Esame Scritto del Primo 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) Dato un campione i.i.d. $X_1, \dots X_n$ (con momento primo finito), un intervallo di fiducia per il valore atteso delle X_i, di livello 95%, è un intervallo I tale che la media campionaria $\bar{X}_n$ delle X_i cade in I con probabilità almeno 95%.
   FALSA: Il valore atteso $\mathbb{E}[X_i]$ cade in I con probabilità almeno 95%, mentre ad esempio per un I.F. per la media di popolazione gaussiana ($I = [\bar{X}_n \pm \sigma q_{1-\alpha/2}/\sqrt{n}]$), $\bar{X}_n$ sta sempre in I.

   (b) Se X e Y sono due v.a. Bernoulli di parametro $p \in (0,1)$, allora $X+Y$ è v.a. binomiale di parametri 2, p.
   FALSA: Se $X = Y$, allora $X + Y = 2X$ assume solo valori 0 e 2 (con probabilità positiva), in particolare non è $B(2,p)$.

   (c) Se due v.a. X e Y (discrete o con densità) hanno la stessa legge, allora $\mathbb{E}[\varphi(X)] = \mathbb{E}[\varphi(Y)]$ per ogni funzione $\varphi : \mathbb{R} \to \mathbb{R}$ limitata.
   VERA: Nel caso discreto, $\mathbb{E}[\varphi(X)] = \sum_x \varphi(x)p(x)$ e la funzione di massa p di X dipende solo dalla legge. Analogamente nel caso con densità usando $\mathbb{E}[\varphi(X)] = \int_{\mathbb{R}} \varphi(x) f(x)\, dx$.

   (d) Se $\mathbb{E}[X] = 0$, allora X è nulla.
   FALSA: Ad esempio $X \sim N(0,1)$ ha media nulla ma non è nulla.

   (e) Dato un campione i.i.d. $X_1, \dots X_n$ (con momento primo finito), $(X_1 + X_2)/2$ è uno stimatore corretto del valore atteso $\mathbb{E}[X_i]$.
   VERA: Per linearità del valore atteso, $\mathbb{E}[(X_1 + X_2)/2] = (\mathbb{E}[X_1] + \mathbb{E}[X_2])/2 = \mathbb{E}[X_1]$.

   (f) Se $(x_1, y_1), \dots (x_n, y_n)$ sono dati bivariati, il loro coefficiente di correlazione r vale ± 1 se e solo se i dati (x_i, y_i) sono allineati (cioè disposti sulla stessa retta).
   VERA: $\inf_{a,b} \sum_i (y_i - ax_i - b)^2$ è nullo se e solo se (x_i, y_i) sono allineati, e d'altra parte $\inf_{a,b} \sum_i (y_i - ax_i - b)^2 = \mathrm{Var}_e(y)(1 - r^2)$ è nullo se e solo se $r = \pm 1$.

2. Aldo e Bruno giocano a carte. In ciascun turno, Aldo estrae due carte distinte da un mazzo di carte napoletane (4 semi, 10 carte per seme). Aldo vince se almeno una carta è del seme "denaro".

   (a) Calcolare la probabilità di vittoria di Aldo in un singolo turno.
   Sia S l'insieme delle carte. Siamo nel caso di un modello uniforme su $\Omega = \{(a,b) \mid a, b \in S,\ a \neq b\}$, che ha cardinalità $40 \cdot 39$. L'evento complementare "nessuna carta di denaro" ha cardinalità $30 \cdot 29$. Quindi la probabilità cercata è

   $$\mathbb{P}(\text{almeno una di denaro}) = 1 - \mathbb{P}(\text{nessuna di denaro}) = 1 - \frac{30 \cdot 29}{40 \cdot 39} = 0.442.$$

   (b) Calcolare la probabilità che, su 5 turni, Aldo vinca in almeno 2 turni.
   Siamo in presenza di uno schema di Bernoulli di 5 prove (i turni), in cui il successo (vittoria di Aldo) ha probabilità $p = 0.442$ (dal punto precedente). In particolare la v.a. X che conta il numero di vittorie di Aldo ha distribuzione $B(5, p)$. La probabilità cercata è

   $$\mathbb{P}(X \geq 2) = 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) = 1 - (1-p)^5 - 5p(1-p)^4 = 0.732.$$

   Il mazzo di carte è stato scelto a caso tra 4 mazzi. Aldo e Bruno vengono però a sapere che uno di questi 4 mazzi è truccato, in modo che tutte le carte di quel mazzo siano di denaro.

