Esame Scritto del Settimo 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 di difficile lettura. 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.
Non è permesso l'uso di note, appunti, manuali o materiale didattico di alcun tipo, al di fuori del formulario e delle tavole statistiche fornite assieme al compito. 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) Data una variabile aleatoria X tale che $\mathbb{E}[X] = 0$, $\mathrm{Var}(X) = 1$, la probabilità dell'evento $\{-a \leq X \leq a\}$ non può mai essere inferiore a $1/a^2$.
   FALSO: per la disuguaglianza di Chebychev,
   $$\mathbb{P}(\{X \notin [-a, a]\}) = \mathbb{P}(|X| > a) \leq \frac{\mathrm{Var}(X)}{a^2} = \frac{1}{a^2},$$
   e quindi $\mathbb{P}(\{X \in [-a, a]\})$ non può essere inferiore a $1 - 1/a^2$.

   (b) Se X, Y sono due variabili aleatorie indipendenti, le variabili e^X, e^Y sono indipendenti.
   VERO: infatti per ogni scelta di intervalli $I, J \subseteq (0, +\infty)$ vale
   $$\begin{aligned}\mathbb{P}(e^X \in I, e^Y \in J) &= \mathbb{P}(X \in \log(I), Y \in \log(J)) \\ &= \mathbb{P}(X \in \log(I))\mathbb{P}(Y \in \log(J)) = \mathbb{P}(e^X \in I)\mathbb{P}(e^Y \in J)\end{aligned}$$
   (in cui $\log(I) = \{x \in \mathbb{R} : e^x \in I\}$).

   (c) Se la somma delle probabilità di due eventi con probabilità non nulla è la probabilità della loro unione, allora gli eventi non possono essere indipendenti.
   VERO: siano infatti $A, B \subseteq \Omega$ tali che
   $$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B),$$
   allora poiché vale in generale
   $$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B),$$
   deve essere $\mathbb{P}(A \cap B) = 0$. Se dunque gli eventi fossero indipendenti, e quindi $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, uno dei due eventi dovrebbe avere probabilità nulla, cosa esclusa per ipotesi.

   (d) Per effettuare il test chi-quadro per la varianza di un campione numeroso è necessario conoscere il valore atteso del campione.
   FALSO: per calcolare la statistica test S_n^2 non è necessario conoscere il valore di $\mathbb{E}[X_i]$, ma solo i valori assunti dal campione.

(e) Il p-value non è la probabilità di un evento, ma una misura della verosimiglianza dei dati che assume valori positivi.
FALSO: il p-value è la probabilità (sotto l'ipotesi nulla) che la statistica test si discosti dall'ipotesi nulla più del suo valore campionario.

(f) Per una variabile discreta X di esiti $x_1, x_2, x_3, \dots$ il valore atteso $\mathbb{E}[X]$ coincide con l'esito x_i che X assume con maggior probabilità.
FALSO: ad esempio il valore atteso di una variabile Bernoulli di parametro $p \in (0,1)$ è p, e non coincide con alcuno degli esiti possibili $0, 1$.

2. Si svolgano i seguenti calcoli.

(a) Data una variabile X con densità esponenziale di parametro λ, mostrare che

$$Y = \lceil X \rceil = \min\{n \in \mathbb{N} : n \geq X\}$$

(l'intero superiore di X, ossia $Y = n$ se $X \in (n-1, n]$) è una variabile discreta con distribuzione geometrica, e determinare il parametro p di tale distribuzione.
Poiché $\lfloor X \rfloor$ assume solo valori interi, Y è una variabile discreta. La funzione di massa è

$$\mathbb{P}(Y = n) = \mathbb{P}(X \in (n-1, n]) = \int_{n-1}^{n} \lambda e^{-\lambda x} = \Big[-e^{-\lambda x}\Big]_{n-1}^{n} = e^{-\lambda(n-1)}(1 - e^{-\lambda}),$$

che corrisponde alla distribuzione geometrica con $p = e^{-\lambda}$.

