

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. 2023–2024

Gianluca Orlando

Indice

1	Tracce	2
	Traccia 17 giugno 2023 - I	3
	Traccia 17 giugno 2024 - II	5
2	Soluzioni	7
	Soluzione 17 giugno 2024 - I	8
	Soluzione 17 giugno 2024 - II	17

1 Tracce

Di seguito le tracce dell'a.a. 2023-2043.

Esame di Probabilità e Statistica [3231]

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: giugno 2024 - turno 1

Data: 17/06/2024

È obbligatorio consegnare la traccia con cognome e nome. In caso contrario, l'esito sarà "RITIRATO". Questa è la traccia n. 120. Scrivere il numero di traccia sullo svolgimento del compito.

Tempo massimo: 2 ore.

Esercizio 1. (6 punti) Una studentessa di Probabilità e Statistica vuole determinare se esiste una relazione tra le ore di studio giornaliere nella preparazione dell'esame e il voto all'esame. Per farlo, ha intervistato alcuni studenti e ha ottenuto i seguenti dati:

ore di studio	0	1	2	3	4	5	6
voto	0	12	11	25	26	30	30

1. Rappresentare i dati in uno scatterplot.
2. Calcolare la retta di regressione lineare (derivando le formule) e disegnarla.
3. Calcolare il coefficiente di correlazione lineare e il coefficiente di determinazione R^2 .

Esercizio 2. (8 punti) Sia (X, Y) un vettore aleatorio discreto con legge congiunta descritta dalla seguente tabella:

	Y	0	1	2
X				
0		0	$\frac{1}{4}$	a
1		$\frac{1}{4}$	0	b
2		c	d	0

Si assuma che:

$$\bullet \mathbb{P}(\{X = 2\}) = \frac{1}{8}, \quad \bullet \mathbb{P}(\{Y = 1\}|\{X = 2\}) = 1, \quad \bullet \mathbb{E}(XY) = \frac{1}{2}.$$

Dopo aver determinato i valori di a, b, c, d , rispondere ai seguenti quesiti:

1. Calcolare la covarianza tra X e Y e stabilire se X e Y sono indipendenti.
2. Calcolare $\text{Var}(X + Y)$.
3. Si supponga di estrarre 14 realizzazioni indipendenti della variabile aleatoria Y . Calcolare la probabilità che l'evento $\{Y = 0\}$ si realizzi almeno 3 volte (3 incluso).

4. Si estraggono tante realizzazioni indipendenti della variabile aleatoria Y . In media, qual è la prima volta in cui si realizza l'evento $\{Y = 0\}$?

Esercizio 3. (7 punti) Una catena di fast food analizza il tempo di servizio dei suoi clienti. I clienti possono ordinare due tipi di menu: il menu A e il menu B. Si assuma che

- Il tempo di servizio per il menu A sia distribuito con legge esponenziale con media 2 minuti.
- Il tempo di servizio per il menu B sia distribuito con legge esponenziale con deviazione standard 3 minuti.
- Il 20% dei clienti ordina il menu A e il 80% il menu B.

Si risponda alle seguenti domande:

1. Si consideri un cliente che ha ordinato il menu A. Qual è la probabilità che il suo tempo di servizio sia inferiore a 3 minuti?
2. Si consideri un cliente che ha ordinato il menu B. Qual è la probabilità che il servizio avvenga esattamente al minuto 1?
3. Si consideri un cliente che ha ordinato il menu B. Ha aspettato 2 minuti e non è ancora stato servito. Sapendo questo fatto, qual è la probabilità che debba aspettare in tutto almeno 5 minuti?
4. Un cliente ha fatto un ordine e ha aspettato un tempo compreso tra 1 e 4 minuti. Qual è la probabilità che abbia ordinato il menu A?

Esercizio 4. (7 punti) Una fabbrica di cereali vuole stimare la media del peso delle confezioni prodotte. Un controllo su un campione di alcune confezioni ha fornito i seguenti pesi in grammi:

500 506 499 507 502 500 496 501

Si supponga che il peso sia distribuito normalmente.

1. Calcolare sui dati un intervallo di confidenza bilaterale al 90% per la media del peso delle confezioni. (N.B.: derivare le formule)
2. Un intervallo di confidenza bilaterale al 91% calcolato sugli stessi dati è più grande o più piccolo di quello calcolato al punto 1? Perché?

Quesito teorico 1. (4 punti) Spiegare in che senso la legge di Poisson approssima la legge binomiale enunciando e dimostrando un teorema.

Quesito teorico 2. (2 punti) Calcolare media e varianza di $X \sim U(a, b)$.

Esame di Probabilità e Statistica [3231]

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: giugno 2024 - turno 2

Data: 17/06/2024

È obbligatorio consegnare la traccia con cognome e nome. In caso contrario, l'esito sarà "RITIRATO". Questa è la traccia n. 120. Scrivere il numero di traccia sullo svolgimento del compito.

Tempo massimo: 2 ore.

Esercizio 1. (6 punti) Si studia il tempo di attesa per il servizio clienti di una banca. I dati vengono raggruppati in classi nella seguente tabella:

intervalli (minuti)	frequenze assolute
[0, 2)	19
[2, 5)	18
[5, 7)	5
[7, 11)	16
[11, 18)	12
[18, 100)	10

1. Rappresentare un istogramma delle densità di frequenze relative.
2. Determinare la classe modale (o le classi modali, se più di una).
3. Calcolare un'approssimazione della media e della varianza dei dati.
4. Calcolare un'approssimazione del 55-esimo percentile.

Esercizio 2. (7 punti) Il numero di errori nelle soluzioni degli esercizi scritte dal docente di Probabilità e Statistica segue una distribuzione di Poisson con una media di 2 errori per soluzione. Si assuma che i numeri di errori in soluzioni di esercizi distinti siano indipendenti.

1. Qual è la probabilità che in una soluzione ci siano almeno 4 errori (inclusi)?
2. Qual è la deviazione standard del numero di errori in una soluzione?
3. Qual è la probabilità che in 5 soluzioni ci siano almeno 4 errori (inclusi)?
4. Consideriamo 5 soluzioni. Abbiamo letto le prime 3 soluzioni e abbiamo individuato almeno 4 errori (inclusi). Sapendo questo fatto, qual è la probabilità che nelle 5 soluzioni ci siano in tutto 6 errori?

