Esame Scritto del Secondo Appello

Tempo a disposizione: 2 ore

La soluzione di ogni esercizio deve essere scritta in modo chiaro e ordinato, e può non essere valutata se la calligrafia è illeggibile. 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. 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. Un call center riceve chiamate da due regioni, che chiamiamo A e B. Il numero di chiamate al minuto dalla regione A ha distribuzione di Poisson di parametro 2, mentre il numero di chiamate al minuto dalla regione B ha distribuzione di Poisson di parametro 3. Supponiamo che il numero di chiamate da A e il numero di chiamate da B siano indipendenti.

   (a) Calcolare la probabilità che, in un dato minuto, il call center riceva almento 2 chiamate da A.
   Detto X il numero di chiamate in un minuto da A, X ha distribuzione di Poisson di parametro 2 e la probabilità cercata è

   $$\mathbb{P}(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - (\frac{2^0}{0!} + \frac{2^1}{1!})e^{-2} \approx 0.594.$$

   (b) Calcolare la probabilità che, in un dato minuto, il call center riceva almeno 2 chiamate *in totale*.
   Siano X e Y il numero di chiamate ricevute rispettivamente da A e da B. Le v.a. X e Y sono indipendenti e hanno distribuzione di Poisson, quindi $X + Y$ ha distribuzione di Poisson di parametro $2 + 3 = 5$. La probabilità cercata è

   $$P(X + Y \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - (\frac{5^0}{0!} + \frac{5^1}{1!})e^{-5} \approx 0.96.$$

   (c) Supponiamo che il numero di chiamate dalla regione A in un dato minuto sia indipendente dal numero di chiamate da A in altri minuti. Calcolare (almeno in modo approssimato) la probabilità che, in un dato intervallo di 2 ore, il call center riceva almeno 260 chiamate da A.
   Sia X_i il numero di chiamate da A nel minuto i-simo, $i = 1, \dots n = 120$; le X_i sono i.i.d. (di momento secondo finito), con $\mathbb{E}[X_i] = \mathrm{Var}(X_i) = 2$. Per il teorema centrale del limite (considerando $n = 120$ abbastanza grande), la probabilità cercata è circa

   $$\mathbb{P}(\sum_i X_i \geq 260) = \mathbb{P}(\frac{\sum_i X_i - 120 \cdot 2}{\sqrt{120 \cdot 2}} \geq \frac{260 - 120 \cdot 2}{\sqrt{120 \cdot 2}} \approx \mathbb{P}(Z \geq 1.29) = 1 - \Phi(1.29) \approx 0.099.$$

2. In uno spazio di probabilità $(\Omega, \mathcal{E}, \mathbb{P})$ si considerano tre eventi A, B, C a due a due indipendenti, di probabilità non nulla e tali che $A \cap B \cap C = \emptyset$.

(a) A, B e C possono essere globalmente indipendenti? (se sì trovare un esempio, se no dimostrarlo).
I tre eventi non possono essere globalmente indipendenti: infatti $0 = \mathbb{P}(\emptyset) \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$ (essendo tutti e tre gli eventi a probabilità non nulla).

(b) Se A, B, C hanno la stessa probabilità, si determini il valore massimo di $\mathbb{P}(A)$ e si esibisca un esempio di spazio di probabilità con eventi A, B, C aventi le proprietà suddette per cui $\mathbb{P}(A)$ è massima.
Indichiamo $x = \mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C)$. Per l'indipendenza a coppie degli eventi si ha:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) = x^2, \quad \text{e similmente} \quad \mathbb{P}(A \cap C) = \mathbb{P}(B \cap C) = x^2.$$

Poichè $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\emptyset) = 0$, vale

$$x = \mathbb{P}(A) \geq \mathbb{P}(A \cap B) \cup \mathbb{P}(A \cap C) = 2x^2,$$

da cui si ha $x \leq 1/2$. Mostriamo con un esempio che $1/2$ è il valore massimo: scegliamo $\Omega = \{1, 2, 3, 4\}$, $\mathcal{E} = \mathcal{P}(\Omega)$, $\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \mathbb{P}(\{3\}) = \mathbb{P}(\{4\}) = 1/4$, $A = \{1, 2\}$, $B = \{2, 3\}$, $C = \{1, 3\}$. Si verifica facilmente che l'esempio soddisfa tutte le richieste. Osserviamo che l'esempio proposto coincide con quello riportato sulle dispense per mostrare che indipendenza a coppie non implica l'indipendenza globale.

