Esame Scritto del Quarto 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.

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

   (a) La funzione di ripartizione

   $$F(t) = \begin{cases} 0 & t < 0 \\ 1/2 & 0 \le t < 1 \\ 1 & t > 1 \end{cases},$$

   essendo differenziabile ovunque ad eccezione dei punti $t = 0, 1$, è la funzione di ripartizione di una variabile con densità data da $F'(t)$.
   FALSO: le funzioni di ripartizione di variabili con densità sono continue, mentre la funzione F proposta ha due discontinuità di salto in $t = 0, 1$.

   (b) Se due variabili aleatorie di Poisson X, Y sono equidistribuite (cioè hanno la stessa legge), hanno lo stesso parametro λ.
   VERO: indicando con λ, λ' i parametri di X, Y, per definizione di equidistribuzione, per ogni intero $k \ge 0$,

   $$e^{-\lambda} \frac{\lambda^k}{k!} = \mathbb{P}(X = k) = \mathbb{P}(Y = k) = e^{-\lambda'} \frac{(\lambda')^k}{k!},$$

   da cui $e^{\lambda - \lambda'} = (\lambda/\lambda')^k$, che per $k = 0$ implica $e^{\lambda-\lambda'} = 1$ e dunque $\lambda = \lambda'$.

   (c) Data una variabile aleatoria non-negativa X, la successione di variabili aleatorie $X_n = X/n$ converge a 0 in probabilità.
   VERO: Poiché la successione di eventi $A_n = \{X > n\epsilon\}$ è decrescente (se $n \ge m$ vale $A_n \subseteq A_m$),

   $$\lim_{n\to\infty} \mathbb{P}(X > n\epsilon) = \lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(\cap_{n=1}^{\infty} A_n) = \mathbb{P}(\emptyset) = 0,$$

   (si può giungere alla stessa conclusione dalle proprietà della funzione di ripartizione). Da ciò si ottiene la convergenza cercata, ossia per ogni $\epsilon > 0$

   $$\lim_{n\to\infty} \mathbb{P}(|X_n - 0| > \epsilon) = 0.$$

(d) Per un campione i.i.d. di variabili aleatorie con densità uniformi sull'intervallo $[0,\theta]$, lo stimatore di massima verosimiglianza del parametro θ è dato dal minimo valore assunto dai dati, $\hat{\theta} = \min(x_1,\dots,x_n)$.
FALSO: lo stimatore di massima verosimiglianza è il massimo, $\hat{\theta} = \max(x_1,\dots,x_n)$, si fa riferimento alle dispense del corso per la dimostrazione.

(e) La somma di due variabili aleatorie di Bernoulli con lo stesso parametro p è una variabile Binomiale di paramentri $n=2$ e p.
FALSO: ad esempio data X variabile di Bernoulli di parametro p, posto $Y = X$ (ancora Bernoulli di parametro p) si ha che $X+Y = 2X$ assume solo due valori, 0 e 2, ma se fosse Binomiale con $n=2$ potrebbe assumere anche valore 1.

(f) Nel test chi-quadro per la varianza per l'ipotesi $H_0)\sigma \le \sigma_0$, la potenza del test $\sigma \mapsto \mathbb{P}_\sigma(C)$ è una funzione crescente di σ.
VERO: infatti vale

$$\mathbb{P}_\sigma(C) = \mathbb{P}_\sigma\left(\sum_{i=1}^n \left(X_i - \bar{X}_n\right)^2 > d\right) = \mathbb{P}_\sigma\left(\frac{\sum_{i=1}^n \left(X_i - \bar{X}_n\right)^2}{\sigma^2} > \frac{d}{\sigma^2}\right) = \mathbb{P}(Q_n > d/\sigma^2),$$

in cui Q_n ha densità $\chi^2(n-1)$, per cui la sua legge non dipende da σ e dunque il membro destro dipende da σ solo nell'argomento. Poichè l'intervallo $(d/\sigma^2, \infty)$ cresce al crescere di σ, così è per la probabilità di Q_n di cadervi, dunque $\mathbb{P}(Q_n > d/\sigma^2)$ è crescente in σ. (Si fa riferimento anche alle note del corso per questo fatto noto).

2. Si consideri la seguente funzione:

$$f(t) = \frac{1}{\pi t}\left(\frac{1}{1+(\log t)^2}\right), \quad t > 0,$$

e con $f(t) = 0$ per $t \le 0$.

(a) Si mostri che la funzione f è una densità di probabilità.
(*Suggerimento:* $\frac{d}{dx}\arctan(x) = \frac{1}{1+x^2}$).
La funzione è chiaramente non negativa. Sostituendo $\log t = s$,

$$\int_{-\infty}^{\infty} f(t)dt = \frac{1}{\pi}\int_0^\infty \frac{1}{1+(\log t)^2}\frac{dt}{t} = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{ds}{1+s^2} = \frac{1}{\pi}[\arctan(s)]_{-\infty}^{\infty} = 1.$$

