Tracce e soluzioni degli esami di

PROBABILITÀ E STATISTICA [3231]

Corso di Studi: Laurea Triennale in Ingegneria Gestionale
Dipartimento di Meccanica, Matematica e Management
Politecnico di Bari

Appelli a.a. 2022–2023

Gianluca Orlando

Indice

1 Tracce

Di seguito le tracce dell'a.a. 2022-2023.

Esame di Probabilità e Statistica [3231]

Esame di Calcolo delle Probabilità e Statistica [2959]

Corso di Studi di Ingegneria Gestionale (D.M.270/04) (L)

Dipartimento di Meccanica, Matematica e Management
Politecnico di Bari

Cognome: ______________________________ Docente: Gianluca Orlando
Nome: ______________________________ Appello: giugno 2023 - turno 1
Matricola: ______________________________ Data: 26/06/2023

Tempo massimo: 2 ore.

Esercizio 1. (6 punti) Si misura il carico di rottura a trazione di alcuni provini di acciaio INOX. Le misurazioni in $MPa = N/mm^2$ sono le seguenti:

792 706 737 580 692 720 711 704 749 725 734 702

1. Determinare i quartili (esclusivi) dei dati.
2. Determinare eventuali dati anomali e sospetti.
3. Tracciare un box plot.

Esercizio 2. (7 punti) Il numero di errori di battitura in una pagina di un libro è distribuito con legge di Poisson. In media, c'è 1 errore di battitura per pagina. Si assumano gli errori in pagine diverse indipendenti.

1. Qual è la probabilità di osservare esattamente 3 errori di battitura in una pagina?
2. Qual è la probabilità di osservare (strettamente) più di 3 errori di battitura nelle prime 4 pagine?
3. Sappiamo che c'è almeno un errore di battitura nelle prime 2 pagine. Tenendo conto di questo, qual è la probabilità che nelle prime 4 pagine ci siano in tutto esattamente 3 errori di battitura?

Esercizio 3. (8 punti) Alice e Bob giocano al seguente gioco. Generano un numero casuale distribuito con legge uniforme nell'intervallo $[0, a]$, dove l'estremo $a > 1$ è da determinare.

1. Calcolare, in funzione del parametro a, la probabilità che il numero generato sia più grande di 1.

Alice e Bob generano tante volte (in modo indipendente) un numero con il procedimento descritto sopra finché viene generato un numero più grande di 1. Se entro la quarta volta (compresa) viene generato un numero più grande di 1, vince Alice. Altrimenti vince Bob.

2. Determinare a in modo che Alice vinca con il 50% di probabilità.

3. In media, qual è la prima volta che viene generato un numero più grande di 1?

Esercizio 4. (7 punti) Si vuole stimare la vita media delle batterie per smartphone prodotte da un'azienda. Si misura la vita delle batterie su un campione casuale e si ottengono le seguenti misurazioni:

$$6.6 \quad 6.7 \quad 7.9 \quad 5.3 \quad 3.2 \quad 1.8 \quad 5.1 \quad 5.9$$

Si supponga che la distribuzione della vita delle batterie sia normale.

1. Calcolare sui dati un intervallo di confidenza bilaterale al 95% per la media.

2. Un intervallo di confidenza calcolato sui dati al 85% sarebbe più grande o più piccolo? (N.B.: Non è richiesto il calcolo esplicito, ma si deve motivare la risposta).

Quesito teorico 1. (4 punti) Siano $X \sim \mathrm{B}(n, p)$ e $Y \sim \mathrm{B}(m, p)$ due variabili aleatorie indipendenti distribuite con legge binomiale. Che legge ha $X + Y$? Motivare la risposta.

Quesito teorico 2. (2 punti) Quale legge continua gode della proprietà di assenza di memoria? Enunciare e dimostrare il risultato.

Esame di Probabilità e Statistica [3231]

Esame di Calcolo delle Probabilità e Statistica [2959]

Corso di Studi di Ingegneria Gestionale (D.M.270/04) (L)

Dipartimento di Meccanica, Matematica e Management
Politecnico di Bari

Cognome: ______________________ Docente: Gianluca Orlando
Nome: ______________________ Appello: giugno 2023 - turno 2
Matricola: ______________________ Data: 26/06/2023

Tempo massimo: 2 ore.

Esercizio 1. (6 punti) Si studia il tempo di vita di una batteria per smartphone. I dati misurati (in anni) vengono raccolti in intervalli e si contano le osservazioni negli intervalli:

intervalli (anni)	frequenza assoluta
$[0,3)$	3
$[3,4)$	4
$[4,5)$	6
$[5,7)$	10
$[7,12)$	8

1. Rappresentare un istogramma delle densità di frequenze relative.
2. Determinare la classe modale.
3. Calcolare un'approssimazione della media e della varianza dei dati.
4. Calcolare un'approssimazione della mediana dei dati.

Esercizio 2. (7 punti) Sia (X_1, X_2) un vettore aleatorio con probabilità congiunta descritta dalla seguente tabella:

X_2 \ X_1	-1	0	1
0	a	$\frac{1}{4}$	$\frac{1}{4}$
1	$\frac{1}{4}$	b	c

dove a, b, c sono parametri da determinare. Si assuma che

- $\mathbb{E}(X_2) = \frac{1}{2}$
- $\mathrm{Var}(X_1 + X_2) = \frac{1}{4}$ (Suggerimento: conviene ricordare la formula per la varianza della somma)

1. Determinare il valore di a utilizzando la condizione $\mathbb{E}(X_2) = \frac{1}{2}$.
2. Determinare il valore di b, c.
3. Stabilire se le variabili aleatorie X_1 e X_2 sono indipendenti per i valori di a, b, c trovati.

Esercizio 3. (8 punti) Il tempo necessario per un certo impiegato allo sportello delle poste per servire un cliente è distribuito con legge esponenziale. In media, il tempo in cui termina di servire il cliente è 7 minuti.

1. Qual è la probabilità che siano necessari più di 10 minuti per servire un cliente?
2. L'impiegato allo sportello inizia a servire un cliente alle 10:00. Passa del tempo, guardiamo l'orologio, sono le 10:10 e l'impiegato non ha ancora terminato di servire il cliente. Qual è la probabilità che il cliente finisca di essere servito dopo le 10:20?
3. Ci sono in tutto 8 clienti da servire. Si assuma che i tempi necessari per servire ciascun cliente siano indipendenti. Qual è la probabilità che l'impiegato termini di servire almeno 3 clienti in più di 10 minuti?

Esercizio 4. (7 punti) Un produttore di acciaio INOX sostiene che il carico di rottura medio del materiale da lui prodotto è 730 $MPa = N/mm^2$. Si vuole stabilire se la media è in realtà più bassa. Si misura il carico di rottura su un campione casuale e si osservano i seguenti risultati:

732	723	729	715	712	721	730	724
749	710	708	691	723	693	734	749
733	715	708	742	728	720	734	708
725	762	729	722	705	730	712	698

La media calcolata sui dati di questo campione risulta essere 722.31. Si supponga che la distribuzione del carico di rottura abbia deviazione standard $\sigma = 20$.

1. I dati sono significativi al 5% per stabilire che la media è effettivamente più bassa di 730?
2. Qual è il più piccolo livello di significatività per cui i dati permettono di affermare che la media è più bassa di 730?

Quesito teorico 1. (3 punti) Calcolare media e varianza di una variabile aleatoria distribuita con legge uniforme in un intervallo.

Quesito teorico 2. (3 punti) Siano $X \sim \mathrm{P}(\lambda)$ e $Y \sim \mathrm{P}(\mu)$ due variabili aleatorie indipendenti distribuite con legge di Poisson. Che legge ha $X + Y$? Motivare la risposta.

Esame di Probabilità e Statistica [3231]

Esame di Calcolo delle Probabilità e Statistica [2959]

Corso di Studi di Ingegneria Gestionale (D.M.270/04) (L)

Dipartimento di Meccanica, Matematica e Management
Politecnico di Bari

Cognome: ____________________	Docente: Gianluca Orlando
Nome: ____________________	Appello: luglio 2023 - turno 1
Matricola: ____________________	Data: 20/07/2023

Tempo massimo: 2 ore.

Esercizio 1. (6 punti) Si analizzano i ritardi di un treno all'arrivo in una stazione in diverse giornate. I dati misurati in minuti sono i seguenti:

$$5 \quad 1 \quad 7 \quad 0 \quad 15 \quad 2 \quad 31 \quad 5$$

1. Determinare i quartili (esclusivi) dei dati.
2. Determinare eventuali dati anomali e sospetti.
3. Tracciare un box plot.

Esercizio 2. (8 punti) Durante l'esame di Probabilità e Statistica, alcune persone dimenticano di portare con sè le tavole delle distribuzioni. Si assuma che, durante un appello, il numero di persone che dimenticano di portare le tavole sia distribuito con una legge di Poisson e che, in media, 6 persone dimentichino le tavole.

1. Viene svolto un appello. Qual è la probabilità che almeno 4 persone dimentichino le tavole?
2. Vengono svolti 3 appelli. Si assumano i tre appelli indipendenti. Qual è la probabilità che nella totalità dei 3 appelli esattamente 20 persone dimentichino le tavole?

N.B.: Nella prossima richiesta, gli appelli non sono indipendenti.

Vengono svolti 2 appelli. Al primo appello il numero di persone che dimenticano di portare le tavole è distribuito con una legge di Poisson con media 6. Per il secondo appello:

- Se al primo appello il numero di persone che dimenticano di portare le tavole è maggiore o uguale a 4, allora il docente comunica che chi non porterà le tavole all'appello successivo avrà un voto più basso. L'effetto è che, nel secondo appello, il numero di persone che dimenticano le tavole è distribuito con una legge di Poisson con media 3.

- Se, invece al primo appello il numero di persone che dimenticano di portare le tavole è strettamente minore di 4, il docente non dice nulla e al secondo appello il numero di persone che dimenticano di portare le tavole è ancora distribuito con legge di Poisson con media 6.

Nelle ipotesi precedenti, rispondere alla seguente richiesta.

3. Sappiamo che al secondo appello esattamente 2 persone hanno dimenticato le tavole. Con che probabilità almeno 4 persone hanno dimenticato le tavole?

Esercizio 3. (7 punti) Devi prendere un mezzo pubblico per arrivare ad un appuntamento. Puoi prendere il bus 13 oppure il bus 17.

- Il bus 13 arriva alla fermata tra le 19:20 e le 19:40 e l'orario di arrivo ha distribuzione uniforme.
- Il bus 17 arriva alla fermata in media alle 19:40, con una varianza di 12 min^2, e l'orario di arrivo ha distribuzione uniforme.

Rispondere ai seguenti quesiti.

1. Calcolare la probabilità che il bus 13 arrivi prima delle 19:25 oppure dopo le 19:35.
2. Calcolare la probabilità che il bus 17 arrivi prima delle 19:38.
3. Calcolare la probabilità che uno dei due bus arrivi prima delle 19:35 (è indifferente se il 13 o il 17). Si assumano i due orari di arrivo indipendenti.

Esercizio 4. (7 punti) Le capacità (in Ah) di 10 batterie sono state registrate come segue:

140 136 150 144 148 152 138 141 143 151

Si assuma che la capacità di una batteria abbia distribuzione normale.

1. Calcolare sui dati un intervallo di confidenza al 95% bilaterale per la varianza della popolazione.
2. Per 8 volte si ripete la misurazione dei dati su un campione della stessa ampiezza di quello di sopra e su ogni campione si calcola sui dati un intervallo di confidenza al 95% bilaterale per la varianza della popolazione con il procedimento del punto 1. Si assumano le misurazioni sui campioni indipendenti. Qual è la probabilità che almeno 6 volte la varianza appartenga all'intervallo calcolato?

Quesito teorico 1. (4 punti) Calcolare la media e la varianza di una variabile aleatoria distribuita con legge Gamma. Che fenomeni descrive la legge Gamma?

Quesito teorico 2. (2 punti) Calcolare la media di una variabile aleatoria distribuita con legge geometrica.

Esame di Probabilità e Statistica [3231]

Esame di Calcolo delle Probabilità e Statistica [2959]

Corso di Studi di Ingegneria Gestionale (D.M.270/04) (L)

Dipartimento di Meccanica, Matematica e Management
Politecnico di Bari

Cognome: ______________________	Docente: Gianluca Orlando
Nome: ______________________	Appello: luglio 2023 - turno 2
Matricola: ______________________	Data: 20/07/2023

Tempo massimo: 2 ore.

Esercizio 1. (6 punti) I seguenti dati mettono in relazione il numero di unità di un bene che sono state ordinate in funzione del prezzo del bene in sei diverse località:

numero ordini	88	112	123	136	158	172
prezzo	50	40	35	30	20	14

1. Rappresentare i dati in uno scatterplot.
2. Determinare (derivando le formule) la retta di regressione lineare e rappresentarla.
3. Determinare il coefficiente di correlazione lineare.

Esercizio 2. (8 punti) La prova scritta di un esame molto difficile è costituita da 4 esercizi. Uno studente si reca all'esame senza aver studiato, sperando di copiare. La probabilità di riuscire a copiare bene un singolo esercizio dipende da quanto è stanco il professore.

- Se il professore è poco stanco, la probabilità di copiare bene un singolo esercizio è del 2%.
- Se il professore è molto stanco, la probabilità di copiare bene un singolo esercizio è del 5%.
- Il professore è molto stanco con il 20% di probabilità.

Si assuma che la copia di un esercizio non influenzi la copia di un altro esercizio. Lo studente supera la prova se riesce a copiare bene almeno 2 esercizi.

1. Il professore annuncia all'inizio della prova: "Oggi sono molto stanco!". Con che probabilità lo studente supera la prova?
2. Il professore non dice nulla all'inizio della prova e non sappiamo se è poco stanco o molto stanco. Finisce l'esame e lo studente supera la prova! Con che probabilità il professore era molto stanco?

Esercizio 3. (8 punti) Devi viaggiare su un treno che percorre il seguente tragitto:

$$A \to B \to C\,.$$

Assumiamo che:

- nella tratta $A \to B$ il treno accumuli un ritardo medio di 2 min e che questo ritardo sia distribuito con legge esponenziale;
- nella tratta $B \to C$ il treno accumuli un ritardo medio di 2 min e che questo ritardo sia distribuito con legge esponenziale;
- il ritardo accumulato nella tratta $A \to B$ e il ritardo accumulato nella tratta $B \to C$ siano indipendenti.

Rispondere ai seguenti quesiti:

1. Qual è la probabilità che nella tratta $A \to B$ il treno accumuli un ritardo compreso tra 4 min e 12 min?
2. In media, qual è il ritardo accumulato nell'intero viaggio $A \to B \to C$? Motivare la risposta.
3. Qual è la varianza del ritardo accumulato nell'intero viaggio $A \to B \to C$? Motivare la risposta.
4. Considerare la probabilità che il ritardo accumulato nell'intero viaggio $A \to B \to C$ sia maggiore di 10 min. Stabilire se questa probabilità è maggiore o minore di 5%.

Esercizio 4. (7 punti) Un'auto è pubblicizzata come avente un consumo medio di benzina inferiore a 4.3 ℓ/100km (litri per 100 km) nella guida in autostrada. Si misura il consumo in un campione e si ottengono i seguenti dati (ℓ/100km):

3.4	3.3	5.5	5.0	3.2	3.9	3.8	4.1	5.3	5.3
3.7	5.7	5.5	5.6	5.7	5.1	4.9	4.5	3.1	5.6
3.1	4.8	5.5	5.1	6.4	3.5	5.6	6.4	4.4	4.6
4.8	3.8	4.2	5.4	3.0	4.2	5.7	6.0	4.3	4.4

La media calcolata sul campione risulta essere 4.7 ℓ/100km. Si sa che la deviazione standard dei consumi è 1.0 ℓ/100km.

1. I dati sono significativi al 10% per non credere alla pubblicità?
2. Assumiamo che la pubblicità dica il vero. Vengono misurati i dati di campioni della stessa ampiezza di quello precedente tante volte e ogni volta viene effettuato un test come nel punto 1. In media, dopo quanti tentativi osserveremo dei dati significativi al 10% per non credere alla pubblicità?

Quesito teorico 1. (4 punti) Che fenomeni descrive la legge di Poisson? Motivare la risposta enunciando e dimostrando un teorema.

Quesito teorico 2. (2 punti) Calcolare la funzione di distribuzione cumulativa di una variabile aleatoria distribuita con legge uniforme in un intervallo.

Esame di Probabilità e Statistica [3231]

Esame di Calcolo delle Probabilità e Statistica [2959]

Corso di Studi di Ingegneria Gestionale (D.M.270/04) (L)

Dipartimento di Meccanica, Matematica e Management
Politecnico di Bari

Cognome: ____________________	Docente: Gianluca Orlando
Nome: ____________________	Appello: settembre 2023 - I
Matricola: ____________________	Data: 05/09/2023

Tempo massimo: 2 ore.

Esercizio 1. (6 punti) Ci si aspetta che la resa percentuale di una reazione chimica sia correlata con la temperatura a cui avviene la reazione. Nella seguente tabella sono rappresentati i dati di un campione:

temperatura ($°C$)	45	55	60	70	85	90
resa (%)	43	48	51	57	66	68

1. Rappresentare i dati in uno scatterplot.
2. Determinare (derivando le formule) la retta di regressione lineare e rappresentarla.
3. Determinare il coefficiente di correlazione lineare.

Esercizio 2. (7 punti) Un'azienda produce *cookies* con gocce di cioccolato. Si assuma che il numero di gocce di cioccolato in un *cookie* sia distribuito con legge di Poisson e che, in media, un *cookie* abbia 10 gocce di cioccolato al suo interno. Si assuma che i numeri di gocce di cioccolato in *cookies* diversi siano indipendenti.

1. Qual è la probabilità che il numero di gocce di cioccolato in un *cookie* sia compreso tra 8 e 12 (estremi inclusi)?
2. Qual è la probabilità che in due *cookies* ci siano in tutto esattamente 30 gocce di cioccolato?

Una confezione di *cookies* contiene 6 *cookies*.

3. Compriamo una confezione. Qual è la probabilità che almeno 3 *cookies* nella confezione contengano ciascuno un numero gocce di cioccolato compreso tra 8 e 12?
4. In media, in una confezione quanti *cookies* contengono un numero gocce di cioccolato compreso tra 8 e 12?

Esercizio 3. (7 punti) Ad un ristorante il tempo di attesa al tavolo dal momento dell'ordine all'arrivo del piatto dipende dal piatto ordinato.

- Se si ordina la pizza, il tempo di attesa è distribuito con legge esponenziale con media 15 minuti.
- Se si ordinano gli spaghetti all'assassina, il tempo di attesa è distribuito con legge esponenziale con una deviazione standard di 30 minuti.
- Il 60% delle persone ordina la pizza, mentre il 40% delle persone ordina gli spaghetti all'assassina.

Al tuo tavolo siete due persone. Si assumano i tempi di preparazione dei vari piatti indipendenti.

1. Tu ordini una pizza. Qual è la probabilità che la pizza sia pronta in un tempo compreso tra i 10 e i 20 minuti?
2. Il/la tuo/a partner ordina spaghetti all'assassina. Qual è la probabilità che gli spaghetti siano pronti dopo 40 minuti?
3. Il servizio porterà al tavolo la pizza e gli spaghetti insieme (quindi il tempo di attesa totale dipenderà da quale preparazione impiega più tempo ad essere pronta). Qual è la probabilità di dover aspettare più di 40 minuti perché arrivino entrambi i piatti?
4. Una persona sola al tavolo a fianco si sta lamentando del servizio perché aspetta da 40 minuti e ancora non è arrivato il suo piatto. Qual è la probabilità che abbia ordinato una pizza?

Esercizio 4. (8 punti) Un responsabile di controllo di qualità misura la distanza tra le piastre degli scambiatori di calore prodotti da un'azienda. Le misurazioni su un campione di 20 scambiatori di calore forniscono una distanza media calcolata sul campione di 4.0 mm e una deviazione standard calcolata sul campione di 0.1 mm. Si assuma che la distribuzione della distanza tra le piastre sia normale.

1. Calcolare sui dati un intervallo di confidenza al 95% bilaterale per la distanza media.
2. Calcolare sui dati un intervallo di confidenza al 90% bilaterale per la varianza della distanza.

Derivare il calcolo degli intervalli di confidenza in entrambi i punti, motivando i passaggi.

Quesito teorico 1. (4 punti) Enunciare e dimostrare la legge dei grandi numeri.

Quesito teorico 2. (2 punti) Spiegare il paradosso della scimmia instancabile di Borel (utilizzare un insieme di 67 caratteri e un testo composto da un numero di caratteri pari al proprio numero di matricola).

Esame di Probabilità e Statistica [3231]

Esame di Calcolo delle Probabilità e Statistica [2959]

Corso di Studi di Ingegneria Gestionale (D.M.270/04) (L)

Dipartimento di Meccanica, Matematica e Management
Politecnico di Bari

Cognome: ____________________ Docente: Gianluca Orlando
Nome: ____________________ Appello: settembre 2023 - II
Matricola: ____________________ Data: 19/09/2023

Tempo massimo: 2 ore.

Esercizio 1. (6 punti) Il numero di veicoli venduti da un importante showroom di auto in un giorno è stato registrato per 10 giorni lavorativi. I dati sono i seguenti:

$$20 \quad 15 \quad 18 \quad 3 \quad 10 \quad 17 \quad 21 \quad 19 \quad 25 \quad 28$$

1. Determinare i quartili (esclusivi) dei dati.
2. Determinare eventuali dati anomali e sospetti.
3. Tracciare un box plot.

Esercizio 2. (7 punti) Si consideri un vettore aleatorio (X_1, X_2) avente funzione di probabilità congiunta descritta dalla seguente tabella:

X_2 \ X_1	1	2	3
1	a_{11}	a_{21}	a_{31}
2	a_{12}	a_{22}	a_{32}
3	a_{13}	a_{23}	a_{33}

con $a_{11}, a_{21}, a_{31}, a_{12}, a_{22}, a_{32}, a_{13}, a_{23}, a_{33} > 0$.

1. Calcolare il range dei possibili valori che possono essere assunti dalla variabile aleatoria $Y = \min\{X_1 + X_2, X_1 \cdot X_2\}$ (valore più piccolo tra $X_1 + X_2$ e $X_1 \cdot X_2$).

Si assuma che i valori nel range di Y siano equiprobabili.

2. Determinare i valori di a_{11}, a_{22}, a_{33}.

Si assuma che:

- $\mathbb{P}(\{X_1 = 2\} \cap \{X_2 = 1\}) = \frac{1}{12}$.

- $\mathbb{P}(\{X_1 = 3\} \cap \{X_2 = 1\}) = \mathbb{P}(\{W < \frac{1}{12}\})$ dove $W \sim \mathrm{U}(0,1)$.
- $\mathbb{P}(\{X_1 = 3\} | \{X_2 = 2\}) = \frac{1}{4}$.

3. Determinare i valori rimanenti $a_{21}, a_{31}, a_{12}, a_{32}, a_{13}, a_{23}$.
4. Stabilire se X_1 e X_2 sono indipendenti.

Esercizio 3. (8 punti) Alice e Bob fanno un gioco. Chiamano un numero verde e attendono che un/a operatore/trice risponda. Il tempo di attesa per la risposta è distribuito con legge esponenziale. Se la risposta avviene entro i 10 minuti vince Alice, altrimenti vince Bob. Supponiamo che il gioco sia equo (Alice vince con il 50% di probabilità).

1. In media, dopo quanto tempo risponde l'operatore/trice?
2. Calcolare la probabilità che la risposta avvenga tra 5 minuti e 10 minuti.
3. Alice e Bob chiamano il numero verde. Hanno aspettato 5 minuti e ancora nessuno ha risposto. Come viene aggiornata la probabilità che Bob vinca? Motivare la risposta.

Alice e Bob giocano 2 volte in sequenza. Appena qualcuno risponde, chiudono la chiamata e richiamano. Si assumano i tempi di attesa indipendenti.

4. Calcolare la varianza della durata dell'intera sequenza di chiamate.
5. Calcolare la probabilità che l'intera sequenza di chiamate duri più di 30 min. (Suggerimento: integrare per parti)

Esercizio 4. (7 punti) Supponiamo di eseguire uno studio immunologico in un campione di individui, studiando la reazione all'antigene nel punto di inoculo. Viene misurato il diametro in *mm* dell'alone cutaneo in un campione, ottenendo i seguenti dati:

$$21 \quad 21 \quad 40 \quad 36 \quad 32 \quad 25 \quad 8 \quad 27$$

Si assuma che la distribuzione del diametro sia normale.

1. Si può stabilire con significatività del 5% che la media del diametro è diversa da 30 *mm*?
2. Si può stabilire con significatività del 5% che la varianza del diametro è superiore a 100 mm^2?

In entrambi i casi, derivare le formule utilizzate per rispondere.

Quesito teorico 1. (4 punti) Enunciare il Teorema del Limite Centrale e spiegare come può essere utilizzato per approssimare la legge binomiale.

Quesito teorico 2. (2 punti) Calcolare la media di una variabile aleatoria distribuita con legge di Poisson.

Esame di Probabilità e Statistica [3231]

Esame di Calcolo delle Probabilità e Statistica [2959]

Corso di Studi di Ingegneria Gestionale (D.M.270/04) (L)

Dipartimento di Meccanica, Matematica e Management
Politecnico di Bari

Cognome: ____________________	Docente: Gianluca Orlando
Nome: ____________________	Appello: novembre 2023
Matricola: ____________________	Data: 06/11/2023

Tempo massimo: 2 ore.

Esercizio 1. (6 punti) È stato misurato il tempo di attesa (in giorni) per un particolare intervento chirurgico su un campione di individui. La distribuzione dei tempi di attesa è riassunta nella seguente tabella:

intervalli (giorni)	frequenza assoluta
$[0, 60)$	1
$[60, 90)$	34
$[90, 100)$	16
$[100, 110)$	18
$[110, 130)$	19
$[130, 200)$	12

1. Rappresentare un istogramma delle densità di frequenze assolute.
2. Determinare la classe modale.
3. Calcolare un'approssimazione della media e della deviazione standard dei dati.
4. Calcolare un'approssimazione della mediana dei dati.

Esercizio 2. (8 punti) Si consideri un vettore aleatorio discreto (X_1, X_2) con $R(X_1) = \{0, 1\}$ e $R(X_2) = \{0, 1\}$. Si assuma che $\mathrm{Cov}(X_1, X_2) = 0$.

1. Dimostrare che X_1 e X_2 sono indipendenti. (N.B.: Il punteggio massimo si ottiene se non si utilizzano le informazioni di sotto.)

