Prova scritta (Appello 27.06.2022)

Tempo a disposizione: 120 minuti

Nome, Cognome e #matricola: ______________________________

Riportare nome e cognome su ogni foglio consegnato. 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. Il risultato di ogni esercizio deve essere riportato nell'apposito spazio su questo foglio (o in forma di risposta o in forma di crocetta). Per le domande a crocette, è necessario dare anche una spiegazione della propria risposta (calcoli o argomento) sul foglio a parte. 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. Ci sono due scatole, contenenti ciascuna 10 schede SD, dello stesso tipo e usate: la prima scatola contiene 2 schede SD difettose e la seconda 4 schede SD difettose. Le altre schede in ogni scatola sono tutte funzionanti.

   (a) Peschiamo due schedine dalla seconda scatola, senza reinserimento. Gli eventi "la prima schedina pescata è difettosa" e "la seconda schedina pescata è difettosa" sono indipendenti?

   a) sì   b) no   c) dati insuff.   d)Altro: ______

   Le due estrazioni sono dipendenti. Per dimostrarlo, si calcola la probabilita congiunta delle due estrazioni e si dimostra che non è uguale al prodotto delle probabilita.

   (b) Gli eventi considerati al punto precedente sono indipendenti anche se la prima schedina è pescata dalla prima scatola e la seconda schedina dalla seconda?

   a) sì   b) no   c) dati insuff.   d)Altro: ______

   Le due estrazioni sono ora indipendenti, siccome si pesca da scatole diverse e la prima estrazione non cambia lo stato della seconda scatola.

   (c) Per effettuare le riparazioni, si prelevano (in blocco) 3 schede SD (senza reinserimento) dalla seconda scatola. Calcolare la probabilità di pescare almeno una schedina difettosa.

   a) 0.125   b) 0.83   c) 0.17   d) 0.875   e)Altro: ______

   Per calcolare la probabilità di pescare almeno una schedina difettosa dalla seconda scatola, consideriamo le seguenti informazioni: La seconda scatola contiene 10 schede SD, di cui 4 sono difettose e 6 sono funzionanti. Dato che stiamo prelevando 3 schede SD senza reinserimento, possiamo utilizzare il concetto di combinazioni. Il numero totale di modi in cui possiamo prelevare 3 schede dalla seconda scatola è dato da $\binom{10}{3}$. Il numero di modi in cui possiamo prelevare 3 schede funzionanti dalla seconda scatola è dato da $\binom{6}{3}$. Quindi, il numero di modi in cui possiamo prelevare solo schede funzionanti dalla seconda scatola è dato dalla combinazione di 3 schede funzionanti tra le 6 disponibili. La probabilità di non pescare schedine difettose dalla seconda scatola può essere calcolata come il rapporto tra questi due valori:

   $$P(\text{nessuna schedina difettosa dalla seconda scatola}) = \frac{\binom{6}{3}}{\binom{10}{3}}$$

   La probabilità di pescare almeno una schedina difettosa dalla seconda scatola è quindi il complemento di questa probabilità:

   $$P(\text{almeno una schedina difettosa dalla seconda scatola}) = 1 - P(\text{nessuna schedina difettosa dalla seconda scatola})$$

   Calcoliamo i coefficienti binomiali:

   $$\binom{6}{3} = \frac{6!}{3! \cdot (6-3)!} = 20$$

   $$\binom{10}{3} = \frac{10!}{3! \cdot (10-3)!} = 120$$

   Sostituendo questi valori, calcoliamo la probabilità:

   $$P(\text{nessuna schedina difettosa dalla seconda scatola}) = \frac{20}{120} = \frac{1}{6}$$

$$P(\text{almeno una schedina difettosa dalla seconda scatola}) = 1 - \frac{1}{6} = \frac{5}{6}$$

Quindi, la probabilità di pescare almeno una schedina difettosa dalla seconda scatola è $\frac{5}{6}$.

