Prova scritta

Appello 06.06.2023

Tempo a disposizione: 120 minuti

La soluzione di ogni esercizio deve essere scritta in maniera pulita e leggibile sugli appositi fogli contrassegnati forniti a inizio prova, spiegando dettagliatamente il ragionamento con argomenti matematicamente validi. I risultati numerici devono essere dati con 4 cifre significative. Punti potranno essere sottratti per inadempienza a questi criteri. Esercizi consegnati su fogli esterni non saranno ritenuti validi. Sono benvenute le frasi che forniscono una spiegazione euristica della strategia che si intende seguire, ma non sono sufficienti, ci si aspetta calcoli espliciti. È permesso l'uso di calcolatrici non programmabili, e di **un foglio formato A4** (scritto fronte-retro, anche al computer) contenente appunti/formule. Un foglio supplementare è permesso per le tavole necessarie.

1. Si consideri un mazzo di carte italiane tradizionali composto da 40 carte (4 semi: denari, coppe, spade, bastoni; e 10 numeri: dall'1 al 7, e tre figure: fante (F), cavallo (C), re (R)).

   (a) Supponiamo di estrarre due carte a caso dal mazzo senza reinserimento. Siano X_1, X_2 il valore della prima e della seconda carta pescata, rispettivamente (le figure valgono 10). Queste variabili aleatorie sono indipendenti?
   no

   a) si   b) no   c) dati insuff.   d) correlazione positiva   e)Altro: ______

   (b) Supponiamo di estrarre due carte a caso dal mazzo senza reinserimento. Qual è la probabilità che la loro somma faccia esattamente 4 (le figure valgono 10)?
   Per calcolare la probabilità che la somma di due carte estratte a caso dal mazzo senza reinserirle faccia esattamente 5, dobbiamo considerare tutte le possibili combinazioni di carte che soddisfano questa condizione.
   Ci sono due casi in cui la somma delle carte è 5:
   a) Estraiamo due volte un 2 (qualsiasi seme)
   b) Estraiamo un asso e un 3 (qualsiasi seme)
   Il numero totale di combinazioni possibili è dato dal prodotto delle possibilità per ciascuna carta estratta. Quindi, considerando che ci sono 4 denari, 4 coppe, 4 spade, 4 bastoni, 3 figure e 7 numeri nel mazzo, otteniamo:

   $$\text{Numero totale di combinazioni} = 40 * 39 = 1560$$

   Il numero totale di combinazioni che danno una somma di 4 è $2 * 4 * 4 + 3 * 4 = 44$. Quindi, la probabilità è data da:

   $$P(\text{somma} = 4) = \frac{44}{1560} = 0.028$$

   a) 0.01   b) 0.018   c) 0.02   d) 0.028   e)Altro: ______

   (c) Supponiamo ora di pescare con reinserimento. Quale è la probabilità di pescare almeno 3 carte di coppe nei primi 4 tentativi?
   Probabilità binomiale. X = numero di carte di coppe pescate

   $$P(X \geq 3) = P(X = 3) - P(X = 4) = (1/4)^4 - 4 * (1/4)^3 (3/4) = 0.051$$

   a) 0.262   b) 0.051   c) 0.046   d) 0.291   e)Altro: ______

   (d) Sempre pescando con reinserimento, sia X il numero di estrazioni necessarie per osservare il primo fante dopo il primo 4.[1] Calcolare $\mathbb{E}(X)$ e $\mathrm{Var}(X)$.

[1] Per esempio, pescando $1, R, 4, 7, F, \ldots$ abbiamo $X = 5$, mentre per $F, 3, 6, C, 4, 4, F, \ldots$ abbiamo $X = 7$

Sia Y_4 la prima estrazione del 4 e Y_F il primo fante, allora $X = Y_4 + Y_F$. Quindi

$$\mathbb{E}(X) = \mathbb{E}(Y_4) + \mathbb{E}(Y_F) = \frac{1}{p_4} + \frac{1}{p_F} = 80$$

mentre per indipendenza di Y_4, Y_F abbiamo

$$Var(X) = Var(Y_4) + Var(Y_F) = \frac{1-p_4}{p_4^2} + \frac{1-p_F}{p_F^2} = 2 * 40 * 39 = 3120$$

