025\end{aligned}$$

I g.d.l. sono $\nu = n - 1 = 4$. Usando la tabella della distribuzione Chi-quadro, il valore critico per 0.025 con 4 g.d.l. é $\chi^2_{0.025} = 11.143$, e $\chi^2_{1-0.025} = \chi^2_{0.975} = 0.484$.
L'intervallo di interesse é

$$\begin{aligned}\frac{(n-1)s^2}{\chi^2_{\alpha/2}} &< \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}\\ \frac{(5-1)\cdot 0.815}{\chi^2_{0.025}} &< \sigma^2 < \frac{(5-1)\cdot 0.815}{\chi^2_{0.975}}\\ \frac{4\cdot 0.815}{11.143} &< \sigma^2 < \frac{4\cdot 0.815}{0.484}\\ 0.29 &< \sigma^2 < 6.74\end{aligned}$$

Poiché 1 cade nell'intervallo di confidenza al 95% si puó concludere che quanto dichiarato dal produttore é plausibile statisticamente.

Compito scritto dell'esame di Statistica e analisi dei dati 20 settembre 2021	Prof. Giuseppe Boccignone	Corso di Laurea
Cognome:	Nome:	Matricola:

Istruzioni

- Il tempo riservato alla prova scritta e' di 2 ore. Durante la prova e' possibile consultare il formulario ed utilizzare la calcolatrice. Non e' possibile consultare libri e appunti.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede piu' passaggi di calcolo, non sara' preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o piu' frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sara' eventualmente rappresentato in quest'ultima forma).

Problemi

Esercizio 1.
Si supponga di avere un mazzo di 40 carte di cui 30 blu e 10 rosse. Si estrae una carta: se esce carta blu si lancia una moneta altrimenti un dado regolare.

(a) Con quale probabilitá esce testa?

Soluzione:
Definiamo i seguenti eventi

- $B =$ "esce carta blu"
- $T =$ "esce testa nel lancio della moneta"
- $D_i =$ "esce faccia i nel lancio del dado" , $i = 1, 2, \cdots, 6$
- $E =$ "esce testa nel gioco"

Osserviamo che

$$E = B \cap T$$

Allora:

$$P(E) = P(B \cap T) = P(T \mid B)P(B) = \frac{1}{2} \cdot \frac{30}{40} = \frac{3}{8}$$

(b) Con quale probabilitá esce il numero 6?

Soluzione:
Definiamo l'evento $F =$ "esce numero 6 nel gioco"
In modo analogo al caso precedente, $F = D_6 \cap \sim B$, dunque:

$$P(F) = P(D_6 \mid \sim B)P(\sim B) = \frac{1}{6} \cdot \frac{10}{40} = \frac{1}{24}$$

Esercizio 2. Un gruppo di bagnanti é costituito per il 65% da persone di carnagione scura (S) e le altre sono di carnagione chiara (C). L'uso inappropriato di creme solari fa sí che si abbia una percentuale di persone ustionate dal sole (U) del 10% se di carnagione scura e del 60% se di carnagione chiara

(a) Sapendo che un bagnante a caso si é ustionato prendendo il sole, con che probabilitá ha carnagione chiara?

Soluzione: (Teorema di Bayes)
I dati del problema ci dicono che

$$P(S) = 0.65$$
$$P(C) = 1 - P(S) = 0.35$$
$$P(U \mid S) = 0.1$$
$$P(U \mid C) = 0.6$$

Vogliamo calcolare $P(C \mid U)$.
Usando Bayes:

$$P(C \mid U) = \frac{P(U \mid C)P(C)}{P(U \mid C)P(C) + P(U \mid S)P(S)} = 0.76. \tag{1}$$

Esercizio 3.
Si effettuano 600 lanci di un dado non truccato.

(a) Calcolare un valore approssimato della probabilitá che il "5" esca un numero di volte compreso tra 94 e 106

Soluzione: (Distribuzione binomiale in approssimazione normale) In ogni lancio la probabilitá che esca il 5 vale $p = 1/6$ (equiprobabilitá di 6 eventi). Poiché $n = 600$ é sufficientemente grande, approssimiamo con una Gaussiana $\mathcal{N}(x \mid \mu = np, \sigma^2 = npq)$ dove

$$\mu = np = 600 \cdot \frac{1}{6} = 100$$

$$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{100 \cdot \frac{1}{6} \cdot \frac{5}{6}} \approx 9.1287$$

Procediamo alla standardizzazione

$$z_1 = \frac{94 - 100}{9.1287} = -0.657$$

$$z_2 = \frac{106 - 100}{9.1287} = 0.657$$

Quindi usando le tabelle della normale ridotta

$$P(94 \leq X \leq 100) = F_Z(0.657) - F_Z(-0.657) = 2F_Z(0.657) - 1 \approx 0.49$$

Esercizio 4.
La quantitá di stoffa per produrre poltrone segue una legge normale. Su un campione casuale di 15 poltrone, si é riscontrato che l'ammontare medio del materiale é $912cm^2$, con una deviazione standard di $64cm^2$.

(a) Qual é l'intervallo di confidenza al 99% per la media della quantitá di materiale?

Soluzione:
Sappiamo che $n = 15, \bar{x} = 912cm^2, s = 64cm^2$. Il campione di taglia $n = 15$ é piccolo e l'intervallo é basato sulla t di Student con $\nu = n - 1 = 14$ gradi di libertá.
Con $\alpha = 0.01$, si ricava dalla tabella della t che il quantile é $t_\alpha = t_{.005} = 2.977$.
L'errore standard é $SE = \frac{\bar{x}}{\sqrt{n}} = \frac{912}{\sqrt{15}} = 16.52473$
L' intervallo di confidenza al 99% per la media é dunque

$$\bar{x} - 2.977 \cdot SE < \mu < \bar{x} + 2.977 \cdot SE,$$

ovvero

$$\mu = 912 \pm 49.2cm^2$$

Esercizio 5.
La misura X del numero di ottani di una benzina ha una deviazione standard di 0.5.