(d) Supponiamo di scegliere a caso, in maniera equiprobabile, una scatola e di prelevare (in blocco, quindi senza reinserimento) da essa 3 schede SD che risultano tutte funzionanti: qual è la probabilità che sia stata scelta la prima scatola?

Per calcolare la probabilità che sia stata scelta la prima scatola, dato che sono state prelevate 3 schede funzionanti, consideriamo le seguenti informazioni: Sappiamo che ci sono due scatole, la prima con 10 schede SD (2 difettose e 8 funzionanti) e la seconda con 10 schede SD (4 difettose e 6 funzionanti). La probabilità di scegliere la prima scatola è 0.5, dato che entrambe le scatole sono equiprobabili. Inoltre, la probabilità di prelevare 3 schede funzionanti dalla prima scatola può essere calcolata come il rapporto tra il numero di modi in cui possiamo prelevare 3 schede funzionanti dalla prima scatola e il numero totale di modi in cui possiamo prelevare 3 schede da entrambe le scatole. Il numero di modi in cui possiamo prelevare 3 schede funzionanti dalla prima scatola è dato da $\binom{8}{3}$, poiché ci sono 8 schede funzionanti nella prima scatola. Il numero totale di modi in cui possiamo prelevare 3 schede da entrambe le scatole è dato da $\binom{10}{3}$. Quindi, la probabilità che sia stata scelta la prima scatola, dato che sono state prelevate 3 schede funzionanti, può essere calcolata come:

$$P(\text{prima scatola scelta} \mid \text{3 schede funzionanti}) = \frac{P(\text{prima scatola scelta}) \cdot P(\text{3 schede funzionanti dalla prima scatola})}{P(\text{3 schede funzionanti})}$$

Poiché entrambe le scatole sono equiprobabili, $P(\text{prima scatola scelta}) = 0.5$. Sostituendo i valori dei coefficienti binomiali:

$$P(\text{3 schede funzionanti dalla prima scatola}) = \frac{\binom{8}{3}}{\binom{10}{3}} = \frac{56}{120} = \frac{7}{15}$$

$$P(\text{3 schede funzionanti}) = \frac{\binom{8}{3} \cdot \binom{6}{0} + \binom{4}{3} \cdot \binom{6}{0}}{\binom{10}{3}} = \frac{1}{2}$$

Sostituendo questi valori, calcoliamo la probabilità:

$$P(\text{prima scatola scelta} \mid \text{3 schede funzionanti}) = \frac{0.5 \cdot \frac{7}{15}}{\frac{1}{2}} = \frac{7}{15}$$

Quindi, la probabilità che sia stata scelta la prima scatola, dato che sono state prelevate 3 schede funzionanti, è $\frac{7}{15}$.

(e) Supponiamo di scegliere a caso, in maniera equiprobabile, una scatola e di continuare a pescare da quella scatola *con reinserimento* finché viene estratta una schedina difettosa. Sia X il numero di estrazioni necessarie per ottenere una schedina difettosa. Calcolare $\mathbb{E}(X)$.

Suggerimento: Può aiutare scrivere X usando una variabile di Bernoulli e delle variabili Y_1, Y_2 che rappresentano il numero di tentativi se si è scelta la scatola 1 o 2.

Siano Y_1 il numero di estrazioni se si sceglie la scatola numero 1 e Y_2 per la 2. Allora $X = Z \cdot Y_1 + (1-Z) \cdot Y_2$ quindi si ha

$$\mathbb{E}(X) = \frac{1}{2}\mathbb{E}(Y_1) + \frac{1}{2}\mathbb{E}(Y_2) = \frac{1}{2}(\frac{1}{p_1} + \frac{1}{p_2})$$

2. Un produttore di salmoni afferma che il peso medio dei suoi salmoni è maggiore del peso medio generale per questa specie di pesce. Per verificare questa affermazione, un gruppo di ricercatori ha raccolto un campione casuale di 8 salmoni dal produttore e ha ottenuto i seguenti dati sul peso (in kg):

7.2 6.8 7.5 7.4 7.1 7.3 7.6 7.3

Si sa che il peso medio dei salmoni nella popolazione generale è di 7.48 kg.