3. Siano $a < b$ due numeri reali: si consideri la funzione

$$F_{a,b}(x) = \begin{cases} 0 & x \leq a, \\ \frac{2}{\pi} \arcsin\left(\sqrt{\frac{x-a}{b-a}}\right) & a < x < b, \\ 1 & b < x. \end{cases}$$

Ricordiamo che $\frac{d}{dx} \arcsin(\sqrt{x}) = \frac{1}{2\sqrt{x(1-x)}}$.

(a) Si mostri che F è la funzione di ripartizione di una variabile aleatoria con densità, e si calcoli la funzione di densità di probabilità associata $f_{a,b}$. Si disegnino approssimativamente i grafici delle due funzioni.
Bisogna controllare che F così definita sia non-decrescente, e che assuma i valori limite 0 e 1 per $x \to \pm\infty$. La seconda richiesta è banale, mentre la non-decrescenza segue dal fatto che tra a e b la funzione è strettamente crescente perchè composizione di funzioni strettamente crescenti, fuori da (a, b) è costante, ed inoltre i limiti per $x \downarrow a$ e $x \uparrow b$ della funzione che definisce $F_{a,b}$ tra a e b sono rispettivamente 0 e 1. In particolare, $F_{a,b}$ è anche continua, ed è differenziabile in ogni punto ad eccezione di a, b: una variabile aleatoria con c.d.f. $F_{a,b}$ ha dunque densità data da

$$f_{a,b}(x) = F'_{a,b}(x) = \begin{cases} 0 & x \leq a, \\ \frac{1}{\pi\sqrt{(x-a)(b-x)}} & a < x < b, \\ 0 & b < x. \end{cases}$$

(b) Si mostri che una variabile aleatoria con densità $f_{a,b}$ ha valore atteso $\frac{a+b}{2}$.
Suggerimento: possono essere utili le sostituzioni $x = a + (b - a)t$ e $t = \sin^2(u)$.
Usando la sostituzione proposta,

$$\mathbb{E}[X] = \frac{1}{\pi} \int_a^b \frac{x dx}{\sqrt{(x-a)(b-x)}} = \frac{1}{\pi} \int_0^1 \frac{a + (b-a)t}{\sqrt{t(1-t)}} dt.$$

Calcoliamo separatamente i due seguenti integrali:

$$\int_0^1 \frac{dt}{\sqrt{t(1-t)}} = \int_0^{\pi/2} \frac{2\sin(u)\cos(u)}{\sqrt{\sin^2(u)\left(1-\sin^2(u)\right)}}du = \int_0^{\pi/2} 2du = \pi,$$

in cui abbiamo sostituito $t = \sin^2(u)$, e con la stessa sostituzione si ha

$$\int_0^1 \frac{tdt}{\sqrt{t(1-t)}} = \int_0^{\pi/2} \frac{2\sin^3(u)\cos(u)}{\sqrt{\sin^2(u)\left(1-\sin^2(u)\right)}}du = \int_0^{\pi/2} 2\sin^2(u)du$$

$$= \int_0^{\pi/2} (1-\cos(2u))du = [u - \sin(u)\cos(u)]_0^{\pi/2} = \frac{\pi}{2}.$$

Combinando i risultati con la formula per $\mathbb{E}[X]$ ottenuta sopra si ha la tesi.

