Esame Scritto: Appello Straordinario

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) Considerando il lancio di una moneta equilibrata e la variabile X che prende valore 1 se esce testa e 0 se croce, e la variabile Y che invece prende valore 0 se esce testa e 1 se croce, le variabili sono di Bernoulli $p = 1/2$ e dunque la loro somma $X + Y$ è di tipo binomiale con $n = 2$ e $p = 1/2$.
   FALSO: la variabile $X + Y$ prende sempre valore 1.

   (b) Una variabile aleatoria X con varianza $\mathrm{Var}(X) = 0$ è tale che esiste un numero $c \in \mathbb{R}$ tale che $\mathbb{P}(X = c) = 1$.
   VERO: in effetti $c = \mathbb{E}[X]$ che è finito perchè X ammette momento secondo essendo la varianza finita (anzi, nulla), e questo perchè dalla disuguaglianza di Chebychev

   $$\mathbb{P}(|X - \mathbb{E}[X]| > \delta) = \frac{1}{\delta^2}\,\mathrm{Var}(X) = 0,$$

   da cui mandando $\delta \to 0$ si ha $\mathbb{P}(X \neq \mathbb{E}[X]) = 1$.

   (c) Considerando una successione $X_1 = X_2 = X_3 = \cdots$ di variabili uguali tra di loro e con momento secondo finito, per la Legge dei Grandi Numeri la quantità $\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$ converge in probabilità a $\mathbb{E}[X]$.
   FALSO: non è verificata l'ipotesi di indipendenza, e in questo caso $\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n) = X_1$.

   (d) Una variabile Gaussiana $N(m, \sigma^2)$ ha tutti i momenti finiti.
   VERO: la condizione di esistenza del momento n-esimo è

   $$\int_{-\infty}^{\infty} |t|^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(t-m)^2/(2\sigma^2)} dt < \infty,$$

   che è verificata per ogni n poichè l'esponenziale negativo decresce più velocemente di ogni potenza, e nello specifico

   $$|t|^n e^{-(t-m)^2/(2\sigma^2)} \leq e^{-c(t-m)^2/(2\sigma^2)}$$

   per $c > 0$ abbastanza piccola.

   (e) Dato un campione statistico $X_1, X_2, \ldots, X_n$ di v.a. indipendenti con momento secondo finito, la funzione $\mathrm{Var}(\bar{X}_n)$ è decrescente in n.
   VERO: infatti per indipendenza $\mathrm{Var}(\bar{X}_n) = \mathrm{Var}(X_1)/n$.

(f) Se il p-value di un test di ipotesi è 0.001, allora l'ipotesi nulla è plausibile per ogni ragionevole livello.
FALSO: si rifiuta l'ipotesi nulla per ogni livello $\alpha > 0.001$, quindi per ogni ragionevole livello.

2. L'ingegnere ed economo Vilfredo Pareto propose di descrivere la distribuzione di ricchezza tra gli individui di una società con una densità di probabilità della forma:

$$f(t) = \begin{cases} \frac{\alpha}{t^{\alpha+1}} & t \geq 1, \\ 0 & t < 1, \end{cases}$$

(dove $\alpha > 0$ è un parametro), ossia una variabile aleatoria con tale densità rappresenta la ricchezza di un individuo estratto uniformemente a caso nella popolazione (assumendo implicitamente che la ricchezza minima sia 1).

(a) Mostrare che la f proposta è una densità di probabilità per ogni $\alpha > 0$ fissato, e determinare quali momenti possiede una variabile X con densità $f_X = f$, al variare di $\alpha > 0$.

La funzione proposta è una densità di probabilità perché è sempre non-negativa e

$$\int_{-\infty}^{\infty} f(t)dt = \int_{1}^{\infty} \frac{\alpha}{t^{\alpha+1}} dt = \left[-\frac{\alpha}{\alpha t^{\alpha}}\right]_{t=1}^{\infty} = 1$$

Con un conto simile si ottiene, per ogni m intero positivo,

$$E[X^m] = \begin{cases} \left[\frac{\alpha}{m-\alpha} t^{m-\alpha}\right]_{t=1}^{\infty} = \frac{\alpha}{\alpha-m}, & m < \alpha \\ [\alpha \log t]_{t=1}^{\infty} = +\infty, & m = \alpha \\ \left[\frac{\alpha}{m-\alpha} t^{m-\alpha}\right]_{t=1}^{\infty} = +\infty, & m > \alpha \end{cases}$$

Quindi X ammette momento di ordine m se e solo se $m < \alpha$.