5. Uno studente legge le soluzioni degli esercizi in sequenza e si blocca quando trova la prima soluzione con almeno 1 errore (incluso). Qual è la probabilità che lo studente si blocchi entro la lettura della terza soluzione (inclusa)?

Esercizio 3. (8 punti) In un centro di assistenza, il tempo necessario per completare un backup di un computer segue una distribuzione esponenziale con un tempo medio di 2 ore.

1. Qual è la probabilità che il backup di un computer duri meno di 1 ora?
2. Qual è la varianza del tempo necessario per completare il backup di un computer?
3. Al centro di assistenza arrivano 16 computer per i quali occorre un backup. Qual è la probabilità che per almeno 3 computer (inclusi) il backup duri più di 3 ore? Si assuma che i tempi di backup dei computer siano indipendenti.
4. Al centro di assistenza arrivano 2 computer per i quali occorre un backup. Il backup del secondo computer inizia non appena il backup del primo computer è completato. Qual è la media del tempo necessario per completare il backup totale dei 2 computer? E la varianza? Si assuma che i tempi di backup dei computer siano indipendenti.
5. Nella situazione del punto 4., qual è la probabilità che il backup totale dei 2 computer sia inferiore a 7 ore?

Esercizio 4. (7 punti) Un'azienda sostiene che la durata media giornaliera degli smartphone che produce è di 12 ore. Un'indagine condotta su alcuni smartphone è volta a mostrare che la durata è in realtà inferiore. Vengono rilevate le seguenti durate (in ore):

11.2	15.2	12.1	10.8	8.6	11.7	13.6	12.7	9.5	11.0
18.2	11.9	14.2	12.5	10.0	11.3	11.4	9.3	10.3	9.5
10.5	13.6	10.8	9.6	13.3	8.9	13.8	12.7	8.9	15.5
11.7	13.2	9.6	10.4	12.2	11.9	9.1	12.2	12.6	9.6

La media calcolata sui dati risulta essere 11.63 ore. È noto che la deviazione standard della popolazione è di 4 ore.

1. È possibile sostenere con significatività 1% che la durata media degli smartphone è in realtà inferiore a 12 ore? (N.B.: derivare le formule)
2. Calcolare il p -value del test.

Quesito teorico 1. (4 punti) Sia $Z \sim \mathcal{N}(0, 1)$. Che legge ha Z^2 ? Motivare la risposta. Siano $Z_1, \dots, Z_n \sim \mathcal{N}(0, 1)$ indipendenti. Che legge ha $\sum_{i=1}^n Z_i^2$? Motivare la risposta.

Quesito teorico 2. (2 punti) Calcolare media e varianza di $X \sim B(n, p)$.

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: _____

Nome: _____

Matricola: _____

Docente: Gianluca Orlando

Appello: giugno 2024 - turno 1

Data: 16/04/2024

Viene usata come riferimento la traccia n. 120.

Esercizio 1. Una studentessa di Probabilità e Statistica vuole determinare se esiste una relazione tra le ore di studio giornaliere nella preparazione dell'esame e il voto all'esame. Per farlo, ha intervistato alcuni studenti e ha ottenuto i seguenti dati:

ore di studio	0	1	2	3	4	5	6
voto	0	12	11	25	26	30	30

1. Rappresentare i dati in uno scatterplot.
2. Calcolare la retta di regressione lineare (derivando le formule) e disegnarla.
3. Calcolare il coefficiente di correlazione lineare e il coefficiente di determinazione R^2 .

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

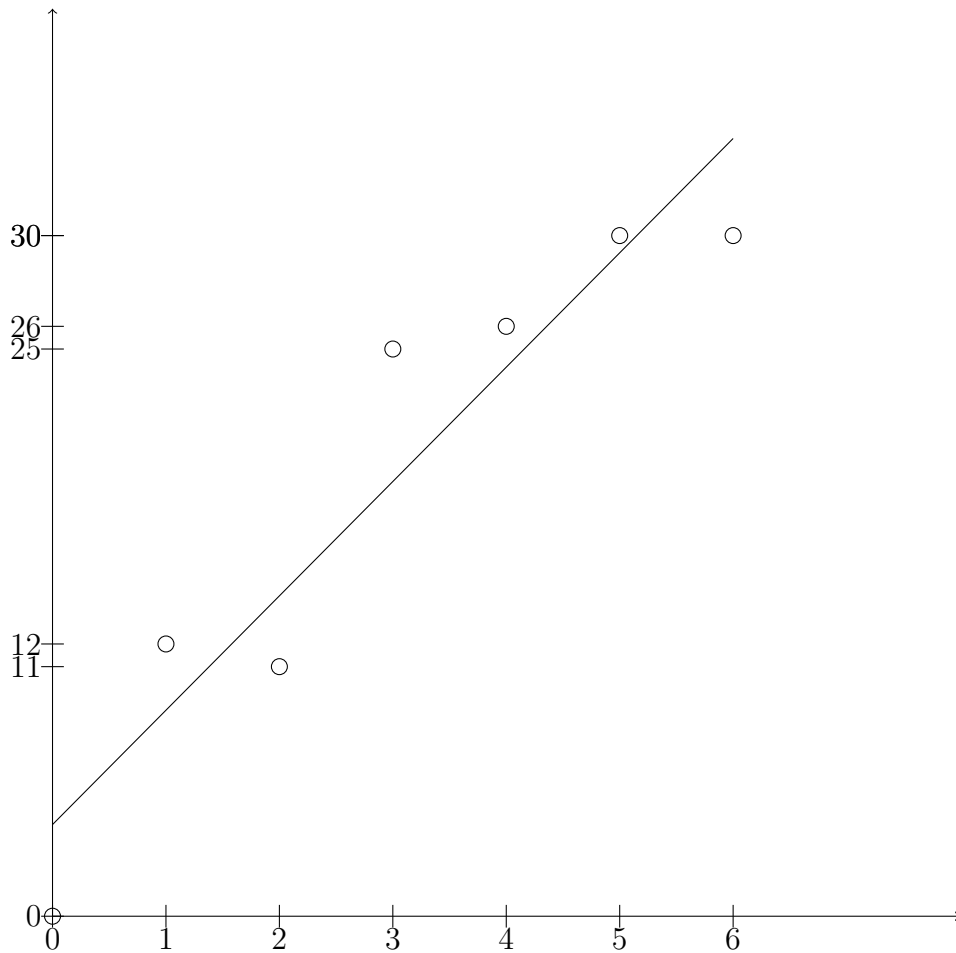


Figura 1: Scatterplot e retta di regressione lineare.