(a) Quante misure indipendenti devono essere fatte per conoscere il numero medio di ottani con un errore massimo di 0.3 e una confidenza del 90%?

Soluzione:
Assumiamo una popolazione normale (media μ incognita e $\sigma = 0.5$ nota) e scriviamo l' IC (bilaterale) per la media μ con confidenza del 90%:

$$\mu = \bar{X} \pm \frac{0.5}{\sqrt{n}} z_{0.95}$$

L'errore richiesto é il semi intervallo $z_{\alpha/2}\dot{S}E.$, con errore standard di $\bar{X}$ $SE = \frac{\sigma}{\sqrt{n}}$, che deve essere al massimo pari a 0.3, dunque

$$e = \frac{0.5}{\sqrt{n}} z_{0.95} \leq 0.3$$

da cui

$$n \geq \frac{0.25}{0.09} z_{0.95}^2 \simeq 7.5$$

Esercizio 6.
Da due popolazioni normalmente distribuite vengono estratti due campioni di taglia 9 e 12. Le varianze campionarie sono 20 e 8.

(a) E' plausibile assumere che le varianze delle popolazioni non siano statisticamente differenti, al livello di significatività 0.05?

Soluzione:
I dati del problema ci dicono che: $n_1 = 9$, $n_2 = 12$, $s_1^2 = 20$, $s_2^2 = 8$.
L'IC per il rapporto delle due varianze, con $\alpha = 0.05$ é:

$$\frac{s_1^2}{s_2^2}\frac{1}{f_{0.025}(\nu_1,\nu_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} f_{0.025}(\nu_2,\nu_1)$$

con $\nu_1 = 8, \nu_2 = 11$, $\frac{s_1^2}{s_2^2} = 2.5$ $f_{0.025}(8,11) = 3.66$, $f_{0.025}(11,8) \simeq 4.25$

$$2.5 \cdot 0.27 < \frac{\sigma_1^2}{\sigma_2^2} < 2.5 \cdot 4.25$$

$$0.68 < \frac{\sigma_1^2}{\sigma_2^2} < 10.6$$

L'intervallo ammette la possibilitá che $\frac{\sigma_1^2}{\sigma_2^2} = 1$, dunque é l'ipotesi è plausibile

Compito scritto dell'esame di Statistica e analisi dei dati 06 settembre 2021	Prof. Giuseppe Boccignone	Corso di Laurea
Cognome:	Nome:	Matricola:

Istruzioni

- Il tempo riservato alla prova scritta e' di 2 ore. Durante la prova e' possibile consultare il formulario ed utilizzare la calcolatrice. Non e' possibile consultare libri e appunti.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede piu' passaggi di calcolo, non sara' preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o piu' frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sara' eventualmente rappresentato in quest'ultima forma).

Problemi

Esercizio 1.

Quando esce di casa al mattino Aldo non prende mai l'ombrello se il cielo é sereno, lo prende sempre se il cielo é nuvoloso e infine, quando il tempo é incerto, decide se prendere o no l'ombrello lanciando una moneta regolare. Secondo i dati storici, nella cittá dove vive Aldo il tempo é sereno (= evento H_1), nuvoloso (= evento H_2) o incerto (= evento H_3) nelle proporzioni $10 : 3 : 2$, rispettivamente.

(a) Nell'arco di un anno (= 365 giorni) quante volte all'incirca Aldo esce con l'ombrello?

Soluzione: Definito l'evento

$$A = \text{Aldo esce con l'ombrello}$$

i dati sono $P(H_1) = 10/15$, $P(H_2) = 3/15$, $P(H_3) = 2/15$, $P(A \mid H_1) = 0$, $P(A \mid H_2) = 1$, P (A | H_3) = 1/2. Allora

$$P(A) = \sum_i P(A \mid H_i)P(H_i) = \frac{4}{15} \simeq 0.27$$

Quindi Aldo esce con l'ombrello mediamente

$$365 \times \frac{4}{15} \simeq 97$$

giorni all'anno.

(b) Se il 30 marzo Aldo é uscito con l'ombrello, qual é la probabilitá che quel giorno fosse nuvoloso?

Soluzione: Se il 30 marzo Aldo é uscito con l'ombrello, la probabilitá che quel giorno fosse nuvoloso vale

$$P(H_2 \mid A) = \frac{P(A \mid H_2)P(H_2)}{P(A)} = \frac{3/15}{4/15} = \frac{3}{4} = 0.75$$

Esercizio 2.

Supponiamo che il 10% dei semi di una certa varietá di pianta siano sterili.

(a) Qual é la probabilitá che germini almeno il 95% dei semi contenuti in una confezione da 20 semi (= evento A)?

Soluzione: Detto X_n il numero di semi sterili in una confezione di n semi, si puó assumere che sia $X_n \sim Bin(n; 0.1)$. Si chiede la probabilitá che una confezione di 20semi contenga al massimo $20 \times 5\% = 1$ seme sterile

$$P(A) = P(X_{20} \leq 1) = \sum_{x=0}^{1} \binom{20}{x} (0.1)^x (0.9)^{20-x} \approx 39.17\%$$

(b) Qual é la probabilitá che germini almeno il 95% dei semi contenuti in una confezione da 100 semi (= evento B)?