Si assuma ulteriormente che: $\mathrm{Var}(X_1) = \frac{1}{4}$ e $\mathbb{P}(\{X_2 = 1\}|\{X_1 = 1\}) = \frac{1}{3}$ Quesiti:

2. Calcolare $\mathbb{P}(\{X_1 \cdot X_2 > 0\})$.
3. Calcolare $\mathbb{P}(\{X_1 + X_2 = 1\})$.

4. Alice e Bob giocano al seguente gioco in cui si susseguono in *lanci*: Alice inizia e *lancia* generando un numero con X_1; Bob segue e *lancia* generando un numero con X_2. Alice e Bob continuano a *lanciare* alternandosi, finché i risultati degli ultimi due *lanci* non sono diversi. In tal caso vince chi dei due ha ottenuto il risultato più grande negli ultimi due lanci. Che probabilità di vincere ha Alice?

Esercizio 3. (8 punti) Si assuma che il ritardo di un singolo treno di un'azienda di trasporti sia distribuito con legge esponenziale. Sono gestiti due tipi di treni: il 20% dei treni sono treni veloci; il resto sono treni regionali. Si assumano i seguenti fatti:

- il ritardo medio di un treno veloce è 5 minuti;
- la probabilità che un treno regionale ritardi meno di $\ln(3^{10})$ minuti è uguale a $\frac{2}{3}$.

Rispondere ai seguenti quesiti:

1. Calcolare la probabilità che il ritardo di un treno veloce sia compreso tra 3 e 7 minuti.
2. Calcolare la deviazione standard del ritardo di un treno regionale.
3. Si consideri un treno qualunque. Qual è la probabilità che ritardi più di 8 minuti?
4. (Difficile) Sei su un treno veloce e devi prendere una coincidenza con un treno regionale. La tabella di marcia prevede che il treno veloce e il treno regionale arrivino alla stazione entrambi alle 14:00. Considerando i possibili ritardi, con che probabilità perderai la coincidenza? Supponi che i ritardi siano indipendenti e che il tempo per passare da un treno all'altro sia trascurabile.

Esercizio 4. (7 punti) Un'azienda farmaceutica produce tranquillanti. Si assuma che la distribuzione della durata della loro efficacia sia normale. Un laboratorio di un ospedale ha sperimentato il farmaco su un campione casuale di 6 pazienti. Il periodo di efficacia del tranquillante per ogni paziente (in ore) è stato il seguente:

$$2.7 \quad 2.8 \quad 3.0 \quad 2.3 \quad 2.3 \quad 2.2\,.$$

1. Calcolare sui dati un intervallo di confidenza bilaterale al 90% per la media.
2. Supponiamo che l'esperimento di sopra venga ripetuto 5 volte (in ogni esperimento vengono misurati i dati di 6 pazienti). Dopo ogni campionamento, viene calcolato un intervallo di confidenza bilaterale al 90% per la media come sopra. Qual è la probabilità che la media vera sia effettivamente negli intervalli calcolati almeno 4 volte?

Quesito teorico 1. (4 punti) Siano $X \sim \mathcal{N}(\mu, \sigma^2)$ e $Y \sim \mathcal{N}(\nu, \tau^2)$ indipendenti. Che legge ha $X + Y$? Dimostrare il risultato.

Quesito teorico 2. (2 punti) Sia $X \sim \mathrm{P}(\lambda)$. Calcolare $\mathbb{E}(X)$.

Esame di Probabilità e Statistica [3231]

Esame di Calcolo delle Probabilità e Statistica [2959]

Corso di Studi di Ingegneria Gestionale (D.M.270/04) (L)

Dipartimento di Meccanica, Matematica e Management
Politecnico di Bari

Cognome: ______________________ Docente: Gianluca Orlando
Nome: ______________________ Appello: gennaio 2024
Matricola: ______________________ Data: 23/01/2024

Tempo massimo: 2 ore.

Esercizio 1. (6 punti) Uno studio esamina la relazione tra l'esposizione al rumore e l'ipertensione. Seguono alcuni dati. I valori x rappresentano il livello di pressione sonora riportato in decibel e i valori y rappresentano l'incremento di pressione sanguigna in mm Hg:

x	60	85	100	70	80	63	90
y	1	5	9	1	3	0	4

1. Rappresentare i dati in uno scatterplot.
2. Determinare (derivando le formule) la retta di regressione lineare e rappresentarla.
3. Determinare il coefficiente di correlazione lineare.

Esercizio 2. (7 punti) Un vettore aleatorio (X_1, X_2) ha funzione di probabilità congiunta rappresentata dalla seguente tabella:

X_2 \ X_1	0	1	2
0	a_{00}	a_{10}	a_{20}
1	a_{01}	a_{11}	a_{21}

Si lavori sotto le seguenti ipotesi:

- Il range è dato da $R(X_1, X_2) = \{(0, 1), (1, 0), (2, 0), (1, 1)\}$.
- $\mathbb{E}(X_1 + X_2) = \frac{3}{2}$.
- Sapendo che si è verificato l'evento $\{X_2 = 0\}$, la probabilità che $\{X_1 = 2\}$ è $\frac{1}{2}$.
- La probabilità che $\{X_1 = 1\}$ è uguale alla probabilità che $\{X_2 = 1\}$.

Si risponda ai seguenti quesiti:

1. Determinare i valori $a_{00}, a_{10}, a_{20}, a_{01}, a_{11}, a_{21}$.
2. Stabilire se X_1 e X_2 sono indipendenti.
3. Calcolare $\mathrm{Var}(X_1 + X_2)$.

Esercizio 3. (8 punti) Un'azienda produce un dispositivo composto da due componenti, A e B. Si assuma che:

- La durata di funzionamento dei componenti A e la durata di funzionamento dei componenti B (in anni) siano variabili aleatorie indipendenti distribuite con legge esponenziale.
- La durata di funzionamento media dei componenti A sia 3 anni.
- La deviazione standard della durata di funzionamento dei componenti B sia 2 anni.

Si risponda ai seguenti quesiti:

1. Qual è la probabilità che un componente A abbia una durata di funzionamento compresa tra 2 e 5 anni?
2. Consideriamo un componente A. Abbiamo osservato che funziona ancora a 2 anni dall'acquisto. Qual è la probabilità che smetta di funzionare dopo 5 anni dall'acquisto? Enunciare e dimostrare la proprietà utilizzata.
3. Qual è la probabilità che un componente di tipo B smetta di funzionare dopo 4 anni?
4. Basta che uno dei due componenti A oppure B del dispositivo si rompa, perché il dispositivo smetta di funzionare. Qual è la durata di funzionamento media dell'intero dispositivo?

Esercizio 4. (7 punti) Un ingegnere sta studiando la durata degli pneumatici per una nuova miscela di gomma. Ha testato un campione di 40 pneumatici fino alla fine del loro ciclo di vita in una prova su strada. La media calcolata sui dati è 60973.82 km. Si assuma che la deviazione standard della popolazione sia 3500 km.

1. Si calcoli sui dati un intervallo di confidenza bilaterale al 90% per la media reale della vita degli pneumatici. (N.B.: ricavare le formule)
2. Si calcoli sui dati un intervallo di confidenza unilaterale destro (limite inferiore di confidenza) al 95% per la media reale della vita degli pneumatici.
3. Supponiamo di svolgere l'esperimento di sopra (misurare un campione di 40 pneumatici) tante volte consecutivamente. In media, quale è la prima volta che la media reale della vita degli pneumatici sia fuori dall'intervallo di confidenza bilaterale al 90%?

Quesito teorico 1. (3 punti) Spiegare in che senso la legge di Poisson approssima la legge binomiale enunciando e dimostrando un teorema.

Quesito teorico 2. (3 punti) Siano $Z_1, \ldots, Z_n \sim \mathcal{N}(0,1)$ indipendenti. Che legge ha $Z_1^2 + \cdots + Z_n^2$? Motivare la risposta.

2 Soluzioni

Di seguito le soluzioni relative alle tracce di sopra.

Esame di Probabilità e Statistica [3231]

Esame di Calcolo delle Probabilità e Statistica [2959]

Corso di Studi di Ingegneria Gestionale (D.M.270/04) (L)

Dipartimento di Meccanica, Matematica e Management
Politecnico di Bari

Cognome: ____________________ Docente: Gianluca Orlando
Nome: ____________________ Appello: giugno 2023 - turno 1
Matricola: ____________________ Data: 26/06/2023

Tempo massimo: 2 ore.

Esercizio 1. (6 punti) Si misura il carico di rottura a trazione di alcuni provini di acciaio INOX. Le misurazioni in $MPa = N/mm^2$ sono le seguenti:

$$792 \quad 706 \quad 737 \quad 580 \quad 692 \quad 720 \quad 711 \quad 704 \quad 749 \quad 725 \quad 734 \quad 702$$

1. Determinare i quartili (esclusivi) dei dati.
2. Determinare eventuali dati anomali e sospetti.
3. Tracciare un box plot.

Soluzione. 1. Per prima cosa ordiniamo i dati:

$$580 \quad 692 \quad 702 \quad 704 \quad 706 \quad 711 \quad 720 \quad 725 \quad 734 \quad 737 \quad 749 \quad 792$$

Abbiamo $n = 12$ dati.

Calcoliamo il primo quartile: $\frac{n+1}{4} = \frac{13}{4} = 3 + 0.25$. Allora

$$Q_1 = (1 - 0.25)x_3 + 0.25x_4 = 0.75 \cdot 702 + 0.25 \cdot 704 = 702.5\,.$$

Calcoliamo il secondo quartile: $(n+1)\frac{2}{4} = 13\frac{2}{4} = 6 + 0.5$. Allora

$$Q_2 = (1 - 0.5)x_6 + 0.5x_7 = 0.5 \cdot 711 + 0.5 \cdot 720 = 715.5\,.$$

Calcoliamo il terzo quartile: $(n+1)\frac{3}{4} = 13\frac{3}{4} = 9 + 0.75$. Allora

$$Q_3 = (1 - 0.75)x_9 + 0.75x_{10} = 0.25 \cdot 734 + 0.75 \cdot 737 = 736.25\,.$$

2. Per determinare i dati anomali e sospetti calcoliamo il range interquartile

$$IQR = Q_3 - Q_1 = 736.25 - 702.5 = 33.75\,.$$

I dati anomali apparterrebbero agli intervalli

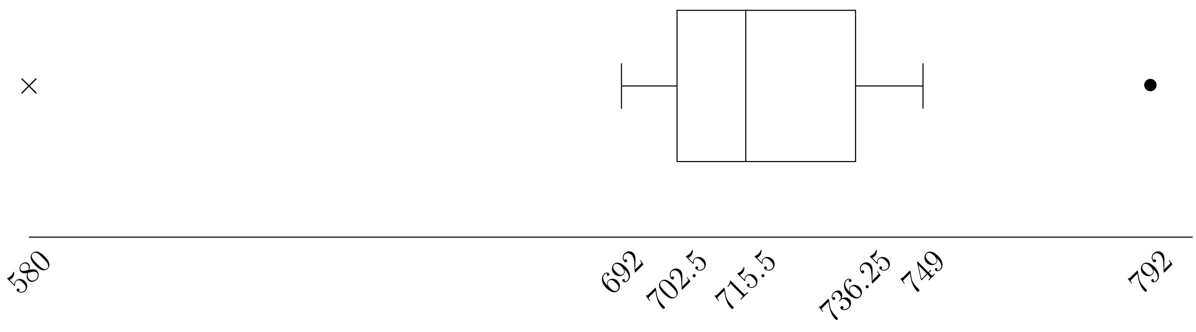
$$(-\infty, Q_1 - 3 \cdot IQR] \cup [Q_3 + 3 \cdot IQR, +\infty) = (-\infty, 601.25] \cup [837.5, +\infty)\,,$$

quindi 580 è un dato anomalo. I dati sospetti appartengono agli intervalli

$$(Q_1-3\cdot IQR, Q_1-1.5\cdot IQR]\cup[Q_3+1.5\cdot IQR, Q_3+3\cdot IQR) = (601.25, 651.875]\cup[786.875, 837.5)\,,$$

quindi 792 è un dato sospetto.

3. Segue il box-plot.



Esercizio 2. (7 punti) Il numero di errori di battitura in una pagina di un libro è distribuito con legge di Poisson. In media, c'è 1 errore di battitura per pagina. Si assumano gli errori in pagine diverse indipendenti.

1. Qual è la probabilità di osservare esattamente 3 errori di battitura in una pagina?
2. Qual è la probabilità di osservare (strettamente) più di 3 errori di battitura nelle prime 4 pagine?
3. Sappiamo che c'è almeno un errore di battitura nelle prime 2 pagine. Tenendo conto di questo, qual è la probabilità che nelle prime 4 pagine ci siano in tutto esattamente 3 errori di battitura?

Soluzione. Consideriamo la variabile aleatoria $X \sim \mathrm{P}(\lambda)$. Ci viene detto che $\mathbb{E}(X) = 1$. Ricordando che $\mathbb{E}(X) = \lambda$, concludiamo che $\lambda = 1$.

1. Calcoliamo

$$\mathbb{P}(\{X = 3\}) = e^{-\lambda}\frac{\lambda^3}{3!} = e^{-1}\frac{1}{6} = 6.13\%\,.$$

2. Consideriamo 4 variabili aleatorie indipendenti $X_1, X_2, X_3, X_4 \sim \mathrm{P}(\lambda)$. Ricordando che la somma di variabili aleatorie di Poisson indipendenti è distribuita con legge di Poisson con parametro dato dalla somma dei parametri, otteniamo che

$$Y = X_1 + X_2 + X_3 + X_4 \sim \mathrm{P}(4\lambda) = \mathrm{P}(4)\,.$$

Possiamo calcolare

$$\begin{aligned}\mathbb{P}(\{Y > 3\}) &= 1 - \mathbb{P}(\{Y \leq 3\}) = 1 - \mathbb{P}(\{Y = 0\}) - \mathbb{P}(\{Y = 1\}) - \mathbb{P}(\{Y = 2\}) - \mathbb{P}(\{Y = 3\}) \\ &= 1 - e^{-4}\Big(1 + 4 + \frac{4^2}{2} + \frac{4^3}{3!}\Big) = 56.65\%\,.\end{aligned}$$

3. Consideriamo

$$Z_1 = X_1 + X_2 \sim \mathrm{P}(2\lambda) = \mathrm{P}(2)$$
$$Z_2 = X_3 + X_4 \sim \mathrm{P}(2\lambda) = \mathrm{P}(2)$$

e osserviamo che $Y = Z_1 + Z_2$ e che Z_1 e Z_2 sono indipendenti. Dobbiamo calcolare

$$\begin{aligned}
\mathbb{P}(\{Y=3\}|\{Z_1 \geq 1\}) &= \frac{\mathbb{P}(\{Z_1+Z_2=3\} \cap \{Z_1 \geq 1\})}{\mathbb{P}(\{Z_1 \geq 1\})} \\
&= \frac{\mathbb{P}(\{Z_2=2\} \cap \{Z_1=1\}) + \mathbb{P}(\{Z_2=1\} \cap \{Z_1=2\}) + \mathbb{P}(\{Z_2=0\} \cap \{Z_1=3\})}{\mathbb{P}(\{Z_1 \geq 1\})} \\
&= \frac{\mathbb{P}(\{Z_2=2\})\mathbb{P}(\{Z_1=1\}) + \mathbb{P}(\{Z_2=1\})\mathbb{P}(\{Z_1=2\}) + \mathbb{P}(\{Z_2=0\})\mathbb{P}(\{Z_1=3\})}{1-\mathbb{P}(\{Z_1=0\})} \\
&= \frac{e^{-2}\cdot\frac{2^2}{2}\cdot e^{-2}\cdot 2 + e^{-2}\cdot 2\cdot e^{-1}\cdot\frac{2^2}{2} + e^{-2}\cdot 1\cdot e^{-2}\cdot\frac{2^3}{3!}}{1-e^{-2}} = 19.77\%\,.
\end{aligned}$$

Esercizio 3. (8 punti) Alice e Bob giocano al seguente gioco. Generano un numero casuale distribuito con legge uniforme nell'intervallo $[0, a]$, dove l'estremo $a > 1$ è da determinare.

1. Calcolare, in funzione del parametro a, la probabilità che il numero generato sia più grande di 1.

Alice e Bob generano tante volte (in modo indipendente) un numero con il procedimento descritto sopra finché non viene generato un numero più grande di 1. Se entro la quarta volta (compresa) viene generato un numero più grande di 1, vince Alice. Altrimenti vince Bob.

2. Determinare a in modo che Alice vinca con il 50% di probabilità.

3. In media, qual è la prima volta che viene generato un numero più grande di 1?

Soluzione. 1. Sia $X \sim \mathrm{U}(0, a)$. Allora

$$\mathbb{P}(\{X \geq 1\}) = \int_1^a \frac{1}{a}\,\mathrm{d}x = \frac{a-1}{a}\,.$$

2. Per descrivere il fenomeno, consideriamo una variabile aleatoria con legge geometrica $Y \sim \mathrm{Geo}(p)$ dove p è la probabilità di un successo, ovvero $p = \frac{a-1}{a}$, come calcolato nel punto precedente. Imponiamo che

$$\begin{aligned}
&50\% = \mathbb{P}(\{Y \leq 4\}) = 1 - \mathbb{P}(\{Y > 4\}) = 1-(1-p)^4 \implies (1-p)^4 = \frac{1}{2} \\
&\implies 1-p = \frac{1}{\sqrt[4]{2}} \implies p = 1 - \frac{1}{\sqrt[4]{2}}\,.
\end{aligned}$$

Segue che

$$\frac{a-1}{a} = p \implies a-1 = pa \implies (1-p)a = 1 \implies a = \frac{1}{1-p} = \sqrt[4]{2} \simeq 1.189\,.$$

3. In media, la prima volta che viene generato un numero più grande di 1 è

$$\mathbb{E}(Y) = \frac{1}{p} = \frac{a}{a-1} \simeq 6.29\,.$$

Esercizio 4. (7 punti) Si vuole stimare la vita media delle batterie per smartphone prodotte da un'azienda. Si misura la vita delle batterie su un campione casuale e si ottengono le seguenti misurazioni:

$$6.6 \quad 6.7 \quad 7.9 \quad 5.3 \quad 3.2 \quad 1.8 \quad 5.1 \quad 5.9$$

Si supponga che la distribuzione della vita delle batterie sia normale.

1. Calcolare sui dati un intervallo di confidenza bilaterale al 95% per la media.

2. Un intervallo di confidenza calcolato sui dati al 85% sarebbe più grande o più piccolo? (N.B.: Non è richiesto il calcolo esplicito, ma si deve motivare la risposta).

Soluzione. Abbiamo un campione casuale $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ con $n = 8$. Vogliamo calcolare un intervallo di confidenza di livello $\beta = 95\%$.

1. Dalla definizione di IC si ha che

$$\beta = \mathbb{P}(\{U_n \le \mu \le V_n\}) \,.$$

Per stimare μ sfrutteremo lo stimatore media campionaria $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Per stimare σ^2 sfrutteremo lo stimatore varianza campionaria $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$. Poiché $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ sono indipendenti, $T_{n-1} = \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \sim t(n-1)$. Allora

$$\begin{aligned}
\beta = \mathbb{P}(\{U_n \le \mu \le V_n\}) &= \mathbb{P}\Big(\Big\{\frac{\overline{X}_n - V_n}{S_n/\sqrt{n}} \le \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \le \frac{\overline{X}_n - U_n}{S_n/\sqrt{n}}\Big\}\Big) \\
&= \mathbb{P}\Big(\Big\{\frac{\overline{X}_n - V_n}{S_n/\sqrt{n}} \le T_{n-1} \le \frac{\overline{X}_n - U_n}{S_n/\sqrt{n}}\Big\}\Big) \\
&= 1 - \mathbb{P}\Big(\Big\{T_{n-1} < \frac{\overline{X}_n - V_n}{S_n/\sqrt{n}}\Big\}\Big) - \mathbb{P}\Big(\Big\{T_{n-1} > \frac{\overline{X}_n - U_n}{S_n/\sqrt{n}}\Big\}\Big) \,.
\end{aligned}$$

Segue che

$$\mathbb{P}\Big(\Big\{T_{n-1} < \frac{\overline{X}_n - V_n}{S_n/\sqrt{n}}\Big\}\Big) + \mathbb{P}\Big(\Big\{T_{n-1} > \frac{\overline{X}_n - U_n}{S_n/\sqrt{n}}\Big\}\Big) = 1 - \beta = \alpha \,.$$

Decidiamo di equipartire α:

$$\mathbb{P}\Big(\Big\{T_{n-1} < \frac{\overline{X}_n - V_n}{S_n/\sqrt{n}}\Big\}\Big) = \mathbb{P}\Big(\Big\{T_{n-1} < \frac{\overline{X}_n - V_n}{S_n/\sqrt{n}}\Big\}\Big) = \frac{\alpha}{2} \,.$$

Definiamo $t_{n-1,\alpha/2}$ come il punto tale che

$$\mathbb{P}(\{T_{n-1} \ge t_{n-1,\alpha/2}) = \frac{\alpha}{2} \,.$$

Scegliendo

$$\begin{aligned}
\frac{\overline{X}_n - U_n}{S_n/\sqrt{n}} &= t_{n-1,\alpha/2} \implies U_n = \overline{X}_n - \frac{S_n}{\sqrt{n}} t_{n-1,\alpha/2} \,, \\
\frac{\overline{X}_n - V_n}{S_n/\sqrt{n}} &= -t_{n-1,\alpha/2} \implies V_n = \overline{X}_n + \frac{S_n}{\sqrt{n}} t_{n-1,\alpha/2} \,,
\end{aligned}$$

si ottiene la condizione che definisce il livello di confidenza. In conclusione

$$\Big[\overline{X}_n - \frac{S_n}{\sqrt{n}} t_{n-1,\alpha/2}, \overline{X}_n + \frac{S_n}{\sqrt{n}} t_{n-1,\alpha/2}\Big]$$

è un intervallo di confidenza bilaterale per μ con livello di confidenza $\beta = 1 - \alpha$.

Calcoliamo l'intervallo di confidenza sui dati

$$\Big[\overline{x}_n - \frac{s_n}{\sqrt{n}} t_{n-1,\alpha/2}, \overline{x}_n + \frac{s_n}{\sqrt{n}} t_{n-1,\alpha/2}\Big]$$

dove

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{8}(6.6 + 6.7 + 7.9 + 5.3 + 3.2 + 1.8 + 5.1 + 5.9) = 5.3125$$

è la media calcolata sui dati e

$$\begin{aligned} s_n^2 &= \frac{1}{n-1}\Big(\sum_{i=1}^{n} x_i^2 - n\overline{x}_n^2\Big) \\ &= \frac{1}{7}\Big(6.6^2 + 6.7^2 + 7.9^2 + 5.3^2 + 3.2^2 + 1.8^2 + 5.1^2 + 5.9^2 - 8 \cdot 5.3125^2\Big) \\ &\simeq 3.9241 \end{aligned}$$

è la varianza calcolata sui dati, da cui $s_n \simeq 1.98$. Dalla tavola della t-Student calcoliamo

$$t_{n-1,\alpha/2} = t_{7,0.05} = 2.365\,.$$

Quindi l'intervallo di confidenza calcolato sui dati è

$$\Big[\overline{x}_n - \frac{s_n}{\sqrt{n}} t_{n-1,\alpha/2}, \overline{x}_n + \frac{s_n}{\sqrt{n}} t_{n-1,\alpha/2}\Big] \simeq [3.66, 6.97]\,.$$

2. L'intervallo di confidenza sarebbe più piccolo in quanto la probabilità che μ sia nell'intervallo $[U_n, V_n]$ sarebbe più piccola. Si può anche osservare dalla tavola della t-Student che per $\alpha = 7.5\%$ si otterrebbe un valore più piccolo di $t_{n-1,\alpha/2}$.

Esame di Probabilità e Statistica [3231]

Esame di Calcolo delle Probabilità e Statistica [2959]

Corso di Studi di Ingegneria Gestionale (D.M.270/04) (L)

Dipartimento di Meccanica, Matematica e Management
Politecnico di Bari

Cognome: ______________________ Docente: Gianluca Orlando
Nome: ______________________ Appello: giugno 2023 - turno 2
Matricola: ______________________ Data: 26/06/2023

Tempo massimo: 2 ore.

Esercizio 1. (6 punti) Si studia il tempo di vita di una batteria per smartphone. I dati misurati (in anni) vengono raccolti in intervalli e si contano le osservazioni negli intervalli:

intervalli (anni)	frequenza assoluta
$[0, 3)$	3
$[3, 4)$	4
$[4, 5)$	6
$[5, 7)$	10
$[7, 12)$	8

1. Rappresentare un istogramma delle densità di frequenze relative.

2. Determinare la classe modale.

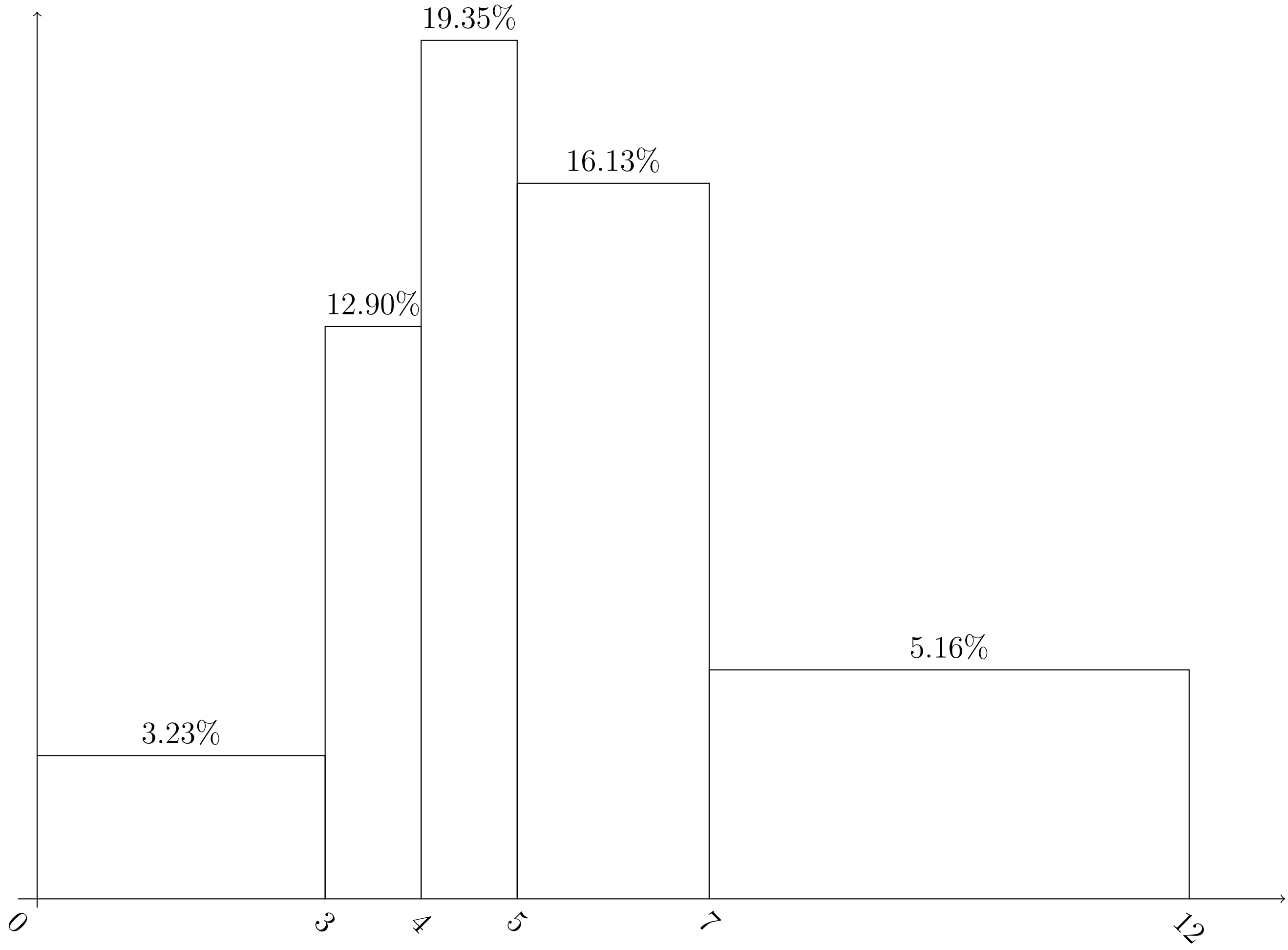
3. Calcolare un'approssimazione della media e della varianza dei dati.

