# Prova scritta (Appello 20.07.2022)

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. Il risultato di ogni esercizio svolto va riportato nell'apposito spazio su questo foglio. 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, formulari e appunti cartacei, ma non di soluzioni di esercizi svolti in classe, né di dispositivi elettronici.

1. Consideriamo l'integrale

$$\int_5^9 \frac{\log(2+\cos(e^x))}{1+x^3}\,dx$$

   (a) Discutere brevemente il metodo di integrazione di Monte Carlo nella sua forma generale e come può essere applicato per l'approssimazione di questo integrale in particolare.
   Vedi corso

   (b) Scrivere i comandi in R la cui esecuzione restituisce, per il metodo discusso sopra, un valore approssimativo di questo integrale.
   $x < -runif(10000, min = 5, max = 9)$
   $1/4 * mean((log(2 + cos(exp(x))))/(1 + x^3))$

2. Si consideri una variabile aleatoria $X$ che rappresenta la proporzione di elementi attivi in un processore prodotto da un'azienda e sia $X_i$ tale proporzione nell'$i$-esimo processore prodotto. Assumiamo che $X$ abbia funzione di densità

$$f_X(x) = \begin{cases} cx^4(1-x^2) & \text{quando } x \in (0,1) \\ 0 & \text{altrimenti} \end{cases}$$

   e che $X_i$ siano *iid* con la stessa distribuzione di $X$. Diciamo che l'$i$-esimo processore è *functional* (F) se $X_i \geq 0.7$, *recoverable* (R) se $X_i \in (0.5, 0.7)$, *compromised* (C) se $X_i \leq 0.5$

   (a) Determinare il valore di $c$ nella definizione di $f_X(x)$

$$1 = c\int_0^1 x^4(1-x^2)dx = c\frac{2}{35} \to c = \frac{35}{2}$$

   (b) Calcolare la probabilità che un processore a caso tra quelli prodotti sia *functional*

$$c\int_{0.7}^1 x^4(1-x^2)dx = c0.035 = 0.6176$$

   (c) Peschiamo 3 processori a caso da quelli prodotti. Calcolare la probabilità che i) si peschi la sequenza ordinata $F, R, C$ e quella che ii) si peschino almeno due processori *functional*

$$\binom{3}{2}0.6176^2(1-0.6176) + 0.6176^3 = 0.4375 + 0.2355 = 0.675$$

3. Il numero di clienti giornalieri di un ristorante ha distribuzione di Poisson con parametro $\lambda_1 = 10$ per i giorni della settimana lavorativa (dal martedì al giovedì) e parametro $\lambda_2 = 16$ per i giorni del finesettimana (venerdì-sabato-domenica). Il ristorante è chiuso il lunedì. Assumiamo che il numero di clienti ogni giorno dell'anno siano mutualmente indipendenti.

(a) Se sappiamo che un certo giorno sono arrivati 7 clienti ma non sappiamo che giorno fosse, quale è la probabilita che tale giorno fosse un giorno del finesettimana?
Sia $X$ il numero di clienti, $Y = 1$ se è un giorno della settimana e $Y = 0$ se il weekend.

$$P(Y = 0|X = 7) = P(X = 7|Y = 0)P(Y = 1)/P(X = 7)$$

Dove $P(X = 7|Y = 0) = e^{-16}16^7 7!$, $P(X = 7|Y = 1) = e^{-10}10^7 7!$, $P(Y = 1) = P(Y = 0) = 1/2$ e

$$P(X = 7) = P(X = 7|Y = 0)P(Y = 0) + P(X = 7|Y = 1)P(Y = 1)$$

Il risultato segue inserendo i valori

(b) Identificare la distribuzione del numero di clienti del ristorante in una settimana (dal lunedì alla domenica)
è la somma di variabili aleatorie di poisson indipendenti e quindi essa stessa una variabile aleatoria di poisson. Il suo parametro $\mu_0 = 10 + 10 + 10 + 17 + 17 + 17$

(c) Approssimare la probabilità che il ristorante abbia almeno 4000 clienti in un anno (si assuma per questo esercizio che un anno è composto da esattamente 52 settimane)
Per il teorema del limite centrale la distribuzione di clienti è approssimativamente normale con $\mu = \sigma^2 = 52 \cdot \mu_0$ (media e varianza di una distribuzione di poisson sono lo stesso numero). Quindi se $V$ è il numero di clienti in un anno, standardizzando

$$P(V < 4000) = P((V - \mu)/\sigma < (4000 - \mu)/\sigma) \approx P(Z < (4000 - \mu)/\sigma)$$

dove $Z$ è una variabile normale standard. Inserendo i valori per calcolare $(4000-\mu)/\sigma$ e consultando la tabella della distribuzione normale standard si orriene la stima desiderata.