(b) Dato un campione $X_1, \dots, X_n$ di variabili geometriche di parametro p, determinare stimatori di p con il metodo di massima verosimiglianza e con il metodo del momento primo e dire se differiscono.
La verosimiglianza è

$$L(p) = L(p; k_1, \dots k_n) = (1-p)^{k_1 + \dots + k_n - n} p^n,$$

dati $a, b > 0$ la funzione $p \mapsto (1-p)^a p^b$ ha massimo nel punto $x^* = \frac{b}{a+b}$, per cui lo stimatore di massimo verosimiglianza è

$$\hat{p} = \frac{n}{k_1 + \dots + k_n}.$$

Per metodo del momento primo si risolve

$$\frac{k_1 + \dots + k_n}{n} = \mathbb{E}[X] = \frac{1}{p},$$

per cui i due stimatori sono uguali.

(c) Data una variabile Z con densità esponenziale di parametro α, determinare se la variabile aleatoria $W = e^Z$ ha densità, se sì calcolarla. Determinare quali momenti possiede la variabile W.
La funzione $h : [0, \infty) \to (1, \infty)$, $h(x) = e^x$ è bigettiva, differenziabile e con inversa $h^{-1}(w) = \log w$, per la formula di cambio di variabile si ha che W ha densità

$$f_W(w) = \begin{cases} f_Z(h^{-1}(w)) |\frac{d}{dw} h^{-1}(w)| & w > 1 \\ 0 & w \leq 1 \end{cases}$$
$$= \begin{cases} \frac{\alpha}{w^{\alpha+1}} & w > 1 \\ 0 & w \leq 1 \end{cases}.$$

Poiché l'integrale
$$\int_1^\infty w^p \frac{\alpha}{w^{\alpha+1}} dw$$
converge se e solo se $\alpha > p$, la variabile W ha momento p-esimo finito se e solo se $p < \alpha$.

3. Il gestore di un negozio è interessato a sapere in quanti giorni il numero di clienti è almeno 2. Negli anni dal 2000 al 2023 il numero di clienti giornalieri è stato compatibile una distribuzione di Poisson di paramentro 3.

   (a) In un giorno a caso tra il 2000 e il 2023, con quale probabilità il negozio ha avuto almeno 2 clienti?
   Detto X il numero di clienti nel dato giorno, la probabilità cercata è
   $$P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - e^{-3} - 3e^{-3} \approx 0.801.$$

Nel 2024 è cambiato il catalogo ed è stato osservato che su 300 giorni in cui il negozio era aperto in 261 ci sono stati almeno 2 clienti.

   (b) Fornire un intervallo di fiducia del 95% per la probabilità che, in un giorno a caso in futuro, ci siano almeno 2 clienti.
   Siamo nel caso intervallo di fiducia I (approssimato) per la probabilità p dell'evento "almeno 2 clienti", caso grandi campioni ($n = 300$, valore della frequenza relativa empirica $\bar{y} = 261/300 = 0.87$, $n\bar{y}(1-\bar{y}) \geq 10$). Detta $\bar{Y}$ la frequenza relativa empirica, l'intervallo I è
   $$I = \left[\bar{Y} \pm q_{0.975} \sqrt{\frac{\bar{Y}(1-\bar{Y})}{n}} \right].$$
   Con il dato $\bar{y} = 0.87$, otteniamo l'intervallo numerico $[0.87 \pm 0.038]$.

   (c) Dai dati forniti, è possibile affermare che il cambio di prodotti ha avuto effetto? Formulare un test di ipotesi di livello 1% (ipotesi nulla: la probabilità di avere almeno 2 clienti giornalieri non è aumentata) e applicarlo ai dati in esame.
   Applichiamo uno Z-test (approssimato) per la probabilità p dell'evento "almeno 2 clienti", caso grandi campioni, ipotesi nulla H_0: $p \leq p_0 = 0.801$. La statistica di test è
   $$Z = \sqrt{n} \frac{\bar{Y} - p_0}{\sqrt{p_0(1-p_0)}}$$
   e la regione critica è $C = \{Z > q_{0.99}\}$. Il dato $\bar{y} = 0.87$ fornisce il valore per la statistica di test $z = 3.03$, che cade nella regione critica. Quindi rifiutiamo H_0: c'è evidenza di una maggiore probabilità di avere almeno 2 clienti giornalieri. Il p-value è 0.0013, quindi c'è evidenza per quasi ogni ragionevole livello.