(b) Data una variabile aleatoria X con densità $f_X = f$, si dimostri che la v.a. $Y = \log X$ ha densità $f_Y(y) = 1/(\pi(1+y^2))$.
La funzione $h(x) = \log x$, $h : (0,\infty) \to \mathbb{R}$ è bigettiva e differenziabile, con inversa $h^{-1}(y) = e^y$ differenziabile. Poiché la variabile X assume valori nell'intervallo $(0,\infty)$, si applica la formula di cambio di variabili che restituisce

$$f_Y(y) = f_X(e^y)|e^y|1_{y\in\mathbb{R}} = \frac{1}{\pi(1+y^2)}$$

(c) Si determini se la variabile Y al punto precedente possiede momento primo finito, e in caso positivo lo si calcoli.
Poiché vale

$$\int_{-\infty}^{\infty} |y| f_Y(y)dy = \frac{2}{\pi}\int_0^\infty \frac{y}{1+y^2}dy = +\infty$$

la variabile Y *non* ha momento primo finito.

3. Si vuole comprendere se il numero di bombe tedesche su un'area generica di Londra durante la Seconda Guerra Mondiale segua una distribuzione di Poisson o meno. Per questo, la superficie di Londra viene divisa in 576 quadrati di uguale area e si conta il numero di bombe cadute in ciascun quadrato. I dati sono riassunti nella seguente tabella:

numero di bombe cadute	0	1	2	3	≥ 4
numero di quadrati	229	211	93	35	8

(ad esempio, in 229 quadrati sono cadute 0 bombe). Inoltre, i dati forniscono media e deviazione standard campionarie, per il numero di bombe in un quadrato, pari rispettivamente a 0.929 e 0.967.

(a) Assumendo che la distribuzione sia di Poisson (precisamente, che i numeri di bombe per ciascun quadrato siano v.a. i.i.d. con distribuzione di Poisson) di parametro λ (che ne è il valore atteso), fornire un intervallo di fiducia di livello 95% per λ.
Siano X = numero di bombe in un quadrato, X_i = numero di bombe nell'i-simo quadrato sono un campione i.i.d. di X. Siamo nel caso I.F. per valore atteso per la v.a. X, varianza non nota e grande numerosità del campione ($n = 576$). L'intervallo di fiducia cercato è quindi

$$[\bar{X} \pm \frac{S_n}{\sqrt{n}} z_{1-\alpha/2}] = [\bar{X} \pm \frac{S}{\sqrt{576}} \cdot 1.96].$$

Inserendo i valori numerici $\bar{x} = 0.929$ e $s = 0.967$, otteniamo l'I.F. numerico $[0.929 \pm 0.079]$.

(b) Formulare ed applicare un test statistico di livello 5% per verificare se i dati sono compatibili con l'ipotesi nulla di probabilità 0.395 che in un quadrato non cadano bombe.
Sia $Y = 1_{X=0}$ la v.a. Bernoulli che considera il successo "no bombe in un quadrato". Siamo nel caso test statistico Z approssimato su una probabilità p (valore atteso di Bernoulli), caso grandi campioni ($np_0 \geq 5$, $n(1-p_0) \geq 5$), ipotesi nulla $H_0 : p = p_0 := 0.395$. La statistica di test, approssimativamente $N(0,1)$ sotto H_0, è

$$Z = \frac{n\bar{Y} - np_0}{\sqrt{np_0(1-p_0)}} = \frac{n\bar{Y} - 227.52}{11.732}$$

e la regione di rifiuto è $C = \{|Z| > z_{1-\alpha/2}\}$. Il valore assunto dalla statistica di test sulla base dei dati ($n\bar{y} = 229$) è $z = 0.13$ e il p-value dei dati è

$$P(|Z| > |z|) = 2(1 - \Phi(0.13)) = 0.8966$$

molto alto, in particolare si accetta l'ipotesi nulla per ogni ragionevole livello.

(c) Formulare ed applicare un test statistico di livello 5% per verificare se la distribuzione di Poisson di parametro (dato) $\lambda = 0.929$ sia compatibile con i dati reali (ipotesi nulla che sia compatibile). Non è necessario eseguire tutti i conti, si può usare il valore 1.024 assunto dalla statistica di test sulla base dei dati.
Siamo nel caso test χ quadrato di adattabilità a una data distribuzione, $H_0 : X \sim P(0.929)$. Calcoliamo quindi le frequenze assolute attese sotto H_0, con la formula $P(X = k) = \frac{0.929^k}{k!} e^{-0.929}$:

numero di bombe cadute	0	1	2	3	≥ 4
frequenze osservate	229	211	93	35	8
frequenze attese	227.5	211.3	98.2	30.4	8.6

La numerosità del campione è grande, in particolare tutte le frequenze attese sono ≥ 5, quindi possiamo usare l'approssimazione χ quadrato. La statistica di test è

$$Q = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

dove O_i, E_i sono rispettivamente le frequenze assolute osservate e attese, e la regione di rifiuto è $C = \{Q > \chi^2_{1-\alpha,m-1} = 9.4877\}$, dove $m = 5$ è il numero di classi. Il valore assunto dalla statistica sulla base dei dati è $q = 1.024$, in particolare non cade nella regione critica: si accetta l'ipotesi nulla. (Il p-value, come si vede dalle tavole, è intorno a 0.9, molto alto, quindi si accetta a ogni ragionevole livello.)