Soluzione: (Distribuzione binomiale in approssimazione poissoniana o Gaussiana) Analogamente alla precedente, si chiede la probabilitá che una confezione di 100semi contenga al massimo $100 \times 5\% = 5$ semi sterili, dunque

$$P(B) = P(X_{100} \leq 5) = \sum_{x=0}^{5} \binom{100}{x} (0.1)^x (0.9)^{100-x}$$

Possiamo in questo caso approssimare sia con Poisson (piccola probabilitá $p = 0.1$) sia con la Normale ($n = 100$)

$$Bin(100; 0.1) \approx Pois(10)$$

$$Bin(100; 0.1) \approx \mathcal{N}(10; 3^2)$$

Con Poisson si ha che

$$P(B) = P(X_{100} \leq 5) \approx \sum_{x=0}^{5} \frac{10^x}{x!} e^{-10} \simeq 6.71\%.$$

Con la normale usando la correzione di continuitá e usando le tabelle della normale ridotta:

$$P(B) = P(X_{100} \leq 5) \approx F\left(\frac{5.5 - 10}{3}\right) \simeq 6.68\%.$$

(si osserva che l'approssimazione normale é lievemente migliore: calcolando esattamente con la Binomiale $P(B) \simeq 5.76\%$)

Esercizio 3. Tra le 23 e le 23:30 lungo un tratto di strada transitano in media 40 veicoli.

(a) Qual è la probabilità di osservare, tra un veicolo e il successivo, un intervallo di tempo minore di 10 sec e uno piú lungo di 5 min?

Soluzione: E' un problema di tempi d'attesa modellabile con la distribuzione esponenziale negativa.
Su 30 min il tempo medio (in secondi) tra due veicoli successivi é $\frac{30 \times 60}{40} = 45sec$ a cui corrisponde un rate di arrivo pari a

$$\lambda = \frac{1}{45} = 0.022 sec^{-1}$$

Per $t = 10$,

$$P(0 \leq T \leq 10) = F(10) = 1 - e^{-0.022 \cdot 10} = 0.199 \approx 20\%$$

Intervalli piú lunghi di $t = 5min = 300sec$ saranno osservati con probabilitá

$$P(T > 300) = 1 - P(0 \leq T \leq 300) = e^{-0.022 \cdot 300} = 1.27 \times 10^{-3} \approx 0.13\%$$

Esercizio 4. Un segnale di intensitá media incognita, trasmesso dalla sorgente S, é raccolto dal ricevitore R con un rumore additivo gaussiano di media nulla e deviazione standard uguale a 2 (pertanto il segnale avrá la stessa distribuzione del rumore, con uguale deviazione standard ma media incognita). Per ridurre l'errore, lo stesso segnale viene inviato 10 volte da S ad R e la media campionaria dei segnali ricevuti é 8.7.

(a) Qual é la confidenza che il segnale trasmesso sia compreso fra 7.0 e 10.4?

Soluzione:
Il segnale X raccolto al ricevitore é variabile aleatoria con distribuzione normale $X \sim \mathcal{N}(\mu, \sigma^2) = \mathcal{N}(\mu, 2^2)$, dunque con μ incognita ma $\sigma = 2$ nota.
Il problema richiede di stimare il livello di confidenza $1-\alpha$ tale che l' IC a due code [7.0,10.4] per la media μ soddisfi

$$P(7.0 < \mu < 10.4) = 1 - \alpha.$$

Poiché X proviene da popolazione normale con μ incognita e σ nota, la stima intervallare, sapendo che $n = 10$ e $\bar{x} = 8.7$, si scrive come

$$P\left(\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

ovvero $\mu = \bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ con confidenza $1 - \alpha$.
In questo caso $z_{\frac{\alpha}{2}}$ é incognito, perché non conosciamo il livello di fiducia α, e va determinato. Assumendo $\mu = \bar{x}$, il semi intervallo di confidenza (simmetrico) ha lunghezza

$$10.4 - \bar{x} = 1.7 = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

da cui $z_{\frac{\alpha}{2}} = \frac{1.7 \cdot \sqrt{10}}{2} \approx 2.7$
Dalla tabella normale $F(z_{\frac{\alpha}{2}}) \approx F(2.7) = 0.9965 = 1 - \frac{\alpha}{2}$.
Per le proprietá della cumulativa standard

$$1 - \alpha = P(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}) = 2F(z_{\frac{\alpha}{2}}) - 1 \approx 2 \cdot 0.9965 - 1 = 0.993 = 99.3\%$$

che rappresenta il livello di confidenza dell'IC a due code [7.0,10.4].

Esercizio 5. Il termostato di un condizionatore d'aria é supposto essere tarato sul valore di soglia $\mu = 25°C$ (cioé l'apparecchio entra in funzione quando la temperatura dell'ambiente supera i $25°C$). Un tecnico misura in 8 occasioni diverse la temperatura X alla quale il termostato scatta, ottenendo i valori seguenti ($°C$):

$$24.6 \;\; 24.8 \;\; 25.2 \;\; 25.4 \;\; 25.5 \;\; 24.0 \;\; 24.7 \;\; 25.3$$

(a) Formulando un adeguato modello per X, stabilire se, al 95% di confidenza, il valore di μ sia plausibile statisticamente

Soluzione:
Assumendo una distribuzione normale della temperatura, $X \sim \mathcal{N}(\mu, \sigma^2)$, poiché la varianza σ^2 é incognita, la stima intervallare di μ al 95% di confidenza si ottiene considerando

$$P\left(\bar{x} - t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right) = 1 - \alpha = 0.95$$

Poiché al 95% di confidenza $\alpha = 0.05$, allora $t_{\frac{\alpha}{2}} = t_{0.025}$, da determinarsi dalla distribuzione di Student con $\nu = n - 1 = 7$ gdl:

$$t_{0.025} = 2.365$$

Dai dati conosciuti si ottengono media e deviazione standard campionaria:

$$\bar{x} \approx 24.93$$

$$s \approx 0.83$$

Pertanto

$$24.93 - 2.365 \times \frac{0.83}{2.64} < \mu < 24.93 + 2.365 \times \frac{0.83}{2.64}$$

ovvero $24.2 < \mu < 25.66$ e dunque il valore $\mu = 25°C$ puó essere ritenuto plausibile.

