CPS I Prova scritta del 22/06/2021.

1. Si consideri la v.a. bidimensionale discreta (X, Y) con densità:

$$p(x, y) = [2^{x+y}(5/2)^y]/[e^7 x!\ y!], \quad x, y \in \mathbb{N} \cup \{0\}$$

(i) Trovare le densità marginali di X e Y. Si tratta di densità note? X e Y sono stocasticamente indipendenti?
(ii) Calcolare $cov(X + 1, Y + 1)$, $E(X - 2Y)$ e $Var(X - 2Y)$.
(iii) Trovare la densità di $Z = X + Y$ e calcolare $P(Z > 3)$.
(iv) Per $h, k \in \mathbb{N} \cup \{0\}$ con $h \geq k$, calcolare $P(X = k|Z = h)$; in particolare calcolare $P(X = 1|Z = 2)$. Condizionatamente a $Z = h$, che cosa si può dire riguardo alla distribuzione di X?

2. Per $\alpha > 0$, si consideri la funzione:

$$f(x, y) = \begin{cases} \alpha(x^2 + xy/2) & \text{se } (x, y) \in [0, 1] \times [0, 2] \\ 0 & \text{altrimenti} \end{cases}$$

(i) Dopo aver determinato la costante α, in modo che f sia la densità di una v.a. bidimensionale assolutamente continua (X, Y), trovare le densità marginali di X e Y e dire se X e Y sono stocasticamente indipendenti.
(ii) Calcolare media e varianza di X e Y e $Cov(X, Y)$.
(iii) Calcolare $P(X \leq Y)$.
(iv) Calcolare $P(Y > 1/2|X < 1/2)$.

3. Si vuole verificare il carico di rottura medio di alcuni cavi. A tale scopo, si eseguono 20 prove di rottura su altrettanti cavi identici, ottenendo carichi di rottura $x_1, \ldots, x_{20}$ con media campionaria pari a $\bar{x}_n = 10.49\ t$ (t = tonnellate). Da considerazioni teoriche, si può assumere che il campione $x_1, \ldots, x_{20}$ abbia deviazione standard pari a $\sigma = 1.2\ t$.
(i) Trovare un intervallo di confidenza a livello $1 - \alpha = 0.99$ per il carico di rottura medio μ.
(ii) Determinare la dimensione minima del campione affinché l'ampiezza dell'intervallo di confidenza a livello 0.99 per la media μ non superi $0.514\ t$.

CPS Soluzioni Prova scritta del 22/06/2021.

1. (i) Osserviamo che $p(x,y)$ può essere scritta come

$$p(x,y) = \frac{2^x 5^y}{x!\; y!} e^{-7}, \quad x, y \in \mathbb{N} \cup \{0\}$$

Pertanto, si ha, per $x = 0, 1, 2, \ldots$:

$$p_X(x) = P(X = x) = \sum_{y=0}^{\infty} \frac{2^x 5^y}{x!\; y!} e^{-7} = e^{-7} \frac{2^x}{x!} \sum_{y=0}^{\infty} \frac{5^y}{y!}$$

$$= e^{-7} \frac{2^x}{x!} e^5 = e^{-2} \frac{2^x}{x!}$$

e quindi $X \sim$ Poisson(2). Analogamente, per $y = 0, 1, 2, \ldots$:

$$p_Y(x) = P(Y = y) = \sum_{x=0}^{\infty} \frac{2^x 5^y}{x!\; y!} e^{-7} = e^{-7} \frac{5^y}{y!} \sum_{x=0}^{\infty} \frac{2^x}{x!}$$

$$= e^{-7} \frac{5^y}{y!} e^2 = e^{-5} \frac{5^y}{y!}$$

e quindi $X \sim$ Poisson(5). Le v.a. X e Y sono stocasticamente indipendenti, essendo $p(x,y) = p_X(x)p_Y(y)$, e quindi $cov(X,Y) = 0$.
(ii) Poiché per ogni a, b, si ha $cov(X + a, Y + b) = E[(X + a - E(X + a))(Y + a - E(Y + b))] = E(X - E(X))(Y - E(Y)] = cov(X,Y)$, allora $cov(X + 1, Y + 1) = cov(X,Y) = 0$. Inoltre, ricordando che, sia la media che la varianza di una v.a. con distribuzione di Poisson di parametro λ sono uguali a λ, si ottiene $E(X - 2Y) = E(X) - 2E(Y) = 2 - 2 \cdot 5 = -8$; inoltre, siccome X e Y sono indipendenti (e quindi anche X e $-2Y$ lo sono), si ha $Var(X - 2Y) = Var(X) + Var(-2Y) = Var(X) + (-2)^2 Var(Y) = 2 + 4 \cdot 5 = 22$.
(iii) Ricordando che la somma di v.a. indipendenti di Poisson, di parametri λ e μ, rispettivamente, ha distribuzione di Poisson di parametro $\lambda + \mu$, si ottiene che $Z \sim$ Poisson(7) e quindi

$$P(Z > 3) = 1 - P(Z = 0) - P(Z = 1) - P(Z = 2) - P(Z = 3)$$

$$= 1 - e^{-7}(1 + 7 + 7^2/2! + 7^3/3!) = 0.918$$

(iii) Per $h \geq k$, si ha

$$P(X = k | Z = h) = \frac{P(X = k, X + Y = h)}{P(X + Y = h)} = \frac{P(X = k, Y = h - k)}{P(Z = h)} =$$

