Prova scritta

Appello 08.06.2022

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. 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 desiderati e tavole necessarie.

1. Si considerino tre variabili aleatorie di Bernoulli X, Y, Z con rispettivi parametri $p_X = 1/3$, $p_Y = 1/5$, $p_Z = 2/3$ Si assuma che

   (a) Calcolare il valore atteso e la varianza di $2X + Y$ e di $X \cdot Z$
   $\mathbb{E}(2X + Y) = 2\mathbb{E}(X) + \mathbb{E}(Y) = 2\frac{1}{3} + \frac{1}{5}$
   $\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Z) = \frac{1}{3} \cdot \frac{2}{3}$
   $Var(2X + Y) = 4Var(X) + Var(Y) = 4\frac{1}{3}\frac{2}{3} + \frac{2}{3}\frac{1}{3}$
   $Var(X \cdot Z) = \mathbb{E}(X^2Z^2) - \mathbb{E}(XZ)^2 = \mathbb{E}(X^2)\mathbb{E}(Z^2) - \mathbb{E}(X)^2\mathbb{E}(Z)^2$

   (b) Calcolare la probabilità che $X = 0$ se $X + Y = 1$
   $\mathbb{P}(X = 0|X + Y = 1) = \mathbb{P}(X = 0, X + Y = 1)/\mathbb{P}(X + Y = 1) = \mathbb{P}(X = 0, Y = 1)/(\mathbb{P}(X = 0, Y = 1) + \mathbb{P}(X = 1, Y = 0)) = \frac{2}{3}\frac{1}{5}/(\frac{2}{3}\frac{1}{5} + \frac{1}{3}\frac{4}{5})$

   (c) Calcolare la funzione di massa congiunta di $X + Y$ e $X \cdot Z$. Verificare se queste variabili aleatorie sono indipendenti.
   Troviamo l'insieme delle immagini range$(X + Y) = \{0, 1, 2\}$, range$(X \cdot Z) = \{0, 1\}$. Quindi la funzione di massa congiunta si puo scrivere su una tabella di due colonne e tre righe. Notiamo che l'elemento $(X + Y, X \cdot Z) = (0, 1)$ ha probabilita 0, che è diversa da $\mathbb{P}(X + Y = 0)\mathbb{P}(X \cdot Z = 1)$ quindi le variabili aleatorie $(X + Y, X \cdot Z)$ non sono indipendenti.

2. Un produttore di pneumatici dice che i suoi prodotti durano almeno il 20% in più della media dei pneumatici in commercio, che è di 50000 (cinquantamila) km. Per verificare questa ipotesi compriamo 10 dei suoi pneumatici e li usiamo finché si consumano. Otteniamo i seguenti risultati (in km)

60900	61500	64400	59500	59580	64265	63400	58500	58000	60200

   Si assuma che la distribuzione della durata dei pneumatici del produttore sia normale, con media μ e varianza σ^2 sconosciute.

   (a) Si calcoli l'intervallo di confidenza bilaterale al 95% per μ.

   Calcoliamo la media empirica e la deviazione standard campionaria $\bar{X}_{10} = 61024$, $\sigma_{10} = 732.26$. Siccome la varianza non è data, la distribuzione di referenza è la distribuzione t, con $\nu = n - 1 = 9$ gradi di liberta. Guardando la tabella osserviamo $t^b_{5\%,9} = 2.262$ e quindi l'intervallo è

   $$61024 \pm 2.262 \cdot 732.26/\sqrt{10}$$

   (b) Si definiscano l'ipotesi nulla $\mathcal{H}_0$ e l'ipotesi alternativa $\mathcal{H}_1$ per il test che ha come obbiettivo quello di confermare il messaggio del produttore.

   $$\mathcal{H}_0 : \mu \leq 1.2\mu_0 = 60000 \qquad \mathcal{H}_1 : \mu > 1.2\mu_0 = 60000$$

(c) Si testi l'ipotesi nulla formulata al punto precedente al livello di significatività del 5%. Che conclusione ci permette di trarre il risultato del test?

Calcoliamo la statistica
$$\frac{\bar{X}_{10} - 1.2 \cdot \mu_0}{\sigma_{10}/\sqrt{10}} = 1.3977$$
e vediamo che è inferiore di $t^u_{5\%,9} = 1.833$, quindi il test non rigetta $\mathcal{H}_0$. Questo significa che non ci sono informazioni necessarie per confermare il messaggio del produttore.

(d) Si calcoli approssimativamente il valore p per il test disucsso ai punti b) e c).

La statistica
$$\frac{\bar{X}_{10} - 1.2 \cdot \mu_0}{\sigma_{10}/\sqrt{10}} = 1.3977$$
corrisponde a $t^u_{10\%,9}$, quindi il valore p è del 10%.

3. Si consideri una serie di variabili aleatorie IID (indipendenti e identicamente distribuite) $X_1, \ldots, X_{100}$ con distribuzione uniforme nell'intervallo (continuo) $[-1, 5] \subset \mathbb{R}$.