4. Calcolare un'approssimazione della mediana dei dati.

Soluzione. 1. Calcoliamo le densità di frequenze relative dividendo le frequenze relative per l'ampiezza degli intervalli.

intervallo	freq. assolute	freq. relative	densità di freq. rel.	freq. cumulate
$[0, 3)$	3	9.68%	3.23%	3
$[3, 4)$	4	12.90%	12.90%	7
$[4, 5)$	6	19.35%	19.35%	13
$[5, 7)$	10	32.26%	16.13%	23
$[7, 12)$	8	25.81%	5.16 %	31

Rappresentiamo le densità di frequenze relative in un istogramma.



2. La classe modale è quella con maggiore densità di frequenza relativa, quindi è l'intervallo $[4,5)$.

3. Per calcolare un'approssimazione della media utilizziamo le frequenze relative ottenute da $p_j = f_j/n$ dove $n = 31$ e i valori centrali $\tilde{v}_j$ degli intervalli

$$\begin{aligned}\overline{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i \simeq \frac{1}{n}\sum_{j=1}^{k} f_j \tilde{v}_j = \sum_{j=1}^{k} p_j \tilde{v}_j \\ &= 9.68\% \cdot 1.5 + 12.90\% \cdot 3.5 + 19.35\% \cdot 4.5 + 32.26\% \cdot 6 + 25.81\% \cdot 9.5 = 5.855\,.\end{aligned}$$

Calcoliamo un'approssimazione della varianza

$$\begin{aligned}s^2 &= \frac{1}{n-1}\Big(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\Big) \simeq \frac{1}{n-1}\Big(\sum_{j=1}^{n} f_j \tilde{v}_j^2 - n\overline{x}^2\Big) = \frac{n}{n-1}\Big(\sum_{j=1}^{n} p_j \tilde{v}_j^2 - \overline{x}^2\Big) \\ &= \frac{31}{30}\Big(9.68\% \cdot 1.5^2 + 12.90\% \cdot 3.5^2 + 19.35\% \cdot 4.5^2 + 32.26\% \cdot 6^2 + 25.81\% \cdot 9.5^2 - 5.855^2\Big) \\ &\simeq 6.55\,.\end{aligned}$$

4. Per calcolare un'approssimazione della mediana dei dati, usiamo le frequenze cumulate. Troviamo l'intervallo I_j tale che $F_j \le \frac{n}{2} = 15.5 < F_{j+1}$. Si tratta dell'intervallo $[5,7)$. Approssimiamo la mediana con

$$Q_2 \simeq a_j + \lambda_j (b_j - a_j)$$

dove

$$\lambda_j = \frac{n/2 - F_j}{F_{j+1} - F_j} = \frac{15.5 - 13}{23 - 13} = 0.25\,.$$

Quindi

$$Q_2 \simeq 5 + 0.25(7-5) = 5.5\,.$$

Esercizio 2. (7 punti) Sia (X_1, X_2) un vettore aleatorio con probabilità congiunta descritta dalla seguente tabella:

$$\begin{array}{c|ccc} X_1 & -1 & 0 & 1 \\ X_2 & & & \\ \hline 0 & a & \frac{1}{4} & \frac{1}{4} \\ 1 & \frac{1}{4} & b & c \end{array}$$

dove a, b, c sono parametri da determinare. Si assuma che

- $\mathbb{E}(X_2) = \frac{1}{2}$
- $\mathrm{Var}(X_1 + X_2) = \frac{1}{4}$ (Suggerimento: conviene ricordare la formula per la varianza della somma)

1. Determinare il valore di a utilizzando la condizione $\mathbb{E}(X_2) = \frac{1}{2}$.
2. Determinare il valore di b, c.
3. Stabilire se le variabili aleatorie X_1 e X_2 sono indipendenti per i valori di a, b, c trovati.

Soluzione. 1. Imponiamo che la somma delle probabilità nella tabella sia 1:

$$a + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + b + c = 1 \implies a + b + c = \frac{1}{4}\,.$$

Imponiamo che $\mathbb{E}(X_2) = \frac{1}{2}$:

$$\frac{1}{2} = \mathbb{E}(X_2) = 1 \cdot \left(\frac{1}{4} + b + c\right) \implies b + c = \frac{1}{4}\,.$$

Da queste due condizioni segue che

$$a + b + c = \frac{1}{4} \implies a + \frac{1}{4} = \frac{1}{4} \implies a = 0\,.$$

2. Imponiamo che $\mathrm{Var}(X_1 + X_2) = \frac{1}{2}$. Ricordiamo la formula per la varianza della somma di due variabili aleatorie:

$$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\mathrm{Cov}(X_1, X_2)$$

Calcoliamo questi termini. Conviene calcolare prima

$$\mathbb{E}(X_1) = -1 \cdot \left(a + \frac{1}{4}\right) + 1 \cdot \left(\frac{1}{4} + c\right) = c\,.$$

Quindi

$$\begin{aligned} \mathrm{Var}(X_1) &= \mathbb{E}(X_1^2) - \mathbb{E}(X_1)^2 = (-1)^2 \cdot \left(a + \frac{1}{4}\right) + 1^2 \cdot \left(\frac{1}{4} + c\right) - c^2 \\ &= \frac{1}{2} + c - c^2 \end{aligned}$$

$$\mathrm{Var}(X_2) = \mathbb{E}(X_2^2) - \mathbb{E}(X_2)^2 = 1 \cdot \Big(\frac{1}{4} + b + c\Big) - \Big(\frac{1}{2}\Big)^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}\,.$$

$$\begin{aligned}\mathrm{Cov}(X_1, X_2) &= \mathbb{E}(X_1 \cdot X_2) - \mathbb{E}(X_1) \cdot \mathbb{E}(X_2)\\ &= -1 \cdot \frac{1}{4} + 1 \cdot c - \mathbb{E}(X_1) \cdot \mathbb{E}(X_2) = -\frac{1}{4} + c - \frac{1}{2}c = \frac{c}{2} - \frac{1}{4}\,.\end{aligned}$$

Segue che

$$\frac{1}{4} = \mathrm{Var}(X_1 + X_2) = \frac{1}{2} + c - c^2 + \frac{1}{4} + c - \frac{1}{2} = \frac{c}{2} - \frac{1}{4} = 2c - c^2 + \frac{1}{4}\,,$$

da cui $2c - c^2 = 0$, cioè $c(2-c) = 0$ e quindi (scartando $c = 2$ che non è ammissibile) $c = 0$, da cui

$$a = 0\,, \quad b = \frac{1}{4}\,, \quad c = 0\,.$$

La tabella completa è

X_2 \ X_1	-1	0	1
0	0	$\frac{1}{4}$	$\frac{1}{4}$
1	$\frac{1}{4}$	$\frac{1}{4}$	0

2. Dai conti del punto precedente abbiamo che $\mathrm{Cov}(X_1, X_2) = -\frac{1}{4}$. Le due variabili aleatorie non possono essere indipendenti perché la condizione di covarianza nulla è una condizione necessaria per l'indipendenza.

Esercizio 3. (8 punti) Il tempo necessario per un certo impiegato allo sportello delle poste per servire un cliente è distribuito con legge esponenziale. In media, il tempo in cui termina di servire il cliente è 7 minuti.

1. Qual è la probabilità che siano necessari più di 10 minuti per servire un cliente?
2. L'impiegato allo sportello inizia a servire un cliente alle 10:00. Passa del tempo, guardiamo l'orologio, sono le 10:10 e l'impiegato non ha ancora terminato di servire il cliente. Qual è la probabilità che il cliente finisca dopo le 10:20?
3. Ci sono in tutto 8 clienti da servire. Si assuma che i tempi necessari per servire ciascun cliente siano indipendenti. Qual è la probabilità che l'impiegato termini di servire almeno 3 clienti in più di 10 minuti?

Soluzione. Consideriamo $X \sim \mathrm{Exp}(\lambda)$ la variabile aleatoria che descrive il tempo allo sportello. Ricordiamo che $\mathbb{E}(X) = \frac{1}{\lambda}$. Quindi $\frac{1}{\lambda} = 7 \implies \lambda = \frac{1}{7}$.

1. Viene chiesto di calcolare

$$\mathbb{P}(\{X > 10\}) = e^{-\lambda 10} = e^{-\frac{10}{7}} = 23.97\%\,.$$

2. Viene chiesto di calcolare

$$\mathbb{P}(\{X > 20\} | \{X > 10\})\,.$$

Utilizzando l'assenza di memoria della legge esponenziale, abbiamo che

$$\mathbb{P}(\{X > 20\}|\{X > 10\}) = \mathbb{P}(\{X > 10\}) = 23.97\% \,.$$

3. Consideriamo l'evento "il cliente viene servito in più di 10 minuti". Questo evento ha probabilità

$$p = \mathbb{P}(\{X > 10\}) = 23.97\% \,.$$

Identifichiamo questo evento come un "successo". Consideriamo una variabile aleatoria con legge binomiale $Y \sim \mathrm{B}(n,p)$ con $n = 8$ e $p = 23.97\%$. Ci viene chiesto di calcolare

$$\mathbb{P}(\{Y \geq 3\}) = 1 - \mathbb{P}(\{Y \leq 2\}) = 1 - \mathbb{P}(\{Y = 0\}) + \mathbb{P}(\{Y = 1\}) + \mathbb{P}(\{Y = 2\})$$
$$= 1 - \binom{8}{0}(23.97\%)^0(1 - 23.97\%)^8 - \binom{8}{1}(23.97\%)^1(1 - 23.97\%)^7 - \binom{8}{2}(23.97\%)^2(1 - 23.97\%)^6$$
$$= 29.60\% \,.$$

Esercizio 4. (7 punti) Un produttore di acciaio INOX sostiene che il carico di rottura medio del materiale da lui prodotto è 730 $MPa = N/mm^2$. Si vuole stabilire se la media è in realtà più bassa. Si misura il carico di rottura su un campione casuale e si osservano i seguenti risultati:

732	723	729	715	712	721	730	724
749	710	708	691	723	693	734	749
733	715	708	742	728	720	734	708
725	762	729	722	705	730	712	698

La media calcolata sui dati di questo campione risulta essere 722.31. Si supponga che la distribuzione del carico di rottura abbia deviazione standard $\sigma = 20$.

1. I dati sono significativi al 5% per stabilire che la media è effettivamente più bassa di 730?
2. Qual è il più piccolo livello di significatività per cui i dati permettono di affermare che la media è più bassa di 730?

Soluzione. Si tratta di un test di ipotesi. Abbiamo un campione casuale $X_1, \ldots, X_n$ con $n = 32$, $\mathbb{E}(X_i) = \mu$ e $\mathrm{Var}(X_i) = \sigma^2$. La distribuzione della popolazione non è nota, la media μ non è nota, la deviazione standard è nota $\sigma = 20$. Il campione è numeroso.

Poniamo $\mu_0 = 730$. Fino a prova contraria, è vera l'ipotesi nulla

$$H_0 : \mu = \mu_0 \,,$$

e ci stiamo chiedendo se i dati sono abbastanza significativi da rifiutare l'ipotesi nulla a favore dell'ipotesi alternativa

$$H_1 : \mu < \mu_0 \,.$$

con livello di significatività $\alpha = 5\%$.

Poiché l'ipotesi alternativa è $H_1 : \mu < \mu_0$, i dati saranno significativi se la media è sufficiente più piccola di μ_0. La regione critica è allora della forma

$$R_c = \{(x_1, \ldots, x_n) \in R(X_1, \ldots, X_n) \ : \ \overline{x}_n < \mu_0 - \delta\} \,,$$

dove $\overline{x}_n = \frac{1}{n}\sum_{i=1}^n x_i$ è la media calcolata sulla realizzazione $x_1, \ldots, x_n$ del campione casuale.

Assumiamo l'ipotesi nulla H_0 vera, cioè $\mu = \mu_0$, ovvero $\mathbb{E}(X_i) = \mu_0$. Consideriamo la media campionaria $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ e utilizziamo la definizione di significatività per ottenere

$$\alpha = \mathbb{P}(\{(X_1, \ldots, X_n) \in R_c\}) = \mathbb{P}(\{\overline{X}_n < \mu_0 - \delta\}) = \mathbb{P}\Big(\Big\{\frac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}} < -\frac{\delta}{\sigma/\sqrt{n}}\Big\}\Big).$$

Non conosciamo la distribuzione della popolazione, ma il campione è numeroso ($n \geq 30$). Per il Teorema del Limite Centrale, si ha che $\frac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}} \overset{n\to+\infty}{\to} Z$ in legge, dove $Z \sim \mathcal{N}(0,1)$. Quindi

$$\alpha = \mathbb{P}\Big(\Big\{\frac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}} < -\frac{\delta}{\sigma/\sqrt{n}}\Big\}\Big) \simeq \mathbb{P}\Big(\Big\{Z < -\frac{\delta}{\sigma/\sqrt{n}}\Big\}\Big) = \mathbb{P}\Big(\Big\{Z > \frac{\delta}{\sigma/\sqrt{n}}\Big\}\Big).$$

Introduciamo il valore z_α tale che

$$\mathbb{P}(\{Z > z_\alpha\}) = \alpha\,.$$

Allora, scegliendo

$$\frac{\delta}{\sigma/\sqrt{n}} = z_\alpha \implies \delta = \frac{\sigma}{\sqrt{n}} z_\alpha\,,$$

otteniamo la condizione che definisce il livello di significatività.

In conclusione, la regione critica è

$$R_c = \Big\{(x_1, \ldots, x_n) \in R(X_1, \ldots, X_n)\ :\ \overline{x}_n < \mu_0 - \frac{\sigma}{\sqrt{n}} z_\alpha\Big\}\,,$$

e decidiamo come segue:

- Se $\overline{x}_n < \mu_0 - \frac{\sigma}{\sqrt{n}} z_\alpha$, i dati sono sufficientemente significativi da rifiutare H_0. L'ipotesi nulla H_0 viene rifiutata (con livello di significatività α).
- Se $\overline{x}_n \geq \mu_0 - \frac{\sigma}{\sqrt{n}} z_\alpha$, i dati non sono sufficientemente significativi da rifiutare. L'ipotesi nulla H_0 non viene rifiutata (con livello di significatività α).

In questo caso è anche possibile calcolare esplicitamente il p-value dei dati. Utilizzando il fatto che la funzione di distribuzione cumulativa della normale standard è strettamente crescente, otteniamo che

$$\begin{aligned} p\text{-value} &= \inf\Big\{\alpha\ :\ \overline{x}_n - \mu_0 < -\frac{\sigma}{\sqrt{n}} z_\alpha\Big\} = \inf\Big\{\alpha\ :\ \frac{\overline{x}_n - \mu_0}{\sigma/\sqrt{n}} < -z_\alpha\Big\} \\ &= \inf\Big\{\alpha\ :\ \Phi\Big(\frac{\overline{x}_n - \mu_0}{\sigma/\sqrt{n}}\Big) < \Phi(-z_\alpha)\Big\} = \inf\Big\{\alpha\ :\ \Phi\Big(\frac{\overline{x}_n - \mu_0}{\sigma/\sqrt{n}}\Big) < \alpha\Big\} \\ &= \Phi\Big(\frac{\overline{x}_n - \mu_0}{\sigma/\sqrt{n}}\Big)\,. \end{aligned}$$

Calcoliamo il p-value utilizzando la tavola della legge normale:

$$p\text{-value} = \Phi\Big(\frac{722.31 - 730}{20/\sqrt{32}}\Big) = \Phi(-2.175) = 1 - \Phi(2.175) = 1 - 0.9852 = 1.48\%\,.$$

Quindi:

1. L'ipotesi nulla è rifiutata con significatività 5%.
2. Il più piccolo livello di significatività per cui i dati permettono di rifiutare l'ipotesi nulla è 1.48%.

Esame di Probabilità e Statistica [3231]

Esame di Calcolo delle Probabilità e Statistica [2959]

Corso di Studi di Ingegneria Gestionale (D.M.270/04) (L)

Dipartimento di Meccanica, Matematica e Management
Politecnico di Bari

Cognome: ______________________ Docente: Gianluca Orlando
Nome: ______________________ Appello: luglio 2023 - turno 1
Matricola: ______________________ Data: 20/07/2023

Tempo massimo: 2 ore.

Esercizio 1. (6 punti) Si analizzano i ritardi di un treno all'arrivo in una stazione in diverse giornate. I dati misurati in minuti sono i seguenti:

$$5 \quad 1 \quad 7 \quad 0 \quad 15 \quad 2 \quad 31 \quad 5$$

1. Determinare i quartili (esclusivi) dei dati.
2. Determinare eventuali dati anomali e sospetti.
3. Tracciare un box plot.

Soluzione. 1. Per prima cosa ordiniamo i dati:

$$0 \quad 1 \quad 2 \quad 5 \quad 5 \quad 7 \quad 15 \quad 31$$

Abbiamo $n = 8$ dati.
Calcoliamo il primo quartile: $\frac{n+1}{4} = \frac{9}{4} = 2 + 0.25$. Allora

$$Q_1 = (1 - 0.25)x_2 + 0.25x_3 = 0.75 \cdot 1 + 0.25 \cdot 2 = 1.25\,.$$

Calcoliamo il secondo quartile: $(n+1)\frac{2}{4} = 9\frac{2}{4} = 4 + 0.5$. Allora

$$Q_2 = (1 - 0.5)x_4 + 0.5x_5 = 0.5 \cdot 5 + 0.5 \cdot 5 = 5\,.$$

Calcoliamo il terzo quartile: $(n+1)\frac{3}{4} = 9\frac{3}{4} = 6 + 0.75$. Allora

$$Q_3 = (1 - 0.75)x_6 + 0.75x_7 = 0.25 \cdot 7 + 0.75 \cdot 15 = 13\,.$$

2. Per determinare i dati anomali e sospetti calcoliamo il range interquartile

$$IQR = Q_3 - Q_1 = 13 - 1.25 = 11.75\,.$$

I dati anomali apparterrebbero agli intervalli

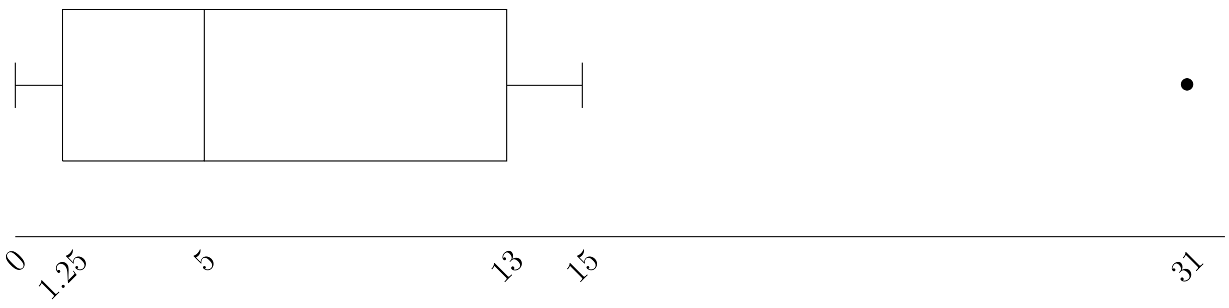
$$(-\infty, Q_1 - 3 \cdot IQR] \cup [Q_3 + 3 \cdot IQR, +\infty) = (-\infty, -34] \cup [48.25, +\infty)\,,$$

quindi non ci sono dati anomali. I dati sospetti appartengono agli intervalli

$$(Q_1 - 3 \cdot IQR, Q_1 - 1.5 \cdot IQR] \cup [Q_3 + 1.5 \cdot IQR, Q_3 + 3 \cdot IQR) = (-34, -16.375] \cup [30.625, 48.25)\,,$$

quindi 31 è un dato sospetto.

3. Segue il box-plot.



Esercizio 2. (8 punti) Durante l'esame di Probabilità e Statistica, alcune persone dimenticano di portare con sè le tavole delle distribuzioni. Si assuma che, durante un appello, il numero di persone che dimenticano di portare le tavole sia distribuito con una legge di Poisson e che, in media, 6 persone dimentichino le tavole.

1. Viene svolto un appello. Qual è la probabilità che almeno 4 persone dimentichino le tavole?

2. Vengono svolti 3 appelli. Si assumano i tre appelli indipendenti. Qual è la probabilità che nella totalità dei 3 appelli esattamente 20 persone dimentichino le tavole?

N.B.: Nella prossima richiesta, gli appelli non sono indipendenti.

Vengono svolti 2 appelli. Al primo appello il numero di persone che dimenticano di portare le tavole è distribuito con una legge di Poisson con media 6. Per il secondo appello:

- Se al primo appello il numero di persone che dimenticano di portare le tavole è maggiore o uguale a 4, allora il docente comunica che chi non porterà le tavole all'appello successivo avrà un voto più basso. L'effetto è che, nel secondo appello, il numero di persone che dimenticano le tavole è distribuito con una legge di Poisson con media 3.

- Se, invece al primo appello il numero di persone che dimenticano di portare le tavole è strettamente minore di 4, il docente non dice nulla e al secondo appello il numero di persone che dimenticano di portare le tavole è ancora distribuito con legge di Poisson con media 6.

Nelle ipotesi precedenti, rispondere alla seguente richiesta.

3. Sappiamo che al secondo appello esattamente 2 persone hanno dimenticato le tavole. Con che probabilità almeno 4 persone hanno dimenticato le tavole?

Soluzione. 1. Consideriamo la variabile aleatoria

$$X_1 = \text{“numero di persone che non portano le tavole in un appello”} \sim \mathrm{P}(\lambda)\,.$$

Ricordiamo che $\mathbb{P}(\{X = k\}) = e^{-\lambda}\frac{\lambda^k}{k!}$. Per la legge di Poisson si ha che $\mathbb{E}(X_1) = \lambda$, si ha che $\lambda = 6$.

Viene chiesto di calcolare

$$\begin{aligned}\mathbb{P}(\{X_1 \geq 4\}) &= 1 - \mathbb{P}(\{X_1 < 4\}) = 1 - \mathbb{P}(\{X_1 \leq 3\}) \\ &= 1 - \mathbb{P}(\{X_1 = 0\}) - \mathbb{P}(\{X_1 = 1\}) - \mathbb{P}(\{X_1 = 2\}) - \mathbb{P}(\{X_1 = 3\}) \\ &= 1 - e^{-\lambda}\frac{\lambda^0}{0!} - e^{-\lambda}\frac{\lambda^1}{1!} - e^{-\lambda}\frac{\lambda^2}{2!} - e^{-\lambda}\frac{\lambda^3}{3!} \\ &= 1 - e^{-\lambda}\Big(\frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!}\Big) \\ &= 1 - e^{-6}\Big(1 + 6 + \frac{6^2}{2} + \frac{6^3}{6}\Big) \simeq 84.88\% \,.\end{aligned}$$

2. Consideriamo tre variabili aleatorie indipendenti

$$X_1 = \text{“numero di persone che non portano le tavole al primo appello”} \sim \mathrm{P}(6)\,.$$

$$X_2 = \text{“numero di persone che non portano le tavole al secondo appello”} \sim \mathrm{P}(6)\,.$$

$$X_3 = \text{“numero di persone che non portano le tavole al terzo appello”} \sim \mathrm{P}(6)\,.$$

Il numero totale di persone che non portano le tavole nei tre appelli è

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

Poiché la somma di variabili aleatorie di Poisson indipendenti è distribuita con legge di Poisson con parametro dato dalla somma dei parametri, otteniamo che

$$X \sim \mathrm{P}(3 \cdot 6) = \mathrm{P}(18)\,.$$

Calcoliamo

$$\mathbb{P}(\{X = 20\}) = e^{-18}\frac{18^{20}}{20!} \simeq 7.98\%\,.$$

3. Consideriamo due variabili aleatorie

$$Y_1 = \text{“numero di persone che non portano le tavole al primo appello”} \sim \mathrm{P}(6)\,.$$

$$Y_2 = \text{“numero di persone che non portano le tavole al secondo appello”}\,.$$

La distribuzione di Y_2 dipende da Y_1, quindi le due variabili non sono indipendenti. Consideriamo due variabili aleatorie ausiliari:

$$Z_2 = \text{“numero di persone che non portano le tavole al secondo appello se } \{Y_1 \geq 4\}\text{”} \sim \mathrm{P}(3)\,.$$

$$W_2 = \text{“numero di persone che non portano le tavole al secondo appello se } \{Y_1 < 4\}\text{”} \sim \mathrm{P}(6)\,.$$

Le ipotesi del problema ci permettono di stabilire che

$$\mathbb{P}(\{Y_2 = k\}|\{Y_1 \geq 4\}) = \mathbb{P}(\{Z_2 = k\})\,,$$

$$\mathbb{P}(\{Y_2 = k\}|\{Y_1 < 4\}) = \mathbb{P}(\{W_2 = k\})\,.$$

Ci viene chiesto di calcolare

$$\begin{aligned}
\mathbb{P}(\{Y_1 \geq 4\}|\{Y_2 = 2\}) &= \frac{\mathbb{P}(\{Y_1 \geq 4\} \cap \{Y_2 = 2\})}{\mathbb{P}(\{Y_2 = 2\})} \\
&= \frac{\mathbb{P}(\{Y_2 = 2\}|\{Y_1 \geq 4\})\mathbb{P}(\{Y_1 \geq 4\})}{\mathbb{P}(\{Y_2 = 2\}|\{Y_1 \geq 4\})\mathbb{P}(\{Y_1 \geq 4\}) + \mathbb{P}(\{Y_2 = 2\}|\{Y_1 < 4\})\mathbb{P}(\{Y_1 < 4\})} \\
&= \frac{\mathbb{P}(\{Z_2 = 2\})\mathbb{P}(\{Y_1 \geq 4\})}{\mathbb{P}(\{Z_2 = 2\})\mathbb{P}(\{Y_1 \geq 4\}) + \mathbb{P}(\{W_2 = 2\})\mathbb{P}(\{Y_1 < 4\})} \\
&= \frac{e^{-3}\frac{3^2}{2!}84.88\%}{e^{-3}\frac{3^2}{2!}84.88\% + e^{-6}\frac{6^2}{2!}(1 - 84.88\%)} \simeq 96.57\% \,.
\end{aligned}$$

Esercizio 3. (7 punti) Devi prendere un mezzo pubblico per arrivare ad un appuntamento. Puoi prendere il bus 13 oppure il bus 17.