2. Supponiamo di essere interessati a confrontare la durata di due tipi di batterie alcaline di marche diverse, chiamate "Marca A" e "Marca B". Per fare ciò, selezioniamo casualmente 7 batterie di ciascuna marca e misuriamo la durata delle batterie in ore. I risultati ottenuti sono i seguenti:

Marca A:	720,	730,	740,	710,	705,	715,	725
Marca B:	708,	703,	693,	718,	713,	698	

Assumiamo che la durata delle batterie di entrambe le marche segua una distribuzione normale, con medie μ_A e μ_B sconosciute e varianze $\sigma_A^2 = 121$ e $\sigma_B^2 = 81$ conosciute.

(a) Calcola l'intervallo di confidenza bilaterale al 95% per la media μ_A.
Per calcolare l'intervallo di confidenza bilaterale al 95% per la media μ_A, possiamo utilizzare la formula dell'intervallo di confidenza per la media di una distribuzione normale quando la deviazione standard è nota. L'intervallo di confidenza è dato da:

$$\bar{X} \pm z\frac{\sigma}{\sqrt{n}}$$

dove $\bar{X}$ rappresenta la media campionaria, z è il valore critico corrispondente al livello di confidenza (95% corrisponde a $z = 1.96$ per una distribuzione normale) e σ è la deviazione standard conosciuta. In questo caso, $\sigma_A = 11$ (radice quadrata di 121) e $n = 7$ (numero di campioni). Sostituendo i valori noti nell'equazione, otteniamo:

$$\bar{X} \pm 1.96\frac{11}{\sqrt{7}}$$

(b) Definire l'ipotesi nulla $\mathcal{H}_0$ e l'ipotesi alternativa $\mathcal{H}_1$ per il test che ha come obiettivo quello di verificare che c'è una differenza significativa nella durata media delle batterie tra le due marche.
L'ipotesi nulla, indicata come H_0, afferma che non vi è una differenza significativa nella durata media delle batterie tra le due marche (cioè $\mathcal{H}_0 \,:\, \mu_A = \mu_B$). L'ipotesi alternativa, indicata come H_1, afferma che vi è una differenza significativa nella durata media delle batterie tra le due marche (cioè $\mathcal{H}_1 \,:\, \mu_A \neq \mu_B$).

(c) Testare l'ipotesi nulla formulata al punto precedente al livello di significatività del 5%. Qual è la conclusione che possiamo trarre dal risultato del test?
Per testare l'ipotesi nulla formulata al punto precedente al livello di significatività del 5%, eseguiamo un test t a due campioni. Calcoliamo prima la statistica di test, che è data da:

$$T = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\frac{{\sigma_A}^2}{n_A} + \frac{{\sigma_B}^2}{n_B}}} = 2.741$$

dove $\bar{X}_A$ e $\bar{X}_B$ sono le medie campionarie delle due marche di batterie, σ_A e σ_B sono le deviazioni standard conosciute delle due marche di batterie, n_A e n_B sono le dimensioni dei rispettivi campioni.
Infine, calcoliamo il valore critico del test utilizzando la distribuzione normale (le variane sono conosciute):

$$z_{95}^b = 1.960$$

e confrontiamo la statistica di test con il valore critico per determinare se rigettare o accettare l'ipotesi nulla. In questo caso la statistica è piu estrema del valore critico e quindi rigettiamo $\mathcal{H}_0$.

(d) Calcola approssimativamente il valore p per il test discusso ai punti b) e c).
il valore p soddisfa $z_p^b = T$, che possiamo scrivere come $p = 1 - 2\Phi(T)$ (il due qui viene dal fatto che si tratta di un test bilatero, quindi dobbiamo guardare entrambe le code). Risolvendo, cioe guardando sulla tabella della distribuzione normale, si trova il valore di p.