(a) Calcola la media campionaria e la deviazione standard campionaria del peso dei salmoni nel campione raccolto.

Per calcolare la media campionaria e la deviazione standard campionaria, utilizziamo i seguenti dati sul peso dei salmoni nel campione:

$$x = \{7.2, 6.8, 7.5, 7.4, 7.1, 7.3, 7.6, 7.3\}$$

La media campionaria può essere calcolata come:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

dove n è la dimensione del campione. In questo caso, $n = 8$. Calcoliamo la media campionaria:

$$\bar{x} = \frac{1}{8}(7.2 + 6.8 + 7.5 + 7.4 + 7.1 + 7.3 + 7.6 + 7.4) = 7.275$$

La deviazione standard campionaria può essere calcolata come:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Calcoliamo la deviazione standard campionaria:

$$s = \sqrt{\frac{1}{7}((7.2 - 7.25)^2 + (6.8 - 7.25)^2 + \ldots + (7.4 - 7.25)^2)} \approx 0.25$$

(b) Formula l'ipotesi nulla ($\mathcal{H}_0$) e l'ipotesi alternativa ($\mathcal{H}_1$) per il test statistico che ha come obbiettivo quello di contraddire il messaggio del produttore.
L'ipotesi nulla (H0) è che il peso medio dei salmoni nel campione sia minore o uguale al peso medio generale per questa specie di pesce, cioè $\mathcal{H}_0 \; : \; \mu \geq \mu_0 = 7.48$ kg. L'ipotesi alternativa (H1) è che il peso medio dei salmoni nel campione sia maggiore del peso medio generale, cioè $\mathcal{H}_1 \; : \; \mu < \mu_0 = 7.48$ kg.

(c) Testare l'ipotesi nulla formulata al punto precedente al livello di significatività del 2%. Quale è la conclusione che questo risultato ci permette di trarre?
Per eseguire il test statistico, utilizzeremo il test t di Student con un livello di significatività del 5Calcoliamo il valore t:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

dove μ è il peso medio generale (7 kg), $\bar{x}$ è la media campionaria (7.25), s è la deviazione standard campionaria (0.232), e n è la dimensione del campione (8).

$$t = \frac{7.275 - 7.48}{0.25/\sqrt{8}} \approx -2.31$$

Consultando la tabella della distribuzione t con 7 gradi di libertà (8-1), troviamo che il valore critico per un livello di significatività del 1% è circa 2.998. Poiché $t > t_{\text{crit}}$, dove t_{crit} è il valore critico, non possiamo rifiutare l'ipotesi nulla (H0).

(d) Calcolare approssimativamente il p-value per il test discusso ai punti precedenti.
guardando la tabella vediamo che il p-value è di circa 2.5%

3. Perfect Pandas SA ha sviluppato giocattoli elettronici. Assumiamo che il numero di ordini ricevuti da Perfect Pandas SA in un'ora abbia una distribuzione di Poisson con parametro $\lambda = 3/2$, indipendente per ogni ora della giornata.

(a) Calcolare la probabilità che Perfect Pandas abbia ricevuto almeno tre ordini nelle prime due ore della giornata.

a) 0.42 b) 0.37 c) 0.22 d) 0.58 e)Altro: ______

il numero di chiamate in due ore ha distribuzione di poisson con parametro $\lambda = 3$ quindi

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = e^{-3}(1 + \frac{3}{1!} + \frac{9}{2!}) = 0.42$$

Assumiamo ora che i giocattoli funzionano correttamente con una probabilità del $p = 0.90$, indipendentemente gli uni dagli altri. I nuovi giocattoli vengono spediti ai negozi di giocattoli di tutto il paese in scatole contenenti 200 unità ciascuna.

