Esame Scritto del Quarto Appello

Tempo a disposizione: 2 ore

Riportare il numero di matricola **all'inizio** di ogni foglio. La soluzione di ogni esercizio deve essere scritta in modo chiaro e ordinato, e può non essere valutata se la calligrafia è illeggibile. Se l'esercizio lo richiede, evidenziare il risultato numerico nella soluzione. Le soluzioni degli esercizi devono essere riportate sul foglio protocollo nell'ordine proposto, la soluzione di ogni esercizio deve iniziare in una nuova pagina.
Le soluzioni devono includere il procedimento dettagliato che porta alle risposte. Risposte corrette non adeguatamente motivate saranno penalizzate.
È permesso l'uso di note, appunti, manuali e materiale didattico. Non è permesso l'uso di dispositivi elettronici ad esclusiva eccezione di una calcolatrice non programmabile. L'infrazione di queste regole o la comunicazione con altri comportano l'annullamento del compito.

1. Per ognuna delle seguenti affermazioni si determini se essa è VERA oppure FALSA, motivando rigorosamente le risposte.

   (a) Dati due eventi A e B disgiunti (e con probabilità positiva), $\mathbb{P}(A \mid B) = \mathbb{P}(A)$.
   FALSA: $\mathbb{P}(A \mid B) = \mathbb{P}(A \cap B)/\mathbb{P}(B) = \mathbb{P}(\emptyset)/\mathbb{P}(B) = 0$.

   (b) Sia X una v.a. con densità, e per ogni $\beta \in (0,1)$, sia r_β il corrispondente quantile. Allora, per ogni $0 < \beta < \gamma < 1$, si ha $\mathbb{P}(r_\beta < X < r_\gamma) = \gamma - \beta$.
   VERA: $\mathbb{P}(r_\beta < X < r_\gamma) = \mathbb{P}(X < r_\gamma) - \mathbb{P}(X \leq r_\beta) = \gamma - \beta$.

   (c) Se $X_1, \ldots, X_n$ sono i.i.d. con media μ e varianza σ^2, allora la media campionaria $\bar{X}_n$ ha anch'essa varianza σ^2.
   FALSA: $\mathrm{Var}(\bar{X}_n) = \sigma^2/n$, non σ^2.

   (d) Se due variabili aleatorie X e Y (limitate) sono indipendenti, allora $\mathrm{cov}(X^2, Y^3) = 0$.
   VERA: L'indipendenza di X e Y implica l'indipendenza di X^2 e Y^3; poiché v.a. indipendenti sono scorrelate, $\mathrm{cov}(X^2, Y^3) = 0$.

   (e) Se $X_1, \ldots, X_n$ sono i.i.d. con distribuzione gaussiana, allora la media campionaria $\bar{X}_n$ ha distribuzione gaussiana per ogni n.
   VERA: Per riproducibilità delle gaussiane, la media di variabili gaussiane indipendenti è ancora gaussiana.

   (f) Sia V_n uno stimatore consistente di un parametro θ (cioè $\lim_{n\to\infty} \mathbb{P}_\theta(|V_n - \theta| \geq \epsilon) = 0$ per ogni $\epsilon > 0$, per ogni θ). Allora V_n^2 è uno stimatore consistente di θ^2.
   VERA: Poiché la funzione $x \mapsto x^2$ è continua, per ogni $\epsilon > 0$ esiste $\delta > 0$ tale che $|x - y| < \delta \Rightarrow |x^2 - y^2| < \epsilon$, o equivalentemente $|x^2 - y^2| \geq \epsilon \Rightarrow |x - y| > \delta$. Quindi

$$\mathbb{P}_\theta(|V_n^2 - \theta^2| \geq \epsilon) \leq \mathbb{P}_\theta(|V_n - \theta| \geq \delta) \to 0$$

   per $n \to \infty$, per ogni θ.

2. Il tempo (in secondi) impiegato da un server per elaborare una richiesta segue una distribuzione esponenziale di parametro 2.

   (a) Calcolare la probabilità che il server impieghi almeno 2 secondi per elaborare la richiesta.
   Sia X il tempo di elaborazione. Poiché $X \sim \mathrm{Exp}(2)$, la probabilità cercata è

   $$\mathbb{P}(X \geq 2) = \int_2^\infty 2e^{-2x}dx = e^{-4} = 0.018.$$

   Quando il server termina di elaborare una richiesta, passa immediatamentente alla successiva. Per ciascuna richiesta, il tempo è (come prima) esponenziale di parametro 2, e i tempi di richieste differenti sono indipendenti.