(b) Stabilire quanto vale al massimo la precisione p (ovvero il reciproco della deviazione standard) del termostato al 95% di confidenza

Soluzione:
Assumendo $\mu = 25°C$ per l'analisi precedente, $X \sim \mathcal{N}(25, \sigma^2)$.
Sappiamo che $p = \frac{1}{\sigma}$, dunque il valore di precisione massima é quello di deviazione standard minima, $p_{max} = \frac{1}{\sigma_{min}}$.
Il problema richiede di determinare

$$P(p < p_{max}) = 1 - \alpha = 0.95$$

Dalla definizione di p questo equivale alla stima unilaterale sul limite inferiore della varianza (σ^2_{min}) :

$$P(p < p_{max}) = P\left(\sqrt{\frac{(n-1)s^2}{\chi^2_\alpha}} < \sigma\right) = 1 - \alpha$$

Ne deriva che al 95% di confidenza il limite superiore di precisione vale $p_{max} = \sqrt{\frac{\chi^2_{0.05}}{(n-1)s^2}}$, dunque, leggendo dalla tabella della distribuzione Chi-quadro, per $\nu = 7$, $\chi^2_{0.05}(7) = 14.067$ e usando s^2 calcolato precedentemente.

$$p < \sqrt{\frac{\chi^2_{0.05}}{(n-1)s^2}} \approx 1.69°C$$

Compito scritto dell'esame di Statistica e analisi dei dati 15 luglio 2021	Prof. Giuseppe Boccignone	Corso di Laurea
Cognome:	Nome:	Matricola:

Istruzioni

- Il tempo riservato alla prova scritta e' di 2 ore. Durante la prova e' possibile consultare il formulario ed utilizzare la calcolatrice. Non e' possibile consultare libri e appunti.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede piu' passaggi di calcolo, non sara' preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o piu' frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sara' eventualmente rappresentato in quest'ultima forma).

Problemi

Esercizio 1. Una recente indagine ha rivelato che il 14% delle segretarie ha dolore al polso. Inoltre, il 6% delle segretarie intervistate ha dolore al polso e al tempo stesso assume regolarmente un farmaco antinfiammatorio.

(a) Qual é la probabilitá che una segretaria che ha dolore al polso assuma regolarmente un farmaco antinfiammatorio?

Soluzione: Definiamo gli eventi seguenti:

$$A = \text{“la segretaria ha il dolore al polso”};$$

$$B = \text{“la segretaria assume il farmaco”}.$$

$$P(A) = 0.14, P(A,B) = 0.06.$$

La probabilitá richiesta é la condizionata:

$$P(B \mid A) = \frac{P(B,A)}{P(A)} = 0.428$$

Esercizio 2.

Un gioco consiste nel lanciare una moneta e nel generare un numero casuale Y tra zero e uno. La vincita é $2Y$ se compare testa e 1 se compare croce.

(a) Calcolare la vincita attesa.

Soluzione: (Valore atteso) Sia T l'evento: “compare testa” definito su un classico spazio di probabilitá $(S, \mathcal{F}, P)$, $S = \{T,C\}$ e sia V la vincita

Usiamo la classica variabile aleatoria indicatrice

$$X_1(\omega) = I_T(\omega) = \begin{cases} 1 & \text{se } \omega \in T \\ \\ 0 & \text{se } \omega \notin T (\omega \in \sim T) \end{cases}. \tag{1}$$

e la VA indicatrice speculare

$$X_2(\omega) = I_{\sim T}(\omega) = \begin{cases} 0 & \text{se } \omega \in T \\ 1 & \text{se } \omega \notin T (\omega \in \sim T) \end{cases}. \tag{2}$$

Allora:

$$V = 2Y I_T - I_{\sim T}$$

Poiché X e I_T sono v.a. indipendenti

$$E[V] = E[2Y I_T - I_{\sim T}] = E[2Y I_T] - E[I_{\sim T}] = 2E[Y]E[I_T] - E[I_{\sim T}]$$

Il valore atteso della VA indicatrice X_1 si calcola facilmente:

$$E[X_1] = 0 \cdot P_X(X_1 = 0) + 1 \cdot P_X(X_1 = 1) = 0 \cdot (1 - P(T)) + 1 \cdot P(T) = P(T) = \frac{1}{2},$$

ovvero, il valor medio della funzione indicatrice é la probabilitá di successo. Analogamente, per $I_{\sim A}$.
Sappiamo che il valor atteso di una VA distribuita uniformemente, $Y \sim Unif(a,b)$ é

$$E[Y] = \frac{b+a}{2}$$

dunque se $Y \sim Unif(0,1)$, $E[Y] = \frac{1}{2}$
Quindi:

$$E[V] = 2 \times \frac{1}{2} \times \frac{1}{2} - \frac{1}{2} = 0$$

Esercizio 3. Il responsabile dell'illuminazione di un palazzo ha a disposizione 3 lampadine di riserva. Ciascuna delle lampadine utilizzate dura in media 200 ore e il responsabile deve aspettare 24 ore perché gli vengano consegnate nuove lampadine di riserva.

(a) Qual é la probabilitá che il responsabile non sia in grado di sostituire le lampadine fulminate?

Soluzione: Il problema sorge se si fulminano piú di tre lampadine entro le 24 ore: in tal caso il responsabile non sará in grado di sostituirle. Dobbiamo dunque determinare la probabilitá congiunta

$$P(\{X_1 \leq 0.24\}, \{X_2 \leq 0.24\}, \{X_3 \leq 0.24\}).$$

dove X_1,X_2,X_3 sono i tempi di attesa della rottura della prima, seconda e terza lampadina.
Si assuma una distribuzione esponenziale negativa dei tempi di attesa e si ragioni, per semplicitá di calcolo in termini di centinaia di ore.
Il tempo medio é pari a $E[X] = 2$ (in centinaia di ore). Il parametro della distribuzione esponenziale é dunque

$$\lambda = \frac{1}{2} = 0.5$$

La probabilitá del tempo di rottura di una singola lampadina entro le 24 ore é

$$P(X \leq 0.24) = 1 - e^{-\lambda \cdot x} = 1 - e^{-0.5 \cdot 0.24} = 1 - 0.887 = 0.113$$

Poiché la rottura di ciascuna lampadina é indipendente dalle altre:

$$P(\{X_1 \leq 0.24\}, \{X_2 \leq 0.24\}, \{X_3 \leq 0.24\}) = P(\{X_1 \leq 0.24\}) \cdot P(\{X_2 \leq 0.24\}) \cdot P(\{X_3 \leq 0.24\}) = (0.113)^3 = 0.0014$$

Esercizio 4. Una variabile aleatoria X ha una distribuzione con media 250 e deviazione standard 20.