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

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

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

$$0 = \partial_a e(a, b) = -2 \sum_{i=1}^n (y_i - ax_i - b)x_i = -2 \sum_{i=1}^n (x_i y_i - ax_i^2 - bx_i),$$

$$0 = \partial_b e(a, b) = -2 \sum_{i=1}^n (y_i - ax_i - b)$$

Dalla seconda equazione segue che

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

Sostituendo nella prima,

$$\begin{aligned}\sum_{i=1}^n (x_i y_i - a x_i^2 - b x_i) = 0 &\implies \sum_{i=1}^n (x_i y_i - a x_i^2 - x_i \bar{y} + a \bar{x} x_i) = 0 \\ &\implies a \left(\sum_{i=1}^n x_i^2 - n \bar{x}^2 \right) = \sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} \\ &\implies a = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}.\end{aligned}$$

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

								somma
x_i	0	1	2	3	4	5	6	21
y_i	0	12	11	25	26	30	30	134
x_i^2	0	1	4	9	16	25	36	91
y_i^2	0	144	121	625	676	900	900	3366
$x_i y_i$	0	12	22	75	104	150	180	543

Pertanto $\bar{x} = 21/7 = 3$ e $\bar{y} = 134/7 = 19.14$. Segue che

$$a = \frac{543 - 7 \cdot 3 \cdot 19.14}{91 - 7 \cdot 3^2} = \frac{141}{28} \simeq 5.04,$$

$$b = 19.14 - 5.04 \cdot 3 = 4.02,$$

ovvero, la retta di regressione lineare ha equazione

$$y = 5.04x + 4.02.$$

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

$$\begin{aligned}\rho_{x,y} &= \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sqrt{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \sqrt{\sum_{i=1}^n y_i^2 - n \bar{y}^2}} = \frac{543 - 7 \cdot 3 \cdot 19.14}{\sqrt{91 - 7 \cdot 3^2} \sqrt{3366 - 7 \cdot 19.14^2}} \\ &= 0.9411.\end{aligned}$$

Il coefficiente di determinazione è $R^2 = \rho_{x,y}^2 = 0.8857$.

Esercizio 2. (8 punti) Sia (X, Y) un vettore aleatorio discreto con legge congiunta descritta dalla seguente tabella:

	Y	0	1	2
X				
0		0	$\frac{1}{4}$	a
1		$\frac{1}{4}$	0	b
2		c	d	0

Si assuma che:

$$\bullet \mathbb{P}(\{X = 2\}) = \frac{1}{8}, \quad \bullet \mathbb{P}(\{Y = 1\} | \{X = 2\}) = 1, \quad \bullet \mathbb{E}(XY) = \frac{1}{2}.$$

Dopo aver determinato i valori di a, b, c, d , rispondere ai seguenti quesiti:

1. Calcolare la covarianza tra X e Y e stabilire se X e Y sono indipendenti.

2. Calcolare $\text{Var}(X + Y)$.
3. Si supponga di estrarre 14 realizzazioni indipendenti della variabile aleatoria Y . Calcolare la probabilità che l'evento $\{Y = 0\}$ si realizzi almeno 3 volte (3 incluso).
4. Si estraggono tante realizzazioni indipendenti della variabile aleatoria Y . In media, qual è la prima volta in cui si realizza l'evento $\{Y = 0\}$?

Soluzione. Per determinare i valori di a, b, c, d usiamo le informazioni fornite. Poiché le probabilità devono sommare a 1, otteniamo che

$$\frac{1}{4} + a + \frac{1}{4} + b + c + d = 1 \implies a + b + c + d = \frac{1}{2}.$$

Dalla condizione $\mathbb{P}(\{X = 2\}) = 1/8$ otteniamo che

$$c + d = \frac{1}{8}.$$

Dalla condizione $\mathbb{P}(\{Y = 1\}|\{X = 2\}) = 1$ otteniamo che

$$\frac{\mathbb{P}(\{Y = 1\} \cap \{X = 2\})}{\mathbb{P}(\{X = 2\})} = 1 \implies \frac{d}{1/8} = 1 \implies d = \frac{1}{8}.$$

Dalla condizione $\mathbb{E}(XY) = 1/2$ otteniamo che

$$\begin{aligned} \frac{1}{2} &= \mathbb{E}(XY) = \sum_{x,y} xy \mathbb{P}(\{X = x\} \cap \{Y = y\}) \\ &= 0 \cdot 0 \cdot 0 + 1 \cdot 0 \cdot \frac{1}{4} + 2 \cdot 0 \cdot c + 0 \cdot 1 \cdot \frac{1}{4} + 1 \cdot 1 \cdot 0 + 2 \cdot 1 \cdot d + 0 \cdot 2 \cdot a + 1 \cdot 2 \cdot b + 2 \cdot 2 \cdot 0 \\ &= 2d + 2b \implies d + b = \frac{1}{4}. \end{aligned}$$

Risolviamo il sistema

$$\begin{cases} a + b + c + d = \frac{1}{2} \\ c + d = \frac{1}{8} \\ d = \frac{1}{8} \\ d + b = \frac{1}{4} \end{cases}$$

ottenendo $a = 1/4, b = 1/8, c = 0, d = 1/8$.

La tabella completa è la seguente:

	Y	0	1	2
X				
0		0	$\frac{1}{4}$	$\frac{1}{4}$
1		$\frac{1}{4}$	0	$\frac{1}{8}$
2		0	$\frac{1}{8}$	0

1. Per calcolare la covarianza utilizziamo la formula

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

Calcoliamo

$$\mathbb{E}(X) = \sum_x x \mathbb{P}(\{X = x\}) = 0 \cdot \left(0 + \frac{1}{4} + \frac{1}{4}\right) + 1 \cdot \left(\frac{1}{4} + 0 + \frac{1}{8}\right) + 2 \cdot \left(0 + \frac{1}{8} + 0\right) = \frac{5}{8},$$

e

$$\mathbb{E}(Y) = \sum_y y \mathbb{P}(\{Y = y\}) = 0 \cdot \left(0 + \frac{1}{4} + 0\right) + 1 \cdot \left(\frac{1}{4} + 0 + \frac{1}{8}\right) + 2 \cdot \left(\frac{1}{4} + \frac{1}{8} + 0\right) = \frac{9}{8}.$$

Segue che

$$\text{Cov}(X, Y) = \frac{1}{2} - \frac{5}{8} \cdot \frac{9}{8} = -\frac{13}{64}.$$

Poiché $\text{Cov}(X, Y) \neq 0$, X e Y non sono indipendenti.

