Esame Scritto del Terzo 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. Per ognuna delle seguenti affermazioni si determini se essa è VERA oppure FALSA, dimostrandola nel primo caso e fornendo un controesempio nel secondo.

   (a) In uno spazio di probabilità, due eventi disgiunti sono indipendenti.
   FALSO: se A, B sono disgiunti e indipendenti,
   $$\mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A \cap B) = \mathbb{P}(\emptyset) = 0$$
   è soddisfatta se e solo se uno dei due eventi è esso stesso vuoto, per cui l'affermazione non è verificata in generale.

   (b) Una variabile aleatoria che assume infiniti valori è una variabile con densità.
   FALSO: ad esempio una variabile di Poisson può assumere qualsiasi valore naturale $(0, 1, 2, \dots)$ ma non è una variabile con densità.

   (c) I momenti $\mathbb{E}[X^n]$ di una variabile aleatoria X di Bernoulli di parametro p sono tutti uguali.
   VERO: infatti vale sempre $X^n = X$, perchè se $X = 0$ allora $X^n = 0^n = 0$, e similmente se $X = 1$ allora $X^n = 1^n = 1$, di conseguenza
   $$\mathbb{E}[X^n] = 1^n\mathbb{P}(X^n = 1) + 0^n\mathbb{P}(X^n = 0) = 1\mathbb{P}(X = 1) + 0\mathbb{P}(X = 0) = \mathbb{E}[X].$$

2. Un certo tipo di pezzi di ricambio viene prodotto da due aziende, che chiamiamo α e β. L'azienda α produce il 70% dei lotti di tali pezzi e l'azienda β il restante 30%. Un pezzo prodotto da α ha probabilità 10% di essere difettoso, mentre un pezzo prodotto da β ha probabilità 5% di essere difettoso. La presenza o meno di difetti per un certo pezzo è indipendente dalla presenza o meno di difetti in altri pezzi.

   (a) Da un lotto di pezzi prodotti da α, estraiamo 5 pezzi. Qual è la probabilità che almeno 2 pezzi siano difettosi? Come cambia la risposta se il lotto è prodotto da β?
   Sia X la v.a. che conta il numero di pezzi difettosi estratti. Sapendo che il lotto viene da α (cioè sotto $P(\cdot \mid \alpha)$), X ha distribuzione binomiale di parametri 5, 0.1. La probabilità cercata è quindi
   $$\begin{aligned} P(X \geq 2 \mid \alpha) &= 1 - P(X = 0 \mid \alpha) - P(X = 1 \mid \alpha) \\ &= 1 - \binom{5}{0} \cdot 0.9^5 - \binom{5}{1} \cdot 0.1 \cdot 0.9^4 = 0.08146. \end{aligned}$$

Se invece il lotto viene da β, X ha distribuzione binomiale di parametri 5, 0.05 e la probabilità cercata è

$$\begin{aligned}P(X \geq 2 \mid \beta) &= 1 - P(X = 0 \mid \beta) - P(X = 1 \mid \beta)\\ &= 1 - \binom{5}{0} \cdot 0.95^5 - \binom{5}{1} \cdot 0.05 \cdot 0.95^4 = 0.0225925.\end{aligned}$$

(b) Scegliamo ora un lotto a caso ed estraimo 5 pezzi. Supponiamo di trovarne almeno 2 difettosi. Qual è la probabilità che il lotto scelto venga da α?
Usiamo il teorema di Bayes:

$$P(\alpha \mid X \geq 2) = \frac{P(X \geq 2 \mid \alpha)P(\alpha)}{P(X \geq 2 \mid \alpha)P(\alpha) + P(X \geq 2 \mid \beta)P(\beta)} = 0.8938,$$

dove abbiamo usato $P(\alpha) = 0.7$.

(c) Nella situazione del punto precedente (scegliamo un lotto a caso, estraiamo 5 pezzi, ne troviamo almeno 2 difettosi), se estraiamo un altro pezzo (diverso dai 5 già estratti), qual è la probabilità che questo sia difettoso?
Sia $Y = 1$ se il nuovo pezzo è difettoso, $= 0$ altrimenti. Usiamo il lemma della partizione:

$$\begin{aligned}P(Y = 1 \mid X \geq 5) &= P(Y = 1 \mid \alpha, X \geq 5)P(\alpha \mid X \geq 5) + P(Y = 1 \mid \beta, X \geq 5)P(\beta \mid X \geq 5)\\ &= P(Y = 1 \mid \alpha)P(\alpha \mid X \geq 5) + P(Y = 1 \mid \beta)(1 - P(\alpha \mid X \geq 5))\\ &= 0.1 \cdot 0.8938 + 0.05 \cdot (1 - 0.8938) = 0.0950;\end{aligned}$$

abbiamo usato che $P(Y = 1 \mid \alpha, X \geq 5) = P(Y = 1 \mid \alpha) = 0.1$ poiché, dato α, X e Y sono indipendenti, e analogamente dato β.