   (c) Se Aldo vince in almeno 2 turni su 5, qual è la probabilità che il mazzo scelto sia truccato?
   Sia $A = \{\text{truccato}\}$, sia X come sopra. Per le ipotesi del testo, $\mathbb{P}(A) = 1/4$. Per il punto precedente $\mathbb{P}(X \geq 2 \mid A^c) = 0.732$. Invece, se il mazzo è truccato, ovviamente Aldo vince sempre, in particolare, $\mathbb{P}(X \geq 2 \mid A) = 1$. Per Bayes, la probabilità cercata è

   $$\mathbb{P}(A \mid X \geq 2) = \frac{\mathbb{P}(X \geq 2 \mid A)\mathbb{P}(A)}{\mathbb{P}(X \geq 2 \mid A)\mathbb{P}(A) + \mathbb{P}(X \geq 2 \mid A^c)\mathbb{P}(A^c)} = \frac{1 \cdot 0.25}{1 \cdot 0.25 + 0.732 \cdot 0.75} = 0.313.$$

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

   $$f_\theta(x) = \theta x^{\theta-1} 1_{(0,1)}(x) = \begin{cases} \theta x^{\theta-1} & \text{se } x \in (0,1) \\ 0 & \text{altrimenti} \end{cases}$$

   con $\theta > 0$ parametro non noto.

   (a) Calcolare (in funzione di θ) valore atteso $\mathbb{E}[X_i]$ e varianza $\mathrm{Var}(X_i)$.
   Poiché $f_\theta = 0$ fuori da un intervallo limitato, esistono finiti tutti i momenti di X_i. Abbiamo, per ogni k intero positivo,

   $$\mathbb{E}[X_i^k] = \int_{\mathbb{R}} x^k f(x)\,dx = \int_0^1 \theta x^{\theta+k-1}\,dx = \frac{\theta}{\theta+k} x^{\theta+k} \Big|_{x=0}^1 = \frac{\theta}{\theta+k}.$$

In particolare $\mathbb{E}[X_i] = \theta/(\theta+1)$ e

$$\mathrm{Var}(X_i) = \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 = \frac{\theta}{\theta+2} - \frac{\theta^2}{(\theta+1)^2}.$$

(b) Calcolare (in funzione di θ) la densità di $Y_i = -\log X_i$, se esiste.
La funzione $h : (0,1) \to (0,\infty)$ data da $h(x) = -\log x$ è C^1, invertibile con inversa $h^{-1}(y) = e^{-y}$ a sua volta C^1. Quindi possiamo applicare la formula di cambio variabile: Y_i ha densità

$$f_Y(y) = f_X(h^{-1}(y))|\frac{dh^{-1}}{dy}(y)|1_{(0,\infty)}(y) = \theta e^{-(\theta-1)y} \cdot |-e^{-\theta y}|1_{(0,\infty)}(y) = \theta e^{-\theta y}1_{(0,\infty)}(y),$$

in particolare Y_i ha distribuzione esponenziale di parametro θ.

(c) Calcolare, se esiste, lo stimatore di massima verosimiglianza di θ.
Dobbiamo massimizzare (in θ) la funzione di verosimiglianza

$$L(\theta; x_1, \dots x_n) = \theta^n \left(\prod_i x_i\right)^{\theta-1} 1_{x_i \in (0,1)\,\forall i}.$$

Equivalentemente, massimizziamo il suo logaritmo (possiamo assumere tutti gli x_i in $(0,1)$, visto che X_i hanno valori in $(0,1)$):

$$\log L(\theta; x_1, \dots x_n) = n\log\theta + (\theta-1)\sum_{i=1}^n \log x_i,$$

$$\frac{d}{d\theta}\log L(\theta; x_1, \dots x_n) = \frac{n}{\theta} + \sum_{i=1}^n \log x_i.$$

In particolare $\frac{d}{d\theta}\log L \geq 0$ se e solo se $\theta \leq \frac{n}{-\sum_{i=1}^n \log x_i}$. Quindi

$$\hat\theta = \frac{1}{-\frac{1}{n}\sum_{i=1}^n \log X_i}$$

è l'unico stimatore di massima verosimiglianza.