Esame Scritto del Quinto 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. 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.
In ROSSO si riportano alcuni errori comuni riscontrati nella correzione.

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

   (a) Il mio telecomando usa due batterie, è un po' che non accendo il televisore e ognuna delle due batterie potrebbe essere scarica (indipendentemente dall'altra) con probabilità 1/2. Provo ad accendere il televisore ed il telecomando funziona ancora, quindi almeno una batteria è carica. La probabilità che lo siano entrambe è 1/3.
   VERO: Siano A, B gli eventi indipendenti "la prima batteria è carica" e "la seconda batteria è carica".

$$\mathbb{P}(\text{“le batterie sono entrambe cariche”}) = \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) = 1/4,$$
$$\mathbb{P}(\text{“almeno una batteria è carica”}) = \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = 3/4,$$

   e la probabilità cercata è

$$\mathbb{P}(A \cap B \mid A \cup B) = \frac{\mathbb{P}(A \cap B \cap (A \cup B))}{\mathbb{P}(A \cup B)} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A \cup B)} = \frac{1}{3}.$$

   La quasi totalità dei candidati non ha tenuto conto dell'informazione nota come condizionamento della probabilità.

   (b) Una variabile aleatoria geometrica di parametro $p \in (0, 1)$ ha tutti i momenti finiti.
   VERO: infatti il momento n-esimo è finito se converge la serie

$$\sum_{k=0}^{\infty} k^n \mathbb{P}(X = k) = p \sum_{k=0}^{\infty} k^n (1-p)^k = p \sum_{k=0}^{\infty} k^n e^{\log(1-p)k},$$

   che è convergente perché $\log(1 - p) < 0$ essendo $0 < 1 - p < 1$, e dunque l'esponenziale negativo prevale su k^n.
   Molti hanno scritto che l'affermazione è vera perché la variabile è limitata, ma questo è FALSO perché una variabile geometrica può prendere valori arbitrariamente grandi.

   (c) Due variabili aleatorie discrete e scorrelate sono indipendenti.
   FALSO: basta prendere X uniforme su $\{-1, 0, 1\}$ e $Y = X^2$, esse sono scorrelate ma non indipendenti.
   Sorprendentemente, molti hanno scritto esplicitamente "variabili scorrelate sono indipendenti", affermazione la cui falsità è stata discussa a lungo a lezione.

(d) Per un campione statistico con legge dipendente da un parametro θ, un intervallo di fiducia è un intervallo numerico $[\theta - d, \theta + d]$ attorno al parametro in cui ci si aspetta di trovare gli esiti del campione con probabilità alta.
FALSO: il parametro θ non è noto, quindi non può essere usato per definire un intervallo di fiducia, che solitamente viene invece centrato in uno stimatore del parametro.

(e) Il prodotto XY di due variabili aleatorie di Bernoulli X e Y è anch'essa una variabile di Bernoulli.
VERO: che siano indipendenti o meno, il prodotto XY può assumere solo valori $0, 1$ perché lo stesso vale per entrambi i fattori, e una variabile aleatoria che assume valori $0, 1$ è di Bernoulli.
Alcuni hanno confuso il risultato sulla somma di variabili di Bernoulli **indipendenti** con questa affermazione, o hanno riportato (senza giustificarlo) il paramentro p della variabile Bernoulli risultante dal prodotto, il quale non si può calcolare senza ipotesi aggiuntive.

(f) Per un test statistico, fissare livello e potenza del test corrisponde a imporre rispettivamente un limite superiore e inferiore alla probabilità di errore di prima specie.
FALSO: fissare il livello corrisponde a porre un limite superiore alla probabilità di errore di prima specie ma *non* è vero che fissare la potenza corrisponde a limitarlo dal basso. In effetti, la potenza misura la probabilità di non commettere un errore di seconda specie, e non viene fissata ma massimizzata.

2. Si svolgano i seguenti calcoli.

(a) Il tempo tra un guasto e il successivo per un grosso macchinario industriale è descritto da una variabile esponenziale di valore atteso 1 anno. Determinare la mediana, ovvero il tempo t tale che il prossimo guasto accadrà prima di t con probabilità $1/2$.
Il valore atteso di una variabile esponenziale di parametro λ è $1/\lambda$, quindi nel nostro caso $\lambda = 1$. Dobbiamo calcolare il valore t per cui

$$\frac{1}{2} = \mathbb{P}(X \leq t) = F_X(t) = 1 - e^{-t},$$

da cui si ricava $t = \log 2 \sim 0.69$ anni.