2. Calcoliamo $\text{Var}(X + Y)$ utilizzando la formula

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y).$$

Utilizzando le formule $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$ e $\text{Var}(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2$, occorre calcolare solo

$$\mathbb{E}(X^2) = \sum_x x^2 \mathbb{P}(\{X = x\}) = 0^2 \cdot \left(0 + \frac{1}{4} + \frac{1}{4}\right) + 1^2 \cdot \left(\frac{1}{4} + 0 + \frac{1}{8}\right) + 2^2 \cdot \left(0 + \frac{1}{8} + 0\right) = \frac{7}{8},$$

e

$$\mathbb{E}(Y^2) = \sum_y y^2 \mathbb{P}(\{Y = y\}) = 0^2 \cdot \left(0 + \frac{1}{4} + 0\right) + 1^2 \cdot \left(\frac{1}{4} + 0 + \frac{1}{8}\right) + 2^2 \cdot \left(\frac{1}{4} + \frac{1}{8} + 0\right) = \frac{15}{8}.$$

Quindi

$$\begin{aligned}\text{Var}(X) &= \frac{7}{8} - \left(\frac{5}{8}\right)^2 = \frac{31}{64}, \\ \text{Var}(Y) &= \frac{15}{8} - \left(\frac{9}{8}\right)^2 = \frac{39}{64}.\end{aligned}$$

Concludiamo che

$$\text{Var}(X + Y) = \frac{31}{64} + \frac{39}{64} + 2 \cdot \left(-\frac{13}{64}\right) = \frac{44}{64}.$$

3. Per calcolare la probabilità che l'evento $\{Y = 0\}$ si realizzi almeno 4 volte (4 incluso) in 14 estrazioni, possiamo utilizzare la distribuzione binomiale. La probabilità di successo è

$$\mathbb{P}(\{Y = 0\}) = \frac{1}{4}.$$

Consideriamo una variabile aleatoria

$$Z = \text{“numero di successi in 14 prove”} \sim B(14, 1/4).$$

Dobbiamo calcolare

$$\begin{aligned}\mathbb{P}(\{Z \geq 4\}) &= 1 - \mathbb{P}(\{Z < 4\}) = 1 - \mathbb{P}(\{Z = 0\}) - \mathbb{P}(\{Z = 1\}) - \mathbb{P}(\{Z = 2\}) - \mathbb{P}(\{Z = 3\}) \\ &= 1 - \left(\frac{3}{4}\right)^{14} - 14 \cdot \frac{1}{4} \cdot \left(\frac{3}{4}\right)^{13} - \binom{14}{2} \cdot \left(\frac{1}{4}\right)^2 \cdot \left(\frac{3}{4}\right)^{12} = 71.89\%\end{aligned}$$

4. Consideriamo la variabile aleatoria

$W = \text{“prima volta nella successione di estrazioni in cui si verifica l'evento } \{Y = 0\}” \sim \text{Geo}(1/4),$

poiché

$$\mathbb{P}(\{Y = 0\}) = \frac{1}{4}.$$

Vogliamo calcolare

$$\mathbb{E}(W) = \frac{1}{1/4} = 4.$$

Esercizio 3. (7 punti) Una catena di fast food analizza il tempo di servizio dei suoi clienti. I clienti possono ordinare due tipi di menu: il menu A e il menu B. Si assuma che

- Il tempo di servizio per il menu A sia distribuito con legge esponenziale con media 2 minuti.
- Il tempo di servizio per il menu B sia distribuito con legge esponenziale con deviazione standard 3 minuti.
- Il 20% dei clienti ordina il menu A e il 80% il menu B.

Si risponda alle seguenti domande:

1. Si consideri un cliente che ha ordinato il menu A. Qual è la probabilità che il suo tempo di servizio sia inferiore a 3 minuti?
2. Si consideri un cliente che ha ordinato il menu B. Qual è la probabilità che il servizio avvenga esattamente al minuto 1?
3. Si consideri un cliente che ha ordinato il menu B. Ha aspettato 2 minuti e non è ancora stato servito. Sapendo questo fatto, qual è la probabilità che debba aspettare in tutto almeno 5 minuti?
4. Un cliente ha fatto un ordine e ha aspettato un tempo compreso tra 1 e 4 minuti. Qual è la probabilità che abbia ordinato il menu A?

Soluzione. Consideriamo le variabili aleatorie

$$X_A = \text{“tempo di servizio per il menu A”} \sim \text{Exp}(\lambda)$$

$$X_B = \text{“tempo di servizio per il menu B”} \sim \text{Exp}(\mu)$$

$$X = \text{“tempo di servizio (senza specificare il menu)”}.$$

Abbiamo che

$$2 = \mathbb{E}(X) = \frac{1}{\lambda} \implies \lambda = \frac{1}{2},$$

$$3 = \sqrt{\text{Var}(X)} = \frac{1}{\mu} \implies \mu = \frac{1}{3}.$$

Quindi $X_A \sim \text{Exp}(1/2)$ e $X_B \sim \text{Exp}(1/3)$.

Consideriamo infine la variabile aleatoria

$$Y = \text{“1 se ordinato il menu A, 0 altrimenti”} \sim \text{Be}(0.2).$$

1. Calcoliamo

$$\mathbb{P}(\{X_A < 3\}) = 1 - \mathbb{P}(\{X_A \geq 3\}) = 1 - e^{-3/2} = 77.69\%.$$

2. Poiché X_B è una variabile continua con densità, la probabilità che il servizio avvenga esattamente al minuto 1 è nulla:

$$\mathbb{P}(\{X_B = 1\}) = 0.$$

3. La probabilità che il cliente debba aspettare in tutto almeno 5 minuti, sapendo che ha già aspettato 2 minuti, si può calcolare utilizzando la proprietà di assenza di memoria della distribuzione esponenziale:

$$\mathbb{P}(\{X_B \geq 5\}|\{X_B \geq 2\}) = \mathbb{P}(\{X_B \geq 3\}) = e^{-3/3} = 36.79\%.$$