4. La media durante il secolo scorso dell'escursione termica[1] media nella città di Roma sia di $13.0C$. Nel corso del 2021 vengono effettuate 365 misurazioni dell'escursione termica di Roma, che vengono salvate in un database. Alcune di queste misurazioni (pescate a caso) sono

13.7 13.5 13.3 12.9 13.0 13.4 13.2 12.8

Si assuma che la distribuzione delle escursioni termiche nel 2021 è normale, con media $\mu$ e deviazione standard $\sigma$ sconosciute.

(a) Si calcoli l'intervallo di confidenza bilaterale al 95% per lo stimatore di $\mu$ dato dalla media empirica dei dati indicati sopra
Abbiamo la media e deviazione standard del database sono $\bar{X} = 13.225$, $S = 0.31$. Siccome un intervallo di confidenza del 95% per la distribuzione $t$ con $\nu = 7$ corrisponde a 2.26 l'intervallo è di

$$13.225 \pm 2.26 \cdot 0.31/\sqrt{8} = (12.977, 13.472)$$

(b) Si formulino l'ipotesi nulla e l'ipotesi alternativa per il test che ha come obbiettivo quello di mostrare che l'escursione termica media è maggiore quella osservata il secolo scorso
Sia $\mu_0 = 13$ la temperatura media nel secolo scorso

$$\mathcal{H}_0 : \mu \leq \mu_0 \qquad \mathcal{H}_1 : \mu > \mu_0$$

(c) Assumiamo che la media e la varianza campionarie del database di 365 misurazioni siano di $\mu_1 = 13.08$, $S_1^2 = (0.5)^2$. Con questi dati si testi l'ipotesi nulla formulata al punto precedente al livello di significatività del 2%. Che tipo di conclusione ci permette di trarre il risultato del test nei confronti di qualcuno che nega il cambiamento climatico?
La statistica per il test t univariato è

$$T = \frac{13.08 - 13}{0.5/\sqrt{365}} = 3 > 2.01 = t^u_{\infty,2\%}$$

Quindi il test rigetta l'ipotesi nulla.

[1] differenza tra temperatura massima e temperatura minima durante una giornata

(d) Assumendo che continuando a fare misurazioni si mantenga la media e la varianza campionarie del punto precedente, quante misurazioni sarebbero necessarie per dimostrare il claim del punto b) al livello di significatività 1%?
Per trovare $n$ basta risolvere

$$T = \frac{13.08 - 13}{0.5/\sqrt{n}} = 2.326 = t^u_{\infty,1\%}$$

quindi $n \approx 210$.

5. Sia dato un campione $X_1 \dots X_n$ dove $X_i$ ha funzione di massa

$$\mathbb{P}_\theta(X = k) = p_X(k|\theta) := \frac{1}{\theta}\left(1 - \frac{1}{\theta}\right)^{k-1} \qquad \text{per ogni } k \in \{1, 2, 3, \dots\}$$

dove $\theta \in (1, \infty)$.

(a) Verificare che $p_X$ sia una distribuzione di probabilità per ogni $\theta > 0$. Di che distribuzione si tratta?
Distribuzione Geometrica: tutte le probabilita sono positive e abbiamo che

$$\sum_{k=1}^{\infty} \mathbb{P}_\theta(X = k) = \frac{1}{\theta} \sum_{k=1}^{\infty} \left(1 - \frac{1}{\theta}\right)^{k-1} = \frac{1}{\theta} \frac{1}{1 - (1 - \frac{1}{\theta})} = 1$$

(b) Calcolare lo stimatore di massima verosimiglianza $\hat{\theta}$ di $\theta$ per un campione $\{k_1, k_2, \dots, k_n\}$ con funzione di massa $p_X$
La funzione di massima verosimiglianza per un campione $\{k_i\}$ di $n$ elementi è:

$$L_\theta(k_i) = \prod_i \frac{1}{\theta}\left(1 - \frac{1}{\theta}\right)^{k_i - 1} = \theta^{-n}\left(1 - \frac{1}{\theta}\right)^{\sum_i k_i}$$

Prendiamo il logaritmo $\log L_\theta(k_i) = -n \log\theta + (\sum_i k_i)\log(1 - 1/\theta)$ e differenziamo pr trovare il massimo:

$$\partial_\theta \log L_\theta(k_i) = -\frac{1}{\theta} + (\sum_i k_i)\frac{1}{\theta^2(1 - 1/\theta)} = 0$$

which is true only if $\theta = \frac{1}{n}\sum_i k_i - 1$

(c) Esiste $n$ tale che questo stimatore non è corretto?
no, è sempre corretto.

---

Formule utili:

$$\text{Per ogni } z \in (0, 1): \qquad \sum_{k=0}^{\infty} z^k = \frac{1}{1 - z}$$