(b) Sia X il numero di giocattoli funzionanti in una scatola. Calcola il valore atteso e la deviazione standard di X, ovvero trova $\mathbb{E}(X)$ e $\mathrm{SD}(X)$.
$\mathbb{E}(X_1) = 180$. $SD(X) = \sqrt{np(1-p)} = \sqrt{18} \approx 4.24264$

(c) Quale percentuale di scatole contiene 180 o più giocattoli funzionanti? (Questo equivale a chiedere quale è la probabilità che una singola scatola contenga 180 o più giocattoli funzionanti.) Usa un'approssimazione normale per stimare la risposta; dovresti utilizzare la correzione di continuità.
La probabilità che una variabile casuale binomiale $(200, 0.9)$ assuma valori maggiori o uguali a 180 può essere approssimata come

$$\mathbb{P}(180 \text{ o più}) \approx 1 - \Phi\left(\frac{179.5 - 180}{\sqrt{18}}\right) = 0.5469$$

4. Per $\theta > 0$, si consideri una popolazione con densità di probabilità data da

$$f_\theta(x) := 1 + \theta x^\theta \mathbb{1}_{(0,1)}(x)$$

da cui peschiamo iid un campione $(X_1, \ldots, X_n)$

(a) Verificare che f_θ è una funzione di densità e calcolare $\mathbb{E}(2X_1 + 3)$ come funzione di θ
Dato che $f_\theta(x)$ rappresenta una densità di probabilità, deve essere integrata su tutto l'intervallo $(0, 1)$ per ottenere il valore di c_θ. Quindi:

$$\int_0^1 c_\theta x^\theta dx = 1$$

Integrando la funzione otteniamo:

$$c_\theta \int_0^1 x^{1/\theta} dx = 1$$

Semplificando l'integrale:

$$c_\theta \left[\frac{x^{1+\theta}}{1+\theta}\right]_0^1 = 1$$
$$c_\theta = 1 + \theta$$

Quindi, è una densita di probabilita.

(b) Identificare lo stimatore di massima verosimiglianza per il parametro θ.
Lo stimatore di massima verosimiglianza (MLE) per il parametro θ può essere ottenuto massimizzando la funzione di verosimiglianza, che è data dal prodotto delle densità di probabilità dei campioni:

$$L(\theta) = \prod_{i=1}^n f_\theta(X_i)$$

Per semplificare i calcoli, prendiamo il logaritmo della funzione di verosimiglianza e massimizziamo il logaritmo della funzione:

$$\log L(\theta) = \sum_{i=1}^n \log f_\theta(X_i)$$

Sostituendo $f_\theta(x)$ con il suo valore:

$$\log L(\theta) = \sum_{i=1}^n \log\left(c_\theta X_i^\theta\right)$$

$$\log L(\theta) = \sum_{i=1}^n (\log c_\theta + \theta \log X_i)$$

$$\log L(\theta) = n \log(1+\theta) + \theta \sum_{i=1}^n \log X_i$$

Deriviamo e mettiamo a 0 in θ Lo stimatore di massima verosimiglianza per il parametro θ è quindi:

$$\hat{\theta}_{\text{MLE}} = -(\frac{1}{n}\sum_{i=1}^n \log X_i)^{-1} - 1$$

(c) Discutere se lo stimatore di θ dato da $\bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$ è corretto[1].
Lo stimatore non è corretto: per $n = 1$ abbiamo

$$\mathbb{E}[\hat{\theta}] = \mathbb{E}[X_1] = \mathbb{E}[X] = 3\theta^{-3}\int_0^{\theta} x \cdot x^2 dx = \frac{3}{4}\theta \neq \theta$$

5. Consideriamo l'integrale

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{\sin(7+e^x)}{(2+\log(x^2+1))}\, dx$$

(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).

[1] questo stimatore può essere diverso da quello calcolato al punto precedente.