(c) Sia X una variabile aleatoria con densità $f_X = f_{a,b}$, siano $c, d \in \mathbb{R}$ fissati. Si determini se la variabile aleatoria $Y = cX + d$ ha densità, e in caso positivo la si calcoli.
Anzitutto, osserviamo che se $c = 0$ la variabile $Y = d$ è costante, dunque è una variabile discreta.
Consideriamo ora il caso $c \neq 0$. La variabile X assume valori nell'intervallo (a, b), e $h(x) = cx + d$ (essendo $c \neq 0$) è una bigezione differenziabile $h : (a, b) \to (ca + d, cb + d)$, con inversa $h^{-1}(y) = \frac{y-d}{c}$. Si applica quindi la formula di cambio di variabile: f_Y è nulla fuori da $(ca + d, cb + d)$ e per y in tale intervallo vale

$$f_Y(y) = f_X(h^{-1}(y))\left|\frac{dh^{-1}}{dy}(y)\right| = \frac{1/c}{\pi\sqrt{(\frac{y-d}{c}-a)(b-\frac{y-d}{c})}} = \frac{1}{\pi\sqrt{(y-ca+d)(cb+d-y)}}.$$

In particolare, Y ha densità $f_Y = f_{ca+d,cb+d}$.

4. Un'azienda afferma che il tempo medio di risposta del servizio clienti (a una richiesta) è inferiore ai 30 secondi. Su un campione di 50 richieste, le risposte hanno una media di 28 secondi. Supponiamo che la distribuzione dei tempi di risposta sia gaussiana con deviazione standard 4 secondi.

(a) Determinare un intervallo di fiducia, di livello 0.99, per il tempo medio di risposta.
Cerchiamo un intervallo di fiducia per la media di una popolazione gaussiana di deviazione standard $\sigma = 0.4$ nota. L'intervallo cercato è ($n = 50$, $\alpha = 0.01$, $q_{0.995} = 2.58$)

$$\left[\bar{X} \pm \frac{\sigma}{\sqrt{n}} q_{1-\alpha/2}\right] = \left[\bar{X} \pm 1.46\right].$$

Con il dato $\bar{x} = 28$, otteniamo l'intervallo numerico

$$[26.54, 29.46].$$

(b) I dati forniscono evidenza *a favore* dell'affermazione dell'azienda? 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 $\sigma = 0.4$ nota; la taglia del campione è $n = 50$. Poiché vogliamo verificare l'evidenza o meno *per* l'affermazione dell'azienda, l'ipotesi nulla è

$H_0 : m \geq 30 (= m_0)$ contro $H_1 : m < 30$. La regione critica di livello $\alpha = 0.01$ ($q_{0.01} = -q_{0.99} = -2.33$) è

$$C = \{\frac{\sqrt{n}}{\sigma}(\bar{X} - m_0) < q_\alpha\} = \{(\bar{X} - 30) < -1.32\}$$

Applichiamo il test al dato $\bar{x} = 28$: i dati cadono nella regione critica C, quindi c'è evidenza, a livello 99%, per l'affermazione dell'azienda.

(c) Applichiamo ora un test di livello α (con gli stessi dati) e supponiamo che, come risultato, non ci sia evidenza *a favore* dell'affermazione dall'azienda (tempo medio inferiore ai 30 secondi). Che cosa possiamo dire del livello α del test?
Poiché accettiamo l'ipotesi nulla $H_0 : m \geq 30$, possiamo dire che $\alpha < \bar{\alpha}$, con $\bar{\alpha}$ il p-value dei risultati. Tale p-value, per $\bar{x} = 28$, si approssima con

$$\mathbb{P}(Z < \frac{\sqrt{n}}{\sigma}(\bar{x} - m_0)) \simeq \mathbb{P}(Z < -3.53) \simeq 0,$$

perchè la Gaussiana standard assegna probabilità numericamente trascurabile al complementare dell'intervallo $[-3, 3]$. Quindi il risultato ha un'altissima significatività statistica a favore dell'affermazione dell'azienda. Di conseguenza, se il test di livello α conclude che non c'è evidenza a favore dell'affermazione dell'azienda, α deve essere estremamente piccolo (condizione poco realistica).

Valori numerici utilizzabili:

$\Phi(0) = 0.5$	$\Phi(1) = 0.8413$	$\Phi(1.29) = 0.9015$	$\Phi(2) = 0.9772$	$\Phi(3) = 0.9987$
$q_{0.9} = 1.282$	$q_{0.95} = 1.645$	$q_{0.975} = 1.960$	$q_{0.99} = 2.326$	$q_{0.995} = 2.576$