Calcolo delle Probabilità e Statistica, Ingegneria Civile e A&T e Informatica
II Prova scritta 2022

Punteggi: **1:** $4 + 3 + 3$; **2:** 4×3; **3:** 4×2.

1. Data la funzione:

$$p(x,y) = \begin{cases} \alpha \ \frac{3^x}{x!} & \text{se } x \in \{0,1,\ldots\};\ y \in \{1,2,3\} \\ 0 & \text{altrimenti,} \end{cases}$$

(i) determinare il valore di $\alpha > 0$ affinché $p(x,y)$ sia la densità di una v.a. discreta bidimensionale (X,Y) e trovare le densità marginali di X e Y.
(ii) Calcolare $cov(X,Y)$ e $Var(X-2Y)$.
(iii) Sia $U = X+Y$. Calcolare la densità condizionale di U dato che $Y=1$.

2. Sia (X,Y) un vettore aleatorio con densità $f(x,y)$ uniforme sul triangolo T di vertici $(0,0),\ (0,1),\ (2,0)$, ovvero:

$$f(x,y) = \begin{cases} 1 & \text{se } (x,y) \in T \\ 0 & \text{altrimenti.} \end{cases}$$

(i) Trovare le densità marginali di X e Y e calcolare $cov(X,Y)$.
Dire se X e Y sono indipendenti.
(ii) Calcolare $P(X > 2Y)$.
(iii) Calcolare $P(X \leq Y | X > 1/2)$.

3. Un campione di 500 monitors viene estratto da una grossa fornitura ed esaminato per rilevare eventuali difetti. Supponiamo che 350 pezzi superino il controllo.
(i) Calcolare un intervallo di confidenza a livello $1 - \alpha = 0.95$ per la percentuale p di monitors della fornitura privi di difetti.
(ii) Siano $X_1, X_2, \ldots$ v.a. di Bernoulli di parametro $\theta = 0.9$, e poniamo
$\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$. Stimare quanto deve essere grande n affinché risulti
$P(|\bar{X}_n - \theta| \geq 0.01) \leq 0.05$. E' sufficiente prendere $n = 3000$?

Soluzioni della II prova scritta di CPS, 2022

1. (i) Affinché $p(x,y)$ sia una densità, deve essere:

$$\sum_{y=1}^{3}\sum_{x=0}^{\infty}\alpha\frac{3^x}{x!}=1,$$

ovvero $3\alpha\cdot e^3=1$, da cui si ricava $\alpha=e^{-3}/3$.
Per le densità marginali, si ha, per $x=0,1,2,\dots$:

$$p_X(x)=P(X=x)=\sum_{y=1}^{3}\alpha\frac{3^x}{x!}=e^{-3}\frac{3^x}{x!}$$

e quindi $X\sim$ Poisson(3). Inoltre, per $y=1,2,3$:

$$p_Y(y)=P(Y=y)=\sum_{x=0}^{\infty}\frac{1}{3}e^{-3}\frac{3^x}{x!}=\frac{1}{3}$$

Pertanto, si riconosce che $Y\sim$ Unif($\{1,2,3\}$). Le v.a. X e Y sono indipendenti, essendo $p(x,y)=p_X(x)p_Y(y)$.
(ii) Siccome X e Y sono indipendenti, risulta $Cov(X,Y)=0$. Inoltre X è indipendente da $2Y$, quindi:

$$Var(X-2Y)=Var(X)+Var(-2Y)=Var(X)+4Var(Y).$$

Ricordando le formule per la media e la varianza di una v.a. di $Poisson(\lambda)$, e per una v.a. uniforme su $\{1,2,\dots,n\}$ (oppure effettuando i calcoli direttamente) si ottiene

$$E(X)=Var(X)=\lambda=3;\ E(Y)=(1+2+3)/3=2,$$
$$Var(Y)=(n^2-1)/12=\frac{2}{3}\ .$$

Pertanto
$$Var(X-2Y)=3+4\cdot 2/3=17/3.$$

(iii) Si noti che $U\in\mathbb{N}\backslash\{0\}$. Quindi, per ogni $k\in\mathbb{N}\backslash\{0\}$ si ha

$$P(U=k|Y=1)=\frac{P(X+Y=k,Y=1)}{P(Y=1)}=\frac{P(X=k-1,Y=1)}{P(Y=1)}$$

$$=\frac{P(X=k-1)P(Y=1)}{P(Y=1)}=e^{-3}\ \frac{3^{k-1}}{(k-1)!}.$$

2. Il triangolo (rettangolo) T in questione ha come cateti il segmento di lunghezza 1 sull'asse delle ordinate, il segmento di lunghezza 2 sull'asse delle ascisse, e come ipotenusa

il segmento ottenuto dall'intersezione con gli assi della retta di equazione $y = 1 - x/2$. Risulta $f(x,y) = \mathbf{1}_T(x,y)$ (l'area di T vale 1).
(i) Poiché per $(x,y) \in T$ deve essere $y \leq 1 - x/2$, si ha, per $x \in (0,2)$:

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)dy = \int_0^{1-x/2} 1 \cdot dy = (1 - x/2),$$

mentre essa vale zero fuori di $(0,2)$.
Dovendo essere $x \leq 2(1-y)$, si ha per $y \in (0,1)$:

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)dx = \int_0^{2(1-y)} 1 \cdot dx = 2(1-y),$$

mentre essa vale zero fuori di $(0,1)$.
Dunque:

$$f_X(x) = (1 - x/2)\mathbf{1}_{(0,2)}(x), \;\; f_Y(y) = 2(1-y)\mathbf{1}_{(0,1)}(y).$$