- Il bus 13 arriva alla fermata tra le 19:20 e le 19:40 e l'orario di arrivo ha distribuzione uniforme.
- Il bus 17 arriva alla fermata in media alle 19:40, con una varianza di 12 min^2, e l'orario di arrivo ha distribuzione uniforme.

Rispondere ai seguenti quesiti.

1. Calcolare la probabilità che il bus 13 arrivi prima delle 19:25 oppure dopo le 19:35.
2. Calcolare la probabilità che il bus 17 arrivi prima delle 19:38.
3. Calcolare la probabilità che uno dei due bus arrivi prima delle 19:35 (è indifferente se il 13 o il 17). Si assumano i due orari di arrivo indipendenti.

Soluzione. Consideriamo le due variabili aleatorie

$$X_{13} = \text{“minuto di arrivo del bus 13”} \sim \mathrm{U}(20, 40)\,,$$

$$X_{17} = \text{“minuto di arrivo del bus 17”} \sim \mathrm{U}(a, b)\,.$$

Ricordiamo che

$$\mathbb{E}(X_{17}) = \frac{a+b}{2}\,, \quad \mathrm{Var}(X_{17}) = \frac{(b-a)^2}{12}\,.$$

Imponiamo che

$$\begin{cases} \dfrac{a+b}{2} = 40 \\ \dfrac{(b-a)^2}{12} = 12 \end{cases} \implies \begin{cases} a+b = 80 \\ b-a = 12 \end{cases} \implies \begin{cases} 2a = 68 \\ 2b = 92 \end{cases} \implies \begin{cases} a = 34 \\ b = 46\,. \end{cases}$$

Quindi $X_{17} \sim \mathrm{U}(34, 46)$.

1. Poiché gli eventi $\{X_{13} \leq 25\}$ e $\{X_{13} \geq 35\}$ sono disgiunti, otteniamo che

$$\begin{aligned}
&\mathbb{P}(\{X_{13} \leq 25\} \cup \{X_{13} \geq 35\}) = \mathbb{P}(\{X_{13} \leq 25\}) + \mathbb{P}(\{X_{13} \geq 35\}) \\
&= \int_{20}^{25} \frac{1}{40-20}\,\mathrm{d}x + \int_{35}^{40} \frac{1}{40-20}\,\mathrm{d}x \\
&= \frac{5}{20} + \frac{5}{20} = \frac{1}{2} = 50\%\,.
\end{aligned}$$

2. Calcoliamo
$$\mathbb{P}(\{X_{17} \leq 38\}) = \int_{34}^{38} \frac{1}{46-34}\,\mathrm{d}x = \frac{4}{12} = \frac{1}{3}\,.$$

3. Viene chiesto di calcolare
$$\begin{aligned}
&\mathbb{P}(\{X_{13} \leq 35\} \cup \{X_{17} \leq 35\}) \\
&= \mathbb{P}(\{X_{13} \leq 35\}) + \mathbb{P}(\{X_{17} \leq 35\}) - \mathbb{P}(\{X_{13} \leq 35\} \cap \{X_{17} \leq 35\}) \\
&= \mathbb{P}(\{X_{13} \leq 35\}) + \mathbb{P}(\{X_{17} \leq 35\}) - \mathbb{P}(\{X_{13} \leq 35\})\mathbb{P}(\{X_{17} \leq 35\}) \\
&= \int_{20}^{35} \frac{1}{20}\,\mathrm{d}x + \int_{34}^{35} \frac{1}{12}\,\mathrm{d}x - \Big(\int_{20}^{35} \frac{1}{20}\,\mathrm{d}x\Big)\Big(\int_{34}^{35} \frac{1}{12}\,\mathrm{d}x\Big) \\
&= \frac{15}{20} + \frac{1}{12} - \frac{15}{20}\cdot\frac{1}{12} \simeq 77.08\%\,.
\end{aligned}$$

In alternativa si può anche calcolare il complementare della probabilità che entrambi i bus arrivino dopo le 19:35, ovvero
$$\begin{aligned}
&\mathbb{P}(\{X_{13} \leq 35\} \cup \{X_{17} \leq 35\}) = 1 - \mathbb{P}(\{X_{13} > 35\} \cap \{X_{17} > 35\}) \\
&= 1 - \mathbb{P}(\{X_{13} > 35\})\mathbb{P}(\{X_{17} > 35\}) = 1 - \frac{5}{20}\cdot\frac{11}{12} = 77.08\%\,.
\end{aligned}$$

Esercizio 4. (7 punti) Le capacità (in Ah) di 10 batterie sono state registrate come segue:

140 136 150 144 148 152 138 141 143 151

Si assuma che la capacità di una batteria abbia distribuzione normale.

1. Calcolare sui dati un intervallo di confidenza al 95% bilaterale per la varianza della popolazione.

2. Per 8 volte si ripete la misurazione dei dati su un campione della stessa ampiezza di quello di sopra e su ogni campione si calcola sui dati un intervallo di confidenza al 95% bilaterale per la varianza della popolazione con il procedimento del punto 1. Si assumano le misurazioni sui campioni indipendenti. Qual è la probabilità che almeno 6 volte la varianza appartenga all'intervallo calcolato?

Soluzione. 1. Abbiamo una popolazione $X \sim \mathcal{N}(\mu, \sigma^2)$ e un campione casuale $X_1, \ldots, X_n$ estratto dalla popolazione con $n = 10$. Dalla definizione di IC si ha che (con $\beta = 95\%$)
$$\beta = \mathbb{P}(\{U_n \leq \sigma^2 \leq V_n\})\,.$$

Per stimare σ^2 sfrutteremo lo stimatore varianza campionaria $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$. Poiché $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ sono indipendenti, $Q_{n-1} = \frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$. Allora
$$\begin{aligned}
\beta &= \mathbb{P}(\{U_n \leq \sigma^2 \leq V_n\}) = \mathbb{P}\Big(\Big\{\frac{(n-1)S_n^2}{V_n} \leq \frac{(n-1)S_n^2}{\sigma^2} \leq \frac{(n-1)S_n^2}{U_n}\Big\}\Big) \\
&= \mathbb{P}\Big(\Big\{\frac{(n-1)S_n^2}{V_n} \leq Q_{n-1} \leq \frac{(n-1)S_n^2}{U_n}\Big\}\Big) \\
&= 1 - \mathbb{P}\Big(\Big\{Q_{n-1} < \frac{(n-1)S_n^2}{V_n}\Big\}\Big) - \mathbb{P}\Big(\Big\{Q_{n-1} > \frac{(n-1)S_n^2}{U_n}\Big\}\Big)\,.
\end{aligned}$$

Segue che

$$\mathbb{P}\Big(\Big\{Q_{n-1} < \frac{(n-1)S_n^2}{V_n}\Big\}\Big) + \mathbb{P}\Big(\Big\{Q_{n-1} > \frac{(n-1)S_n^2}{U_n}\Big\}\Big) = 1-\beta = \alpha\,.$$

Decidiamo di equipartire α:

$$\mathbb{P}\Big(\Big\{Q_{n-1} < \frac{(n-1)S_n^2}{V_n}\Big\}\Big) = \mathbb{P}\Big(\Big\{Q_{n-1} > \frac{(n-1)S_n^2}{U_n}\Big\}\Big) = \frac{\alpha}{2}\,,$$

da cui

$$\mathbb{P}\Big(\Big\{Q_{n-1} > \frac{(n-1)S_n^2}{U_n}\Big\}\Big) = \frac{\alpha}{2}\,, \quad \mathbb{P}\Big(\Big\{Q_{n-1} \geq \frac{(n-1)S_n^2}{V_n}\Big\}\Big) = 1-\frac{\alpha}{2}\,.$$

Definiamo $\chi^2_{n-1,\alpha/2}$ e $\chi^2_{n-1,1-\alpha/2}$ come i punti tali che

$$\mathbb{P}(\{Q_{n-1} \geq \chi^2_{n-1,\alpha/2}\}) = \frac{\alpha}{2}\,, \quad \mathbb{P}(\{Q_{n-1} \geq \chi^2_{n-1,1-\alpha/2}\}) = 1-\frac{\alpha}{2}\,.$$

Scegliendo

$$\frac{(n-1)S_n^2}{U_n} = \chi^2_{n-1,\alpha/2} \implies U_n = \frac{(n-1)S_n^2}{\chi^2_{n-1,\alpha/2}}\,,$$

$$\frac{(n-1)S_n^2}{V_n} = \chi^2_{n-1,1-\alpha/2} \implies V_n = \frac{(n-1)S_n^2}{\chi^2_{n-1,1-\alpha/2}}\,,$$

si ottiene la condizione desiderata. In conclusione

$$\Big[\frac{(n-1)S_n^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)S_n^2}{\chi^2_{n-1,1-\alpha/2}}\Big]$$

è un intervallo di confidenza bilaterale per σ^2 con livello di confidenza $\beta = 1-\alpha$.

Per calcolarlo sui dati, calcoliamo

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{10}(140+136+\cdots+143+151) = 144.3\,,$$

$$s_n^2 = \frac{1}{n-1}\Big(\sum_{i=1}^{n} x_i^2 - n\overline{x}_n^2\Big) = \frac{1}{9}(140^2+\cdots+151^2-10\cdot 144.3^2) \simeq 32.23\,.$$

Calcoliamo l'intervallo di confidenza sui dati

$$\Big[\frac{(n-1)s_n^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)s_n^2}{\chi^2_{n-1,1-\alpha/2}}\Big] \simeq \Big[\frac{9\cdot 32.23}{\chi^2_{9,0.025}}, \frac{9\cdot 32.23}{\chi^2_{9,0.975}}\Big] \simeq \Big[\frac{9\cdot 32.23}{19.023}, \frac{9\cdot 32.23}{2.700}\Big]$$
$$\simeq [15.25, 107.43]\,.$$

2. Consideriamo come successo l'evento "l'intervallo di confidenza contiene la varianza della popolazione". Questo evento ha probabilità 95%. In tutto si fanno 8 tentativi, quindi

$$Y = \text{"numero di successi in 8 tentativi"} \sim \mathrm{B}(8, 95\%)\,.$$

Viene chiesto di calcolare

$$\mathbb{P}(\{Y \geq 6\}) = \mathbb{P}(\{Y=6\}) + \mathbb{P}(\{Y=7\}) + \mathbb{P}(\{Y=8\})$$
$$= \binom{8}{6}(95\%)^6(5\%)^2 + \binom{8}{7}(95\%)^7(5\%)^1 + \binom{8}{8}(95\%)^8(5\%)^0$$
$$\simeq 99.42\%\,.$$

Esame di Probabilità e Statistica [3231]

Esame di Calcolo delle Probabilità e Statistica [2959]

Corso di Studi di Ingegneria Gestionale (D.M.270/04) (L)

Dipartimento di Meccanica, Matematica e Management
Politecnico di Bari

Cognome: ____________________ Docente: Gianluca Orlando
Nome: ____________________ Appello: luglio 2023 - turno II
Matricola: ____________________ Data: 20/07/2023

Tempo massimo: 2 ore.

Esercizio 1. (6 punti) I seguenti dati mettono in relazione il numero di unità di un bene che sono state ordinate in funzione del prezzo del bene in sei diverse località:

numero ordini	88	112	123	136	158	172
prezzo	50	40	35	30	20	14

1. Rappresentare i dati in uno scatterplot.
2. Determinare (derivando le formule) la retta di regressione lineare e rappresentarla.
3. Determinare il coefficiente di correlazione lineare.

Soluzione. Soluzione. 1. Segue lo scatterplot (con la retta di regressione lineare):

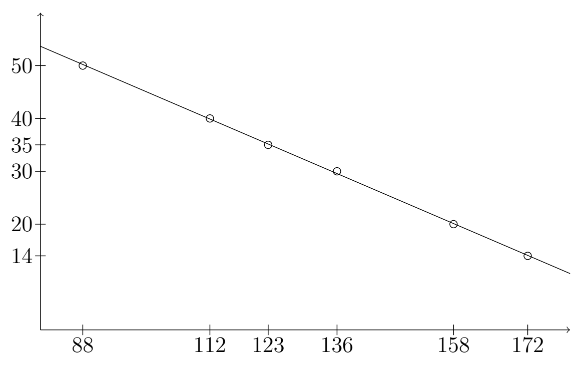


Figura 1: Scatterplot e retta di regressione lineare.

2. Denotiamo con $(x_1, y_1), \ldots, (x_n, y_n)$, $n = 6$, i dati del campione. Cerchiamo la retta di equazione

$$y = ax + b$$

che meglio approssima i dati, utilizzando il metodo dei minimi quadrati. Vogliamo minimizzare l'errore

$$r(a,b) = \sum_{i=1}^{n} (y_i - ax_i - b)^2 .$$

Imponiamo che il gradiente rispetto ad (a,b) sia nullo, ovvero,

$$\begin{aligned} 0 = \partial_a r(a,b) &= -2\sum_{i=1}^{n}(y_i - ax_i - b)x_i = -2\sum_{i=1}^{n}(x_i y_i - ax_i^2 - bx_i) , \\ 0 = \partial_b r(a,b) &= -2\sum_{i=1}^{n}(y_i - ax_i - b) \end{aligned}$$

Dalla seconda equazione segue che

$$nb = \sum_{i=1}^{n}(y_i - ax_i) \implies b = \overline{y} - a\overline{x} .$$

Sostituendo nella prima,

$$\begin{aligned} \sum_{i=1}^{n}(x_i y_i - ax_i^2 - bx_i) = 0 &\implies \sum_{i=1}^{n}(x_i y_i - ax_i^2 - x_i\overline{y} + a\overline{x}x_i) = 0 \\ &\implies a\Big(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\Big) = \sum_{i=1}^{n} x_i y_i - n\overline{x}\,\overline{y} \\ &\implies a = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{x}\,\overline{y}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2} . \end{aligned}$$

Completiamo la tabella con i valori necessari a calcolare a e b:

							somma
x_i	88	112	123	136	158	172	789
y_i	50	40	35	30	20	14	189
x_i^2	7744	12544	15129	18496	24964	29584	108461
y_i^2	2500	1600	1225	900	400	196	6821
$x_i y_i$	4400	4480	4305	4080	3160	2408	22833

Pertanto $\overline{x} = 789/6 = 131.5$ e $\overline{y} = 189/6 = 31.5$. Segue che

$$a = \frac{22833 - 6\cdot 131.5\cdot 31.5}{108461 - 6\cdot 131.5^2} = \frac{-2020.5}{4707.5} \simeq -0.43 ,$$
$$b = 31.5 + 0.43\cdot 131.5 \simeq 88.05 ,$$

ovvero, la retta di regressione lineare ha equazione

$$y = -0.43x + 88.05 .$$

3. Per calcolare il coefficiente di correlazione lineare usiamo la formula

$$\begin{aligned} \rho_{x,y} &= \frac{\sum_{i=1}^{n} x_i y_i - n\overline{x}\,\overline{y}}{\sqrt{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}\sqrt{\sum_{i=1}^{n} y_i^2 - n\overline{y}^2}} = \frac{22833 - 6\cdot 131.5\cdot 31.5}{\sqrt{108461 - 6\cdot 131.5^2}\sqrt{6821 - 6\cdot 31.5^2}} \\ &= \frac{-2020.5}{\sqrt{4707.5}\sqrt{867.5}} \simeq -0.9998 . \end{aligned}$$

Esercizio 2. (7 punti) La prova scritta di un esame molto difficile è costituita da 4 esercizi. Uno studente si reca all'esame senza aver studiato, sperando di copiare. La probabilità di riuscire a copiare bene un singolo esercizio dipende da quanto è stanco il professore.

- Se il professore è poco stanco, la probabilità di copiare bene un singolo esercizio è del 2%.
- Se il professore è molto stanco, la probabilità di copiare bene un singolo esercizio è del 5%.
- Il professore è molto stanco con il 20% di probabilità.

Si assuma che la copia di un esercizio non influenzi la copia di un altro esercizio. Lo studente supera la prova se riesce a copiare bene almeno 2 esercizi.

1. Il professore annuncia all'inizio della prova: "Oggi sono molto stanco!". Con che probabilità lo studente supera la prova?
2. Il professore non dice nulla all'inizio della prova e non sappiamo se è poco stanco o molto stanco. Finisce l'esame e lo studente supera la prova! Con che probabilità il professore era molto stanco?

Soluzione. Consideriamo le variabili aleatorie

$$X = \text{“numero di esercizi copiati bene”}$$

e poiché dobbiamo contare il numero di successi in 4 prove indipendenti (la copia di un esercizio non influenza la copia di un altro esercizio)

$$Y = \text{“numero di esercizi copiati bene se il professore è poco stanco”} \sim \mathrm{B}(4, 2\%)$$

$$Z = \text{“numero di esercizi copiati bene se il professore è molto stanco”} \sim \mathrm{B}(4, 5\%)\,.$$

Ricordiamo che per una binomiale $\mathrm{B}(n, p)$, la probabilità di assumere il valore k è $\binom{n}{k}p^k(1-p)^{n-k}$. Inoltre, introduciamo

$$W \sim \mathrm{Be}(20\%)\,,$$

dove $W = 1$ se il professore è molto stanco e $W = 0$ se il professore è poco stanco.

Dalle ipotesi del problema,

$$\mathbb{P}(\{X = k\}|\{W = 0\}) = \mathbb{P}(\{Y = k\})\,,$$
$$\mathbb{P}(\{X = k\}|\{W = 1\}) = \mathbb{P}(\{Z = k\})\,.$$

1. Sappiamo che si è realizzato l'evento $\{W = 1\}$. Allora dobbiamo calcolare

$$\begin{aligned}
&\mathbb{P}(\{X \geq 2\}|\{W = 1\}) = \mathbb{P}(\{Z \geq 2\}) = \mathbb{P}(\{Z = 2\}) + \mathbb{P}(\{Z = 3\}) + \mathbb{P}(\{Z = 4\}) \\
&= \binom{4}{2}(5\%)^2(95\%)^2 + \binom{4}{3}(5\%)^3(95\%)^1 + \binom{4}{4}(5\%)^4(95\%)^0 \\
&= 1.40\%\,.
\end{aligned}$$

2. Sappiamo che lo studente ha superato la prova, cioè si è realizzato l'evento $\{X \geq 2\}$. Calcoliamo allora

$$\begin{aligned}
&\mathbb{P}(\{W = 1\}|\{X \geq 2\}) = \frac{\mathbb{P}(\{W = 1\} \cap \{X \geq 2\})}{\mathbb{P}(\{X \geq 2\})} \\
&= \frac{\mathbb{P}(\{X \geq 2\}|\{W = 1\})\mathbb{P}(\{W = 1\})}{\mathbb{P}(\{X \geq 2\}|\{W = 1\})\mathbb{P}(\{W = 1\}) + \mathbb{P}(\{X \geq 2\}|\{W = 0\})\mathbb{P}(\{W = 0\})} \\
&= \frac{\mathbb{P}(\{X \geq 2\}|\{W = 1\})\mathbb{P}(\{W = 1\})}{\mathbb{P}(\{X \geq 2\}|\{W = 1\})\mathbb{P}(\{W = 1\}) + \mathbb{P}(\{X \geq 2\}|\{W = 0\})\mathbb{P}(\{W = 0\})} \\
&= \frac{\mathbb{P}(\{Z \geq 2\})\mathbb{P}(\{W = 1\})}{\mathbb{P}(\{Z \geq 2\})\mathbb{P}(\{W = 1\}) + \mathbb{P}(\{Y \geq 2\})\mathbb{P}(\{W = 0\})} \\
&= \frac{1.40\% \cdot 20\%}{1.40\% \cdot 20\% + 0.23\% 80\%} \simeq 60.34\%\,,
\end{aligned}$$

dove abbiamo usato che

$$\begin{aligned}
&\mathbb{P}(\{Y \geq 2\}) = \mathbb{P}(\{Y = 2\}) + \mathbb{P}(\{Y = 3\}) + \mathbb{P}(\{Y = 4\}) \\
&= \binom{4}{2}(2\%)^2(98\%)^2 + \binom{4}{3}(2\%)^3(98\%)^1 + \binom{4}{4}(2\%)^4(98\%)^0 \\
&= 0.23\%\,.
\end{aligned}$$

Esercizio 3. (8 punti) Devi viaggiare su un treno che percorre il seguente tragitto:

$$A \to B \to C\,.$$

Assumiamo che:

- nella tratta $A \to B$ il treno accumuli un ritardo medio di 2 min e che questo ritardo sia distribuito con legge esponenziale;
- nella tratta $B \to C$ il treno accumuli un ritardo medio di 2 min e che questo ritardo sia distribuito con legge esponenziale;
- il ritardo accumulato nella tratta $A \to B$ e il ritardo accumulato nella tratta $B \to C$ siano indipendenti.

Rispondere ai seguenti quesiti:

1. Qual è la probabilità che nella tratta $A \to B$ il treno accumuli un ritardo compreso tra 4 min e 12 min?
2. In media, qual è il ritardo accumulato nell'intero viaggio $A \to B \to C$? Motivare la risposta.
3. Qual è la varianza del ritardo accumulato nell'intero viaggio $A \to B \to C$? Motivare la risposta.
4. Considerare la probabilità che il ritardo accumulato nell'intero viaggio $A \to B \to C$ sia maggiore di 10 min. Stabilire se questa probabilità è maggiore o minore di 5%.

Soluzione. Consideriamo le due variabili aleatorie

$$X_1 = \text{“ritardo accumulato nella tratta } A \to B\text{”} \sim \mathrm{Exp}(\lambda)\,,$$

$$X_2 = \text{“ritardo accumulato nella tratta } B \to C\text{”} \sim \mathrm{Exp}(\lambda)\,.$$

In entrambi i casi si ha che $\mathbb{E}(X_1) = \mathbb{E}(X_2) = \frac{1}{\lambda} = 2$ quindi $\lambda = \frac{1}{2}$.

Ricordiamo che la densità di una legge esponenziale con parametro λ è $\lambda e^{-\lambda x}$.

1. Viene chiesto di calcolare

$$\begin{aligned}
\mathbb{P}(\{4 \leq X_1 \leq 12\}) &= \int_4^{12} \lambda e^{-\lambda x}\,\mathrm{d}x = \Big[-e^{-\lambda x}\Big]_4^{12} \\
&= e^{-10\lambda} - e^{-30\lambda} = e^{-\frac{4}{2}} - e^{-\frac{12}{2}} = e^{-2} - e^{-6} \simeq 13.29\%\,.
\end{aligned}$$

2. Il ritardo accumulato nell'intero viaggio è $X_1 + X_2$. Per la linearità del valore atteso si ha che $\mathbb{E}(X_1 + X_2) = \mathbb{E}(X_1) + \mathbb{E}(X_2) = \frac{1}{\lambda} + \frac{1}{\lambda} = \frac{2}{\lambda} = 4$.

3. Poiché X_1 e X_2 sono indipendenti, si ha che

$$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) = \frac{1}{\lambda^2} + \frac{1}{\lambda^2} = \frac{2}{\lambda^2} = 2 \cdot 4 = 8\,.$$

4. Dobbiamo calcolare

$$\mathbb{P}(\{X_1 + X_2 \geq 10\})\,.$$

Occorre ricordare che la somma di esponenziali indipendenti è distribuita come una legge Gamma, ovvero $X_1 + X_2 \sim \mathrm{Gamma}(2, \lambda)$, che ha densità

$$\frac{\lambda^2}{\Gamma(2)} x e^{-\lambda x} = \lambda^2 x e^{-\lambda x}\,,$$

dove abbiamo usato il fatto che

$$\Gamma(n) = (n-1)\Gamma(n-1) \implies \Gamma(2) = 1 \cdot \Gamma(1) = 1\,.$$

Quindi, integrando per parti,

$$\begin{aligned}\mathbb{P}(\{X_1 + X_2 \geq 10\}) &= \int_{10}^{+\infty} \lambda^2 x e^{-\lambda x}\,\mathrm{d}x = \Big[-\lambda x e^{-\lambda x}\Big]_{10}^{+\infty} + \int_{10}^{+\infty} \lambda e^{-\lambda x}\,\mathrm{d}x \\ &= 10\lambda e^{-10\lambda} + e^{-10\lambda} = (10\lambda + 1)e^{-10\lambda} = (10/2 + 1)e^{-5} = 6e^{-5} \simeq 4\%\,,\end{aligned}$$

quindi è minore del 5%.

In alternativa, si può ricordare che $\Gamma(\frac{n}{2}, \frac{1}{2}) = \chi^2(n)$, quindi $\Gamma(2, \frac{1}{2}) = \chi^2(4)$. Utilizzando la tavola della chi-quadro troviamo che una $Q_4 \sim \chi^2(4)$ soddisfa

$$\mathbb{P}(\{Q_4 > 9.488\}) = 5\%\,,$$

quindi la probabilità di essere maggiore di 10 è più piccola del 5%.

Esercizio 4. (7 punti) Un'auto è pubblicizzata come avente un consumo medio di benzina inferiore a 4.3 ℓ/100km (litri per 100 km) nella guida in autostrada. Si misura il consumo in un campione e si ottengono i seguenti dati (ℓ/100km):

3.4	3.3	5.5	5.0	3.2	3.9	3.8	4.1	5.3	5.3
3.7	5.7	5.5	5.6	5.7	5.1	4.9	4.5	3.1	5.6
3.1	4.8	5.5	5.1	6.4	3.5	5.6	6.4	4.4	4.6
4.8	3.8	4.2	5.4	3.0	4.2	5.7	6.0	4.3	4.4

La media calcolata sul campione risulta essere 4.7 ℓ/100km. Si sa che la deviazione standard dei consumi è 1.0 ℓ/100km.