3. Si consideri la seguente funzione,

$$f_\sigma(x) = \frac{1}{2\sigma} \exp\left(-\frac{|x|}{\sigma}\right),$$

dipendente da un parametro $\sigma > 0$.

(a) Si mostri che f_σ è una densità di probabilità e che la funzione di ripartizione relativa è data da

$$F_X(t) = \begin{cases} e^{t/\sigma}/2 & \text{per } t \leq 0 \\ 1 - e^{-t/\sigma}/2 & \text{per } t > 0 \end{cases}.$$

La funzione è chiaramente positiva, per mostrare che è una densità di probabilità basta mostrare che integra ad 1:

$$\frac{1}{2\sigma} \int_{-\infty}^{\infty} \exp\left(-\frac{|x|}{\sigma}\right) dx = \frac{1}{2} \int_{-\infty}^{\infty} e^{-|y|} dy = \int_0^{\infty} e^{-y} dy = 1,$$

in cui abbiamo usato la sostituzione $y = x/\sigma$. La c.d.f. si calcola in modo simile: partendo dall'uguaglianza

$$F_X(t) = \int_{-\infty}^{t} f_\sigma(x)dx = \frac{1}{2\sigma} \int_{-\infty}^{t} \exp\left(-\frac{|x|}{\sigma}\right) dx,$$

distinguiamo i due casi $t \leq 0$ e $t \geq 0$:

$$(t \leq 0): \qquad \frac{1}{2\sigma}\int_{-\infty}^{t} \exp\left(-\frac{|x|}{\sigma}\right)dx = \frac{1}{2\sigma}\int_{-t}^{\infty} \exp\left(-\frac{x}{\sigma}\right)dx = \frac{1}{2}\int_{-t/\sigma}^{\infty} e^{-y}dy = \frac{e^{t/\sigma}}{2},$$

$$\begin{aligned}(t \geq 0): \qquad \frac{1}{2\sigma}\int_{-\infty}^{t} \exp\left(-\frac{|x|}{\sigma}\right)dx &= \frac{1}{2\sigma}\int_{-\infty}^{0} \exp\left(-\frac{|x|}{\sigma}\right)dx + \frac{1}{2\sigma}\int_{0}^{t} \exp\left(-\frac{x}{\sigma}\right)dx \\ &= \frac{1}{2\sigma}\int_{0}^{\infty} \exp\left(-\frac{x}{\sigma}\right)dx + \frac{1}{2\sigma}\int_{0}^{t} \exp\left(-\frac{x}{\sigma}\right)dx \\ &= \frac{1}{2}\int_{0}^{\infty} e^{-y}dy + \frac{1}{2}\int_{0}^{t/\sigma} e^{-y}dy = \frac{2-e^{-t/\sigma}}{2}.\end{aligned}$$

(b) Si mostri che una variabile aleatoria con densità f_σ ha valore atteso 0 e varianza $2\sigma^2$.
Valore atteso e varianza sono calcolati con la stessa sostituzione di cui sopra: il valore atteso è dato da

$$\mathbb{E}[X] = \frac{1}{2\sigma}\int_{-\infty}^{\infty} xe^{-\frac{|x|}{\sigma}}dx = \frac{1}{2}\int_{-\infty}^{\infty} \sigma y e^{-|y|}dy = \frac{\sigma}{2}\int_{-\infty}^{\infty} ye^{-|y|}dy$$

in cui l'ultimo integrale è nullo perché l'integranda è una funzione dispari. Per la varianza,

$$\mathrm{Var}(X) = \frac{1}{2\sigma}\int_{-\infty}^{\infty} x^2e^{-\frac{|x|}{\sigma}}dx = \frac{\sigma^2}{2}\int_{-\infty}^{\infty} y^2e^{-|y|}dy = \sigma^2\int_{0}^{\infty} y^2e^{-y}dy = 2\sigma^2$$

in cui il calcolo dell'integrale al membro di destra è una semplice integrazione per parti.

(c) Data X una variabile aleatoria con densità f_σ, si mostri che la variabile $Z = |X|$ è esponenziale di parametro $1/\sigma$.