Le v.a. X e Y non sono indipendenti, essendo $f(x,y) \neq f_X(x)f_Y(y)$.
Risulta:

$$E(X) = \int_0^2 x \cdot (1 - x/2)dx = \left[x^2/2 - x^3/6\right]_0^2 = 2/3$$

$$E(Y) = \int_0^1 y \cdot 2(1-y)dy = 2\left[y^2/2 - y^3/3\right]_0^1 = 1/3$$

$$E(XY) = \int_0^2 xdx \int_0^{1-x/2} ydy = \ldots = 1/6$$

Quindi

$$cov(X.Y) = E(XY) - E(X)E(Y) = \frac{1}{6} - \frac{2}{3} \cdot \frac{1}{3} = -\frac{1}{18} \; .$$

(ii) Si ha:

$$P(X > 2Y) = \int\int_E f(x,y)dxdy$$

ove

$$E = \{(x,y) : (x,y) \in T, \; y < x/2\} = E_1 \cup E_2,$$

e E_1 è il triangolo rettangolo di vertici $(0,0), \; (1,0), \; (1,1/2)$, mentre E_2 è il triangolo rettangolo di vertici $(1,0), \; (2,0), \; (1,1/2)$. Allora:

$$\int\int_E f(x,y)dxdy = \int_0^1 dx \int_0^{x/2} dy + \int_1^2 dx \int_0^{1-x/2} dy$$

$$= \int_0^1 dx \; x/2 + \int_1^2 (1 - x/2)dx = 1/4 + 1/4 = 1/2$$

(iii) Si ha:

$$P(X \leq Y | X > 1/2) = \frac{P(1/2 < X \leq Y)}{P(X > 1/2)}$$

Il numeratore è dato dall'area del triangolo delimitato dalla retta di eq. $y = x$, dalla retta $x = 1/2$, dalla retta $y = 1 - x/2$. Pertanto:

$$P(X \leq Y | X > 1/2) = \frac{\int_{1/2}^{2/3} dx \int_x^{1-x/2} dy}{\int_{1/2}^{2}(1 - x/2)dx} = \ldots = \frac{\frac{1}{48}}{\frac{9}{16}} = \frac{1}{27}$$

3. (i) Per $n = 500$, sia $X_i,\ i = 1, .., n$ la v.a. che vale 1 se l'i–esimo monitor è privo di difetti, e 0 altrimenti. Le v.a. X_i sono indipendenti e Bernoulliane di parametro p incognito (quindi anche la media $\mu = p$ e la varianza $\sigma^2 = p(1-p)$ sono incognite); $X = X_1 + X_2 + .. + X_n$ rappresenta il numero totale di monitors privi di difetti. Il problema fornisce per la media campionaria di monitors non difettosi il valore $\bar{x} = (x_1 + x_2 + .. + x_n)/n = \frac{350}{500} = 0.7$. 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 $n = 500,\ \bar{x} = 0.7$ e σ è incognito. Da $1 - \alpha = 0.95$ segue $1 - \alpha/2 = 0.975$, e quindi dalla tavola dei valori di Φ si ricava $\phi_{1-\frac{\alpha}{2}} = 1.96$. Sostituendo in $(*)$, si ottiene l'intervallo:

$$I(\sigma) = \left[0.7 - \frac{\sigma}{10\sqrt{5}} \cdot 1.96,\ 0.7 + \frac{\sigma}{10\sqrt{5}} \cdot 1.96\right].$$

Poiché $\sigma = \sqrt{p(1-p)} \leq 1/2,\ \forall p \in [0, 1]$, l'intervallo $I(\sigma)$ è certamente contenuto in

$$I = \left[0.7 - \frac{0.5}{10\sqrt{5}} \cdot 1.96,\ 0.7 + \frac{0.5}{10\sqrt{5}} \cdot 1.96\right] = [0.6562, 0.7438]\ ,$$

che è l'intervallo di confidenza cercato per la percentuale di monitors non fifettosi. Si noti che l'intervallo trovato ha un'ampiezza $\geq$ di quella che si sarebbe trovata se σ fosse stata nota, avendo dovuto fare una maggiorazione.

(ii) Si ha:

$$P\left(|\bar{X}_n - \theta| \geq 0.01\right) = P\left(|X_1 + \ldots X_n - n\theta| \geq 0.01 \cdot n\right) =$$
$$= P\left(\left|\frac{X_1 + \ldots X_n - n\theta}{\sigma\sqrt{n}}\right| \geq \frac{0.01 \cdot n}{\sigma\sqrt{n}}\right)$$

e, siccome $\sigma^2 = 0.9 \cdot 0.1 = 0.09$, ovvero $\sigma = 0.3$, per l'approssimazione normale questa probabilità vale circa

$$P\left(|W| \geq \frac{0.01 \cdot \sqrt{n}}{0.3}\right) = 1 - P\left(|W| \leq \sqrt{n}/30\right) = 2(1 - \Phi\left(\sqrt{n}/30\right)),$$

dove abbiamo indicato con W una v.a. Gaussiana standard. Affinché l'ultima quantità ottenuta sia ≤ 0.05, deve aversi $\Phi\left(\sqrt{n}/30\right) \geq 0.975 \approx \Phi(1.96)$. Dunque deve essere $\sqrt{n}/30 \geq 1.96$, ovvero $n \geq 3458$, e quindi non è sufficiente prendere $n = 3000$.