(b) Il numero di gocce di cioccolato in un biscotto di una nota azienda segue una distribuzione di Poisson. L'azienda desidera che ci siano almeno due gocce in un biscotto con probabilità almeno pari al 99%. Determinare se un valore atteso (del numero di gocce) pari a 6 assicura che questo accada.
Il valore atteso di una variabile di Poisson di parametro λ è λ. Dobbiamo imporre che una variabile di Poisson X soddisfi:

$$0.99 \leq \mathbb{P}(X \geq 2) = 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) = 1 - e^{-\lambda} - \lambda e^{-\lambda},$$

ma per $\lambda = 6$ il membro destro vale 0.982, quindi la condizione proposta non è sufficiente.

(c) Xavier e Yannick hanno un dado a sei facce a testa, e li lanciano finchè non esce il 6. Calcolare la probabilità che impieghino lo stesso numero di lanci.
Indichiamo con X, Y le variabili aleatorie che contano il numero di lanci necessari rispettivamente a Xavier e Yannick per vedere il primo 6. Sono variabili geometriche indipendenti di parametro $p = 1/6$. L'evento "servono lo stesso numero di lanci" si riscrive come unione disgiunta infinita:

$$\{X = Y\} = \{X = 1, Y = 1\} \cup \{X = 2, Y = 2\} \cup \{X = 3, Y = 3\} \cup \dots,$$

quindi calcoliamo

$$\mathbb{P}(X=Y)=\sum_{n=1}^{\infty}\mathbb{P}(X=n,Y=n)=\sum_{n=1}^{\infty}\mathbb{P}(X=n)\mathbb{P}(Y=n)$$
$$=\sum_{n=1}^{\infty}(1-p)^{n-1}p\cdot(1-p)^{n-1}p=p^2\sum_{n=0}^{\infty}(1-p)^{2n}=\frac{p^2}{1-(1-p)^2}=\frac{1/36}{1-25/36}=\frac{1}{11}\sim 0.09.$$

3. Si terrà un referendum e viene effettuato un sondaggio per stimare la percentuale di persone per il SÌ. Su 800 intervistati 420 dichiarano che voteranno SÌ.

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

$$[\bar{X}\pm q_{1-\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}]=[\bar{X}\pm 1.96\sqrt{\frac{\bar{X}(1-\bar{X})}{800}}]$$

   Sostituendo il valore $\bar{x}=420/800=0.525$, otteniamo l'intervallo numerico $[0.525\pm 0.0346]$.

   (b) Formulare un test per l'ipotesi nulla "almeno il 50% voterà NO" e applicarlo ai dati del campione con livello 0.01.
   L'ipotesi nulla è $H_0: p\le 0.5 =: p_0$, contro $H_1: p>0.5$. 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

$$C=\{\sqrt{n}\frac{\bar{X}-p_0}{\sqrt{p_0(1-p_0)}}<-q_{1-\alpha}\}=\{\sqrt{800}\frac{\bar{X}-0.5}{0.5}>2.33\}.$$

   Sostituendo il valore $\bar{x}=0.525$, troviamo

$$\sqrt{800}\frac{\bar{X}-0.5}{0.5}=1.41<2.33$$

   quindi accettiamo H_0: non c'è evidenza che più della metà della popolazione sia favorevole. Possiamo anche calcolare il p-value (chiamiamo Z una v.a. normale standard):

$$P(Z>1.41)=1-\Phi(1.41)=0.0793,$$

   che risulta più alto del livello 0.01, a conferma che accettiamo H_0.

   (c) Se vogliamo che l'intervallo di fiducia (per la percentuale di favorevoli), sempre a livello 0.95, abbia semiampiezza al massimo del 3%, 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 ≤ 0.03 e troviamo

$$n\ge\left(\frac{q_{1-\alpha/2}}{2\cdot 0.03}\right)^2=1067.11,$$

   quindi $n\ge 1068$.