Osserviamo anzitutto che la funzione $x \mapsto x^2$ non è bigettiva su $\mathbb{R}$ (ovvero l'insieme dei valori che può assumere X), dunque non si può applicare la formula di cambio di variabile.
Calcoliamo la funzione di ripartizione. Osserviamo anzitutto che se $t < 0$ allora $F_Z(t) = \mathbb{P}(Z \leq t) = \mathbb{P}(|X| \leq t) = 0$, perchè $|X|$ non può mai essere negativo. Per $t \geq 0$ calcoliamo

$$\begin{aligned}F_Z(t) &= \mathbb{P}(Z \leq t) = \mathbb{P}(|X| \leq t) = \mathbb{P}(-t \leq X \leq t) \\ &= \mathbb{P}(X \leq t) - \mathbb{P}(X \leq -t) = F_X(t) - F_X(-t) \\ &= 1 - e^{-t/\sigma}/2 - e^{-t/\sigma}/2 = 1 - e^{-t/\sigma}\end{aligned}$$

in cui abbiamo usato la formula trovata sopra per F_X.

4. Un'azienda afferma che almeno il 60% dei suoi clienti fa più di un acquisto l'anno. Per verificare questa affermazione, si considera un campione di 300 clienti, trovando che 175 di loro hanno effettuato più di un acquisto in un anno.

(a) Fornire un intervallo di fiducia di livello 95% per la percentuale di clienti con più di un acquisto l'anno.
Siamo nel caso di popolazione Bernoulli di parametro p, con campione grande. Usiamo dunque l'approssimazione normale: detta $\bar{X}$ la frequenza relativa campionaria (dei clienti con più di un acquisto l'anno), l'intervallo di fiducia cercato è (con $1-\alpha = 0.95$)

$$[\bar{X} \pm q_{1-\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}] = [\bar{X} \pm 1.96\sqrt{\frac{\bar{X}(1-\bar{X})}{300}}]$$

Sostituendo il valore $\bar{x} = 175/300 = 0.5833$, otteniamo l'intervallo numerico $[0.5833 \pm 0.0558]$.

(b) Se vogliamo che l'intervallo di fiducia (sempre a livello 95%) abbia precisione (cioè metà della sua ampiezza) al massimo del 2%, quando grande deve essere il campione?
La semi-ampiezza dell'intervallo di fiducia è $q_{1-\alpha/2}\sqrt{\bar{X}(1-\bar{X})/n}$. Poiché essa non è nota a priori (prima di effettuare le misurazioni), usando $\max_{x\in[0,1]} x(1-x) = 1/4$ possiamo stimare la semi-ampiezza con

$$q_{1-\alpha/2}\frac{1}{2\sqrt{n}}.$$

Imponiamo quindi che la quantità sopra sia ≤ 0.02 e troviamo

$$n \geq \frac{q^2_{1-\alpha/2}}{2 \cdot 0.02^2} = 4802.$$

(c) Formulare un test di ipotesi, a livello 1%, per verificare se l'affermazione dell'azienda è plausibile o meno, e applicarlo ai dati del problema (175 clienti con più di un acquisto, sul campione di 300 clienti).
L'ipotesi nulla (affermazione dell'azienda) è $H_0 : p \geq 0.6 =: p_0$, contro $H_1 : p < 0.6$. Siamo quindi nel caso di test unilatero di livello $\alpha = 0.01$ per media di popolazione Bernoulli, con campione grande. Detta $\bar{X}$ la frequenza relativa campionaria, la regione critica è quindi

$$C = \{\sqrt{n}\frac{\bar{X} - p_0}{\sqrt{p_0(1-p_0)}} < -q_{1-\alpha}\} = \{\sqrt{300}\frac{\bar{X} - 0.6}{\sqrt{0.6 * 0.4}} < -2.33\}.$$

Sostituendo il valore $\bar{x} = 0.5833$, troviamo

$$\sqrt{300}\frac{\bar{X} - 0.6}{\sqrt{0.6 \cdot 0.4}} = -0.5904 > -2.33$$

quindi accettiamo H_0: l'affermazione dell'azienda è plausibile a livello 0.01. Possiamo anche calcolare il p-value (chiamiamo Z una v.a. normale standard):

$$P(Z < -0.5904) = 1 - \Phi(0.5904) = 0.2776,$$

(che risulta più alto del livello, a conferma che accettiamo H_0) ottenendo un valore piuttosto alto: l'affermazione dell'azienda sembra abbastanza plausibile in generale.

Valori numerici utilizzabili:

$$\Phi(0.59) = 0.7224, \quad \Phi(1.18) = 0.8810, \quad \Phi(1.77) = 0.9616, \quad \Phi(2.36) = 0.9908,$$
$$q_{0.95} = 1.64, \quad q_{0.975} = 1.96, \quad q_{0.99} = 2.33, \quad q_{0.995} = 2.58,$$
$$t_{60,0.95} = 1.67, \quad t_{60,0.975} = 2.00, \quad t_{60,0.99} = 2.39, \quad t_{60,0.995} = 2.66$$