1. I dati sono significativi al 10% per non credere alla pubblicità?

2. Assumiamo che la pubblicità dica il vero. Vengono misurati i dati di campioni della stessa ampiezza di quello precedente tante volte e ogni volta viene effettuato un test come nel punto 1. In media, dopo quanti tentativi osserveremo dei dati significativi al 10% che non ci fanno credere alla pubblicità?

Soluzione. 1. Abbiamo una popolazione $X\colon \Omega \to \mathbb{R}$ (non necessariamente normale) con $\mathbb{E}(X) = \mu$ e $\mathrm{Var}(X) = \sigma^2$ e un campione casuale $X_1, \dots, X_n$ estratto dalla popolazione con $n = 40 \geq 30$. La varianza $\sigma^2 = 1$ è nota. Dobbiamo impostare il seguente test di ipotesi

$$H_0 : \mu \leq \mu_0\,, \quad H_1 : \mu > \mu_0\,,$$

con livello di significatività $\alpha = 10\%$, dove $\mu_0 = 4.3$.

Poiché l'ipotesi alternativa è $H_1 : \mu > \mu_0$, i dati saranno significativi se la media è sufficientemente più grande di μ_0. La regione critica è allora della forma

$$R_c = \{(x_1, \dots, x_n) \in R(X_1, \dots, X_n) \ : \ \overline{x}_n > \mu_0 + \delta\}\,,$$

dove $\overline{x}_n = \frac{1}{n}\sum_{i=1}^n x_i$ è la media calcolata sulla realizzazione $x_1, \dots, x_n$ del campione casuale.

Assumiamo l'ipotesi nulla H_0 vera, cioè $\mu \leq \mu_0$.

Se $\mu \leq \mu_0$, non è detto che $\mathbb{E}(X_i) = \mu_0$. Il valore della media della popolazione è μ, quindi $\mathbb{E}(X_i) = \mu$. Tuttavia, poiché $\mu \leq \mu_0$

$$\overline{X}_n > \mu_0 + \delta \implies \overline{X}_n > \mu + \delta\,.$$

Utilizziamo questo per il calcolo della regione critica.

Utilizzando la media campionaria $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$,

$$\begin{aligned}\mathbb{P}(\{(X_1, \dots, X_n) \in R_c\}) &= \mathbb{P}(\{\overline{X}_n > \mu_0 + \delta\}) \leq \mathbb{P}(\{\overline{X}_n > \mu + \delta\})\\ &= \mathbb{P}\Big(\Big\{\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} > \frac{\delta}{\sigma/\sqrt{n}}\Big\}\Big)\,,\end{aligned}$$

Non conosciamo la distribuzione della popolazione, ma il campione è numeroso ($n \geq 30$). Per il Teorema del Limite Centrale, si ha che $\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \overset{n\to+\infty}{\to} Z$ in legge, dove $Z \sim \mathcal{N}(0,1)$. Quindi

$$\mathbb{P}\Big(\Big\{\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} > \frac{\delta}{\sigma/\sqrt{n}}\Big\}\Big) \simeq \mathbb{P}\Big(\Big\{Z > \frac{\delta}{\sigma/\sqrt{n}}\Big\}\Big)\,.$$

Introduciamo il valore z_α tale che

$$\mathbb{P}(\{Z > z_\alpha\}) = \alpha\,.$$

Allora, scegliendo

$$\frac{\delta}{\sigma/\sqrt{n}} = z_\alpha \implies \delta = \frac{\sigma}{\sqrt{n}} z_\alpha\,,$$

otteniamo (approssimativamente)

$$\mathbb{P}(\{(X_1, \dots, X_n) \in R_c\}) \leq \alpha\,,$$

(ovvero la probabilità di commettere un errore del I tipo è approssimativamente più piccola di α).

In conclusione, la regione critica è

$$R_c = \Big\{(x_1, \dots, x_n) \in R(X_1, \dots, X_n) \ : \ \overline{x}_n > \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha\Big\}\,,$$

e decidiamo come segue:

- Se $\overline{x}_n > \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha$, i dati sono sufficientemente significativi da rifiutare H_0. L'ipotesi nulla H_0 viene rifiutata (con livello di significatività α).

- Se $\overline{x}_n \leq \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha$, i dati non sono sufficientemente significativi da rifiutare. L'ipotesi nulla H_0 non viene rifiutata (con livello di significatività α).

Calcoliamo $z_{0.1} \simeq 1.285$ utilizzando le tavole e quindi

$$\mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha = 4.4 + \frac{1}{\sqrt{40}} 1.285 \simeq 4.6\,.$$

Poiché

$$\overline{x}_n = 4.7 > 4.6\,,$$

i dati sono significativi al 10% per rifiutare l'ipotesi nulla.

2. Consideriamo una legge geometrica $Y \sim \mathrm{Geo}(p)$ dove p è la probabilità di un successo, ovvero la probabilità di rifiutare l'ipotesi nulla quando questa è vera. Quindi $p = \alpha = 10\%$. Per la legge geometrica si sa che

$$\mathbb{E}(Y) = \frac{1}{p} = 10\,.$$

Esame di Probabilità e Statistica [3231]

Esame di Calcolo delle Probabilità e Statistica [2959]

Corso di Studi di Ingegneria Gestionale (D.M.270/04) (L)

Dipartimento di Meccanica, Matematica e Management
Politecnico di Bari

Cognome: ______________________ Docente: Gianluca Orlando
Nome: ______________________ Appello: settembre 2023 - I
Matricola: ______________________ Data: 05/09/2023

Tempo massimo: 2 ore.

Esercizio 1. (6 punti) Ci si aspetta che la resa percentuale di una reazione chimica sia correlata con la temperatura a cui avviene la reazione. Nella seguente tabella sono rappresentati i dati di un campione:

temperatura ($°C$)	45	55	60	70	85	90
resa (%)	43	48	51	57	66	68

1. Rappresentare i dati in uno scatterplot.
2. Determinare (derivando le formule) la retta di regressione lineare e rappresentarla.
3. Determinare il coefficiente di correlazione lineare.

Soluzione. 1. Segue lo scatterplot (con la retta di regressione lineare):

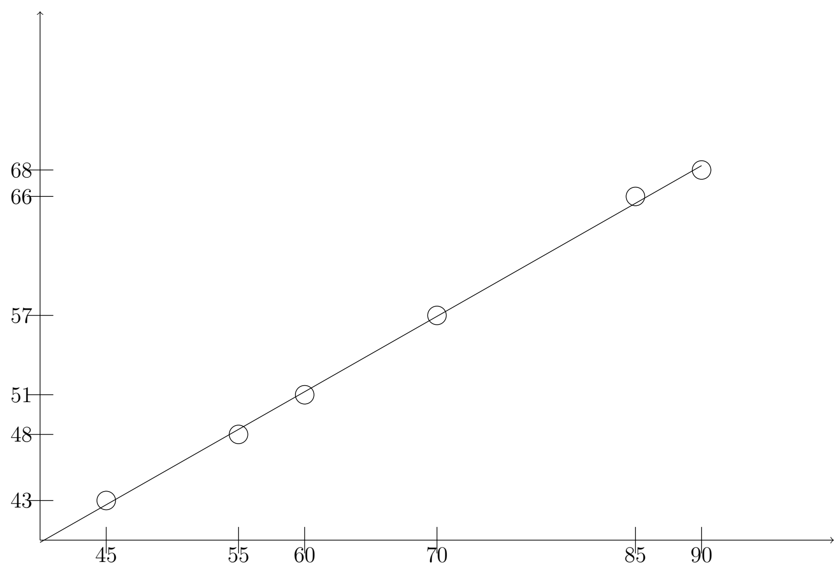


Figura 1: Scatterplot e retta di regressione lineare.

2. Denotiamo con $(x_1, y_1), \dots, (x_n, y_n)$, $n = 6$, i dati del campione. Cerchiamo la retta di equazione

$$y = ax + b$$

che meglio approssima i dati, utilizzando il metodo dei minimi quadrati. Vogliamo minimizzare l'errore

$$r(a, b) = \sum_{i=1}^{n} (y_i - ax_i - b)^2 .$$

Imponiamo che il gradiente rispetto ad (a, b) sia nullo, ovvero,

$$\begin{aligned} 0 = \partial_a r(a, b) &= -2 \sum_{i=1}^{n} (y_i - ax_i - b)x_i = -2 \sum_{i=1}^{n} (x_i y_i - ax_i^2 - bx_i) , \\ 0 = \partial_b r(a, b) &= -2 \sum_{i=1}^{n} (y_i - ax_i - b) \end{aligned}$$

Dalla seconda equazione segue che

$$nb = \sum_{i=1}^{n} (y_i - ax_i) \implies b = \overline{y} - a\overline{x} .$$

Sostituendo nella prima,

$$\begin{aligned} \sum_{i=1}^{n} (x_i y_i - ax_i^2 - bx_i) = 0 &\implies \sum_{i=1}^{n} (x_i y_i - ax_i^2 - x_i\overline{y} + a\overline{x}x_i) = 0 \\ &\implies a\Big(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2 \Big) = \sum_{i=1}^{n} x_i y_i - n\overline{x}\,\overline{y} \\ &\implies a = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{x}\,\overline{y}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2} . \end{aligned}$$

Completiamo la tabella con i valori necessari a calcolare a e b:

							somma
x_i	45	55	60	70	85	90	405
y_i	43	48	51	57	66	68	333
x_i^2	2025	3025	3600	4900	7225	8100	28875
y_i^2	1849	2304	2601	3249	4356	4624	18983
$x_i y_i$	1935	2640	3060	3990	5610	6120	23355

Pertanto $\overline{x} = 405/6 = 67.5$ e $\overline{y} = 55.5$. Segue che

$$\begin{aligned} a &= \frac{23355 - 6 \cdot 67.5 \cdot 55.5}{28875 - 6 \cdot 67.5^2} = \frac{877.5}{1537.5} \simeq 0.57 , \\ b &= 55.5 - 0.57 \cdot 67.5 \simeq 17.03 , \end{aligned}$$

ovvero, la retta di regressione lineare ha equazione

$$y = 0.57x + 17.03 .$$

3. Per calcolare il coefficiente di correlazione lineare usiamo la formula

$$\begin{aligned} \rho_{x,y} &= \frac{\sum_{i=1}^{n} x_i y_i - n\overline{x}\,\overline{y}}{\sqrt{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}\sqrt{\sum_{i=1}^{n} y_i^2 - n\overline{y}^2}} = \frac{23355 - 6 \cdot 67.5 \cdot 55.5}{\sqrt{28875 - 6 \cdot 67.5^2}\sqrt{18983 - 6 \cdot 55.5^2}} \\ &= \frac{877.5}{\sqrt{1537.5}\sqrt{501.5}} \simeq 0.9993 . \end{aligned}$$

Esercizio 2. (7 punti) Un'azienda produce *cookies* con gocce di cioccolato. Si assuma che il numero di gocce di cioccolato in un *cookie* sia distribuito con legge di Poisson e che, in media, un *cookie* abbia 10 gocce di cioccolato al suo interno. Si assuma che i numeri di gocce di cioccolato in *cookies* diversi siano indipendenti.

1. Qual è la probabilità che il numero di gocce di cioccolato in un *cookie* sia compreso tra 8 e 12 (estremi inclusi)?

2. Qual è la probabilità che in due *cookies* ci siano in tutto esattamente 30 gocce di cioccolato?

Una confezione di *cookies* contiene 6 *cookies*.

3. Compriamo una confezione. Qual è la probabilità che almeno 3 *cookies* nella confezione contengano ciascuno un numero gocce di cioccolato compreso tra 8 e 12?

4. In media, in una confezione quanti *cookies* contengono un numero gocce di cioccolato compreso tra 8 e 12?

Soluzione. Consideriamo la variabile aleatoria

$$X = \text{“numero di gocce di cioccolato in un } \textit{cookie}\text{”} \sim \mathrm{P}(\lambda)\,.$$

Ricordiamo che

$$\mathbb{E}(X) = \lambda$$

per una variabile distribuita con legge di Poisson, quindi $\lambda = 10$ dalle ipotesi del problema. Ricordiamo che

$$\mathbb{P}(\{X = k\}) = e^{-\lambda}\frac{\lambda^k}{k!}\,.$$

1. Ci viene chiesto di calcolare

$$\begin{aligned}
&\mathbb{P}(\{8 \leq X \leq 12\})\\
&= \mathbb{P}(\{X = 8\}) + \mathbb{P}(\{X = 9\}) + \mathbb{P}(\{X = 10\}) + \mathbb{P}(\{X = 11\}) + \mathbb{P}(\{X = 12\})\\
&= e^{-10}\Big(\frac{10^8}{8!} + \frac{10^9}{9!} + \frac{10^{10}}{10!} + \frac{10^{11}}{11!} + \frac{10^{12}}{12!}\Big) \simeq 57.13\%
\end{aligned}$$

2. Consideriamo la variabile aleatoria

$$Y = \text{“numero di gocce di cioccolato nel secondo } \textit{cookie}\text{”} \sim \mathrm{P}(\lambda)\,.$$

La traccia dice che X e Y sono indipendenti. Quindi, per una proprietà delle variabili aleatorie distribuite con legge di Poisson, abbiamo che la somma è ancora distribuita con legge di Poisson con parametro dato dalla somma dei parametri, ovvero $X + Y \sim \mathrm{P}(2\lambda) = \mathrm{P}(20)$. Possiamo allora calcolare

$$\mathbb{P}(\{X + Y = 30\}) = e^{-20}\frac{20^{30}}{30!} = 0.8\%\,.$$

3. Consideriamo come successo l'evento "il numero di gocce di cioccolato nel *cookie* è compreso tra 8 e 12". Questo successo ha probabilità

$$p = \mathbb{P}(\{8 \leq X \leq 12\}) = 57.13\%\,.$$

Consideriamo la variabile aleatoria

$$Z = \text{“numero di successi in 6 tentativi”} \sim \mathrm{B}(6, p)\,.$$

Questa è distribuita come una legge binomiale perché i tentativi sono indipendenti.
Ci viene chiesto di calcolare

$$\begin{aligned}\mathbb{P}(\{Z \geq 3\}) &= 1 - \mathbb{P}(\{Z \leq 2\}) = 1 - \mathbb{P}(\{Z = 0\}) - \mathbb{P}(\{Z = 1\}) - \mathbb{P}(\{Z = 2\}) \\ &= 1 - \binom{6}{0} p^0 (1-p)^6 - \binom{6}{1} p^1 (1-p)^5 - \binom{6}{2} p^2 (1-p)^4 = 77.88\%\,.\end{aligned}$$

4. Il valore atteso della binomiale è $\mathbb{E}(Z) = np = 6 \cdot 57.13\% \simeq 3.43$.

Esercizio 3. (7 punti) Ad un ristorante il tempo di attesa al tavolo dal momento dell'ordine all'arrivo del piatto dipende dal piatto ordinato.

- Se si ordina la pizza, il tempo di attesa è distribuito con legge esponenziale con media 15 minuti.
- Se si ordinano gli spaghetti all'assassina, il tempo di attesa è distribuito con legge esponenziale con una deviazione standard di 30 minuti.
- Il 60% delle persone ordina la pizza, mentre il 40% delle persone ordina gli spaghetti all'assassina.

Al tuo tavolo siete due persone. Si assumano i tempi di preparazione dei vari piatti indipendenti.

1. Tu ordini una pizza. Qual è la probabilità che la pizza sia pronta in un tempo compreso tra i 10 e i 20 minuti?
2. Il/la tuo/a partner ordina spaghetti all'assassina. Qual è la probabilità che gli spaghetti siano pronti dopo 40 minuti?
3. Il servizio porterà al tavolo la pizza e gli spaghetti insieme (quindi il tempo di attesa totale dipenderà da quale preparazione impiega più tempo ad essere pronta). Qual è la probabilità di dover aspettare più di 40 minuti perché arrivino entrambi i piatti?
4. Una persona sola al tavolo a fianco si sta lamentando del servizio perché aspetta da 40 minuti e ancora non è arrivato il suo piatto. Qual è la probabilità che abbia ordinato una pizza?

Soluzione. Consideriamo le seguenti variabili aleatorie:

$$X_{\text{pizza}} = \text{“tempo di attesa se si ordina la pizza”} \sim \mathrm{Exp}(\lambda)$$

$$X_{\text{spagh}} = \text{“tempo di attesa se si ordinano gli spaghetti all'assassina”} \sim \mathrm{Exp}(\mu)$$

$$X = \text{“tempo di attesa (senza conoscere l'ordine)”}$$

$$Y \sim \mathrm{Be}(60\%)\,,\ \text{dove } \{Y = 1\} \text{ se viene ordinata pizza}$$

1. Ricordiamo che per la legge esponenziale si ha $\mathbb{E}(X_{\text{pizza}}) = \frac{1}{\lambda} = 15$ da cui segue che $\lambda = \frac{1}{15}$. Viene chiesto di calcolare

$$\begin{aligned}\mathbb{P}(\{10 \leq X_{\text{pizza}} \leq 20\}) &= \int_{10}^{20} \lambda e^{-\lambda x}\, \mathrm{d}x = \Big[-e^{-\lambda x} \Big]_{x=10}^{x=20} = e^{-10\lambda} - e^{-20\lambda} \\ &= e^{-\frac{10}{15}} - e^{-\frac{20}{15}} \simeq 24.98\%\,.\end{aligned}$$

2. Ricordiamo che per la legge esponenziale si ha $\mathrm{Var}(X_{\text{spagh}}) = \frac{1}{\mu^2}$. La traccia fornisce il valore della deviazione standard, quindi $\frac{1}{\mu^2} = 30^2$, da cui $\mu = \frac{1}{30}$. Possiamo allora calcolare

$$\mathbb{P}(\{X_{\text{spagh}} \geq 40\}) = e^{-40\mu} = e^{-\frac{40}{30}} \simeq 26.36\% \,.$$

3. Il tempo di attesa sarà dato dal massimo tra X_{pizza} e X_{spagh}, poiché bisogna aspettare che entrambi i piatti siano pronti. Ci viene chiesto di calcolare

$$\mathbb{P}(\{\max\{X_{\text{pizza}}, X_{\text{spagh}}\} \geq 40\}) \,.$$

Conviene calcolare la probabilità dell'evento complementare. Infatti $\max\{X_{\text{pizza}}, X_{\text{spagh}}\} < 40$ quando entrambi $X_{\text{pizza}} < 40$ e $X_{\text{spagh}} < 40$. Quindi, usando l'indipendenza

$$\begin{aligned}
&\mathbb{P}(\{\max\{X_{\text{pizza}}, X_{\text{spagh}}\} \geq 40\}) = 1 - \mathbb{P}(\{\max\{X_{\text{pizza}}, X_{\text{spagh}}\} < 40\}) \\
&= 1 - \mathbb{P}(\{X_{\text{pizza}} < 40\} \cap \{X_{\text{spagh}} < 40\}) = 1 - \mathbb{P}(\{X_{\text{pizza}} < 40\})\mathbb{P}(\{X_{\text{spagh}} < 40\}) \\
&= 1 - (1 - e^{-\frac{40}{15}})(1 - e^{-\frac{40}{30}}) \simeq 31.48\% \,.
\end{aligned}$$

4. Sappiamo che si è verificato l'evento $X \geq 40$. Calcoliamo

$$\begin{aligned}
&\mathbb{P}(\{Y = 1\} | \{X \geq 40\}) = \frac{\mathbb{P}(\{Y = 1\} \cap \{X \geq 40\})}{\mathbb{P}(\{X \geq 40\})} = \frac{\mathbb{P}(\{X \geq 40\} | \{Y = 1\})\mathbb{P}(\{Y = 1\})}{\mathbb{P}(\{X \geq 40\})} \\
&= \frac{\mathbb{P}(\{X_{\text{pizza}} \geq 40\})\mathbb{P}(\{Y = 1\})}{\mathbb{P}(\{X \geq 40\} | \{Y = 1\})\mathbb{P}(\{Y = 1\}) + \mathbb{P}(\{X \geq 40\} | \{Y = 0\})\mathbb{P}(\{Y = 0\})} \\
&= \frac{\mathbb{P}(\{X_{\text{pizza}} \geq 40\})\mathbb{P}(\{Y = 1\})}{\mathbb{P}(\{X_{\text{pizza}} \geq 40\})\mathbb{P}(\{Y = 1\}) + \mathbb{P}(\{X_{\text{spagh}} \geq 40\})\mathbb{P}(\{Y = 0\})} \\
&= \frac{e^{-\frac{40}{15}} 60\%}{e^{-\frac{40}{15}} 60\% + e^{-\frac{40}{30}} 40\%} \simeq 28.34\% \,.
\end{aligned}$$

Esercizio 4. (8 punti) Un responsabile di controllo di qualità misura la distanza tra le piastre degli scambiatori di calore prodotti da un'azienda. Le misurazioni su un campione di 20 scambiatori di calore forniscono una distanza media calcolata sul campione di 4.0 mm e una deviazione standard calcolata sul campione di 0.1 mm. Si assuma che la distribuzione della distanza tra le piastre sia normale.

1. Calcolare sui dati un intervallo di confidenza al 95% bilaterale per la distanza media.
2. Calcolare sui dati un intervallo di confidenza al 90% bilaterale per la varianza della distanza.

Derivare il calcolo degli intervalli di confidenza in entrambi i punti, motivando i passaggi.

Soluzione. Abbiamo a che fare con una popolazione $X \sim \mathcal{N}(\mu, \sigma^2)$ e un campione casuale $X_1, \ldots, X_n$ estratto dalla popolazione con $n = 20$. La media μ e la varianza σ^2 non sono note.
1. Dalla definizione di IC si ha che

$$\beta = \mathbb{P}(\{U_n \leq \mu \leq V_n\}) \,.$$

dove $\beta = 95\%$. Per stimare μ sfrutteremo lo stimatore media campionaria $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Per stimare σ^2 sfrutteremo lo stimatore varianza campionaria $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$. Poiché

$X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ sono indipendenti, $T_{n-1} = \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \sim t(n-1)$ ha distribuzione t-Student con $n-1 = 19$ gradi di libertà. Allora

$$\begin{aligned}
\beta = \mathbb{P}(\{U_n \le \mu \le V_n\}) &= \mathbb{P}\Big(\Big\{\frac{\overline{X}_n - V_n}{S_n/\sqrt{n}} \le \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \le \frac{\overline{X}_n - U_n}{S_n/\sqrt{n}}\Big\}\Big) \\
&= \mathbb{P}\Big(\Big\{\frac{\overline{X}_n - V_n}{S_n/\sqrt{n}} \le T_{n-1} \le \frac{\overline{X}_n - U_n}{S_n/\sqrt{n}}\Big\}\Big) \\
&= 1 - \mathbb{P}\Big(\Big\{T_{n-1} < \frac{\overline{X}_n - V_n}{S_n/\sqrt{n}}\Big\}\Big) - \mathbb{P}\Big(\Big\{T_{n-1} > \frac{\overline{X}_n - U_n}{S_n/\sqrt{n}}\Big\}\Big).
\end{aligned}$$

Segue che

$$\mathbb{P}\Big(\Big\{T_{n-1} < \frac{\overline{X}_n - V_n}{S_n/\sqrt{n}}\Big\}\Big) + \mathbb{P}\Big(\Big\{T_{n-1} > \frac{\overline{X}_n - U_n}{S_n/\sqrt{n}}\Big\}\Big) = 1 - \beta = \alpha\,.$$

Decidiamo di equipartire α:

$$\mathbb{P}\Big(\Big\{T_{n-1} < \frac{\overline{X}_n - V_n}{S_n/\sqrt{n}}\Big\}\Big) = \mathbb{P}\Big(\Big\{T_{n-1} < \frac{\overline{X}_n - V_n}{S_n/\sqrt{n}}\Big\}\Big) = \frac{\alpha}{2}\,.$$

Definiamo $t_{n-1,\alpha/2}$ come il punto tale che

$$\mathbb{P}(\{T_{n-1} \ge t_{n-1,\alpha/2}) = \frac{\alpha}{2}\,.$$

Scegliendo

$$\begin{aligned}
\frac{\overline{X}_n - U_n}{S_n/\sqrt{n}} &= t_{n-1,\alpha/2} \implies U_n = \overline{X}_n - \frac{S_n}{\sqrt{n}} t_{n-1,\alpha/2}\,, \\
\frac{\overline{X}_n - V_n}{S_n/\sqrt{n}} &= -t_{n-1,\alpha/2} \implies V_n = \overline{X}_n + \frac{S_n}{\sqrt{n}} t_{n-1,\alpha/2}\,,
\end{aligned}$$

si ottiene la condizione che definisce l'IC. In conclusione

$$\Big[\overline{X}_n - \frac{S_n}{\sqrt{n}} t_{n-1,\alpha/2}, \overline{X}_n + \frac{S_n}{\sqrt{n}} t_{n-1,\alpha/2}\Big]$$

è un intervallo di confidenza bilaterale per μ con livello di confidenza $\beta = 1 - \alpha$.