3. Un call center riceve chiamate da due regioni, che chiamiamo A e B. Il numero di chiamate al minuto dalla regione A ha distribuzione di Poisson di parametro $\lambda_A = 2$, mentre il numero di chiamate al minuto dalla regione B ha distribuzione di Poisson di parametro $\lambda_B = 3$. Supponiamo che il numero di chiamate da A e il numero di chiamate da B siano indipendenti.

(a) Calcolare la probabilità che, in un dato minuto, il call center riceva almento 2 chiamate da A.

Detto X il numero di chiamate in un minuto da A, X ha distribuzione di Poisson di parametro 2 e la probabilità cercata è

$$\mathbb{P}(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - (\frac{2^0}{0!} - \frac{2^1}{1!})e^{-2} \approx 0.594.$$

a) 0.594 b) 0.323 c) 0.271 d) 0.729 e)Altro: ______

(b) Calcolare la probabilità che, in un dato minuto, il call center riceva almeno 2 chiamate *in totale*.

Siano X e Y il numero di chiamate ricevute rispettivamente da A e da B. Le v.a. X e Y sono indipendenti e hanno distribuzione di Poisson, quindi $X+Y$ ha distribuzione di Poisson di parametro $2+3=5$. La probabilità cercata è

$$P(X+Y \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - (\frac{5^0}{0!} + \frac{5^1}{1!})e^{-5} \approx 0.96.$$

(c) Calcolare la probabilità che se sono arrivate esattamente due chiamate al call center durante l'ultimo minuto, entrambe siano arrivate dalla regione B.
Qui abbiamo per Bayes

$$P(X = 2|X+Y = 2) = \frac{P(X+Y = 2|X = 2)P(X = 2)}{P(X+Y = 2)} = \frac{P(Y = 0)P(X = 2)}{P(X+Y = 2)}\frac{e^{-2}e^{-3}3^2/2!}{e^{-5}5^2/2!} = 9/25$$

(d) Supponiamo che il numero di chiamate dalla regione A in un dato minuto sia indipendente dal numero di chiamate da A in altri minuti. Calcolare (almeno in modo approssimato) la probabilità che, in un dato intervallo di 2 ore, il call center riceva almeno 260 chiamate da A.

Sia X_i il numero di chiamate da A nel minuto i-simo, $i = 1, \dots n = 120$; le X_i sono i.i.d. (di momento secondo finito), con $\mathbb{E}[X_i] = \mathrm{Var}(X_i) = 2$. Per il teorema centrale del limite (considerando $n = 120$ abbastanza grande), la probabilità cercata è circa

$$\mathbb{P}(\sum_i X_i \geq 260) = \mathbb{P}(\frac{\sum_i X_i - 120 \cdot 2}{\sqrt{120 \cdot 2}} \geq \frac{260 - 120 \cdot 2}{\sqrt{120 \cdot 2}} \approx \mathbb{P}(Z \geq 1.29) = 1 - \Phi(1.29) \approx 0.099.$$

4. Si consideri la seguente funzione,

$$f_\theta(x) = \frac{2}{\theta} x \exp\left(-\frac{x^2}{\theta}\right) \mathbb{1}_{(0,\infty)}(x),$$

dipendente dal parametro $\theta > 0$.

(a) Si mostri che, per ogni $\theta > 0$, f_θ è una densità di probabilità di una variabile aleatoria X, e si calcoli $\mathbb{P}(X \leq x)$ in funzione di x. **Suggerimento:** Sostituzione $x^2/\theta = u$

Abbiamo

$$\begin{aligned}
P(X \leq x) &= \int_{-\infty}^{x} f_\theta(t)dt \\
&= \int_{0}^{x} \frac{2t}{\theta} \exp\left(-\frac{t^2}{\theta}\right) dt \\
&\text{(effettuiamo una sostituzione con } u = \frac{t^2}{\theta}) \\
&= [-\exp(-u)]_{-\infty}^{\frac{x^2}{\theta}} \\
&= 1 - \exp\left(-\frac{x^2}{\theta}\right)
\end{aligned}$$

quindi

$$\int_{-\infty}^{\infty} f_\theta(x)dx = 1$$

e la funzione è positiva, quindi è una densita.