(a) Determinare le probabilitá :

- $P(210 < X < 290)$
- $P(220 < X < 280)$

Soluzione: La distribuzione di X é ignota e si usa la disuguaglianza di Chebychev nella forma

$$P_X(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}.$$

con $\mu = 250, \sigma = 20$.
Entrambi gli intervalli hanno come punto centrale la media della distribuzione. Per esempio, nel primo caso il punto centrale vale $(210 + 290)/2 = 250$
Si trova il raggio dell'intorno di 250 che é $(290 - 210)/2 = 40$. Quindi l'intervallo é 250 ± 40
Si esprime il raggio come un multiplo della deviazione standard:

$$40 = k \times 20$$

da cui $k = 2$ (il raggio 40 é il doppio della deviazione standard).
Usando Chebychev:

$$P(210 < X < 290) \geq 1 - \frac{1}{k^2} = 1 - \frac{1}{4} = 0.75$$

Analogamente:

$$P(220 < X < 280) \geq 1 - \frac{1}{k^2} = 1 - \frac{1}{1.5^2} = 0.555$$

(b) Determinare le stesse probabilitá sapendo che X segue una legge normale $X \sim \mathcal{N}(\mu, \sigma)$

Soluzione: Usando le tavole della Gaussiana standard:

$$P(210 < X < 290) = P(-2 < Z < 2) = 2(0.9772) - 1 = 0.9544$$

$$P(220 < X < 280) = P(-1.5 < Z < 1.5) = 2(0.9332) - 1 = 0.8664$$

Esercizio 5. Due tipi di soluzioni chimiche sono state provate per misurarne il pH. L'analisi di 6 campioni della prima soluzione ha mostrato un pH medio di 7.52 con deviazione standard campionaria di 0.032; su 5 campioni della seconda si sono misurati un pH medio di 7.49 e deviazione standard campionaria di 0.024

(a) Si verifichi se sia ragionevole assumere che le varianze delle due popolazioni sono uguali al livello di significatività 0.05

Soluzione:
Sotto l'ipotesi di distribuzione normale, consideriamo l'intervallo di confidenza

$$\frac{s_1^2}{s_2^2}\frac{1}{f_{\frac{\alpha}{2}}(\nu_1,\nu_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2}\frac{1}{f_{1-\frac{\alpha}{2}}(\nu_1,\nu_2)}$$

Questo è l'intervallo per cui $P(f_{1-\frac{\alpha}{2}}(\nu_1,\nu_2) < F < f_{\frac{\alpha}{2}}(\nu_1,\nu_2) = 1-\alpha$ con $\alpha = 0.05$ e dove F è la statistica $F = \frac{\sigma_2^2 s_1^2}{\sigma_1^2 s_2^2}$
Semplicemente, nel caso in cui $\sigma_1^2 = \sigma_2^2$ è sufficiente verificare che

$$f_{1-\frac{\alpha}{2}}(\nu_1,\nu_2) < \frac{s_1^2}{s_2^2} < f_{\frac{\alpha}{2}}(\nu_1,\nu_2)$$

Sappiamo che $n_1 = 6, s_1 = 0.032, n_2 = 5, s_2 = 0.024$
Il rapporto delle varianze campionarie è

$$\frac{s_1^2}{s_2^2} = \frac{0.032^2}{0.024^2} = 1.78$$

I g.d.l sono $\nu_1 = n_1 - 1 = 5$, $\nu_2 = n_2 - 1 = 4$, dunque, dalle tavole

$$f_{1-\frac{\alpha}{2}}(\nu_1,\nu_2) = f_{0.975}(5{,}4) = \frac{1}{f_{0.025}(4{,}5)} = \frac{1}{7.39} = 0.14$$

$$f_{\frac{\alpha}{2}}(\nu_1,\nu_2) = f_{0.025}(5{,}4) = 9.36$$

Non possiamo dunque rigettare l'ipotesi che $\sigma_1^2 = \sigma_2^2$.

(b) Si stabilisca se i valori medi di pH delle due soluzioni possano essere ritenuti uguali con livello di significatività 0.05

Soluzione:
Abbiamo che $\bar{x}_1 = 7.51$, $\bar{x}_2 = 7.49$.
Sempre nell'ipotesi normale, le varianze della popolazione sono incognite, ma al punto precedente abbiamo inferito che plausibilmente sono uguali.
L'intervallo di confidenza per la differenza fra le medie della popolazione é dunque

$$(\bar{x}_1 - \bar{x}_2) - t_{\frac{\alpha}{2}} s_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\frac{\alpha}{2}} s_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

Nell'ipotesi che $\mu_1 = \mu_2$ avremo che per la statistica

$$T = \frac{\bar{x}_1 - \bar{x}_2}{s_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

deve valere

$$-t_{\frac{\alpha}{2}} < T < t_{\frac{\alpha}{2}}$$

I gdl della distribuzione sono

$$\nu = n_1 + n_2 - 2 = 9$$

La varianza pooled é

$$s_P^2 = \frac{5 \cdot 0.032^2 + 4 \cdot 0.024^2}{6+5-2} = 0.000825$$

Quindi

$$T = \frac{7.52 - 7.49}{\sqrt{0.000825(\frac{1}{6} + \frac{1}{5})}} = 1.72$$

Per $\alpha = 0.05$, $t_{\frac{\alpha}{2}} = t_{0.025} = 2.262$, dunque

$$-2.262 < T < 2.262$$

e concludiamo che la differenza fra le due medie non é statisticamente significativa al livello di significatività 0.05.