Calcoliamo la realizzazione dell'IC sui dati. La traccia fornisce la media calcolata sui dati $\overline{x}_n = 4.0$ e la deviazione standard calcolata sui dati $s_n = 0.1$. Calcoliamo, utilizzando la tabella,

$$t_{n-1,\alpha/2} = t_{19,0.025} \simeq 2.093\,.$$

Segue che l'IC calcolato sui dati è

$$\Big[4.0 - \frac{0.1}{\sqrt{20}} 2.093, 4.0 + \frac{0.1}{\sqrt{20}} 2.093\Big] \simeq [3.95, 4.05]\,.$$

2. Dalla definizione di IC si ha che

$$\beta = \mathbb{P}(\{U_n \le \sigma^2 \le V_n\})\,,$$

con $\beta = 90\%$. Per stimare σ^2 sfrutteremo lo stimatore varianza campionaria $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$. Poiché $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ sono indipendenti, $Q_{n-1} = \frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$ ha distribuzione chi-quadro con $n-1 = 19$ gradi di libertà. Allora

$$\begin{aligned}
\beta = \mathbb{P}(\{U_n \le \sigma^2 \le V_n\}) &= \mathbb{P}\Big(\Big\{\frac{(n-1)S_n^2}{V_n} \le \frac{(n-1)S_n^2}{\sigma^2} \le \frac{(n-1)S_n^2}{U_n}\Big\}\Big) \\
&= \mathbb{P}\Big(\Big\{\frac{(n-1)S_n^2}{V_n} \le Q_{n-1} \le \frac{(n-1)S_n^2}{U_n}\Big\}\Big) \\
&= 1 - \mathbb{P}\Big(\Big\{Q_{n-1} < \frac{(n-1)S_n^2}{V_n}\Big\}\Big) - \mathbb{P}\Big(\Big\{Q_{n-1} > \frac{(n-1)S_n^2}{U_n}\Big\}\Big).
\end{aligned}$$

Segue che

$$\mathbb{P}\Big(\Big\{Q_{n-1} < \frac{(n-1)S_n^2}{V_n}\Big\}\Big) + \mathbb{P}\Big(\Big\{Q_{n-1} > \frac{(n-1)S_n^2}{U_n}\Big\}\Big) = 1-\beta = \alpha\,.$$

Decidiamo di equipartire α:

$$\mathbb{P}\Big(\Big\{Q_{n-1} < \frac{(n-1)S_n^2}{V_n}\Big\}\Big) = \mathbb{P}\Big(\Big\{Q_{n-1} > \frac{(n-1)S_n^2}{U_n}\Big\}\Big) = \frac{\alpha}{2}\,,$$

da cui

$$\mathbb{P}\Big(\Big\{Q_{n-1} > \frac{(n-1)S_n^2}{U_n}\Big\}\Big) = \frac{\alpha}{2}\,, \quad \mathbb{P}\Big(\Big\{Q_{n-1} \geq \frac{(n-1)S_n^2}{V_n}\Big\}\Big) = 1-\frac{\alpha}{2}\,.$$

Definiamo $\chi^2_{n-1,\alpha/2}$ e $\chi^2_{n-1,1-\alpha/2}$ come i punti tali che

$$\mathbb{P}(\{Q_{n-1} \geq \chi^2_{n-1,\alpha/2}\}) = \frac{\alpha}{2}\,, \quad \mathbb{P}(\{Q_{n-1} \geq \chi^2_{n-1,1-\alpha/2}\}) = 1-\frac{\alpha}{2}\,.$$

Scegliendo

$$\frac{(n-1)S_n^2}{U_n} = \chi^2_{n-1,\alpha/2} \implies U_n = \frac{(n-1)S_n^2}{\chi^2_{n-1,\alpha/2}}\,,$$

$$\frac{(n-1)S_n^2}{V_n} = \chi^2_{n-1,1-\alpha/2} \implies V_n = \frac{(n-1)S_n^2}{\chi^2_{n-1,1-\alpha/2}}\,,$$

si ottiene la condizione che definisce l'IC. In conclusione

$$\Big[\frac{(n-1)S_n^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)S_n^2}{\chi^2_{n-1,1-\alpha/2}}\Big]$$

è un intervallo di confidenza bilaterale per σ^2 con livello di confidenza $\beta = 1-\alpha$.

Per calcolare la realizzazione dell'IC sui dati, usiamo la varianza campionaria $s_n^2 = 0.1^2 = 0.01$ e i valori delle tavole

$$\chi^2_{19,0.05} = 30.144\,, \quad \chi^2_{19,0.095} = 10.117\,.$$

Segue che

$$\Big[\frac{19\cdot 0.01}{30.144}, \frac{19\cdot 0.01}{10.117}\Big] \simeq [0.0063, 0.0188]\,.$$

Esame di Probabilità e Statistica [3231]

Esame di Calcolo delle Probabilità e Statistica [2959]

Corso di Studi di Ingegneria Gestionale (D.M.270/04) (L)

Dipartimento di Meccanica, Matematica e Management
Politecnico di Bari

Cognome: ______________________
Nome: ______________________
Matricola: ______________________

Docente: Gianluca Orlando
Appello: settembre 2023 - II
Data: 19/09/2023

Tempo massimo: 2 ore.

Esercizio 1. (6 punti) Il numero di veicoli venduti da un importante showroom di auto in un giorno è stato registrato per 10 giorni lavorativi. I dati sono i seguenti:

$$20 \quad 15 \quad 18 \quad 3 \quad 10 \quad 17 \quad 21 \quad 19 \quad 25 \quad 28$$

1. Determinare i quartili (esclusivi) dei dati.
2. Determinare eventuali dati anomali e sospetti.
3. Tracciare un box plot.

Soluzione. 1. Per prima cosa ordiniamo i dati:

$$3 \quad 10 \quad 15 \quad 17 \quad 18 \quad 19 \quad 20 \quad 21 \quad 25 \quad 28$$

Abbiamo $n = 10$ dati.

Calcoliamo il primo quartile: $\frac{n+1}{4} = \frac{11}{4} = 2 + 0.75$. Allora

$$Q_1 = (1 - 0.75)x_2 + 0.75x_3 = 0.25 \cdot 10 + 0.75 \cdot 15 = 13.75\,.$$

Calcoliamo il secondo quartile: $(n+1)\frac{2}{4} = 11\frac{2}{4} = 5 + 0.5$. Allora

$$Q_2 = (1 - 0.5)x_5 + 0.5x_6 = 0.5 \cdot 18 + 0.5 \cdot 19 = 18.5\,.$$

Calcoliamo il terzo quartile: $(n+1)\frac{3}{4} = 11\frac{3}{4} = 8 + 0.25$. Allora

$$Q_3 = (1 - 0.25)x_8 + 0.25x_9 = 0.75 \cdot 21 + 0.25 \cdot 25 = 22\,.$$

2. Per determinare i dati anomali e sospetti calcoliamo il range interquartile

$$IQR = Q_3 - Q_1 = 22 - 13.75 = 8.25\,.$$

I dati anomali apparterrebbero agli intervalli

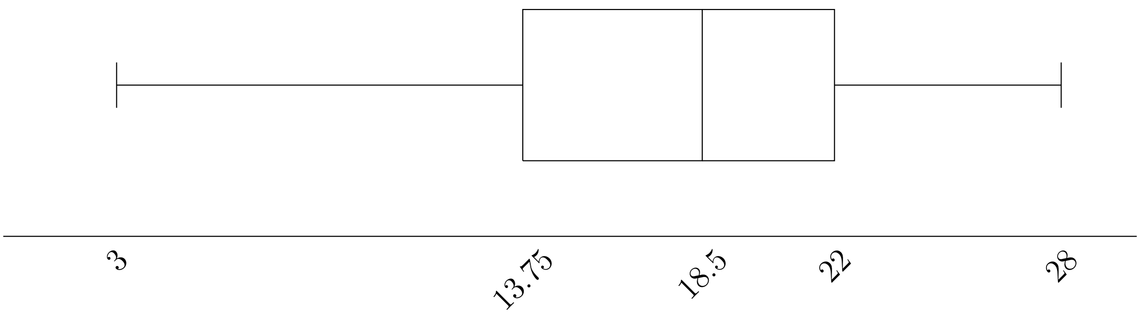
$$(-\infty, Q_1 - 3 \cdot IQR] \cup [Q_3 + 3 \cdot IQR, +\infty) = (-\infty, -11] \cup [46.75, +\infty)\,,$$

quindi non ci sono dati anomali. I dati sospetti appartengono agli intervalli

$$(Q_1 - 3 \cdot IQR, Q_1 - 1.5 \cdot IQR] \cup [Q_3 + 1.5 \cdot IQR, Q_3 + 3 \cdot IQR) = (-11, 1.375] \cup [34.375, 46.75)\,,$$

quindi non ci sono dati sospetti.

3. Segue il box-plot.



Esercizio 2. (7 punti) Si consideri un vettore aleatorio (X_1, X_2) avente funzione di probabilità congiunta descritta dalla seguente tabella:

$X_2 \backslash X_1$	1	2	3
1	a_{11}	a_{21}	a_{31}
2	a_{12}	a_{22}	a_{32}
3	a_{13}	a_{23}	a_{33}

con $a_{11}, a_{21}, a_{31}, a_{12}, a_{22}, a_{32}, a_{13}, a_{23}, a_{33} > 0$.

1. Calcolare il range dei possibili valori che possono essere assunti dalla variabile aleatoria $Y = \min\{X_1 + X_2, X_1 \cdot X_2\}$ (valore più piccolo tra $X_1 + X_2$ e $X_1 \cdot X_2$).

Si assuma che i valori nel range di Y siano equiprobabili.

2. Determinare i valori di a_{11}, a_{22}, a_{33}.

Si assuma che:

- $\mathbb{P}(\{X_1 = 2\} \cap \{X_2 = 1\}) = \frac{1}{12}$.
- $\mathbb{P}(\{X_1 = 3\} \cap \{X_2 = 1\}) = \mathbb{P}(\{W < \frac{1}{12}\})$ dove $W \sim \mathrm{U}(0, 1)$.
- $\mathbb{P}(\{X_1 = 3\} | \{X_2 = 2\}) = \frac{1}{4}$.

3. Determinare i valori rimanenti $a_{21}, a_{31}, a_{12}, a_{32}, a_{13}, a_{23}$.

4. Stabilire se X_1 e X_2 sono indipendenti.

Soluzione. 1. Per calcolare il range della variabile aleatoria Y, scriviamo tutti i possibili valori che possono essere assunti da Y.

$$X_1 = 1, X_2 = 1 \implies X_1 + X_2 = 2, X_1 \cdot X_2 = 1 \implies Y = 1\,,$$

$$X_1 = 1, X_2 = 2 \implies X_1 + X_2 = 3, X_1 \cdot X_2 = 2 \implies Y = 2\,,$$
$$X_1 = 2, X_2 = 1 \implies X_1 + X_2 = 3, X_1 \cdot X_2 = 2 \implies Y = 2\,,$$

$$X_1 = 1, X_2 = 3 \implies X_1 + X_2 = 4, X_1 \cdot X_2 = 3 \implies Y = 3\,,$$
$$X_1 = 3, X_2 = 1 \implies X_1 + X_2 = 4, X_1 \cdot X_2 = 3 \implies Y = 3\,,$$

$$X_1 = 2, X_2 = 2 \implies X_1 + X_2 = 4, X_1 \cdot X_2 = 4 \implies Y = 4\,,$$

$$X_1 = 2, X_2 = 3 \implies X_1 + X_2 = 5, X_1 \cdot X_2 = 6 \implies Y = 5\,,$$
$$X_1 = 3, X_2 = 2 \implies X_1 + X_2 = 5, X_1 \cdot X_2 = 6 \implies Y = 5\,,$$

$$X_1 = 3, X_2 = 3 \implies X_1 + X_2 = 6, X_1 \cdot X_2 = 9 \implies Y = 6\,,$$

Quindi $R(Y) = \{1, 2, 3, 4, 5, 6\}$ (osserviamo che ciascuno di questi valori ha probabilità positiva).

2. Scrivendo per quali valori di X_1, X_2 vengono assunti i valori di Y, otteniamo che

$$\begin{aligned}
\mathbb{P}(\{Y = 1\}) &= a_{11} \\
\mathbb{P}(\{Y = 2\}) &= a_{12} + a_{21} \\
\mathbb{P}(\{Y = 3\}) &= a_{13} + a_{31} \\
\mathbb{P}(\{Y = 4\}) &= a_{22} \\
\mathbb{P}(\{Y = 5\}) &= a_{23} + a_{32} \\
\mathbb{P}(\{Y = 6\}) &= a_{33}\,.
\end{aligned}$$

Poiché questi eventi sono equiprobabili, otteniamo che

$$a_{11} = a_{12} + a_{21} = a_{13} + a_{31} = a_{22} = a_{23} + a_{32} = a_{33} = b\,,$$

dove b è da determinare. Ricordiamo che è necessario avere

$$a_{11} + a_{12} + a_{21} + a_{13} + a_{31} + a_{22} + a_{23} + a_{32} + a_{33} = 1$$

da cui segue che

$$6b = 1 \implies b = \frac{1}{6}\,.$$

Aggiorniamo la tabella:

X_2 \ X_1	1	2	3
1	$1/6$	a_{21}	a_{31}
2	a_{12}	$1/6$	a_{32}
3	a_{13}	a_{23}	$1/6$

dove

$$a_{12} + a_{21} = a_{13} + a_{31} = a_{23} + a_{32} = \frac{1}{6}\,.$$

3. Imponiamo la prima condizione

$$\frac{1}{12} = \mathbb{P}(\{X_1 = 2\} \cap \{X_2 = 1\}) = a_{21}$$

e utilizziamo

$$a_{12} + a_{21} = \frac{1}{6} \implies a_{12} + \frac{1}{12} = \frac{1}{6} \implies a_{12} = \frac{1}{12}\,.$$

Imponiamo la seconda condizione, ricordando che

$$\mathbb{P}(\{W < \tfrac{1}{12}\}) = \int_0^{\frac{1}{12}} 1 \, \mathrm{d}x = \frac{1}{12}\,.$$

Otteniamo

$$\frac{1}{12} = \mathbb{P}(\{W < \frac{1}{12}\}) = \mathbb{P}(\{X_1 = 3\} \cap \{X_2 = 1\}) = a_{31}\,.$$

Utilizziamo

$$a_{13} + a_{31} = \frac{1}{6} \implies a_{13} + \frac{1}{12} = \frac{1}{6} \implies a_{13} = \frac{1}{12}\,.$$

Imponiamo l'ultima condizione:

$$\frac{1}{4} = \mathbb{P}(\{X_1 = 3\}|\{X_2 = 2\}) = \frac{\mathbb{P}(\{X_1 = 3\} \cap \{X_2 = 2\})}{\mathbb{P}(\{X_2 = 2\})} = \frac{a_{32}}{a_{12} + a_{22} + a_{32}} = \frac{a_{32}}{\frac{1}{12} + \frac{1}{6} + a_{32}}$$

$$\implies a_{32} = \frac{1}{4}\Big(\frac{1}{4} + a_{32}\Big) = \frac{1}{16} + \frac{1}{4}a_{32} \implies \frac{3}{4}a_{32} = \frac{1}{16} \implies a_{32} = \frac{1}{12}\,.$$

Utilizziamo

$$a_{23} + a_{32} = \frac{1}{6} \implies a_{23} + \frac{1}{12} = \frac{1}{6} \implies a_{23} = \frac{1}{12}\,.$$

Aggiorniamo la tabella:

X_2 \ X_1	1	2	3
1	1/6	1/12	1/12
2	1/12	1/6	1/12
3	1/12	1/12	1/6

3. Osserviamo che

$$\mathbb{P}(\{X_1 = 3\}) = a_{31} + a_{32} + a_{33} = \frac{1}{12} + \frac{1}{12} + \frac{1}{6} = \frac{1}{3}$$

e, invece,

$$\mathbb{P}(\{X_1 = 3\}|\{X_2 = 2\}) = \frac{1}{4}\,.$$

Quindi l'evento $\{X_2 = 2\}$ influenza il valore assunto da X_1. Le due variabili non sono indipendenti.

Esercizio 3. (8 punti) Alice e Bob fanno un gioco. Chiamano un numero verde e attendono che un/a operatore/trice risponda. Il tempo di attesa per la risposta è distribuito con legge esponenziale. Se la risposta avviene entro i 10 minuti vince Alice, altrimenti vince Bob. Supponiamo che il gioco sia equo (Alice vince con il 50% di probabilità).

1. In media, dopo quanto tempo risponde l'operatore/trice?

2. Calcolare la probabilità che la risposta avvenga tra 5 minuti e 10 minuti.

3. Alice e Bob chiamano il numero verde. Hanno aspettato 5 minuti e ancora nessuno ha risposto. Come viene aggiornata la probabilità che Bob vinca? Motivare la risposta.

Alice e Bob giocano 2 volte in sequenza. Appena qualcuno risponde, chiudono la chiamata e richiamano. Si assumano i tempi di attesa indipendenti.

4. Calcolare la varianza della durata dell'intera sequenza di chiamate.

5. Calcolare la probabilità che l'intera sequenza di chiamate duri più di 30 min. (Suggerimento: integrare per parti)

Soluzione. Consideriamo la variabile aleatoria

$$X = \text{“tempo di attesa per una chiamata”} \sim \mathrm{Exp}(\lambda)\,.$$

Ci viene detto che il gioco è equo, quindi

$$\frac{1}{2} = \mathbb{P}(\{X < 10\}) = \int_0^{10} \lambda e^{-\lambda x}\,\mathrm{d}x = \Big[-e^{-\lambda x}\Big]_0^{10} = 1 - e^{-10\lambda}\,.$$

Da questa equazione possiamo ricavare λ:

$$e^{-10\lambda} = \frac{1}{2} \implies -10\lambda = \ln\left(\frac{1}{2}\right) = -\ln 2 \implies \lambda = \frac{\ln 2}{10} \simeq 6.93 \cdot 10^{-2}\,.$$

1. Calcoliamo
$$\mathbb{E}(X) = \frac{1}{\lambda} = \frac{10}{\ln 2} \simeq 14.43\,.$$

2. Calcoliamo
$$\begin{aligned}\mathbb{P}(\{5 \le X \le 10\}) &= \int_5^{10} \lambda e^{-\lambda x}\,\mathrm{d}x = \Big[-e^{-\lambda x}\Big]_5^{10} = e^{-5\lambda} - e^{-10\lambda} = e^{-5\frac{\ln 2}{10}} - e^{-10\frac{\ln 2}{10}} \\ &= e^{-\frac{\ln 2}{2}} - e^{-\ln 2} = \frac{1}{\sqrt{2}} - \frac{1}{2} \simeq 20.71\%\,.\end{aligned}$$

3. Utilizziamo l'assenza di memoria della legge geometrica per calcolare
$$\mathbb{P}(\{X > 10\} | \{X > 5\}) = \mathbb{P}(\{X > 5\}) = e^{-5\lambda} = \frac{1}{\sqrt{2}} \simeq 70.71\%\,.$$

4. Consideriamo due variabili aleatorie indipendenti $X_1, X_2 \sim \mathrm{Exp}(\lambda)$. Il tempo dell'intera sequenza è $X_1 + X_2$. Poiché X_1 e X_2 sono indipendenti, abbiamo che

$$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) = \frac{1}{\lambda^2} + \frac{1}{\lambda^2} = \frac{2}{\lambda^2} \simeq 416.27\,.$$

5. Il tempo totale $X_1 + X_2$ è distribuito con legge Gamma$(2, \lambda)$, quindi ha densità

$$f_{X_1+X_2}(x) = \begin{cases} \frac{\lambda^2}{\Gamma(2)} x^{2-1} e^{-\lambda x}\,, & \text{se } x > 0\,, \\ 0\,, & \text{altrimenti.} \end{cases}$$

Ricordiamo che $\Gamma(2) = (2-1)! = 1$. Possiamo allora calcolare, integrando per parti,

$$\begin{aligned}
\mathbb{P}(\{X_1 + X_2 > 30\}) &= \int_{30}^{+\infty} \lambda^2 x e^{-\lambda x} \, \mathrm{d}x \\
&= \Big[-\lambda x e^{-\lambda x} \Big]_{30}^{+\infty} + \int_{30}^{+\infty} \lambda e^{-\lambda x} \, \mathrm{d}x \\
&= 30\lambda e^{-30\lambda} + e^{-30\lambda} = (30\lambda + 1) e^{-30\lambda} \\
&= (\ln 8 + 1) \frac{1}{8} = 38.49\% \,.
\end{aligned}$$

Esercizio 4. (7 punti) Supponiamo di eseguire uno studio immunologico in un campione di individui, studiando la reazione all'antigene nel punto di inoculo. Viene misurato il diametro in mm dell'alone cutaneo in un campione, ottenendo i seguenti dati:

$$21 \quad 21 \quad 40 \quad 36 \quad 32 \quad 25 \quad 8 \quad 27$$

Si assuma che la distribuzione del diametro sia normale.

1. Si può stabilire con significatività del 5% che la media del diametro è diversa da 30 mm?
2. Si può stabilire con significatività del 5% che la varianza del diametro è superiore a 100 mm^2?

In entrambi i casi, derivare le formule utilizzate per rispondere.

Soluzione. 1. Si tratta di una popolazione $X \sim \mathcal{N}(\mu, \sigma^2)$ da cui è estratto un campione casuale $X_1, \ldots, X_n$ con $n = 8$. La media μ e la varianza σ^2 sono incognite.

Sia $\mu_0 = 30$. Consideriamo il seguente test di ipotesi

$$H_0 : \mu = \mu_0 \,, \quad H_1 : \mu \neq \mu_0 \,,$$

con livello di significatività $\alpha = 5\%$.

Poiché l'ipotesi alternativa è $H_1 : \mu \neq \mu_0$, i dati saranno significativi se la media è sufficientemente distante da μ_0. La regione critica è allora della forma

$$R_c = \{(x_1, \ldots, x_n) \in R(X_1, \ldots, X_n) \ : \ |\overline{x}_n - \mu_0| > \delta\} \,,$$

dove $\overline{x}_n = \frac{1}{n}\sum_{i=1}^n x_i$ è la media calcolata sulla realizzazione $x_1, \ldots, x_n$ del campione casuale.

Assumiamo l'ipotesi nulla H_0 vera, cioè $\mu = \mu_0$. Quindi $X_1, \ldots, X_n \sim \mathcal{N}(\mu_0, \sigma^2)$. Non è nota σ^2, quindi verrà stimata dalla varianza campionaria $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$, dove $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ è la media campionaria. Dalla definizione di livello di significatività:

$$\alpha = \mathbb{P}(\{(X_1, \ldots, X_n) \in R_c\}) = \mathbb{P}(\{|\overline{X}_n - \mu_0| > \delta\}) = \mathbb{P}\Big(\Big\{ \frac{|\overline{X}_n - \mu_0|}{S_n/\sqrt{n}} > \frac{\delta}{S_n/\sqrt{n}} \Big\}\Big) \,.$$

Poiché $X_1, \ldots, X_n \sim \mathcal{N}(\mu_0, \sigma^2)$ indipendenti, si ha che $T_{n-1} = \frac{\overline{X}_n - \mu_0}{S_n/\sqrt{n}} \sim t(n-1)$. Quindi, per la simmetria della t-Student

$$\begin{aligned}
\alpha &= \mathbb{P}\Big(\Big\{|T_{n-1}| > \frac{\delta}{S_n/\sqrt{n}}\Big\}\Big) = \mathbb{P}\Big(\Big\{T_{n-1} < -\frac{\delta}{S_n/\sqrt{n}}\Big\}\Big) + \mathbb{P}\Big(\Big\{T > \frac{\delta}{S_n/\sqrt{n}}\Big\}\Big) \\
&= 2\mathbb{P}\Big(\Big\{T_{n-1} > \frac{\delta}{S_n/\sqrt{n}}\Big\}\Big) \,,
\end{aligned}$$

da cui

$$\mathbb{P}\Big(\Big\{T_{n-1} > \frac{\delta}{S_n/\sqrt{n}}\Big\}\Big) = \frac{\alpha}{2}\,.$$

Introduciamo il valore $t_{n-1,\alpha/2}$ tale che

$$\mathbb{P}(\{T_{n-1} > t_{n-1,\alpha/2}\}) = \frac{\alpha}{2}\,.$$

Allora, scegliendo

$$\frac{\delta}{S_n/\sqrt{n}} = t_{n-1,\alpha/2} \implies \delta = \frac{S_n}{\sqrt{n}} t_{n-1,\alpha/2}\,,$$

otteniamo la condizione che definisce la significatività.

In conclusione, la regione critica è

$$R_c = \Big\{(x_1,\ldots,x_n) \in R(X_1,\ldots,X_n)\ :\ |\overline{x}_n - \mu_0| > \frac{s_n}{\sqrt{n}} t_{n-1,\alpha/2}\Big\}\,.$$

Calcoliamo

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{8}(21+21+40+36+32+25+8+27) = 26.25$$

e

$$\begin{aligned} s_n^2 &= \frac{1}{n-1}\Big(\sum_{i=1}^{n} x_i^2 - n\overline{x}_n^2\Big) = \frac{1}{7}(21^2+21^2+40^2+36^2+32^2+25^2+8^2+27^2-8\cdot 26.25^2) \\ &= 101.07\,. \end{aligned}$$

Dalle tavole otteniamo

$$t_{n-1,\alpha/2} = t_{7,0.025} \simeq 2.365\,.$$

Quindi

$$|\overline{x}_n - \mu_0| = |26.25 - 30| = 3.75\,, \quad \frac{s_n}{\sqrt{n}} t_{n-1,\alpha/2} = \frac{\sqrt{101.07}}{\sqrt{8}} 2.365 \simeq 8.41\,.$$

I dati non sono significativi al 5% per stabilire che la media è diversa da 30.

2. Sia $\sigma_0^2 = 100$. Consideriamo il seguente test di ipotesi

$$H_0 : \sigma^2 \le \sigma_0^2\,, \quad H_1 : \sigma^2 > \sigma_0^2\,,$$

con livello di significatività $\alpha = 5\%$.

Poiché l'ipotesi alternativa è $H_1 : \sigma^2 > \sigma_0^2$, i dati saranno significativi se la varianza è sufficientemente più grande di σ_0^2. La regione critica è allora della forma

$$R_c = \{(x_1,\ldots,x_n) \in R(X_1,\ldots,X_n)\ : s_n^2 > \delta\sigma_0^2\}\,,$$

dove $s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x}_n)^2$ e $\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$ sono la varianza e la media calcolata sulla realizzazione $x_1,\ldots,x_n$ del campione casuale.

Assumiamo l'ipotesi nulla H_0 vera, cioè $\sigma^2 \le \sigma_0^2$. Dalla definizione di livello di significatività e utilizzando la varianza campionaria $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$,

$$\begin{aligned} \mathbb{P}(\{(X_1,\ldots,X_n) \in R_c\}) &= \mathbb{P}(\{S_n^2 > \delta\sigma_0^2\}) \le \mathbb{P}(\{S_n^2 > \delta\sigma^2\}) \\ &= \mathbb{P}\Big(\Big\{\frac{(n-1)S_n^2}{\sigma^2} > (n-1)\delta\Big\}\Big)\,. \end{aligned}$$

Poiché $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ indipendenti, si ha che $Q_{n-1} = \frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$. Introduciamo il valore $\chi^2_{n-1,\alpha}$ tale che

$$\mathbb{P}(\{Q_{n-1} > \chi^2_{n-1,\alpha}\}) = \alpha\,.$$

Allora, scegliendo

$$(n-1)\delta = \chi^2_{n-1,\alpha} \implies \delta = \frac{\chi^2_{n-1,\alpha}}{n-1}\,,$$

otteniamo

$$\mathbb{P}(\{(X_1, \dots, X_n) \in R_c\}) \leq \alpha\,,$$

(ovvero la probabilità di commettere un errore del I tipo è più piccola di α).

In conclusione, la regione critica è

$$R_c = \left\{(x_1, \dots, x_n) \in R(X_1, \dots, X_n) \ :\ s_n^2 > \frac{\chi^2_{n-1,\alpha}}{n-1}\sigma_0^2\right\},$$

Dalle tavole otteniamo

$$\chi^2_{n-1,\alpha} = \chi^2_{7,0.05} \simeq 14.067\,.$$

I dati forniscono i seguenti valori

$$s_n^2 = 101.07\,, \qquad \frac{\chi^2_{n-1,\alpha}}{n-1}\sigma_0^2 = \frac{14.067}{7}100 \simeq 200.96\,.$$

I dati non sono significativi al 5% per stabilire che la varianza è più grande di 100.