4. La probabilità che un cliente abbia ordinato il menu A, sapendo che ha aspettato un tempo compreso tra 1 e 4 minuti, si può calcolare utilizzando la formula di Bayes:

$$\begin{aligned}\mathbb{P}(\{Y = 1\}|\{1 \leq X \leq 4\}) &= \frac{\mathbb{P}(\{1 \leq X \leq 4\} \cap \{Y = 1\})}{\mathbb{P}(\{1 \leq X \leq 4\})} \\ &= \frac{\mathbb{P}(\{1 \leq X \leq 4\}|\{Y = 1\})\mathbb{P}(\{Y = 1\})}{\mathbb{P}(\{1 \leq X \leq 4\})} \\ &= \frac{\mathbb{P}(\{1 \leq X_A \leq 4\})\mathbb{P}(\{Y = 1\})}{\mathbb{P}(\{1 \leq X \leq 4\})}.\end{aligned}$$

Calcoliamo i tre termini:

$$\mathbb{P}(\{1 \leq X_A \leq 4\}) = \int_1^4 \frac{1}{2}e^{-x/2} dx = e^{-1/2} - e^{-2} = 47.12\%.$$

$$\mathbb{P}(\{Y = 1\}) = 20\%.$$

Utilizzando il teorema della probabilità totale:

$$\begin{aligned}\mathbb{P}(\{1 \leq X \leq 4\}) &= \mathbb{P}(\{1 \leq X \leq 4\} \cap \{Y = 1\}) + \mathbb{P}(\{1 \leq X \leq 4\} \cap \{Y = 0\}) \\ &= \mathbb{P}(\{1 \leq X_A \leq 4\})\mathbb{P}(\{Y = 1\}) + \mathbb{P}(\{1 \leq X_B \leq 4\})\mathbb{P}(\{Y = 0\}) \\ &= (e^{-1/2} - e^{-2})20\% + (e^{-1/3} - e^{-4/3})80\% = 45.66\%.\end{aligned}$$

Concludiamo che

$$\mathbb{P}(\{Y = 1\}|\{1 \leq X \leq 4\}) = \frac{47.12\% \cdot 20\%}{45.66\%} = 20.64\%.$$

Esercizio 4.(7 punti) Una fabbrica di cereali vuole stimare la media del peso delle confezioni prodotte. Un controllo su un campione di alcune confezioni ha fornito i seguenti pesi in grammi:

500 506 499 507 502 500 496 501

Si supponga che il peso sia distribuito normalmente.

1. Calcolare sui dati un intervallo di confidenza bilaterale al 90% per la media del peso delle confezioni. (N.B.: derivare le formule)
2. Un intervallo di confidenza bilaterale al 91% calcolato sugli stessi dati è più grande o più piccolo di quello calcolato al punto 1? Perché?

Soluzione. 1. La popolazione è descritta da una variabile aleatoria $X \sim \mathcal{N}(\mu, \sigma^2)$. I parametri μ e σ^2 non sono noti. Dalla popolazione viene estratto un campione X_1, \dots, X_n di $n = 8$ osservazioni. 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 $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Per stimare σ^2 sfrutteremo lo stimatore varianza campionaria $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$. Poiché $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ sono indipendenti, $T_{n-1} = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim t(n-1)$. Allora

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

Segue che

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

Decidiamo di equipartire α :

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

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

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

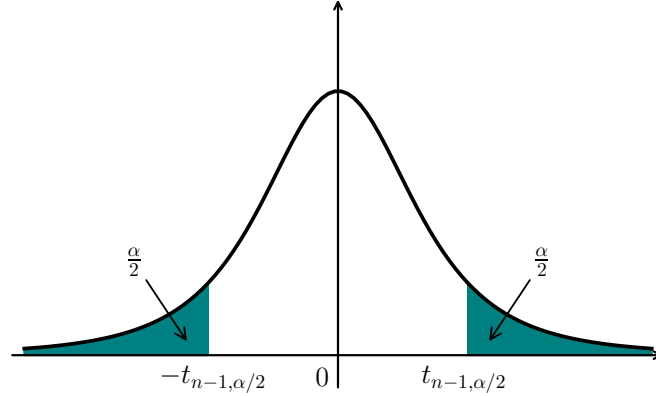


Figura 2: Definizione di $t_{n-1, \alpha/2}$.

Scegliendo

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

si ottiene la condizione che definisce l'intervallo di confidenza. In conclusione

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

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

Calcoliamo l'intervallo di confidenza sui dati. Per farlo, calcoliamo

$$\bar{x}_n = \frac{1}{8}(500 + 506 + 499 + 507 + 502 + 500 + 496 + 501) = 501.375,$$

$$s_n^2 = \frac{1}{7}(500^2 + 506^2 + 499^2 + 507^2 + 502^2 + 500^2 + 496^2 + 501^2 - 8 \cdot 501.375^2) = 13.125$$
$$\implies s_n = \sqrt{13.125} = 3.623.$$

Infine, per $\beta = 90\%$ abbiamo che $\alpha/2 = 0.05$. Dalla tabella della t di Student, $t_{7,0.05} = 1.895$. Pertanto l'intervallo di confidenza è

$$\left[501.375 - \frac{3.623}{\sqrt{8}} \cdot 1.895, 501.375 + \frac{3.623}{\sqrt{8}} \cdot 1.895 \right] = [498.95, 503.80].$$

Un intervallo di confidenza al 91% sarebbe più grande di quello al 90% poiché $\alpha/2$ sarebbe più piccolo e di conseguenza il quantile $t_{7,\alpha/2}$ sarebbe più grande.

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: giugno 2024 - turno 2

Data: 17/06/2024

Viene usata come riferimento la traccia n. 120.