Compito scritto dell'esame di Statistica e analisi dei dati 30 giugno 2021	Prof. Giuseppe Boccignone	Corso di Laurea
Cognome:	Nome:	Matricola:

Istruzioni

- Il tempo riservato alla prova scritta e' di 2 ore. Durante la prova e' possibile consultare il formulario ed utilizzare la calcolatrice. Non e' possibile consultare libri e appunti.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede piu' passaggi di calcolo, non sara' preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o piu' frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sara' eventualmente rappresentato in quest'ultima forma).

Problemi

Esercizio 1. In un recente sondaggio sul sindaco di una certa cittá, il 62% dei rispondenti ha fiducia nel sindaco. Le donne costituiscono il 53% del campione, e tra queste il 46% ha fiducia nel sindaco.

(a) Si seleziona a caso una persona tra quelle intervistate. Qual é la probabilitá che la persona selezionata sia maschio?

Soluzione: Esercizio banale se si filtrano i dati che non servono. Gli unici eventi che ci interessano sono

- $D =$ "rispondente donna"
- $M =$ "rispondente maschio"

Sappiamo che $P(D) = 0.53$. Dunque:

$$P(M) = 1 - P(D) = 1 - 0.53 = 0.47$$

Esercizio 2. La tabella riporta le probabilitá congiunte di un insieme di nuove aziende che operano nel settore del

Table 1: Tabella aziende

Crescita	Nord-Est	Sud	Centro	Nord-Ovest
Bassa	0.04	0.12	0.14	0.19
Media	0.05	0.08	0.06	0.12
Alta	0.03	0.05	0.08	0.04

commercio via internet, classificate per regione di ubicazione e prospettiva di crescita.

(a) Se l'azienda ha una crescita attesa media o alta, qual é la probabilitá che sia ubicata nel Nord-Ovest?

Soluzione:
Indichiamo con M l'evento "l'azienda ha una crescita media", con A l'evento "l'azienda ha una crescita alta" e con NO l'evento "azienda ubicata nel Nord-Ovest".
Si tratta di determinare la probabilitá condizionata $P(NO \mid (M \cup A))$:

$$P(NO \mid (M \cup A)) = \frac{P(NO \cap (M \cup A)}{P(M \cup A)} \tag{1}$$

Essendo A, M disgiunti:

$$P(M \cup A) = P(M) + P(A)$$

Dalla tabella abbiamo le seguenti probabilitá marginali. :

$$P(M) = 0.05 + 0.08 + 0.06 + 0.12 = 0.31$$

$$P(A) = 0.03 + 0.05 + 0.08 + 0.04 = 0.20$$

Dunque

$$P(M \cup A) = P(M) + P(A) = 0.31 + 0.20 = 0.51.$$

Usando la proprieta' distributiva dell'intersezione rispetto all' unione, la disgiunzione di A, M, e leggendo in tabella

$$P(NO \cap M) = 0.12$$

e

$$P(NO \cap A) = 0.04$$

si ottiene:

$$P(NO \mid (M \cup A)) = \frac{P(NO \cap M) + P(NO \cap A)}{P(M \cup A)} = \frac{0.12 + 0.04}{0.51} = 0.31. \tag{2}$$

Esercizio 3. Una fabbrica di cioccolato promuove una campagna pubblicitaria per la vendita di un nuovo tipo di uova pasquali, caratterizzate dal contenuto della sorpresa, costituita da bracciali di ottone. Su una partita di 1000 uova, 5 di esse contengono dei bracciali d'oro, del valore di 100000 euro cadauno. Le uova contenenti bracciali d'oro, a loro volta, vengono inserite in modo casuale in scatole da 20, e vendute ai negozianti.

(a) Qual é la probabilitá che un negoziante che acquista una scatola di uova, venda al pubblico una o piú uova contenenti bracciali d'oro?

Soluzione:
La risposta richiederebbe l'uso della distribuzione binomiale $Bin(n,p)$, con parametri $n = 20$ e $p = 5/1000 = 0.005$. In tal caso, il numero atteso di bracciali d'oro per scatola é dato da $np = 20 \times 0.005 = 0.1$. Tuttavia, in questo caso la distribuzione di Poisson di parametro $\mu = np$ puó essere una buona approssimazione della binomiale, essendo $p = 0.005$ molto piccolo, e semplifica il calcolo della probabilitá di avere $1,2,\cdots$ uova contenenti bracciali d'oro. Pertanto:

$$P(X \geq 1) = 1 - P(X = 0) = 1 - Pois(X = 0) = 1 - \frac{0.1^0}{0!} e^{(-0,1)} = 1 - 0.904837418 = 0.09516258$$

(con la Binomiale si otterrebbe 0.0953895197)

Esercizio 4. Una variabile aleatoria X, con valori nell'intervallo $1 \leq x \leq 2$, ha densitá

$$f_X(x) = \frac{k}{x^2}$$

.

(a) Calcolare valore atteso $E[X]$ e deviazione standard σ_X

Soluzione: Si calcola la costante k utilizzando la proprietá di normalizzazione:

$$1 = \int_1^2 f_X(x)dx = \int_1^2 \frac{k}{x^2}dx = k\left[-\frac{1}{x}\right]_1^2 = k(-\frac{1}{2} + 1) = \frac{k}{2},$$

da cui $k = 2$

$$E[X] = \int_1^2 x\frac{2}{x^2}dx = 2\log 2 \approx 1.386$$

$$\sigma_X^2 = E\left[X^2\right] - E[X]^2 = E\left[X^2\right] - 4\log^2 2 = 2\int_1^2 dx - 4\log^2 2 = 2(1 - 2\log^2 2) \approx 0.078$$

$$\sigma_X = \sqrt{0.078} \approx 0.28.$$

Esercizio 5. Una popolazione assume un farmaco per mantenere il valor medio del battito cardiaco (battito/min, bpm) tra 85 e 95 bpm. Il bpm medio rilevato su un campione di 49 soggetti é pari a 90 bpm. Da dati storici, si sa che la variazione del bpm della popolazione segue una legge normale con deviazione standard $\sigma = 10$.