(b) Data una variabile X con densità $f_X = f$, si consideri $Y = \log X$. Determinare se anche Y ha densità, e in caso affermativo calcolarla.
X è supportata su $(1, +\infty)$ e la funzione $h : (1, +\infty) \to (0, +\infty)$, $h(x) = \log x$, è C^1, invertibile con inversa $h^{-1}(y) = e^y$ anch'essa C^1. Quindi possiamo applicare la formula di cambio variabili: Y ha densità

$$f_Y(y) = f_X(h^{-1}(y))|\frac{d}{dy}h^{-1}(y)|1_{(0,+\infty)}(y) = \alpha e^{-(\alpha+1)y}|e^y|1_{(0,+\infty)}(y) = \alpha e^{-\alpha y}1_{(0,+\infty)}(y)$$

(dove $1_{(0,+\infty)} = 1$ se $y \in (0, +\infty)$, $= 0$ altrimenti). In particolare, riconosciamo che Y ha densità esponenziale di parametro α.

Supponiamo ora di voler determinare statisticamente il parametro α, e quindi misuriamo la ricchezza di n individui estratti a caso, ottenendo esiti $x_1, \ldots, x_n \geq 1$. Assumiamo di sapere a priori che $\alpha > 1$.

(c) Ricordare la definizione di stima con il metodo del momento primo, e determinare il relativo stimatore per il parametro α.
Rimandiamo alle note del corso per la definizione, la condizione da imporre è

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \mathbb{E}[X] = \frac{\alpha}{\alpha - 1},$$

da cui determiniamo lo stimatore

$$\bar{\alpha} = \frac{\bar{x}}{\bar{x}-1}.$$

3. Un certo algoritmo può essere eseguito correttamente su una macchina solo se il suo tempo di esecuzione non supera i 60 secondi. Secondo un report, il tempo di esecuzione dell'algoritmo (misurato in secondi) segue una distribuzione gaussiana di media 50 e deviazione standard 10.

    a) Secondo il report, qual è la probabilità che l'algoritmo venga eseguito correttamente (cioè il suo tempo di esecuzione non superi i 60 secondi)?
    Detto X il tempo di elaborazione, X ha distribuzione $N(50, 10^2)$. Quindi la probabilità cercata è (chiamando $Z = (X-50)/10$, v.a. normale standard)

    $$P(X \leq 60) = P(Z \leq \frac{60-50}{10}) = \Phi(1) = 0.8413.$$

L'algoritmo viene testato in un esperimento. Su 100 prove, esso viene eseguito correttamente in esattamente 80 prove.

   b) Sulla base dei dati dell'esperimento, fornire un intervallo di fiducia, di livello 95%, per la probabilità che l'algoritmo venga eseguito correttamente.
   Siamo nel caso di intervallo di fiducia per la media di popolazione di Bernoulli, caso grandi campioni. Chiamiamo X_i la v.a. che rapprensenta l'esecuzione corretta all'i-sima prova ($X_i = 1$ se l'algoritmo viene eseguito correttamente, $= 0$ altrimenti), per $i = 1, \ldots n = 100$; chiamiamo $\bar{X}_n$ la media campionaria (la frequenza relativa delle esecuzioni corrette), p la probabilità di esecuzione corretta. L'intervallo cercato per p per $\alpha = 0.05$ è quindi

   $$\left[\bar{X}_n \pm q_{1-\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right] = \left[\bar{X}_{100} \pm 0.196 \cdot \sqrt{\bar{X}(1-\bar{X})}\right].$$

   Applicato al dato $\bar{x}_{100} = 80/100 = 0.8$, otteniamo $[0.8 \pm 0.0784]$.

   c) Sulla base dei dati dell'esperimento, l'affermazione fornita dal report è plausibile? Formulare ed applicare un opportuno test di ipotesi di livello 5%.
   L'affermazione del report ci fornisce (per il punto (a)) $p = 0.8413$, mentre l'esperimento con le 100 prove ci fornisce una stima per p. Testiamo quindi l'affermazione $p = 0.8413$ con un test di ipotesi sulla media di popolazione Bernoulli, caso grandi campioni. Le ipotesi sono $H_0 : p = 0.8413 (= p_0)$, $H_1 : p \neq 0.8413$. La statistica del test è

   $$Z = \sqrt{n}\frac{\bar{X}_n - p_0}{\sqrt{p_0(1-p_0)}} = 27.368 \cdot (\bar{X}_n - 0.8413),$$

   con distribuzione approssimativamente gaussiana standard sotto H_0, la regione critica è

   $$C = \{|Z| > q_{1-\alpha/2} = 1.96\}.$$

   Applichiamo il test: il dato $\bar{x}_{100} = 0.8$: otteniamo $z = 1.13$ che non cade nella regione critica. L'affermazione del report è quindi plausibile.

Valori numerici utilizzabili:

$$\Phi(1) = 0.8413, \quad \Phi(2) = 0.9773, \quad \Phi(3) = 0.9987,$$
$$\Phi(0.56) = 0.7123, \quad \Phi(1.13) = 0.8708, \quad \Phi(2.26) = 0.9881,$$
$$q_{0.95} = 1.64, \quad q_{0.975} = 1.96, \quad q_{0.99} = 2.33, \quad q_{0.995} = 2.58,$$
$$t_{0.95,99} = 1.66, \quad t_{0.975,99} = 1.98, \quad t_{0.99,99} = 2.36, \quad t_{0.995,99} = 2.63.$$