Esercizio 1. (6 punti) Si studia il tempo di attesa per il servizio clienti di una banca. I dati vengono raggruppati in classi nella seguente tabella:

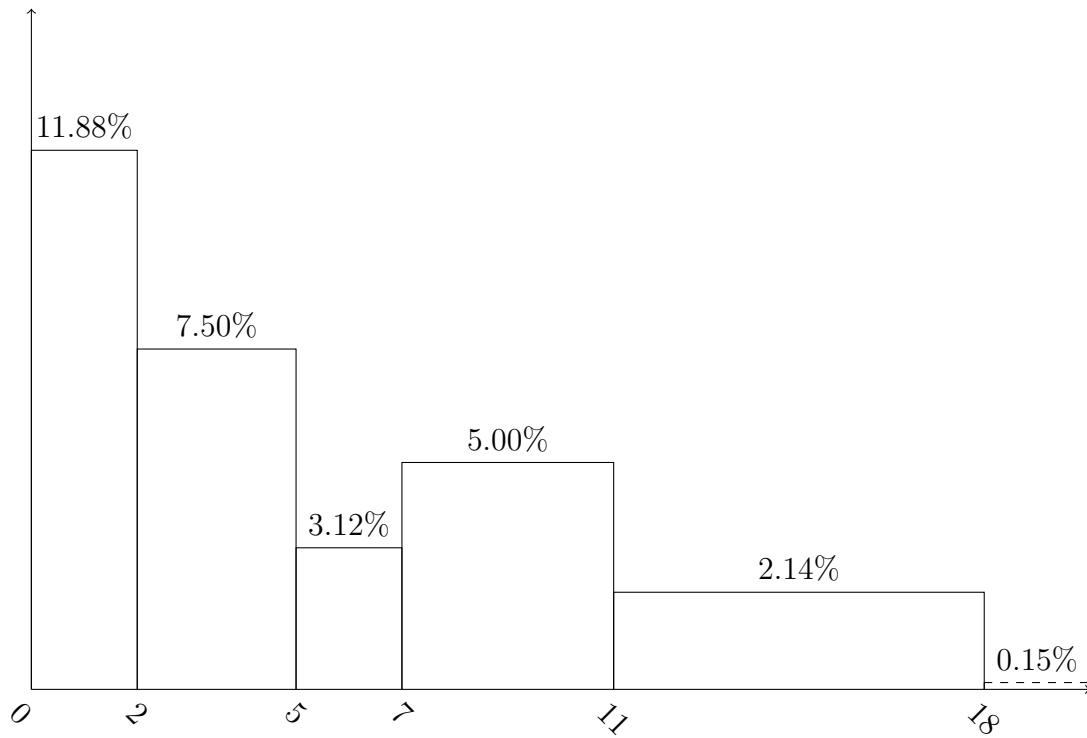
intervalli (minuti)	frequenze assolute
[0, 2)	19
[2, 5)	18
[5, 7)	5
[7, 11)	16
[11, 18)	12
[18, 100)	10

1. Rappresentare un istogramma delle densità di frequenze relative.
2. Determinare la classe modale (o le classi modali, se più di una).
3. Calcolare un'approssimazione della media e della varianza dei dati.
4. Calcolare un'approssimazione del 55-esimo percentile.

Soluzione. 1. Completiamo la tabella:

intervallo	freq. assolute	freq. relative	densità di freq. rel.	freq. cumulate
[0, 2)	19	23.75%	11.88%	19
[2, 5)	18	22.50%	7.50%	37
[5, 7)	5	6.25%	3.12%	42
[7, 11)	16	20.00%	5.00%	58
[11, 18)	12	15.00%	2.14%	70
[18, 100)	10	12.50%	0.15%	80

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 $[0, 2)$.

3. Per calcolare un'approssimazione della media utilizziamo le frequenze relative ottenute da $p_j = f_j/n$ dove $n = 80$ e i valori centrali \tilde{v}_j degli intervalli

$$\begin{aligned}\bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i \simeq \frac{1}{n} \sum_{j=1}^k f_j \tilde{v}_j = \sum_{j=1}^k p_j \tilde{v}_j \\ &= 23.75\% \cdot 1 + 22.50\% \cdot 3.5 + 6.25\% \cdot 6 + 20.00\% \cdot 9 + 15.00\% \cdot 14.5 + 12.50\% \cdot 59 = 12.75.\end{aligned}$$

Calcoliamo un'approssimazione della varianza

$$\begin{aligned}s^2 &= \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) \simeq \frac{1}{n-1} \left(\sum_{j=1}^k f_j \tilde{v}_j^2 - n\bar{x}^2 \right) = \frac{n}{n-1} \left(\sum_{j=1}^k p_j \tilde{v}_j^2 - \bar{x}^2 \right) \\ &= \frac{80}{79} \left(23.75\% \cdot 1^2 + 22.50\% \cdot 3.5^2 + 6.25\% \cdot 6^2 + 20.00\% \cdot 9^2 + 15.00\% \cdot 14.5^2 + 12.50\% \cdot 59^2 \right. \\ &\quad \left. - 12.75^2 \right) = 329.66.\end{aligned}$$

4. Per calcolare un'approssimazione del 55-esimo percentile dei dati, usiamo le frequenze cumulate. Troviamo l'intervallo I_j tale che $F_j \leq 55\%n = 44 < F_{j+1}$. Si tratta dell'intervallo $[7, 11)$. Approssimiamo la mediana con

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

dove

$$\lambda_j = \frac{55\%n - F_{j-1}}{F_j - F_{j-1}} = \frac{44 - 42}{58 - 42} = 0.125.$$

Quindi

$$P_{55} \simeq 7 + 0.125 \cdot (11 - 7) = 7.5.$$

Esercizio 2. (7 punti) Il numero di errori nelle soluzioni degli esercizi scritte dal docente di Probabilità e Statistica segue una distribuzione di Poisson con una media di 2 errori per soluzione. Si assuma che i numeri di errori in soluzioni di esercizi distinti siano indipendenti.

1. Qual è la probabilità che in una soluzione ci siano almeno 4 errori (inclusi)?
2. Qual è la deviazione standard del numero di errori in una soluzione?
3. Qual è la probabilità che in 5 soluzioni ci siano almeno 4 errori (inclusi)?
4. Consideriamo 5 soluzioni. Abbiamo letto le prime 3 soluzioni e abbiamo individuato almeno 4 errori (inclusi). Sapendo questo fatto, qual è la probabilità che nelle 5 soluzioni ci siano in tutto 6 errori?
5. Uno studente legge le soluzioni degli esercizi in sequenza e si blocca quando trova la prima soluzione con almeno 1 errore (incluso). Qual è la probabilità che lo studente si blocchi entro la lettura della terza soluzione (inclusa)?