   (b) Calcolare la probabilità che il server impieghi almeno sei minuti (= 360 secondi) per elaborare 700 richieste.
   Sia $X_1, \ldots, X_{700}$ il tempo impiegato per ciascuna richiesta. Le X_i sono i.i.d. $\mathrm{Exp}(2)$, quindi hanno media $\mu = 1/2$ e varianza $\sigma^2 = 1/4$. Il tempo totale è $\sum X_i = 700\bar{X}_{700}$, e per il TCL:

   $$\sqrt{700}\left(\bar{X}_{700} - \frac{1}{2}\right)/\frac{1}{2} \approx N(0,1).$$

   La probabilità cercata è

   $$\begin{aligned}\mathbb{P}\left(\sum X_i \geq 360\right) &= \mathbb{P}\left(Z \geq \sqrt{700}\cdot\frac{\frac{360}{700}-\frac{1}{2}}{1/2}\right)\\ &= \mathbb{P}(Z \geq 0.76) = 1 - \Phi(0.76) = 0.2236.\end{aligned}$$

   (c) Trovare un valore r tale che, con probabilità 99%, il server abbia elaborato 700 richieste in al massimo r secondi.
   Usando l'approssimazione normale di cui sopra, cerchiamo r tale che

   $$\mathbb{P}\left(\sum X_i \leq r\right) = \mathbb{P}\left(Z \leq \sqrt{700}\cdot\frac{\frac{r}{700}-\frac{1}{2}}{1/2} =: a(r)\right) = 0.99.$$

   Quindi, imponendo $a(r) = z_{0.99} = 2.33$, otteniamo

   $$r = 700\left(\frac{1}{2} + \frac{1}{2}\cdot\frac{2.33}{\sqrt{700}}\right) = 380.82.$$

3. Sia $X_1, \ldots X_n$ un campione i.i.d. di densità

   $$f_\theta(x) = \frac{6}{\theta}\cdot\frac{x}{\theta}\cdot\left(1-\frac{x}{\theta}\right)1_{(0,\theta)}(x) = \begin{cases}\frac{6}{\theta}\cdot\frac{x}{\theta}\cdot\left(1-\frac{x}{\theta}\right) & \text{se } x \in (0,\theta)\\ 0 & \text{altrimenti}\end{cases}$$

   con $\theta > 0$ parametro non noto.

(a) Calcolare la funzione di ripartizione F_θ delle X_i.
La funzione di ripartizione F si ottiene come $F_\theta(x) = \int_{-\infty}^{x} f_\theta(y)\,dy$. Poiché $f_\theta = 0$ fuori da $(0,\theta)$, $F_\theta(x) = 0$ per $x < 0$, $F_\theta(x) = 1$ per $x \geq \theta$. Per $x \in [0,\theta)$, abbiamo (ad esempio con il cambio di variabili $y = x/\theta$)

$$F_\theta(x) = \frac{6}{\theta^2}\left(\int_0^x t\,dt - \frac{1}{\theta}\int_0^x t^2\,dt\right) = \frac{3x^2}{\theta^2} - \frac{2x^3}{\theta^3}$$

(b) Calcolare la densità di X_i/θ.
Sia $Y_i = X_i/\theta$. La trasformazione $h(x) = x/\theta$ è C^1 e invertibile con inversa $h^{-1}(y) = \theta y$. Applicando la formula del cambio di variabile:

$$f_Y(y) = f_X(h^{-1}(y)) \cdot \left|\frac{d}{dy}h^{-1}(y)\right| \cdot 1_{(0,1)}(y) = f_X(\theta y)\cdot\theta\cdot 1_{(0,1)}(y) = 6y(1-y)\cdot 1_{(0,1)}(y).$$

(c) Dimostrare che (per dati x_i strettamente positivi) esiste un unico stimatore di massima verosimiglianza di θ (non è richiesto di calcolarlo).
La funzione di verosimiglianza è (supponendo $x_i > 0$ per ogni i):

$$L(\theta; x_1,\dots,x_n) = \prod_{i=1}^n f_\theta(x_i) = \left(\frac{6}{\theta^3}\right)^n \cdot \prod_{i=1}^n x_i\cdot(\theta - x_i)\, 1_{\theta > \max_i x_i}.$$

Bisogna dunque verificare che L ha un unico punto di massimo in $(\max_i x_i, +\infty)$ (essendo L strettamente positiva su $(\max_i x_i, +\infty)$, questo sarà anche l'unico punto di massimo su tutto $\mathbb{R}$). Possiamo equivalentemente massimizzare $\log L$ (su $(\max_i x_i, +\infty)$):

$$\log L(\theta) = n\log 6 - 3n\log\theta + \sum_{i=1}^n \log x_i + \sum_{i=1}^n \log(\theta - x_i).$$

La funzione $\log L$ è C^1 in θ per $\theta > \max_i x_i$, con $\lim_{\theta\to\max_i x_i}\log L(\theta) = \lim_{\theta\to\infty}\log L(\theta) = -\infty$. Quindi esiste almeno un punto di massimo, e in ogni punto di massimo la derivata di $\log L$ si annulla. Resta da mostrare l'unicità del punto di massimo, per questo dimostriamo che la derivata di $\log L$ si annulla in un solo punto. Abbiamo

$$\frac{d}{d\theta}\log L(\theta) = \frac{n}{\theta}\left(-3 + \frac{1}{n}\sum_{i=1}^n \frac{1}{1 - x_i/\theta}\right) =: \frac{n}{\theta}g(\theta),$$

in particolare la derivata si annulla se e solo se g si annulla. Ora g è continua, strettamente decrescente (su $(\max_i x_i, +\infty)$), e con limiti $+\infty$ per $\theta \to \max_i x_i$ e -2 per $\theta \to +\infty$. Dunque ammette un solo punto in cui si annulla. Questo conclude la dimostrazione.