(a) Si calcoli la probabilità che non ci siano valori negativi tra $X_1, \ldots, X_{10}$.
per indipendenza $\mathbb{P}(X_1 > 0, \ldots, X_{10} > 0) = \mathbb{P}(X_i > 0)^1 0 = (\frac{5}{6})^{10}$

(b) Si calcoli la probabilità dell'evento "non ci sono valori negativi tra $X_1, \ldots, X_{10}$ e che non ci sono valori maggiori di 3 tra $X_6, \ldots, X_{13}$".
per indipendenza $\mathbb{P}(X_1 > 0, \ldots, X_{10} > 0, X_6 < 3, \ldots, X_{13} < 3) = \mathbb{P}(X_i > 0)^5 \mathbb{P}(X_i \in (0,3))^5 \mathbb{P}(X_i < 3)^3 = (\frac{5}{6})^5 (\frac{3}{6})^5 (\frac{4}{6})^3$

(c) Si calcoli il valore atteso e la varianza di $Y = \sum_{i=1}^{100} X_i$
$\mathbb{E}(\sum_{i=1}^{100} X_i) = 100\mathbb{E}(X_i) = 200$
$Var(\sum_{i=1}^{100} X_i) = 100 Var(X_i) = 100 \cdot \frac{1}{12}(6^2) = 300$

(d) Calcolare approssimativamente il valore di $\mathbb{P}(Y < 170)$.
Per il teorema del limite centrale $\mathbb{P}(Y < 170) = \mathbb{P}(\frac{Y - \mathbb{E}(Y)}{SD(Y)} < \frac{170 - \mathbb{E}(Y)}{SD(Y)}) \approx \Phi((170 - 200)/(10\sqrt{3})$ e il risultato si trova sulla tabella

4. Sia $(X_1, X_2, \ldots, X_n)$ un campione estratto da una popolazione di densità
$$f_\theta(x) = c_\theta \mathbf{1}_{[\theta,1]}(x),$$
dove $\theta \in (0, 1)$ è un parametro e $\mathbb{1}_{[\theta,1]}(x)$ rappresenta la funzione indicatrice sull'insieme $[\theta, 1]$.

(a) Determinare il valore di c_θ e di $\mathbb{E}(X_1)$ per ogni $\theta \in [0, 1)$.
$1 = c_\theta \int_{-\infty}^{\infty} \mathbf{1}_{[\theta,1]}(x)dx = c_\theta \int_\theta^1 dx$ quindi $c_\theta = \frac{1}{1-\theta}$.
$\mathbb{E}(X_i) = (1 + \theta)/2$

(b) Determinare uno stimatore $\hat{\theta}_n$ di massima verosimiglianza per θ.
si osserva che la funzione di verosimiglianza $L(x|\theta) = \prod_i \frac{1}{1-\theta} \mathbf{1}_{[\theta,1]}(x_i) = \frac{1}{(1-\theta)^n} \prod_i \mathbf{1}_{[0,x_i]}(\theta) = \frac{1}{(1-\theta)^n} \mathbf{1}_{[0,\min_i x_i]}(\theta)$ è crescente in $\theta \in (0, \min_i x_i)$ e 0 altrove, quindi il massimo si trova in $\hat{\theta} = \min_i x_i$

(c) Verificare se lo stimatore
$$\tilde{\theta}(X_1, \ldots, X_n) = \max\left\{\left(-1 + 2\frac{1}{n}\sum_{i=1}^n X_i\right), 0\right\}$$
è corretto per $n = 1$
Per $n = 1$ abbiamo $\tilde{\theta}(X_1, \ldots, X_n) = \max\{(-1 + 2X_1), 0\}$ ma
$$\mathbb{E}(\max\{(-1 + 2X_1), 0\}) = \int_{-\infty}^{\infty} \max\{(-1 + 2x), 0\} c_\theta \mathbf{1}_{[\theta,1]}(x)$$

e quindi per esempio per $\theta = 1/3$

$$\begin{aligned}\mathbb{E}(\max\{(-1+2X_1),0\}) &= \int_{1/3}^{1} \max\{(-1+2x),0\}\, c_\theta dx \\ &= \int_{1/2}^{1} (-1+2x)\frac{1}{1-1/3} dx = \frac{1}{/}4\frac{3}{2} \neq \frac{1}{3} = \theta\end{aligned}$$

quindi non è corretto

(d) Verificare se $\tilde{\theta}(X_1, \ldots, X_n)$ definito al punto precedente è consistente.
Per ogni $\theta \in (0,1)$ abbiamo, per ogni $\epsilon > 0$

$$0 = \lim_{n\to\infty} \mathbb{P}(|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}(X_i)| > \epsilon)$$

dove $\mathbb{E}(X_i) = (1+\theta)/2$ e quindi

$$0 = \lim_{n\to\infty} \mathbb{P}(|\frac{1}{n}\sum_{i=1}^{n} X_i - (1+\theta)/2| > \epsilon) = \lim_{n\to\infty} \mathbb{P}(|-1+\frac{2}{n}\sum_{i=1}^{n} X_i - \theta| > \epsilon')$$

per $\epsilon' = 2\epsilon$. Di conseguenza, per ϵ' sufficientemente piccolo

$$0 = \lim_{n\to\infty} \mathbb{P}(|\max\left\{-1+\frac{2}{n}\sum_{i=1}^{n} X_i, 0\right\} - \theta| > \epsilon') = \lim_{n\to\infty} \mathbb{P}(|\tilde{\theta} - \theta| > \epsilon')$$

e lo stimatore è consistente

5. Consideriamo l'integrale

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

(a) Discutere brevemente il metodo di integrazione di Monte Carlo e come può essere applicato per l'approssimazione di questo integrale.
Per la discussione del metodo di MonteCarlo si vedano le note del corso. In questo caso si puo applicare il metodo di monte carlo scegliendo $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ (la distribuzione normale standard) e $g(x) = e^{\frac{3}{2}} \frac{\log(7+e^x)}{\sqrt{2\pi}(2+\sin(x))}$

(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).
I comandi sono:
x < − rnorm(10000, mean=0, sd=1)
mean(exp(3/2)*log(7 + exp(x))/(2+sin(x)))