Soluzione. Consideriamo la variabile aleatoria

$$X = \text{“numero di errori in una soluzione”} \sim P(\lambda).$$

Poiché

$$\lambda = \mathbb{E}(X) = 2 \implies X \sim P(2).$$

1. La probabilità che in una soluzione ci siano almeno 4 errori è

$$\begin{aligned}\mathbb{P}(\{X \geq 4\}) &= 1 - \mathbb{P}(X < 4) = 1 - \mathbb{P}(X \leq 3) \\ &= 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) - \mathbb{P}(X = 2) - \mathbb{P}(X = 3) \\ &= 1 - e^{-2} \left(\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!} \right) = 14.30\%.\end{aligned}$$

2. La deviazione standard del numero di errori in una soluzione è

$$\sigma = \sqrt{\text{Var}(X)} = \sqrt{\lambda} = \sqrt{2} = 1.41.$$

3. Per calcolare questa probabilità, consideriamo le variabili aleatorie X_1, X_2, X_3, X_4, X_5 indipendenti e identicamente distribuite con $X_i \sim P(2)$. La variabile X_i rappresenta il numero di errori nella i -esima soluzione. Poiché la somma di variabili aleatorie di Poisson indipendenti è ancora una variabile aleatoria di Poisson, la variabile aleatoria $Y = X_1 + X_2 + X_3 + X_4 + X_5$ ha distribuzione $P(5 \cdot 2) = P(10)$. La probabilità che in 5 soluzioni ci siano almeno 4 errori è quindi

$$\begin{aligned}\mathbb{P}(\{Y \geq 4\}) &= 1 - \mathbb{P}(Y < 4) = 1 - \mathbb{P}(Y \leq 3) \\ &= 1 - \mathbb{P}(Y = 0) - \mathbb{P}(Y = 1) - \mathbb{P}(Y = 2) - \mathbb{P}(Y = 3) \\ &= 1 - e^{-10} \left(\frac{10^0}{0!} + \frac{10^1}{1!} + \frac{10^2}{2!} + \frac{10^3}{3!} \right) = 98.97\%.\end{aligned}$$

4. Consideriamo le variabili aleatorie

$$Y = X_1 + X_2 + X_3 \sim P(6) \quad \text{e} \quad Z = X_4 + X_5 \sim P(4).$$

La probabilità che nelle 5 soluzioni ci siano in tutto 6 errori, sapendo che nelle prime 3 soluzioni ci sono almeno 4 errori, è, utilizzando l'indipendenza tra Y e Z ,

$$\begin{aligned}\mathbb{P}(\{Y + Z = 6\}|\{Y \geq 4\}) &= \frac{\mathbb{P}(\{Y + Z = 6\} \cap \{Y \geq 4\})}{\mathbb{P}(\{Y \geq 4\})} \\ &= \frac{\mathbb{P}(\{Y = 4\} \cap \{Z = 2\}) + \mathbb{P}(\{Y = 5\} \cap \{Z = 1\}) + \mathbb{P}(\{Y = 6\} \cap \{Z = 0\})}{1 - \mathbb{P}(\{Y < 4\})} \\ &= \frac{\mathbb{P}(\{Y = 4\})\mathbb{P}(\{Z = 2\}) + \mathbb{P}(\{Y = 5\})\mathbb{P}(\{Z = 1\}) + \mathbb{P}(\{Y = 6\})\mathbb{P}(\{Z = 0\})}{1 - \mathbb{P}(\{Y = 0\}) - \mathbb{P}(\{Y = 1\}) - \mathbb{P}(\{Y = 2\}) - \mathbb{P}(\{Y = 3\})} \\ &= \frac{e^{-6} \frac{6^4}{4!} e^{-4} \frac{4^2}{2!} + e^{-6} \frac{6^5}{5!} e^{-4} \frac{4^1}{1!} + e^{-6} \frac{6^6}{6!} e^{-4} \frac{4^0}{0!}}{1 - e^{-6} \left(\frac{6^0}{0!} + \frac{6^1}{1!} + \frac{6^2}{2!} + \frac{6^3}{3!} \right)} = 4.04\%.\end{aligned}$$

5. Consideriamo la variabile aleatoria

$$Y = \text{"prima soluzione con almeno 1 errore"} \sim \text{Geo}(p),$$

dove p è la probabilità che una soluzione abbia almeno 1 errore, ovvero

$$p = \mathbb{P}(\{X \geq 1\}) = 1 - \mathbb{P}(\{X = 0\}) = 1 - e^{-2} = 86.47\%.$$

La probabilità che lo studente si blocchi entro la lettura della terza soluzione è

$$\mathbb{P}(\{Y \leq 3\}) = 1 - \mathbb{P}(\{Y > 3\}) = 1 - (1 - p)^3 = 1 - (1 - 86.47\%)^3 = 99.75\%.$$

Esercizio 3. (8 punti) In un centro di assistenza, il tempo necessario per completare un backup di un computer segue una distribuzione esponenziale con un tempo medio di 2 ore.

1. Qual è la probabilità che il backup di un computer duri meno di 1 ora?
2. Qual è la varianza del tempo necessario per completare il backup di un computer?
3. Al centro di assistenza arrivano 16 computer per i quali occorre un backup. Qual è la probabilità che per almeno 3 computer (inclusi) il backup duri più di 3 ore? Si assuma che i tempi di backup dei computer siano indipendenti.
4. Al centro di assistenza arrivano 2 computer per i quali occorre un backup. Il backup del secondo computer inizia non appena il backup del primo computer è completato. Qual è la media del tempo necessario per completare il backup totale dei 2 computer? E la varianza? Si assuma che i tempi di backup dei computer siano indipendenti.
5. Nella situazione del punto 4., qual è la probabilità che il backup totale dei 2 computer sia inferiore a 7 ore?

Soluzione. Consideriamo la variabile aleatoria

$$X = \text{"tempo necessario per completare un backup"} \sim \text{Exp}(\lambda).$$

Poiché

$$\frac{1}{\lambda} = \mathbb{E}(X) = 2 \implies \lambda = \frac{1}{2} \implies X \sim \text{Exp}\left(\frac{1}{2}\right).$$

1. La probabilità che il backup di un computer duri meno di 1 ora è

$$\mathbb{P}(\{X < 1\}) = 1 - \mathbb{P}(X \geq 1) = 1 - e^{-\frac{1}{2}} = 39.35\%.$$