(essendo X e Y indipendenti)

$$\left[e^{-2} \frac{2^k}{k!} \cdot e^{-5} \frac{5^{h-k}}{(h-k)!} \right] \Big/ \left[e^{-7} \frac{7^h}{h!} \right]$$

$$= \frac{2^k 5^{h-k}}{7^h} \cdot \frac{h!}{k!(h-k)!} = \binom{h}{k} \frac{2^k}{7^k} \cdot \frac{5^{h-k}}{7^{h-k}} = \binom{h}{k} \left(\frac{2}{7}\right)^k \left(\frac{5}{7}\right)^{h-k}$$

Pertanto, condizionatamente a $Z = h$, la v.a. X ha distribuzione binomiale di parametri h e $2/7$.
Per $k = 1$ e $h = 2$, si ottiene

$$P(X = 1 | Z = 2) = \binom{2}{1} \frac{2}{7} \cdot \frac{5}{7} = 2 \cdot \frac{10}{49} = \frac{20}{49} = 0.4081$$

2. Si ha:

$$\int_0^1 dx \int_0^2 dy(x^2 + xy/2) = 7/6,$$

per cui $k = \frac{6}{7}$.
(i) Per $x \in (0, 1)$ si ha:

$$f_X(x) = \frac{6}{7}\int_0^2 (x^2 + xy/2)dy = \frac{6}{7}(2x^2 + x),$$

e 0 altrimenti. Analogamente, per $y \in (0, 2)$, si ha:

$$f_Y(y) = \frac{6}{7}\int_0^1 (x^2 + xy/2)dx = \frac{6}{7}(1/3 + y/4),$$

e 0 altrimenti. Ovviamente, X e Y non sono indipendenti, in quanto $f(x, y) \neq f_X(x)f_Y(y)$.
(ii) Si ha:

$$E(X) = \frac{6}{7}\int_0^1 x(2x^2 + x)dx = 5/7,$$

$$E(Y) = \frac{6}{7}\int_0^2 (y/3 + y^2/4)dy = 8/7,$$

mentre

$$E(XY) = \frac{6}{7}\int_0^1 dx \int_0^2 dy\ xy(x^2 + xy/2) = \frac{17}{21}\ .$$

Pertanto $cov(X, Y) = E(XY) - E(X)E(Y) = \frac{17}{21} - \frac{5}{7}\frac{8}{7} = -\frac{1}{147}$.
(iii) Si ha:

$$P(X > Y) = \frac{6}{7}\int_0^1 dx \int_0^x (x^2 + xy/2)dy = \frac{15}{56},$$

e quindi $P(X < Y) = 1 - \frac{15}{56} = \frac{41}{56}$.
(iv) Si ha:

$$P(Y > 1/2|X < 1/2) = [P(Y > 1/2, X < 1/2)]/P(X < 1/2);$$

il numeratore vale

$$\int_0^{1/2} dx \int_{1/2}^2 dy \frac{6}{7}(x^2 + xy/2) = \frac{69}{448},$$

ed il denominatore è

$$\int_0^{1/2} dx \frac{6}{7}(2x^2 + x) = \frac{5}{28};$$

Quindi, effettuando i calcoli:

$$P(Y > 1/2|X < 1/2) = 69/80 = 0.8625$$

3. (i) Un intervallo I di confidenza a livello $1-\alpha$ per la media incognita di una distribuzione avente varianza σ^2, è:

$$I = \left[\bar{x} - \frac{\sigma}{\sqrt{n}}\phi_{1-\alpha/2}, \bar{x} + \frac{\sigma}{\sqrt{n}}\phi_{1-\alpha/2}\right] \qquad (*)$$

dove $\bar{x}$ è la media campionaria, e ϕ_β è il quantile della Gaussiana standard, tale che $\Phi(\phi_\beta) = \beta$. Nel caso in esame, si ha $n = 20$, $\bar{x} = 10.49$ e $\sigma = 1.2$ Da $1 - \alpha = 0.99$ segue $1 - \alpha/2 = 0.995$, e quindi dalla tavola dei valori di Φ si ricava $\phi_{1-\frac{\alpha}{2}} = 2.57$. Sostituendo in $(*)$, si ottiene l'intervallo:

$$I = \left(10.49 - \frac{1.2 \cdot 2.57}{\sqrt{20}},\ 10.49 + \frac{1.2 \cdot 2.57}{\sqrt{20}}\right) = (9.79, 11.18).$$

(ii) L'ampiezza dell'intervallo di confidenza è $a = \frac{2 \cdot 1.2}{\sqrt{n}} \cdot 2.57$; imponendo che $a \leq 0.514$ si ottiene una disequazione nell'incognita n che, risolta, fornisce $n \geq 144$.