Esame Scritto del Quinto Appello

Tempo a disposizione: 2 ore

Riportare la matricola **all'inizio** di ogni foglio. La soluzione di ogni esercizio deve iniziare in una nuova pagina ed essere scritta in modo chiaro e ordinato: non verrà valutata se la calligrafia è illeggibile. Si riporti 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. Non è permesso l'uso di dispositivi elettronici ad esclusiva eccezione di una calcolatrice non programmabile.

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

   (a) Lo stimatore $\frac{1}{n-1}\sum_i (X_i - \bar{X})^2$ ha denominatore $n-1$ per far sì che esso sia consistente.

   FALSO: il denominatore è scelto in modo che lo stimatore sia corretto, e quest'ultimo sarebbe consistente anche con denominatore n per la Legge dei Grandi numeri e il fatto che $\frac{n-1}{n} \to 1$ per $n \to \infty$.

   (b) Due eventi A e B di probabilità positiva sono indipendenti se e solo se $P(A \mid B) = P(B)$.

   FALSO: A e B sono indipendenti se e solo se $P(A \mid B) = P(A)$.

   (c) Per una variabile aleatoria X con funzione di ripartizione $F(x) = P(X \leq x)$, abbiamo $P(a < X \leq b) = \int_a^b F(x)dx$.
   FALSO: $P(a < X \leq b) = F(b) - F(a)$.

   (d) Un intervallo di fiducia ha come estremi due variabili aleatorie e può non contenere il parametro stimando.
   VERO: un intervallo di fiducia è dato da una coppia di variabili aleatorie $I = [a(\omega), b(\omega)]$ date da opportune funzioni del campione aleatorio, e contiene il parametro θ con una certa probabilità, determinata dal livello di fiducia.

   (e) Imporre il livello di un test statistico fissa un limite alla probabilità di commettere errori di prima specie, ma non dà vincoli su quelli di seconda.
   VERO: la prima parte dell'enunciato è la definizione di livello, mentre va ricordato che è la potenza del test a quantificare la probabilità di commettere errori di seconda specie.

   (f) Considerando per fissare le idee lo Z-test, all'aumentare della numerosità del campione la regione di accettazione si riduce, ossia è più facile che l'ipotesi nulla sia rifiutata.
   VERO: l'ampiezza della regione di accettazione è $2\sigma q_{1-\alpha/2}/\sqrt{n}$, dunque è decrescente nella numerosità n.

2. La velocità del vento in un certo punto è descritta da un vettore aleatorio (X_1, X_2) in cui X_1, X_2 sono variabili i.i.d. $N(0,1)$. Di conseguenza, il modulo quadro della velocità $Z = X_1^2 + X_2^2$ ha distribuzione $\chi^2(2)$ (chi-quadro a 2 gradi di libertà) di cui ricordiamo la densità

$$f_Z(t) = \begin{cases} \frac{1}{2}e^{-t/2} & \text{se } t > 0 \\ 0 & \text{se } t \leq 0 \end{cases}$$

(a) Calcolare media e varianza di Z.
Usando la definizione,

$$\mathbb{E}[Z] = \int_0^\infty \frac{t}{2} e^{-t/2} dt = 2,$$
$$\mathrm{Var}(Z) = \int_0^\infty \frac{(t-2)^2}{2} e^{-t/2} dt = \int_0^\infty \frac{t^2}{2} e^{-t/2} dt - 2\int_0^\infty t e^{-t/2} dt + 2\int_0^\infty e^{-t/2} dt$$
$$= 8 - 8 + 4 = 4.$$

(b) Si consideri ora una variabile aleatoria R con densità

$$f_R(t) = \begin{cases} te^{-t^2/2} & \text{se } t > 0 \\ 0 & \text{se } t \leq 0 \end{cases}$$

Calcolandone la densità, mostrare che la variabile R^2 ha la stessa distribuzione di Z (di conseguenza R ha la distribuzione di $\sqrt{Z}$, il modulo della velocità del vento).
Ricorrendo alla formula di cambio di variabile, poichè $x \mapsto x^2$ è bigettiva e differenziabile sul supporto della p.d.f. $t > 0$, con inversa $x \mapsto \sqrt{x}$,

$$f_{R^2}(y) = f_R(\sqrt{y})|\frac{d}{dy}\sqrt{y}| = \sqrt{y}e^{-(\sqrt{y})^2/2}\frac{1}{2\sqrt{y}} = \frac{1}{2}e^{-y/2}, \quad y > 0.$$

(c) Si calcoli il valore atteso del modulo della velocità del vento $\sqrt{Z}$.
Per quanto visto sopra, e ricordando che variabili equidistribuite hanno tutti i momenti uguali, vale

$$\mathbb{E}[\sqrt{Z}] = \mathbb{E}[R] = \int_0^\infty t \cdot te^{-t^2/2} dt = \sqrt{\frac{\pi}{2}},$$

in cui l'ultimo passaggio segue dal fatto che una Gaussiana standard ha varianza 1.