Esame di Probabilità e Statistica [3231]

Esame di Calcolo delle Probabilità e Statistica [2959]

Corso di Studi di Ingegneria Gestionale (D.M.270/04) (L)

Dipartimento di Meccanica, Matematica e Management
Politecnico di Bari

Cognome: ____________________ Docente: Gianluca Orlando
Nome: ____________________ Appello: novembre 2023
Matricola: ____________________ Data: 06/11/2023

Tempo massimo: 2 ore.

Esercizio 1. (6 punti) È stato misurato il tempo di attesa (in giorni) per un particolare intervento chirurgico su un campione di individui. La distribuzione dei tempi di attesa è riassunta nella seguente tabella:

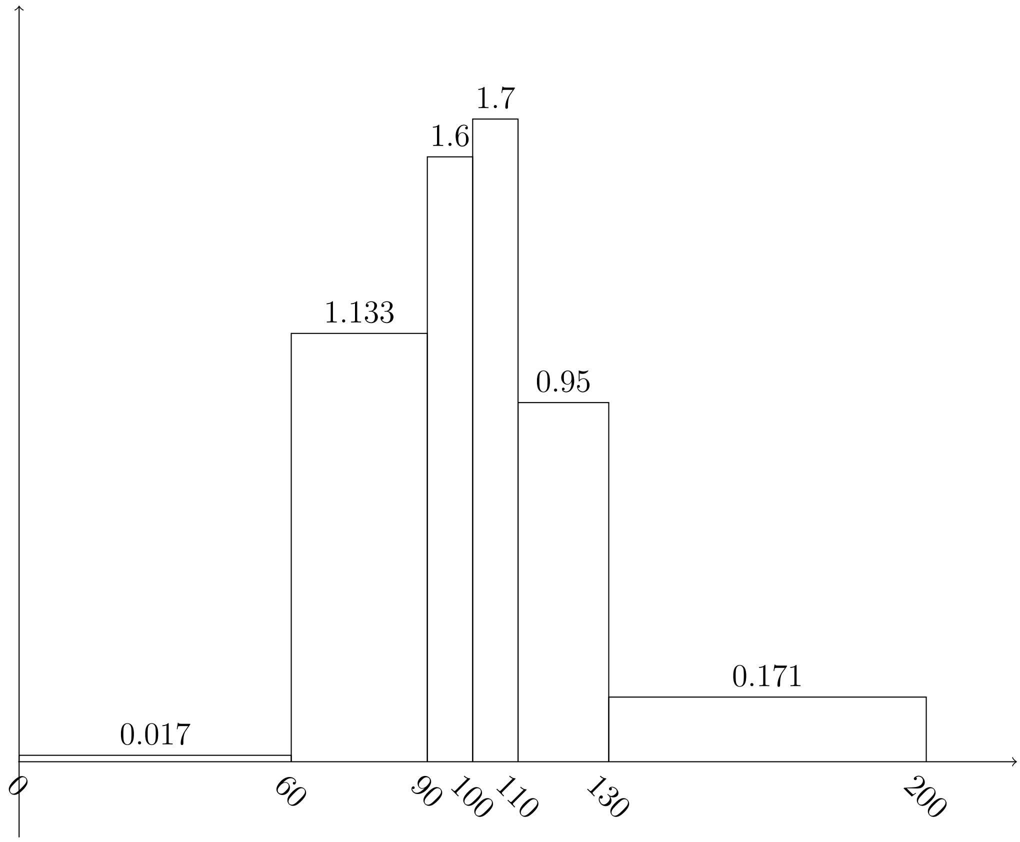
intervalli (giorni)	frequenza assoluta
$[0, 60)$	1
$[60, 90)$	34
$[90, 100)$	16
$[100, 110)$	18
$[110, 130)$	19
$[130, 200)$	12

1. Rappresentare un istogramma delle densità di frequenze assolute.
2. Determinare la classe modale.
3. Calcolare un'approssimazione della media e della deviazione standard dei dati.
4. Calcolare un'approssimazione della mediana dei dati.

Soluzione. 1. Completiamo la tabella con tutte le informazioni necessarie. Ricordiamo che la densità di frequenza assoluta si ottiene rapportando le frequenze assolute all'ampiezza degli intervalli.

intervallo	freq. assolute	freq. relative	densità di freq. ass.	freq. cumulate
$[0, 60)$	1	1%	0.017	1
$[60, 90)$	34	34%	1.133	35
$[90, 100)$	16	16%	1.6	51
$[100, 110)$	18	18%	1.7	69
$[110, 130)$	19	19%	0.95	88
$[130, 200)$	12	12%	0.171	100

Rappresentiamo le densità di frequenze assolute in un istogramma.



2. La classe modale è quella con maggiore densità di frequenza assoluta, quindi è l'intervallo $[100, 110)$.

3. Per calcolare un'approssimazione della media utilizziamo le frequenze relative ottenute da $p_j = f_j/n$ dove $n = 100$ e i valori centrali $\tilde{v}_j$ degli intervalli

$$\begin{aligned}\overline{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i \simeq \frac{1}{n}\sum_{j=1}^{k} f_j\tilde{v}_j = \sum_{j=1}^{k} p_j\tilde{v}_j \\ &= 1\% \cdot 30 + 34\% \cdot 75 + 16\% \cdot 95 + 18\% \cdot 105 + 19\% \cdot 120 + 12\% \cdot 165 \simeq 102.5\,.\end{aligned}$$

Calcoliamo un'approssimazione della varianza

$$\begin{aligned}s^2 &= \frac{1}{n-1}\Big(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\Big) \simeq \frac{1}{n-1}\Big(\sum_{j=1}^{n} f_j\tilde{v}_j^2 - n\overline{x}^2\Big) = \frac{n}{n-1}\Big(\sum_{j=1}^{n} p_j\tilde{v}_j^2 - \overline{x}^2\Big) \\ &= \frac{100}{99}\Big(1\% \cdot 30^2 + 34\% \cdot 75^2 + 16\% \cdot 95^2 + 18\% \cdot 105^2 + 19\% \cdot 120^2 + 12\% \cdot 165^2 - 102.5^2\Big) \\ &= \frac{100}{99}(11353 - 10506.25) \simeq 855.30\,,\end{aligned}$$

da cui

$$s \simeq 29.25\,.$$

4. Per calcolare un'approssimazione della mediana dei dati, usiamo le frequenze cumulate. Troviamo l'intervallo I_j tale che $F_j \leq \frac{n}{2} = 50 < F_{j+1}$. Si tratta dell'intervallo $[90, 100)$. Approssimiamo la mediana con

$$Q_2 \simeq a_j + \lambda_j(b_j - a_j)$$

dove

$$\lambda_j = \frac{n/2 - F_j}{F_{j+1} - F_j} = \frac{50 - 35}{51 - 35} = 0.9375\,.$$

Quindi

$$Q_2 \simeq 90 + 0.9375(100 - 90) = 99.375\,.$$

Esercizio 2. (8 punti) Si consideri un vettore aleatorio discreto (X_1, X_2) con $R(X_1) = \{0, 1\}$ e $R(X_2) = \{0, 1\}$. Si assuma che $\mathrm{Cov}(X_1, X_2) = 0$.

1. Dimostrare che X_1 e X_2 sono indipendenti. (N.B.: Il punteggio massimo si ottiene se non si utilizzano le informazioni di sotto.)

Si assuma ulteriormente che:

- $\mathrm{Var}(X_1) = \frac{1}{4}$.
- $\mathbb{P}(\{X_2 = 1\}|\{X_1 = 1\}) = \frac{1}{3}$.

Quesiti:

2. Calcolare $\mathbb{P}(\{X_1 \cdot X_2 > 0\})$.
3. Calcolare $\mathbb{P}(\{X_1 + X_2 = 1\})$.
4. Alice e Bob giocano al seguente gioco in cui si susseguono in *lanci*: Alice inizia e *lancia* generando un numero con X_1; Bob segue e *lancia* generando un numero con X_2. Alice e Bob continuano a *lanciare* alternandosi, finché i risultati degli ultimi due *lanci* non sono diversi. In tal caso vince chi dei due ha ottenuto il risultato più grande negli ultimi due lanci. Esempi:

$$X_1 = 1 \text{ (Alice)} \to X_2 = 1 \text{ (Bob)} \to X_1 = 0 \text{ (Alice)} \to \text{Fine: Bob vince.}$$

$$X_1 = 0 \text{ (Alice)} \to X_2 = 0 \text{ (Bob)} \to X_1 = 0 \text{ (Alice)} \to X_2 = 1 \text{ (Bob)} \to \text{Fine: Bob vince.}$$

Che probabilità di vincere ha Alice?

Soluzione. Scriviamo una tabella per la funzione di probabilità congiunta:

X_2 \ X_1	0	1
0	a_{00}	a_{10}
1	a_{01}	a_{11}

1. Poiché $R(X_1) = \{0, 1\}$ e $R(X_2) = \{0, 1\}$, entrambe sono variabili aleatorie distribuite con legge di Bernoulli, quindi $X_1 \sim \mathrm{Be}(p)$ e $X_2 \sim \mathrm{Be}(q)$ per certi p e q. Ricordando che $\mathbb{E}(X_1) = p$ e $\mathbb{E}(X_2) = q$, usiamo che $\mathrm{Cov}(X_1, X_2) = 0$ per dedurre

$$0 = \mathrm{Cov}(X_1, X_2) = \mathbb{E}(X_1 \cdot X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2) = a_{11} - pq \implies pq = a_{11}\,.$$

Otteniamo che

$$p = \mathbb{P}(\{X_1 = 1\}) = a_{10} + a_{11} \implies a_{10} = p - a_{11} = p - pq = p(1 - q)\,.$$

Analogamente

$$q = \mathbb{P}(\{X_2 = 1\}) = a_{01} + a_{11} \implies a_{01} = q - a_{11} = q - pq = q(1-p)\,.$$

Infine,

$$1 - p = \mathbb{P}(\{X_1 = 0\}) = a_{00} + a_{01} = a_{00} + q - pq \implies a_{00} = 1 - p - q + pq = (1-p)(1-q)\,.$$

Possiamo completare la tabella:

X_2 \ X_1	0	1
0	$(1-p)(1-q)$	$p(1-q)$
1	$(1-p)q$	pq

e osserviamo che $\mathbb{P}(\{X_1 = i\} \cap \{X_2 = j\}) = \mathbb{P}(\{X_1 = i\})\mathbb{P}(\{X_2 = j\})$ per ogni $i, j \in \{0, 1\}$.

2. Ricordiamo che per una Bernoulli $\mathrm{Var}(X_1) = p(1-p)$. Quindi

$$p - p^2 = \frac{1}{4} \implies p^2 - p + \frac{1}{4} = 0 \implies \left(p - \frac{1}{2}\right)^2 = 0 \implies p = \frac{1}{2}\,.$$

Usiamo la seconda informazione. Si ha che

$$\frac{1}{3} = \mathbb{P}(\{X_2 = 1\} | \{X_1 = 1\}) = \frac{\mathbb{P}(\{X_2 = 1\} \cap \{X_1 = 1\})}{\mathbb{P}(\{X_1 = 1\})} = \frac{pq}{p} = q\,.$$

(Chi riesce a dimostrare il punto 1., può usarlo qui per dire che $\mathbb{P}(\{X_2 = 1\}|\{X_1 = 1\})) = \mathbb{P}(\{X_2 = 1\}) = q$.) Abbiamo la tabella completa:

X_2 \ X_1	0	1
0	$\frac{1}{3}$	$\frac{1}{3}$
1	$\frac{1}{6}$	$\frac{1}{6}$

Possiamo calcolare

$$\mathbb{P}(\{X_1 \cdot X_2 > 0\}) = \mathbb{P}(\{X_1 \cdot X_2 = 1\}) = \frac{1}{6}\,.$$

3. Si ha che

$$\mathbb{P}(\{X_1 + X_2 = 1\}) = \mathbb{P}(\{X_1 = 1\} \cap \{X_2 = 0\}) + \mathbb{P}(\{X_1 = 0\} \cap \{X_2 = 1\}) = \frac{1}{3} + \frac{1}{6} = \frac{1}{2}\,.$$

4. Consideriamo l'evento

$$A = \text{“Alice vince”}$$

e la sequenza di variabili aleatorie

$$Y_n = \text{“esito dell'}n\text{-esimo } \textit{lancio}\text{”}$$

Distinguiamo i casi in base al risultato iniziale. Supponiamo di sapere che al primo lancio $Y_1 = 1$. Se Bob ottiene $Y_2 = 0$ (con probabilità $\mathbb{P}(\{X_2 = 0\}) = 1 - q = \frac{2}{3}$), Alice vince automaticamente; se Bob ottiene $Y_2 = 1$ (con probabilità $\mathbb{P}(\{X_2 = 1\}) = q = \frac{1}{3}$), tocca nuovamente a Alice. Se Alice ottiene $Y_3 = 0$ (con probabilità $\mathbb{P}(\{X_1 = 0\}) = 1 - p = \frac{1}{2}$), perde;

se Alice ottiene $Y_3 = 1$ (con probabilità $\mathbb{P}(\{X_1 = 1\}) = p = \frac{1}{2}$), il gioco si è resettato e siamo nella situazione precedente! Scriviamo questa relazione ricorsiva:

$$\begin{aligned}
\mathbb{P}(A|\{Y_1 = 1\}) &= (1-q) + q\Big(p\mathbb{P}(A|\{Y_1 = 1\})\Big) \\
&= \frac{2}{3} + \frac{1}{3}\Big(\frac{1}{2}\mathbb{P}(A|\{Y_1 = 1\})\Big) \\
&\implies \frac{5}{6}\mathbb{P}(A|\{Y_1 = 1\}) = \frac{2}{3} \\
&\implies \mathbb{P}(A|\{Y_1 = 1\}) = \frac{4}{5}\,.
\end{aligned}$$

Supponiamo di sapere che al primo lancio $Y_1 = 0$. Se Bob ottiene $Y_2 = 1$ (con probabilità $\mathbb{P}(\{X_2 = 1\}) = q = \frac{1}{3}$), Alice perde automaticamente; se Bob ottiene $Y_2 = 0$ (con probabilità $\mathbb{P}(\{X_2 = 0\}) = 1 - q = \frac{2}{3}$), tocca nuovamente a Alice. Se Alice ottiene $Y_3 = 1$ (con probabilità $\mathbb{P}(\{X_1 = 1\}) = p = \frac{1}{2}$), vince; se Alice ottiene $Y_3 = 0$ (con probabilità $\mathbb{P}(\{X_1 = 0\}) = 1-p = \frac{1}{2}$), il gioco si è resettato e siamo nella situazione precedente! Scriviamo questa relazione ricorsiva:

$$\begin{aligned}
\mathbb{P}(A|\{Y_1 = 0\}) &= (1-q)\Big(p + (1-p)\mathbb{P}(A|\{Y_1 = 0\})\Big) \\
&= \frac{2}{3}\Big(\frac{1}{2} + \frac{1}{2}\mathbb{P}(A|\{Y_1 = 0\})\Big) \\
&\implies \frac{2}{3}\mathbb{P}(A|\{Y_1 = 0\}) = \frac{1}{3} \\
&\implies \mathbb{P}(A|\{Y_1 = 0\}) = \frac{1}{2}\,.
\end{aligned}$$

Utilizziamo la formula della probabilità totale per concludere che

$$\mathbb{P}(A) = \mathbb{P}(A|\{Y_1 = 1\})\mathbb{P}(\{Y_1 = 1\}) + \mathbb{P}(A|\{Y_1 = 0\})\mathbb{P}(\{Y_1 = 0\}) = \frac{4}{5} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{13}{20}\,.$$

Esercizio 3. (8 punti) Si assuma che il ritardo di un singolo treno di un'azienda di trasporti sia distribuito con legge esponenziale. Sono gestiti due tipi di treni: il 20% dei treni sono treni veloci; il resto sono treni regionali. Si assumano i seguenti fatti:

- il ritardo medio di un treno veloce è 5 minuti;
- la probabilità che un treno regionale ritardi meno di $\ln(3^{10})$ (logaritmo naturale di tre elevato alla 10) minuti è uguale a $\frac{2}{3}$.

Rispondere ai seguenti quesiti:

1. Si consideri un treno veloce. Calcolare la probabilità che il ritardo sia compreso tra 3 e 7 minuti.
2. Calcolare la deviazione standard del ritardo di un treno regionale.
3. Si consideri un treno qualunque. Qual è la probabilità che ritardi più di 8 minuti?
4. (Difficile) Sei su un treno veloce e devi prendere una coincidenza con un treno regionale ad una certa stazione. La tabella di marcia prevede che il treno veloce arrivi e il treno regionale arrivino alla stazione entrambi alle 14:00. Considerando i possibili ritardi, con che probabilità perderai la coincidenza? Supponi che i ritardi siano indipendenti e che il tempo per passare da un treno all'altro sia trascurabile.

Soluzione. Consideriamo tre variabili aleatorie

$$X_{\text{vel}} = \text{“ritardo di un treno veloce”} \sim \text{Exp}(\lambda)\,,$$

$$X_{\text{reg}} = \text{“ritardo di un treno regionale”} \sim \text{Exp}(\mu)\,,$$

$$Y = \text{“tipo di treno”} \sim \text{Be}(20\%)\,,$$

dove $Y = 1$ se il treno è veloce, $Y = 0$ se il treno è regionale.

Per determinare λ, ricordiamo che $\mathbb{E}(X_{\text{vel}}) = \frac{1}{\lambda}$, quindi $\frac{1}{\lambda} = 5$, da cui $\lambda = \frac{1}{5}$.

Per determinare μ, chiamiamo $a = \ln(3^{10})$ e imponiamo

$$\frac{2}{3} = \mathbb{P}(\{X_{\text{reg}} \le a\}) = \int_0^a \mu e^{-\mu x}\,\mathrm{d}x = \Big[- e^{-\mu x} \Big]_{x=0}^{x=a} = 1 - e^{-\mu\, a} \implies e^{-\mu\, a} = 1 - \frac{2}{3} = \frac{1}{3}$$

$$\implies -\mu\, a = \ln\left(\frac{1}{3}\right) \implies \mu = \frac{\ln(3)}{a} = \frac{\ln(3)}{\ln(3^{10})} = \frac{1}{10}\,.$$

1. Calcoliamo

$$\mathbb{P}(\{3 \le X_{\text{vel}} \le 7\}) = \int_3^7 \lambda e^{-\lambda x}\,\mathrm{d}x = \Big[e^{-\lambda x} \Big]_{x=3}^{x=7} = e^{-3\lambda} - e^{-7\lambda} = e^{-\frac{3}{5}} - e^{-\frac{7}{5}} \simeq 30.22\%\,.$$

2. Ricordiamo che $\text{Var}(X_{\text{reg}}) = \frac{1}{\lambda^2}$, quindi la deviazione standard è $\sqrt{\text{Var}(X_{\text{reg}})} = \frac{1}{\lambda} = 10$.
3. Consideriamo la variabile aleatoria

$$X = \text{“ritardo di un treno”}\,.$$

Per rispondere al quesito, utilizziamo il teorema della probabilità totale e utilizziamo il fatto di conoscere la distribuzione per i treni veloci e i treni regionali:

$$\begin{aligned}
\mathbb{P}(\{X \ge 8\}) &= \mathbb{P}(\{X \ge 8\}|\{Y = 1\})\mathbb{P}(\{Y = 1\}) + \mathbb{P}(\{X \ge 8\}|\{Y = 0\})\mathbb{P}(\{Y = 0\}) \\
&= \mathbb{P}(\{X_{\text{vel}} \ge 8\})\mathbb{P}(\{Y = 1\}) + \mathbb{P}(\{X_{\text{reg}} \ge 8\})\mathbb{P}(\{Y = 0\}) \\
&= e^{-8\lambda}\, 20\% + e^{-8\mu}\, 80\% = e^{-\frac{8}{5}}\, 20\% + e^{-\frac{8}{10}}\, 80\% \simeq 39.98\%\,.
\end{aligned}$$

4. Si perde la coincidenza se il ritardo del treno veloce è maggiore del ritardo del treno regionale, ovvero $\{X_{\text{vel}} > X_{\text{reg}}\}$. Per l'indipendenza, la densità congiunta è data da $f_{(X_{\text{vel}},X_{\text{reg}})}(x_1, x_2) = \lambda\mu e^{-\lambda x_1} e^{-\mu x_2}$ per $x_1 > 0$ e $x_2 > 0$, zero altrimenti. Allora, integrando sul dominio normale $\{x_1 > x_2,\ x_1 > 0,\ x_2 > 0\}$, otteniamo

$$\begin{aligned}
\mathbb{P}(\{X_{\text{vel}} > X_{\text{reg}}\}) &= \iint_{\{x_1 > x_2\}} f_{(X_{\text{vel}},X_{\text{reg}})}(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 \\
&= \int_0^{+\infty} \Big(\int_0^{x_1} \lambda\mu e^{-\lambda x_1} e^{-\mu x_2}\,\mathrm{d}x_2 \Big)\,\mathrm{d}x_1 \\
&= \int_0^{+\infty} \lambda e^{-\lambda x_1} \Big(\int_0^{x_1} \mu e^{-\mu x_2}\,\mathrm{d}x_2 \Big)\,\mathrm{d}x_1 = \\
&= \int_0^{+\infty} \lambda e^{-\lambda x_1} \Big(1 - e^{-\mu x_1} \Big)\,\mathrm{d}x_1 \\
&= \int_0^{+\infty} \lambda e^{-\lambda x_1}\,\mathrm{d}x_1 - \int_0^{+\infty} \lambda e^{-(\lambda+\mu)x_1}\,\mathrm{d}x_1 \\
&= 1 - \frac{\lambda}{\lambda + \mu} = \frac{\mu}{\lambda + \mu} = \frac{\frac{1}{10}}{\frac{1}{5} + \frac{1}{10}} = \frac{1}{3}\,.
\end{aligned}$$

Esercizio 4. (7 punti) Un'azienda farmaceutica produce tranquillanti. Si assuma che la distribuzione della durata della loro efficacia sia normale. Un laboratorio di ricerca di un ospedale ha sperimentato il farmaco su un campione casuale di 6 pazienti. Il periodo di efficacia del tranquillante per ogni paziente (in ore) è stato il seguente:

$$2.7 \quad 2.8 \quad 3.0 \quad 2.3 \quad 2.3 \quad 2.2 \,.$$

1. Calcolare sui dati un intervallo di confidenza bilaterale al 90% per la media del periodo di efficacia.

2. Supponiamo che l'esperimento di sopra venga svolto 5 volte (in ogni esperimento vengono misurati i dati di 6 pazienti). Dopo ogni campionamento, viene calcolato un intervallo di confidenza bilaterale al 90% per la media come sopra. Qual è la probabilità che la media vera sia effettivamente nell'intervallo calcolato almeno 4 volte?

Soluzione. 1. Abbiamo un campione casuale $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ con $n = 6$. Sia μ che σ^2 sono incognite.

Dalla definizione di IC si ha che, ponendo $\beta = 90\%$,

$$\beta = \mathbb{P}(\{U_n \le \mu \le V_n\}) \,.$$

Per stimare μ sfrutteremo lo stimatore media campionaria $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Per stimare σ^2 sfrutteremo lo stimatore varianza campionaria $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$. Poiché $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ sono indipendenti, $T_{n-1} = \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \sim t(n-1)$. Allora

$$\begin{aligned}
\beta = \mathbb{P}(\{U_n \le \mu \le V_n\}) &= \mathbb{P}\Big(\Big\{\frac{\overline{X}_n - V_n}{S_n/\sqrt{n}} \le \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \le \frac{\overline{X}_n - U_n}{S_n/\sqrt{n}}\Big\}\Big) \\
&= \mathbb{P}\Big(\Big\{\frac{\overline{X}_n - V_n}{S_n/\sqrt{n}} \le T_{n-1} \le \frac{\overline{X}_n - U_n}{S_n/\sqrt{n}}\Big\}\Big) \\
&= 1 - \mathbb{P}\Big(\Big\{T_{n-1} < \frac{\overline{X}_n - V_n}{S_n/\sqrt{n}}\Big\}\Big) - \mathbb{P}\Big(\Big\{T_{n-1} > \frac{\overline{X}_n - U_n}{S_n/\sqrt{n}}\Big\}\Big) \,.
\end{aligned}$$

Segue che

$$\mathbb{P}\Big(\Big\{T_{n-1} < \frac{\overline{X}_n - V_n}{S_n/\sqrt{n}}\Big\}\Big) + \mathbb{P}\Big(\Big\{T_{n-1} > \frac{\overline{X}_n - U_n}{S_n/\sqrt{n}}\Big\}\Big) = 1 - \beta = \alpha = 10\% \,.$$

Decidiamo di equipartire α:

$$\mathbb{P}\Big(\Big\{T_{n-1} < \frac{\overline{X}_n - V_n}{S_n/\sqrt{n}}\Big\}\Big) = \mathbb{P}\Big(\Big\{T_{n-1} < \frac{\overline{X}_n - V_n}{S_n/\sqrt{n}}\Big\}\Big) = \frac{\alpha}{2} = 5\% \,. \tag{1}$$

Definiamo $t_{n-1,\alpha/2}$ come il punto tale che

$$\mathbb{P}(\{T_{n-1} \ge t_{n-1,\alpha/2}) = \frac{\alpha}{2} \,.$$

Scegliendo

$$\begin{aligned}
\frac{\overline{X}_n - U_n}{S_n/\sqrt{n}} &= t_{n-1,\alpha/2} \implies U_n = \overline{X}_n - \frac{S_n}{\sqrt{n}} t_{n-1,\alpha/2} \,, \\
\frac{\overline{X}_n - V_n}{S_n/\sqrt{n}} &= -t_{n-1,\alpha/2} \implies V_n = \overline{X}_n + \frac{S_n}{\sqrt{n}} t_{n-1,\alpha/2} \,,
\end{aligned}$$

si ottiene (1). In conclusione

$$\left[\overline{X}_n - \frac{S_n}{\sqrt{n}} t_{n-1,\alpha/2}, \overline{X}_n + \frac{S_n}{\sqrt{n}} t_{n-1,\alpha/2}\right]$$

è un intervallo di confidenza bilaterale per μ con livello di confidenza $\beta = 1 - \alpha$. Calcoliamolo sui dati. Dalla tavola della t-Student otteniamo che

$$t_{n-1,\alpha/2} = t_{5,0.05} = 2.015\,.$$

Calcoliamo media e varianza sul campione di dati:

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{6}(2.7 + 2.8 + 3.0 + 2.3 + 2.3 + 2.2) = 2.55\,,$$

$$\begin{aligned} s_n &= \sqrt{\frac{1}{n-1}\Big(\sum_{i=1}^{n} x_i^2 - n\overline{x}_n^2\Big)} = \sqrt{\frac{1}{5}\Big(2.7^2 + 2.8^2 + 3.0^2 + 2.3^2 + 2.3^2 + 2.2^2 - 6 \cdot 2.55^2\Big)} \\ &= \sqrt{0.107} \simeq 0.33\,. \end{aligned}$$

Segue che

$$\left[\overline{x}_n - \frac{s_n}{\sqrt{n}} t_{n-1,\alpha/2}, \overline{x}_n + \frac{s_n}{\sqrt{n}} t_{n-1,\alpha/2}\right] \simeq [2.27, 2.82]\,.$$

2. Consideriamo una variabile aleatoria

$$Y = \text{“numero di volte che la media è nell'IC al 90\% su 5 tentativi”} \sim \mathrm{B}(5, 90\%)\,.$$

Allora

$$\mathbb{P}(\{Y \geq 4\}) = \mathbb{P}(\{Y = 4\}) + \mathbb{P}(\{Y = 5\}) = \binom{5}{4}(90\%)^4 10\% + \binom{5}{5}(90\%)^5 = 91.854\%\,.$$

Esame di Probabilità e Statistica [3231]

Esame di Calcolo delle Probabilità e Statistica [2959]

Corso di Studi di Ingegneria Gestionale (D.M.270/04) (L)

Dipartimento di Meccanica, Matematica e Management
Politecnico di Bari

Cognome: ______________________	Docente: Gianluca Orlando
Nome: ______________________	Appello: gennaio 2024
Matricola: ______________________	Data: 23/01/2024

Tempo massimo: 2 ore.