2. La varianza del tempo necessario per completare il backup di un computer è

$$\text{Var}(X) = \frac{1}{\lambda^2} = 2^2 = 4.$$

3. Per calcolare questa probabilità, consideriamo la variabile aleatoria

$$Y = \text{“numero dei 16 computer per i quali il backup dura più di 3 ore”} \sim B(16, p),$$

dove

$$p = \mathbb{P}(\{X > 3\}) = e^{-\frac{3}{2}} = 22.31\%.$$

La probabilità che per almeno 3 computer il backup duri più di 3 ore è

$$\begin{aligned} \mathbb{P}(\{Y \geq 3\}) &= 1 - \mathbb{P}(\{Y < 3\}) = 1 - \mathbb{P}(\{Y = 0\}) - \mathbb{P}(\{Y = 1\}) - \mathbb{P}(\{Y = 2\}) \\ &= 1 - \binom{16}{0} (22.31\%)^0 (77.69\%)^{16} - \binom{16}{1} (22.31\%)^1 (77.69\%)^{15} - \binom{16}{2} (22.31\%)^2 (77.69\%)^{14} \\ &= 72.72\%. \end{aligned}$$

4. Consideriamo le variabili aleatorie

$$X_1 = \text{“tempo necessario per completare il backup del primo computer”} \sim \text{Exp}\left(\frac{1}{2}\right)$$

$$X_2 = \text{“tempo necessario per completare il backup del secondo computer”} \sim \text{Exp}\left(\frac{1}{2}\right).$$

Per calcolare il valore atteso, sfruttiamo la linearità del valore atteso per ottenere che

$$\mathbb{E}(X_1 + X_2) = \mathbb{E}(X_1) + \mathbb{E}(X_2) = 2 + 2 = 4.$$

Per calcolare la varianza, sfruttiamo l'indipendenza delle variabili aleatorie per ottenere che

$$\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2) = 4 + 4 = 8.$$

5. La variabile aleatoria $X_1 + X_2$ ha distribuzione Gamma(2, 1/2), quindi ha densità (ricordiamo che $\Gamma(2) = 1! = 1$)

$$f(x) = \begin{cases} \left(\frac{1}{2}\right)^2 x e^{-\frac{1}{2}x} & \text{per } x > 0 \\ 0 & \text{per } x \leq 0. \end{cases}$$

La probabilità che il backup totale dei 2 computer sia inferiore a 7 ore è quindi, integrando per parti,

$$\begin{aligned} \mathbb{P}(\{X_1 + X_2 < 7\}) &= \int_0^7 \left(\frac{1}{2}\right)^2 x e^{-\frac{1}{2}x} dx = \left[-\frac{1}{2} x e^{-\frac{1}{2}x}\right]_0^7 + \int_0^7 \frac{1}{2} e^{-\frac{1}{2}x} dx \\ &= -\frac{7}{2} e^{-\frac{7}{2}} + \left[-e^{-\frac{1}{2}x}\right]_0^7 = -\frac{7}{2} e^{-\frac{7}{2}} + 1 - e^{-\frac{7}{2}} \\ &= 1 - \frac{9}{2} e^{-\frac{7}{2}} = 86.41\%. \end{aligned}$$

Esercizio 4. (7 punti) (7 punti) Un'azienda sostiene che la durata media giornaliera degli smartphone che produce è di 12 ore. Un'indagine condotta su alcuni smartphone è volta a mostrare che la durata è in realtà inferiore. Vengono rilevate le seguenti durate (in ore):

11.2	15.2	12.1	10.8	8.6	11.7	13.6	12.7	9.5	11.0
18.2	11.9	14.2	12.5	10.0	11.3	11.4	9.3	10.3	9.5
10.5	13.6	10.8	9.6	13.3	8.9	13.8	12.7	8.9	15.5
11.7	13.2	9.6	10.4	12.2	11.9	9.1	12.2	12.6	9.6

La media calcolata sui dati risulta essere 11.63 ore. È noto che la deviazione standard della popolazione è di 4 ore.

1. È possibile sostenere con significatività 1% che la durata media degli smartphone è in realtà inferiore a 12 ore? (N.B.: derivare le formule)
2. Calcolare il p -value del test.

Soluzione. Si deve impostare un test di ipotesi. La popolazione è descritta da una variabile aleatoria X con media $\mathbb{E}(X) = \mu$ e varianza $\text{Var}(X) = \sigma^2 = 4^2$. La legge di X non è nota. Dalla popolazione viene estratto un campione X_1, \dots, X_n con $n = 40 > 30$. Osserviamo che il campione è numeroso. Sia $\mu_0 = 12$. Il test di ipotesi è il seguente:

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

con livello di significatività $\alpha = 1\%$.

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

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

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

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

$$\alpha = \mathbb{P}(\{(X_1, \dots, X_n) \in R_c\}) = \mathbb{P}(\{\bar{X}_n < \mu_0 - \delta\}) = \mathbb{P}\left(\left\{\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} < -\frac{\delta}{\sigma/\sqrt{n}}\right\}\right).$$

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

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

Introduciamo il valore z_α tale che

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

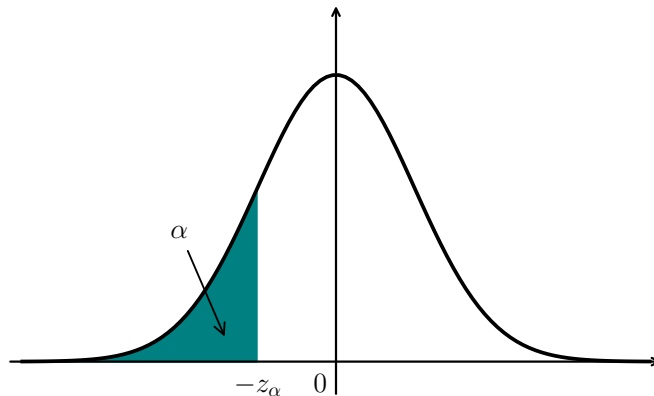


Figura 1: Coda sinistra della legge normale con probabilità α .

Allora, scegliendo

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

otteniamo la condizione desiderata sulla probabilità di errore del primo tipo.

In conclusione, la regione critica è

$$R_c = \left\{ (x_1, \dots, x_n) \in R(X_1, \dots, X_n) : \bar{x}_n < \mu_0 - \frac{\sigma}{\sqrt{n}} z_\alpha \right\},$$

e decidiamo come segue:

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

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

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

Calcoliamo il p -value dei dati:

$$\Phi\left(\frac{11.63 - 12}{4/\sqrt{40}}\right) = \Phi(-0.58) = 1 - \Phi(0.58) = 1 - 71.90\% = 28.1\%.$$

Poiché $1\% < 28.1\%$, non possiamo rifiutare l'ipotesi nulla H_0 con significatività 1% .