3. Si vuole misurare la temperatura (in gradi Celsius) di un certo solido. Si sa che l'errore di misurazione del termometro usato ha distribuzione gaussiana di media pari alla temperatura del solido e di deviazione standard pari a 1 grado Celsius. Su 100 misurazioni, risulta una temperatura media di 32.5 gradi Celsius.

(a) Fornire un intervallo di fiducia, di livello 0.95, per la temperatura del solido.
Stiamo cercando un intervallo di fiducia per la media di una popolazione gaussiana, deviazione standard nota $\sigma = 1$. L'invervallo di fiducia cercato è ($n = 16$, $\alpha = 0.05$)

$$\left[\bar{X} \pm \frac{\sigma}{\sqrt{n}} q_{1-\alpha/2}\right] = \left[\bar{X} \pm 0.196\right].$$

Inserendo il valore $\bar{x} = 32.5$, troviamo $[32.304, 32.696]$.

(b) I dati forniscono evidenza *contro* l'ipotesi che il solido abbia temperatura *non superiore* a 32 gradi? Formulare un opportuno test di ipotesi di livello 0.01 e applicarlo ai dati del campione.
Siamo in presenza di un test di ipotesi sulla media m di una popolazione gaussiana,

con deviazione standard $\sigma = 1$ nota. L'ipotesi nulla è $H_0 : m \leq 32(= m_0)$, contro $H_1 : m > 32$. La regione critica di livello $\alpha = 0.01$ è ($n = 100$)

$$C = \{\frac{\sqrt{n}}{\sigma}(\bar{X} - m_0) > q_{1-\alpha}\} = \{(\bar{X} - 32) > 0.233\}$$

Applichiamo il test al dato $\bar{x} = 32.5$: i dati cadono nella regione critica C, quindi c'è evidenza, a livello 0.01, per una temperatura superiore a 32 gradi.
In alternativa, si può calcolare il p-value relativo a $\bar{x} = 32.5$, che è (chiamando Z una v.a. normale standard)

$$\bar{\alpha} = P\left(Z > \frac{\sqrt{n}}{\sigma}(\bar{x} - m_0)\right) = 1 - \Phi(5) \approx 0,$$

quindi c'è evidenza a ogni livello ragionevole per una temperatura superiore a 32 gradi.

(c) Supponiamo ora di avere un nuovo termometro, le cui misurazioni abbiano distribuzione gaussiana, con media sempre pari alla temperatura del solido e con deviazione standard pari a 0.5 (contro 1 del precedente). Con questo nuovo termometro, possiamo fare solo 50 misurazioni (contro 100 del precedente). Quale dei due termometri produce una stima più precisa? cioè, fissato un livello di fiducia, quale dei due termometri produce un intervallo di fiducia più stretto (cioè con minore ampiezza)?
La semi-ampiezza di un intervallo di fiducia di livello α è $\frac{\sigma}{\sqrt{n}} q_{1-\alpha/2}$. Quindi, nel caso delle misurazioni con il termometro originario ($\sigma = 1$, $n = 100$), la semi-ampiezza è $0.1 \cdot q_{1-\alpha/2}$, mentre nel caso delle misurazioni con il nuovo termometro ($\sigma = 0.5$, $n = 50$), la semi-ampiezza è $0.071 \cdot q_{1-\alpha/2}$, inferiore a quella del termometro originario. La stima con il nuovo termometro è dunque più precisa.

Valori numerici utilizzabili:

$$q_{0.95} = 1.64, \quad q_{0.975} = 1.96, \quad q_{0.99} = 2.33, \quad q_{0.995} = 2.58,$$
$$t_{0.95,99} = 1.66, \quad t_{0.975,99} = 1.98, \quad t_{0.99,99} = 2.36, \quad t_{0.995,99} = 2.63.$$