(b) Dato un campione $\{X_1, \dots, X_n\}$ estratto iid da $f_\theta(x)$, calcolare lo stimatore di massima verosimiglianza del parametro θ.

$$\begin{aligned}
\mathcal{L}(\theta; x_1, \dots, x_n) &= \prod_{i=1}^{n} f_\theta(x_i) \\
&= \prod_{i=1}^{n} \left(\frac{2x_i}{\theta} \exp\left(-\frac{x_i^2}{\theta}\right)\right) \\
&= \frac{2^n}{\theta^n} \exp\left(-\frac{1}{\theta}\sum_{i=1}^{n} x_i^2\right) \prod_{i=1}^{n} x_i
\end{aligned}$$

Applicando il logaritmo naturale ad entrambi i lati dell'equazione, otteniamo:

$$\begin{aligned}
\log(\mathcal{L}(\theta; x_1, \dots, x_n)) &= \log\left(\frac{2^n}{\theta^n} \exp\left(-\frac{1}{\theta}\sum_{i=1}^{n} x_i^2\right) \prod_{i=1}^{n} x_i\right) \\
&= n\log(2) - n\log(\theta) - \frac{1}{\theta}\sum_{i=1}^{n} x_i^2 + \log\left(\prod_{i=1}^{n} x_i\right) \\
&= n\log(2) - n\log(\theta) - \frac{1}{\theta}\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} \log(x_i)
\end{aligned}$$

Per trovare lo stimatore di massima verosimiglianza, dobbiamo massimizzare questa equazione rispetto a θ. Possiamo fare ciò derivando rispetto a θ e impostando la derivata uguale a zero:

$$\begin{aligned}
\frac{\partial}{\partial\theta} \log(\mathcal{L}(\theta; x_1, \dots, x_n)) &= -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i^2 \\
&= \frac{1}{\theta^2}\left(\sum_{i=1}^{n} x_i^2 - n\theta\right) = 0
\end{aligned}$$

Da cui otteniamo l'equazione:

$$\sum_{i=1}^{n} x_i^2 - n\theta = 0$$

Quindi lo stimatore di massima verosimiglianza del parametro θ è dato da:

$$\hat{\theta}_{\mathrm{ML}} = \frac{1}{n} \sum_{i=1}^{n} x_i^2$$

(c) Si tratta di uno stimatore corretto?
Dobbiamo calcolare il valore atteso dello stimatore e confrontarlo con il valore vero del parametro θ. In altre parole, dobbiamo verificare se $\mathbb{E}[\hat{\theta}_{\mathrm{ML}}] = \theta$.

$$\begin{aligned}\mathbb{E}[\hat{\theta}_{\mathrm{ML}}] &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} x_i^2\right] \\ &= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[x_i^2]\end{aligned}$$

Dal momento che stiamo considerando un campione estratto da una distribuzione con densità $f_\theta(x)$, possiamo calcolare il valore atteso di x_i^2 come:

$$\mathbb{E}[x_i^2] = \int_{-\infty}^{\infty} x_i^2 f_\theta(x)\, dx$$

Sostituendo la densità di probabilità $f_\theta(x) = 2x/\theta \exp\left(-x^2/\theta\right)$ nel calcolo del valore atteso, otteniamo:

$$\mathbb{E}[x_i^2] = \int_0^{\infty} x^2 \left(\frac{2x}{\theta} \exp\left(-\frac{x^2}{\theta}\right)\right) dx = \theta$$

quindi lo stimatore è corretto.

5. Consideriamo l'integrale

$$\int_{-3}^{5} \frac{\log(7+e^x)}{(2+\sin(x))}\, dx$$

(a) Discutere brevemente il metodo di integrazione di Monte Carlo e come può essere applicato per l'approssimazione di questo integrale.

(b) Scrivere i comandi in R la cui esecuzione restituisce, per il metodo discusso sopra, un valore approssimativo di questo integrale.
Nota: potrebbero essere utili i comandi `sin(·)` (il seno), `log(·)` (il logaritmo naturale) e `exp(·)` (la funzione esponenziale).