Esercizio 1. (6 punti) Uno studio esamina la relazione tra l'esposizione al rumore e l'ipertensione. Seguono alcuni dati. I valori x rappresentano il livello di pressione sonora riportato in decibel e i valori y rappresentano l'incremento di pressione sanguigna in mm Hg:

x	60	85	100	70	80	63	90
y	1	5	9	1	3	0	4

1. Rappresentare i dati in uno scatterplot.
2. Determinare (derivando le formule) la retta di regressione lineare e rappresentarla.
3. Determinare il coefficiente di correlazione lineare.

1. Segue lo scatterplot (con la retta di regressione lineare):

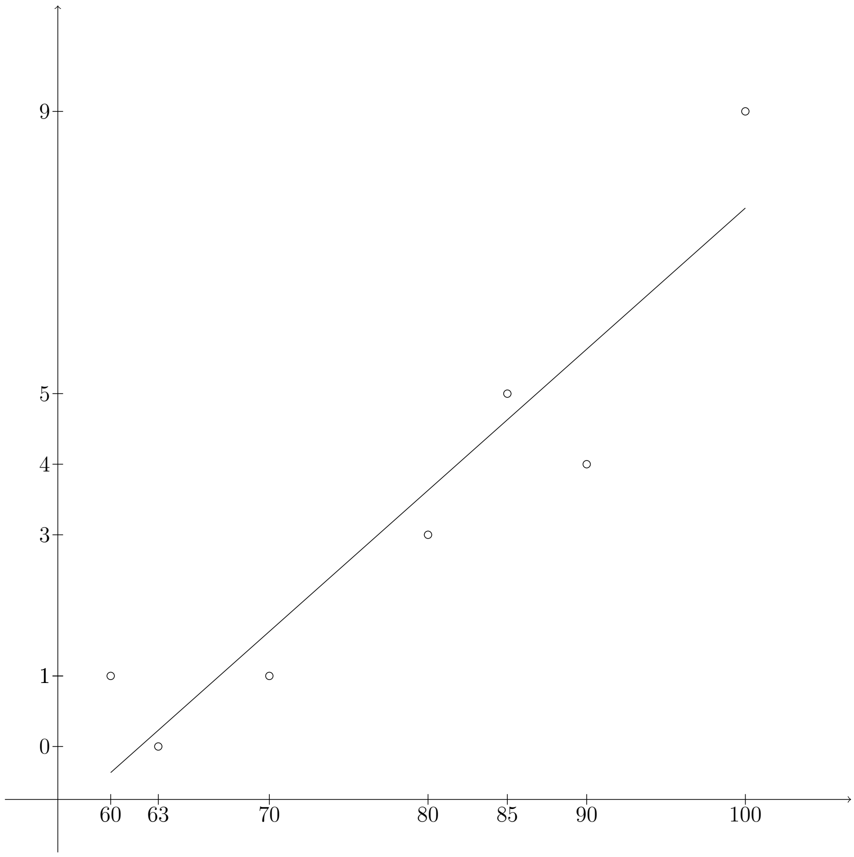


Figura 1: Scatterplot e retta di regressione lineare.

2. Denotiamo con $(x_1, y_1), \ldots, (x_n, y_n)$, $n = 7$, i dati del campione. Cerchiamo la retta di equazione

$$y = ax + b$$

che meglio approssima i dati, utilizzando il metodo dei minimi quadrati. Vogliamo minimizzare l'errore

$$r(a, b) = \sum_{i=1}^{n} (y_i - ax_i - b)^2 .$$

Imponiamo che il gradiente rispetto ad (a, b) sia nullo, ovvero,

$$\begin{aligned}
0 &= \partial_a r(a, b) = -2 \sum_{i=1}^{n} (y_i - ax_i - b) x_i = -2 \sum_{i=1}^{n} (x_i y_i - a x_i^2 - b x_i) , \\
0 &= \partial_b r(a, b) = -2 \sum_{i=1}^{n} (y_i - ax_i - b)
\end{aligned}$$

Dalla seconda equazione segue che

$$nb = \sum_{i=1}^{n} (y_i - ax_i) \implies b = \overline{y} - a\overline{x} .$$

Sostituendo nella prima,

$$\sum_{i=1}^{n}(x_iy_i - ax_i^2 - bx_i) = 0 \implies \sum_{i=1}^{n}(x_iy_i - ax_i^2 - x_i\overline{y} + a\overline{x}x_i) = 0$$

$$\implies a\Big(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\Big) = \sum_{i=1}^{n} x_iy_i - n\overline{x}\,\overline{y}$$

$$\implies a = \frac{\sum_{i=1}^{n} x_iy_i - n\overline{x}\,\overline{y}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}\,.$$

Completiamo la tabella con i valori necessari a calcolare a e b:

								somma
x_i	60	85	100	70	80	63	90	548
y_i	1	5	9	1	3	0	4	23
x_i^2	3600	7225	10000	4900	6400	3969	8100	44194
y_i^2	1	25	81	1	9	0	16	133
x_iy_i	60	425	900	70	240	0	360	2055

Pertanto $\overline{x} = 548.00/7 = 78.29$ e $\overline{y} = 23/7 \simeq 3.29$. Segue che:

$$a = \frac{2055.00 - 7 \cdot 78.29 \cdot 3.29}{44194.00 - 7 \cdot 78.29^2} = \frac{251.98}{1288.73} \simeq 0.20\,,$$

$$b = 3.29 - 0.20 \cdot 78.29 \simeq= -12.37\,,$$

ovvero, la retta di regressione lineare ha equazione

$$y = 0.20x - 12.37\,.$$

3. Per calcolare il coefficiente di correlazione lineare usiamo la formula

$$\rho_{x,y} = \frac{\sum_{i=1}^{n} x_iy_i - n\overline{x}\,\overline{y}}{\sqrt{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}\sqrt{\sum_{i=1}^{n} y_i^2 - n\overline{y}^2}} = \frac{251.98}{\sqrt{44194.00 - 7 \cdot 78.29^2}\sqrt{133.00 - 7 \cdot 3.29^2}}$$
$$= \frac{251.98}{\sqrt{1288.73}\sqrt{57.23}} \simeq 0.93\,.$$

Esercizio 2. (7 punti) Un vettore aleatorio (X_1, X_2) ha funzione di probabilità congiunta rappresentata nella seguente tabella:

X_2 \ X_1	0	1	2
0	a_{00}	a_{10}	a_{20}
1	a_{01}	a_{11}	a_{21}

Si lavori sotto le seguenti ipotesi:

- Il range è dato da $R(X) = \{(0,1),(0,2),(1,0),(1,1)\}$.
- $\mathbb{E}(X_1 + X_2) = \frac{3}{2}$.
- Sapendo che si è verificato l'evento $\{X_2 = 0\}$, la probabilità che $\{X_1 = 2\}$ è $\frac{1}{2}$.
- La probabilità che $\{X_1 = 1\}$ è uguale alla probabilità che $\{X_2 = 1\}$.

Si risponda ai seguenti quesiti:

1. Determinare i valori $a_{00}, a_{10}, a_{20}, a_{01}, a_{11}, a_{21}$.
2. Stabilire se X_1 e X_2 sono indipendenti.
3. Calcolare $\mathrm{Var}(X_1 + X_2)$.

Soluzione. 1. Imponiamo tutte le condizioni. Dalle informazioni sul range del vettore aleatorio sappiamo che non possono essere assunti i valori $(0,0)$ e $(2,1)$, quindi

$$0 = \mathbb{P}(\{X_1 = 0, X_2 = 0\}) = a_{00}\,,$$

$$0 = \mathbb{P}(\{X_1 = 2, X_2 = 1\}) = a_{21}\,.$$

Imponiamo inoltre la condizione che la somma di tutte le probabilità sia 1, cioè

$$a_{00} + a_{10} + a_{20} + a_{01} + a_{11} + a_{21} = 1 \implies a_{10} + a_{20} + a_{01} + a_{11} = 1\,. \tag{1}$$

Imponiamo la condizione sul valore atteso. Ricordando la linearità del valore atteso:

$$\frac{3}{2} = \mathbb{E}(X_1 + X_2) = \mathbb{E}(X_1) + \mathbb{E}(X_2)$$

Calcoliamo separatamente i due valori attesi:

$$\begin{aligned}\mathbb{E}(X_1) &= 0 \cdot \mathbb{P}(\{X_1 = 0\}) + 1 \cdot \mathbb{P}(\{X_1 = 1\}) + 2 \cdot \mathbb{P}(\{X_1 = 2\})\\ &= 1 \cdot \Big(a_{10} + a_{11}\Big) + 2 \cdot \Big(a_{20} + a_{21}\Big)\\ &= a_{10} + a_{11} + 2a_{20}\,.\end{aligned}$$

$$\begin{aligned}\mathbb{E}(X_2) &= 0 \cdot \mathbb{P}(\{X_2 = 0\}) + 1 \cdot \mathbb{P}(\{X_2 = 1\})\\ &= 1 \cdot \Big(a_{01} + a_{11} + a_{21}\Big)\\ &= a_{01} + a_{11}\,.\end{aligned}$$

Quindi, utilizzando la condizione (1),

$$\begin{aligned}&\frac{3}{2} = \mathbb{E}(X_1 + X_2) = \mathbb{E}(X_1) + \mathbb{E}(X_2) = a_{10} + a_{11} + a_{20} + a_{01} + a_{11} + a_{20} = 1 + a_{11} + a_{20}\\ &\quad\implies a_{11} + a_{20} = \frac{1}{2}\,.\end{aligned}$$

Sostituendo nella (1),

$$a_{10} + a_{20} + a_{01} + a_{11} = 1 \implies a_{10} + a_{01} = \frac{1}{2}\,.$$

Imponiamo la condizione sulla probabilità condizionata. Le ipotesi stabiliscono che

$$\begin{aligned}&\frac{1}{2} = \mathbb{P}(\{X_1 = 2\}|\{X_2 = 0\}) = \frac{\mathbb{P}(\{X_1 = 2, X_2 = 0\})}{\mathbb{P}(\{X_2 = 0\})} = \frac{a_{20}}{a_{10} + a_{20}}\\ &\quad\implies a_{20} = \frac{1}{2}a_{10} + \frac{1}{2}a_{20} \implies a_{20} = a_{10}\,.\end{aligned}$$

Imponiamo l'ultima condizione

$$\mathbb{P}(\{X_1 = 1\}) = \mathbb{P}(\{X_2 = 1\}) \implies a_{10} + a_{11} = a_{01} + a_{11} \implies a_{10} = a_{01}\,.$$

Raccogliamo le condizioni trovate:

$$\begin{cases} a_{00} = 0 \\ a_{21} = 0 \\ a_{11} + a_{20} = \dfrac{1}{2} \\ a_{10} + a_{01} = \dfrac{1}{2} \\ a_{20} = a_{10} \\ a_{10} = a_{01} \end{cases} \implies \begin{cases} a_{00} = 0 \\ a_{21} = 0 \\ a_{11} = \dfrac{1}{4} \\ a_{10} = \dfrac{1}{4} \\ a_{20} = \dfrac{1}{4} \\ a_{01} = \dfrac{1}{4} \end{cases}$$

La tabella completa è allora

X_2 \ X_1	0	1	2
0	0	$\frac{1}{4}$	$\frac{1}{4}$
1	$\frac{1}{4}$	$\frac{1}{4}$	0

2. Osserviamo che la seconda ipotesi fornisce l'informazione sulla probabilità condizionata

$$\mathbb{P}(\{X_1 = 2\}|\{X_2 = 0\}) = \frac{1}{2}\,.$$

D'altro canto

$$\mathbb{P}(\{X_1 = 2\}) = \frac{1}{4}\,.$$

Quindi l'evento $\{X_2 = 0\}$ condiziona l'evento $\{X_1 = 2\}$ e le variabili aleatorie non sono indipendenti.

2. Per calcolare la varianza della somma ricordiamo la formula

$$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\mathrm{Cov}(X_1, X_2)\,.$$

Calcoliamo questi tre termini:

$$\mathrm{Var}(X_1) = \mathbb{E}(X_1^2) - \mathbb{E}(X_1)^2 = \frac{1}{2} + 4 \cdot \frac{1}{4} - \left(\frac{1}{2} + 2 \cdot \frac{1}{2}\right)^2 = \frac{1}{2}$$

$$\mathrm{Var}(X_2) = \mathbb{E}(X_2^2) - \mathbb{E}(X_2)^2 = \left(\frac{1}{4} + \frac{1}{4}\right) - \left(\frac{1}{4} + \frac{1}{4}\right)^2 = \frac{1}{4}$$

$$\mathrm{Cov}(X_1, X_2) = \mathbb{E}(X_1 X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2) = 1 \cdot \frac{1}{4} - 1 \cdot \frac{1}{2} = -\frac{1}{4}\,.$$

Concludiamo che

$$\mathrm{Var}(X_1 + X_2) = \frac{1}{2} + \frac{1}{4} - 2\frac{1}{4} = \frac{1}{4}\,.$$

Esercizio 3. (8 punti) Un'azienda produce un dispositivo composto da due componenti, A e B. Si assuma che:

- La durata di funzionamento dei componenti A e la durata di funzionamento dei componenti B (in anni) siano variabili aleatorie indipendenti distribuite con legge esponenziale.

- La durata di funzionamento media dei componenti A sia 3 anni.

- La deviazione standard della durata di funzionamento dei componenti B sia 2 anni.

Si risponda ai seguenti quesiti:

1. Qual è la probabilità che un componente A abbia una durata di funzionamento compresa tra 2 e 5 anni?
2. Consideriamo un componente A. Abbiamo osservato che funziona ancora a 2 anni dall'acquisto. Qual è la probabilità che smetta di funzionare dopo 5 anni dall'acquisto? Enunciare e dimostrare la proprietà utilizzata.
3. Qual è la probabilità che un componente di tipo B smetta di funzionare dopo 4 anni?
4. Basta che uno dei due componenti A oppure B del dispositivo si rompa, perché il dispositivo smetta di funzionare. Qual è la durata di funzionamento media dell'intero dispositivo?

Soluzione. Consideriamo le due variabili aleatorie indipendenti

$$X_A = \text{“durata di funzionamento del componente } A\text{”} \sim \mathrm{Exp}(\lambda)\,,$$

$$X_B = \text{“durata di funzionamento del componente } B\text{”} \sim \mathrm{Exp}(\mu)\,.$$

Sappiamo che il valore atteso di X_A è 3, quindi

$$\mathbb{E}(X_A) = \frac{1}{\lambda} = 3 \implies \lambda = \frac{1}{3}\,.$$

Inoltre la deviazione standard di X_B è 2, quindi

$$\mathrm{Var}(X_B) = \frac{1}{\mu^2} \implies \sqrt{\mathrm{Var}(X_B)} = \frac{1}{\mu} = 2 \implies \mu = \frac{1}{2}\,.$$

1. Ci viene chiesto di calcolare

$$\mathbb{P}(\{2 \le X_A \le 5\}) = \int_2^5 \lambda e^{-\lambda x}\ \mathrm{d}x = \Big[-e^{-\lambda x}\Big]_2^5 = e^{-2\lambda} - e^{-5\lambda} = e^{-2/3} - e^{-5/3} \simeq 32.45\%\,.$$

2. Utilizziamo la proprietà di assenza di memoria dell'esponenziale per calcolare

$$\mathbb{P}(\{X_A > 5\}|\{X_A > 2\}) = \mathbb{P}(\{X_A > 3\}) = e^{-3\lambda} = e^{-3/3} = 36.79\%\,.$$

3. Calcoliamo

$$\mathbb{P}(\{X_B > 4\}) = e^{-4/2} = 13.53\%\,.$$

4. Il dispositivo smette di funzionare quando uno dei due componenti si rompe. Quindi la durata di funzionamento del dispositivo è data dal minimo tra le durate $\min\{X_A, X_B\}$. Calcoliamone il valore atteso, utilizzando l'indipendenza per stabilire che la densità congiunta

è data da $f_{(X_A,X_B)}(x_1,x_2) = \lambda\mu e^{-\lambda x_1}e^{-\mu x_2}$:

$$
\begin{aligned}
&\mathbb{E}(\min\{X_A,X_B\}) = \int_0^{+\infty}\Big(\int_0^{+\infty}\min\{x_1,x_2\}f_{(X_A,X_B)}(x_1,x_2)\ \mathrm{d}x_1\Big)\,\mathrm{d}x_2\\
&= \int_0^{+\infty}\Big(\int_0^{x_2}x_1\lambda\mu e^{-\lambda x_1}e^{-\mu x_2}\ \mathrm{d}x_1 + \int_{x_2}^{+\infty}x_2\lambda\mu e^{-\lambda x_1}e^{-\mu x_2}\ \mathrm{d}x_1\Big)\,\mathrm{d}x_2\\
&= \int_0^{+\infty}\Big(\mu e^{-\mu x_2}\int_0^{x_2}x_1\lambda e^{-\lambda x_1}\ \mathrm{d}x_1 + x_2\mu e^{-\mu x_2}\int_{x_2}^{+\infty}\lambda e^{-\lambda x_1}\ \mathrm{d}x_1\Big)\,\mathrm{d}x_2\\
&= \int_0^{+\infty}\Big(\mu e^{-\mu x_2}\Big[-x_1e^{-\lambda x_1}\Big]_0^{x_2} + \mu e^{-\mu x_2}\int_0^{x_2}e^{-\lambda x_1}\ \mathrm{d}x_1 + x_2\mu e^{-\mu x_2}\Big[-e^{-\lambda x_1}\Big]_{x_2}^{+\infty}\Big)\,\mathrm{d}x_2\\
&= \int_0^{+\infty}\Big(-\mu x_2e^{-(\lambda+\mu)x_2} + \mu e^{-\mu x_2}\Big[-\frac{1}{\lambda}e^{-\lambda x_1}\Big]_0^{x_2} + x_2\mu e^{-\mu x_2}e^{-\lambda x_2}\Big)\,\mathrm{d}x_2\\
&= \int_0^{+\infty}\Big(-\mu x_2e^{-(\lambda+\mu)x_2} + \mu e^{-\mu x_2}\Big[-\frac{1}{\lambda}e^{-\lambda x_1}\Big]_0^{x_2} + x_2\mu e^{-\mu x_2}e^{-\lambda x_2}\Big)\,\mathrm{d}x_2\\
&= \int_0^{+\infty}\frac{\mu}{\lambda}\Big(e^{-\mu x_2} - e^{-(\mu+\lambda)x_2}\Big)\,\mathrm{d}x_2 = \frac{1}{\lambda} - \frac{\mu}{\lambda}\frac{1}{\mu+\lambda} = \frac{1}{\mu+\lambda} = \frac{6}{5}\,.
\end{aligned}
$$

In alternativa si può osservare che, per l'indipendenza,

$$
\begin{aligned}
\mathbb{P}(\{\min\{X_A,X_B\} > t\}) &= \mathbb{P}(\{X_A > t\}\cap\{X_B > t\}) = \mathbb{P}(\{X_A > t\})\mathbb{P}(\{X_B > t\}) = e^{-\lambda t}e^{-\mu t}\\
&= e^{-(\lambda+\mu)t}\,.
\end{aligned}
$$

Questo è sufficiente a stabilire che $\min\{X_A,X_B\} \sim \mathrm{Exp}(\lambda+\mu)$ (perché?). Concludiamo immediatamente che $\mathbb{E}(\min\{X_A,X_B\}) = \frac{1}{\lambda+\mu}$.

Esercizio 4. (7 punti) Un ingegnere sta studiando la durata degli pneumatici per una nuova miscela di gomma. Ha testato un campione di 40 pneumatici fino alla fine del loro ciclo di vita in una prova su strada. La media calcolata sui dati è 60973.82 km. Si assuma che la deviazione standard della popolazione sia 3500 km.

1. Si calcoli sui dati un intervallo di confidenza bilaterale al 90% per la media reale della vita degli pneumatici. (N.B.: ricavare le formule)
2. Si calcoli sui dati un intervallo di confidenza unilaterale destro (limite inferiore di confidenza) al 95% per la media reale della vita degli pneumatici.
3. Supponiamo di svolgere l'esperimento di sopra (misurare un campione di 40 pneumatici) tante volte consecutivamente. In media, quale è la prima volta che la media reale della vita degli pneumatici sia fuori dall'intervallo di confidenza bilaterale al 90%?

Soluzione. 1. Si tratta di una popolazione $X\colon\Omega\to\mathbb{R}$ (non necessariamente normale) con $\mathbb{E}(X) = \mu$ e $\mathrm{Var}(X) = \sigma^2$ e un campione casuale $X_1,\dots,X_n$ estratto dalla popolazione con $n = 40 \geq 30$. La varianza $\sigma^2 = 3500^2$ è nota. Vogliamo calcolare un IC bilaterale con livello di confidenza $\beta = 90\%$.

Dalla definizione di IC si ha che

$$\beta = \mathbb{P}(\{U_n \leq \mu \leq V_n\})\,.$$

Per stimare μ sfrutteremo lo stimatore media campionaria $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Poiché il campione $X_1, \dots, X_n$ è numeroso ($n \geq 30$) si può applicare il Teorema del Limite Centrale per concludere che $\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$ è approssimata da $Z \sim \mathcal{N}(0,1)$. Allora

$$\begin{aligned}\beta = \mathbb{P}(\{U_n \leq \mu \leq V_n\}) &= \mathbb{P}\Big(\Big\{\frac{\overline{X}_n - V_n}{\sigma/\sqrt{n}} \leq \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq \frac{\overline{X}_n - U_n}{\sigma/\sqrt{n}}\Big\}\Big)\\ &\simeq \mathbb{P}\Big(\Big\{\frac{\overline{X}_n - V_n}{\sigma/\sqrt{n}} \leq Z \leq \frac{\overline{X}_n - U_n}{\sigma/\sqrt{n}}\Big\}\Big)\\ &= 1 - \mathbb{P}\Big(\Big\{Z < \frac{\overline{X}_n - V_n}{\sigma/\sqrt{n}}\Big\}\Big) - \mathbb{P}\Big(\Big\{Z > \frac{\overline{X}_n - U_n}{\sigma/\sqrt{n}}\Big\}\Big).\end{aligned}$$

Segue che

$$\mathbb{P}\Big(\Big\{Z < \frac{\overline{X}_n - V_n}{\sigma/\sqrt{n}}\Big\}\Big) + \mathbb{P}\Big(\Big\{Z > \frac{\overline{X}_n - U_n}{\sigma/\sqrt{n}}\Big\}\Big) = 1 - \beta = \alpha\,.$$

Decidiamo di equipartire α:

$$\mathbb{P}\Big(\Big\{Z < \frac{\overline{X}_n - V_n}{\sigma/\sqrt{n}}\Big\}\Big) = \mathbb{P}\Big(\Big\{Z < \frac{\overline{X}_n - V_n}{\sigma/\sqrt{n}}\Big\}\Big) = \frac{\alpha}{2}\,.$$

Definiamo $z_{\alpha/2}$ come il punto tale che

$$\mathbb{P}(\{Z \geq z_{\alpha/2}\}) = \frac{\alpha}{2}\,.$$

Scegliendo

$$\begin{aligned}\frac{\overline{X}_n - U_n}{\sigma/\sqrt{n}} &= z_{\alpha/2} \implies U_n = \overline{X}_n - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\,,\\ \frac{\overline{X}_n - V_n}{\sigma/\sqrt{n}} &= -z_{\alpha/2} \implies V_n = \overline{X}_n + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\,,\end{aligned}$$

si ottiene la condizione che definisce l'IC. In conclusione

$$\Big[\overline{X}_n - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}, \overline{X}_n + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\Big]$$

è un intervallo di confidenza bilaterale per μ con livello di confidenza $\beta = 1 - \alpha$. Calcoliamolo sui dati.

$$\overline{x}_n = 60973.82\,.$$

Dalla tavola della legge normale otteniamo che

$$z_{\alpha/2} = z_{0.05} = 1.645\,.$$

Segue che

$$\begin{aligned}\Big[\overline{x}_n - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}, \overline{x}_n + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\Big] &= \Big[60973.82 - \frac{3500}{\sqrt{40}} 1.645, 60973.82 + \frac{3500}{\sqrt{40}} 1.645\Big]\\ &= \Big[60063.48, 61884.16\Big]\,.\end{aligned}$$

2. Un intervallo di confidenza unilaterale destro è dato da

$$\Big[\overline{X}_n - \frac{\sigma}{\sqrt{n}} z_{\alpha}, +\infty\Big)\,.$$

Calcoliamo l'intervallo sui dati,

$$\left[\overline{x}_n - \frac{\sigma}{\sqrt{n}} z_\alpha, +\infty\right) = \left[60063.48, +\infty\right).$$

3. Consideriamo la variabile aleatoria

$$Y = \text{"prima volta che la media della vita degli pneumatici è fuori dall'IC bilaterale al 90\%"}.$$

Abbiamo che $Y \sim \text{Geo}(10\%)$. Segue che

$$\mathbb{E}(Y) = \frac{1}{10\%} = 10.$$