(a) Si verifichi se, relativamente al bpm medio, il farmaco possa ritenersi efficace con livello di confidenza al 90% e 95%

Soluzione:

Sappiamo che $n = 49$ e $\bar{x} = 90$.

L' IC al 90% si ottiene per

$$\begin{aligned}(100)(1-\alpha) &= 90\\ 1-\alpha &= 0.9\\ \alpha &= 0.1\\ \alpha/2 &= 0.05\end{aligned}$$

Il valore critico $z_{0.05}$ é il valore che lascia un'area a destra pari a 0.05 (a sinistra pari a $1 - 0.05 = 0.95$).

Usando la tabella della Normale standard lo z-value che lascia a destra 0.05 é $z_{0.05} = 1.64$.

L'IC cercato é:

$$90 - 1.64\frac{10}{\sqrt{49}} < \mu < 90 + 1.64\frac{10}{\sqrt{49}}$$
$$87.66 < \mu < 92.34$$

Analogamente, per l'IC al 95%, $\alpha = 0.05$ e $z_{\alpha/2} = 1.96$.

Pertanto,

$$90 - 1.96\frac{10}{\sqrt{49}} < \mu < 90 + 1.96\frac{10}{\sqrt{49}}$$
$$87.2 < \mu < 92.8$$

Sulla base di tali risultati il farmaco si puó ritenere efficace.

Esercizio 6. Un produttore di batterie per auto garantisce che il suo prodotto dura in media 3 anni con una deviazione standard di 1 anno. Si campionano casualmente 5 batterie e ne risulta che abbiano una durata di $1.9, 2.4, 3.0, 3.5, 4.2$ anni.

(a) Assumendo che la vita della popolazione di batterie sia distribuita con legge normale, si tragga una conclusione con un livello di confidenza al 95% sull'asserzione del produttore che $\sigma = 1$.

Soluzione:

Utilizziamo l'intervallo di confidenza

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

con $\sigma^2 = \sigma = 1$

Determiniamo la varianza del campione s^2 con

$$s^2 = \frac{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}{n(n-1)}$$

Per $n = 5$:

$$\sum_{i=1}^5 x_i = 1.9 + 2.4 + 3.0 + 3.5 + 4.2 = 15$$

e

$$\sum_{i=1}^5 x_i^2 = 1.9^2 + 2.4^2 + 3.0^2 + 3.5^2 + 4.2^2 = 48.26$$

Pertanto

$$\begin{aligned}s^2 &= \frac{n\sum_{i=1}^5 x_i^2 - (\sum_{i=1}^5 x_i)^2}{n(n-1)}\\ &= \frac{5\cdot 48.26 - (15)^2}{5(5-1)}\\ &= \frac{16.3}{20}\\ &= 0.815\end{aligned}$$

Determiniamo $\alpha/2$. Da $95\% = 100(1-\alpha)\%$, si ha:

$$\begin{aligned}(100)(1-\alpha) &= 95\\ 1-\alpha &= 0.95\\ \alpha &= 0.05\\ \alpha/2 &= 0.025\end{aligned}$$

I g.d.l. sono $\nu = n - 1 = 4$. Usando la tabella della distribuzione Chi-quadro, il valore critico per 0.025 con 4 g.d.l. é $\chi^2_{0.025} = 11.143$, e $\chi^2_{1-0.025} = \chi^2_{0.975} = 0.484$.
L'intervallo di interesse é

$$\begin{gathered}\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}\\ \frac{(5-1)\cdot 0.815}{\chi^2_{0.025}} < \sigma^2 < \frac{(5-1)\cdot 0.815}{\chi^2_{0.975}}\\ \frac{4\cdot 0.815}{11.143} < \sigma^2 < \frac{4\cdot 0.815}{0.484}\\ 0.29 < \sigma^2 < 6.74\end{gathered}$$

Poiché 1 cade nell'intervallo di confidenza al 95% si puó concludere che quanto dichiarato dal produttore é plausibile statisticamente.

Compito scritto dell'esame di Statistica e analisi dei dati 15 giugno 2021	Prof. Giuseppe Boccignone	Firma leggibile dello studente
Cognome:	Nome:	Matricola:

Istruzioni

- Il tempo riservato alla prova scritta e' di 2 ore. Durante la prova e' possibile consultare il formulario ed utilizzare la calcolatrice. Non e' possibile consultare libri e appunti.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede piu' passaggi di calcolo, non sara' preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o piu' frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sara' eventualmente rappresentato in quest'ultima forma).

Problemi

Esercizio 1. Da uno scaffale che contiene 5 romanzi, 3 libri di poesie e 1 dizionario vengono scelti a caso 3 libri

(a) Qual è la probabilità di selezionare 1 dizionario?

Soluzione: Ci sono

$$\#(possibili) = C_{9,3} = \binom{9}{3} = 84$$

modi equiprobabili di scegliere 3 libri su 9.
Il numero di modi favorevoli è il numero di modi in cui 1 dei dizionari e 2 non-dizionari possono essere selezionati:

$$\#(favorevoli) = C_{1,1} \cdot C_{8,2} = \binom{1}{1}\binom{8}{2} = 1 \cdot 28 = 28$$

Dunque:

$$\frac{\#(favorevoli)}{\#(possibili)} = \frac{28}{84} \approx 0.33$$

Esercizio 2. Una catena di negozi di vernici produce e vende solo vernice al lattice e semilucida. Sulla base delle vendite a lungo termine, la probabilità che un cliente acquisti vernice al lattice è di 0.75. Di quelli che acquistano vernice al lattice, il 60% acquista anche i rulli. Solo il 30% degli acquirenti di vernice semilucida, invece, acquistano rulli.

(a) Un acquirente selezionato a caso acquista un rullo e un barattolo di vernice. Qual è la probabilità che la vernice sia al lattice?

Soluzione:
Indicando con L l'evento "un cliente acquista vernice al lattice", si ha

$$P(L) = 0.75$$

Di conseguenza per l'evento S "un cliente acquista vernice semilucida", si ha

$$P(S) = 1 - 0.75 = 0.25$$

Definito R l'evento "il cliente acquista un rullo", si ha che

$$P(R \mid L) = 0.60, P(R \mid S) = 0.30$$

La percentuale di acquirenti di un rullo

$$P(R) = P(R \mid L)P(L) + P(R \mid S)P(S) = 0.6 \cdot 0.75 + 0.3 \cdot 0.25 = 0.525.$$

La probabilità che un acquirente di un rullo acquisti anche un barattolo di vernice al lattice è la probabilità a posteriori $P(L \mid R)$ calcolabile con la regola di Bayes:

$$P(L \mid R) = \frac{P(R \mid L)P(L)}{P(R)} = \frac{0.6 \times 0.75}{0.525} \approx 0.857. \tag{1}$$

Esercizio 3. Sia X una variabile aleatoria con densità

$$f(x) = \begin{cases} x, & \text{per } 0 < x < 1 \\ 2 - x, & \text{per } 1 \leq x < 2 \\ 0, & \text{altrimenti} \end{cases}$$

Determinare la probabilità che X assuma valori

(a) tra 0.6 e 1.2

(b) maggiori di 1.8

Soluzione: Abbiamo che

$$\begin{aligned} P(0.6 < X < 1.2) &= \int_{0.6}^{1.2} f(x)\,dx = \int_{0.6}^{1} x\,dx + \int_{1}^{1.2} (2-x)\,dx \\ &= \frac{x^2}{2}\Big|_{0.6}^{1} + \left(2x - \frac{x^2}{2}\right)\Big|_{1}^{1.2} \\ &= \frac{1}{2} - \frac{0.36}{2} - \frac{3.36}{2} + \frac{3}{2} = 0.5 \end{aligned}$$

e

$$P(X > 1.8) = \int_{1.8}^{2} f(x)\,dx = \int_{1.8}^{2} (2-x)\,dx = \left(2x - \frac{x^2}{2}\right)\Big|_{1.8}^{2} = \frac{0.04}{2} = 0.02$$

Esercizio 4. Durante un esperimento di laboratorio, il numero medio di particelle radioattive che passano attraverso un contatore in 1 millisecondo è 4.

(a) Qual è la probabilità che 6 particelle entrino nel contatore in un dato millisecondo?

Soluzione: Usando la distribuzione di Poisson con $x = 6$ e $\lambda t = 4$ e facendo riferimento alle Tavole Statistiche abbiamo:

$$p(6; 4) = \frac{e^{-4}(4)^6}{6!} = \sum_{x=0}^{6} p(x; 4) - \sum_{x=0}^{5} p(x; 4) = 0.8893 - 0.7851 = 0.1042$$

Esercizio 5. Un distributore di bibite è regolato in modo che eroghi una media di 200 millilitri per bicchiere. Se la quantità di bevanda è normalmente distribuita con una deviazione standard pari a 15 millilitri.

(a) qual è la probabilità che un bicchiere conterrà più di 224 millilitri?

(b) quante tazze probabilmente trabboccheranno se si usano tazze da 230 millilitri per le prossime 1000 bevande?

Soluzione:
Sia X la quantità di bevanda distribuita con $\mu = 200$ e $\sigma = 15$.
(a) $P(X > 224)$ (?)
Innanzitutto, bisogna trovare il valore z corrispondente a $x = 224$

$$z = \frac{x - \mu}{\sigma} = \frac{224 - 200}{15} = 1.6$$

Dunque

$$\begin{aligned} P(X > 224) &= P(Z > 1.6) \\ &= 1 - P(Z > 1.6) \\ &= 1 - F(1.6) \\ &= 1 - 0.9452 \\ &= 0.0548 \end{aligned}$$

(b)
Standardizziamo $x = 230$

$$z = \frac{230 - 200}{15} = 2$$

Dunque

$$P(X > 230) = P(Z > 2) = 1 - P(Z < 2) = 1 - F(2) = 1 - 0.9772 = 0.0228$$

Per conoscere la quantità prevista di tazze che traboccheranno nei prossimi 1000 bicchieri, adottiamo la proprietà binomiale

$$\begin{aligned} E[X] &= n \cdot p \\ &= 1000 \cdot 0.0228 \\ &= 22.8 \approx 23 \end{aligned}$$

Esercizio 6. I punteggi di un test d'ingresso posto alle matricole di un college negli ultimi cinque anni sono distribuiti con legge normale a media $\mu = 74$ e varianza $\sigma^2 = 8$.

(a) Si può considerare ancora $\sigma^2 = 8$ un valore di varianza valido se un campione casuale di 25 studenti che eseguono il test d'ingresso quest'anno ottenesse un valore di $S^2 = 15$?

Soluzione: Se S^2 è la varianza di un campione casuale di dimensione n preso da una popolazione normale avente varianza σ^2, allora la statistica:

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

ha una distribuzione chi-quadro con $v = n - 1$ gradi di libertà.
Quindi,

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{24 \cdot S^2}{8} = 3 \cdot S^2$$

ha una distribuzione chi-quadro con 24 gradi di libertà.
Verificheremo se $\sigma^2 = 8$ è un'assunzione ragionevole determinando la probabilità

$$P(S^2 > 15) = P(\chi^2 > 3 \cdot 15) = P(\chi^2 > 45) < P(\chi^2 > 42.980) \approx 0.01$$

quindi, abbiamo trovato che sotto l'ipotesi $\sigma^2 = 8$, c'è meno di 1% di possibilità che $P(S^2 > 15)$, quindi l'ipotesi non è ragionevole.