Tutorato di Probabilità 1, VIII
a.a. 2003/2004

Esercizio 1 Siano X e Y v.a. correlate ($\mathrm{Cov}(X,Y) \neq 0$) e sia $y = g(x)$ l'equazione della retta di regressione di Y rispetto a X:

$$g(x) = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}\Big(x - \mathbb{E}(X)\Big) + \mathbb{E}(Y).$$

Indichiamo con ψ la funzione inversa di g: $\psi(y) = g^{-1}(y)$. Si noti che ψ è lineare, quindi l'equazione $x = \psi(y)$ definisce una retta. È vero (sempre, qualche volta, mai) che ψ è la retta di regressione di X rispetto a Y?

Esercizio 2 Siano X e Y due v.a. indipendenti, $X,Y \sim \mathrm{Po}(\lambda)$. Siano α, β costanti reali.

a) Calcolare[1] la covarianza tra X e $Z^{\alpha,\beta} = \alpha X + \beta Y$. Stabilire il tipo di dipendenza tra X e $Z^{\alpha,\beta}$ al variare di $\alpha, \beta \in \mathbb{R}$.

b) Disegnare la retta di regressione di X e $Z^{\alpha,\beta}$.

Esercizio 3 Un dado viene lanciato due volte. Sia X il risultato del primo lancio e Y il massimo risultato ottenuto nei due lanci.

a) Calcolare[2] la distribuzione congiunta e le distribuzioni marginali di X e Y.

b) Calcolare $\mathrm{Cov}(X,Y)$. Che tipo di dipendenza c'è tra X e Y?

c) Disegnare la retta di regressione di X e Y.

d) Stimare il numero di volte in cui occorre lanciare un dado affinché con probabilità minore di 0.2 la media aritmetica dei risultati disti da 3.5 (cioè dal valore atteso) per più di 0.1.

Esercizio 4

a) Siano X e Y due v.a. bernoulliane di parametro p e $\hat{p}$ rispettivamente. Dimostrare che X e Y sono indipendenti se e solo se $\mathrm{Cov}(X,Y) = 0$. Generalizzare tale proprietà per X e Y v.a. che possono assumere solo due valori: $E_X = \{x_1, x_2\}$ e $E_Y = \{y_1, y_2\}$, con $x_1 \neq x_2$ e $y_1 \neq y_2$. [Sugg.: Per la seconda parte, si può tenere conto delle v.a. ausiliarie $U = \frac{X - x_1}{x_2 - x_1}$ e $V = \frac{Y - y_1}{y_2 - y_1}$.]

b) Siano X e Y due v.a. a valori in $E_X = \{-2, 1\}$ e $E_Y = \{-1, 0, 2\}$ rispettivamente, con distribuzione congiunta descritta tramite la seguente tabella:

	$X = -2$	$X = 1$	
$Y = -1$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$
$Y = 0$	$\frac{1}{12}$	$\frac{5}{12}$	$\frac{1}{2}$
$Y = 2$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$
	$\frac{1}{3}$	$\frac{2}{3}$	1

Mostrare che $\mathrm{Cov}(X,Y) = 0$ e che X e Y non sono indipendenti. Dedurre che, in generale, covarianza nulla **non** implica indipendenza.

---

[1] Sugg.: non utilizzare la distribuzione congiunta di X e $Z^{\alpha,\beta}$; piuttosto, si consideri che $Z^{\alpha,\beta} = \alpha X + \beta Y$...

[2] Si ricorda che

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}, \quad \sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}, \quad \sum_{k=1}^{n} k^3 = \left(\frac{n(n+1)}{2}\right)^2.$$

Soluzioni

1) Posto $y = g(x)$, risolvendo rispetto a x (il che si può fare perché $\mathrm{Cov}(X,Y) \neq 0$) si ottiene:

$$x = \frac{\mathrm{Var}(X)}{\mathrm{Cov}(X,Y)}\Big(y - \mathbb{E}(Y)\Big) + \mathbb{E}(X),$$

e quindi $\psi(y) = \frac{\mathrm{Var}(X)}{\mathrm{Cov}(X,Y)}\Big(y - \mathbb{E}(Y)\Big) + \mathbb{E}(X)$. Ora, l'equazione della retta di regressione di X rispetto a Y è

$$x = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)}\Big(y - \mathbb{E}(Y)\Big) + \mathbb{E}(X).$$

Allora $x = \psi(y)$ è la retta di regressione richiesta se e solo se

$$\frac{\mathrm{Var}(X)}{\mathrm{Cov}(X,Y)} = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)}$$

che equivale a

$$|\,\mathrm{Cov}(X,Y)\,| = \sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}.$$

Quindi $x = \psi(y) \equiv g^{-1}(y)$ è l'equazione della retta di regressione richiesta se e solo se vale il caso limite della disuguaglianza di Cauchy-Schwartz[3].

2) a) Usando le proprietà tipiche della covarianza, si ha

$$\mathrm{Cov}(X, Z^{\alpha,\beta}) = \mathrm{Cov}(X, \alpha\, X + \beta\, Y) = \alpha \mathrm{Cov}(X,X) + \beta \mathrm{Cov}(X,Y) = \alpha \mathrm{Var}(X)$$

perché X e Y sono indipendenti, quindi non correlate: $\mathrm{Cov}(X,Y) = 0$. Ora, basta ricordare che la varianza di una $\mathrm{Po}(\lambda)$ è $\lambda (> 0)$:

$$\mathrm{Cov}(X, Z^{\alpha,\beta}) = \alpha\,\lambda \text{ quindi: } \begin{cases} \text{se } \alpha > 0 \text{ allora c'è dipendenza } \textbf{positiva} \\ \text{se } \alpha < 0 \text{ allora c'è dipendenza } \textbf{negativa} \\ \text{se } \alpha = 0 \text{ allora c'è dipendenza } \textbf{nulla} \end{cases}$$

Si noti, in particolare, che il tipo di dipendenza **non** dipende da β. Inoltre, se $\alpha = 0$ allora $Z^{\alpha,\beta} = Z^{0,\beta} = \beta\, Y$, quindi X e $Z^{0,\beta}$ non sono solo incorrelate ma anche indipendenti.

b) L'equazione della retta di regressione sul piano (x, z) è

$$z = \frac{\mathrm{Cov}(X, Z^{\alpha,\beta})}{\mathrm{Var}(X)}\Big(x - \mathbb{E}(X)\Big) + \mathbb{E}(Z^{\alpha,\beta}).$$

Poiché $\mathrm{Cov}(X, Z^{\alpha,\beta}) = \alpha\,\lambda$, $\mathbb{E}(X) = \lambda = \mathbb{E}(Y)$ e $\mathbb{E}(Z^{\alpha,\beta}) = \alpha\mathbb{E}(X) + \beta\mathbb{E}(Y) = (\alpha+\beta)\lambda$, la retta di regressione qui diviene $z = \alpha\, x + \beta\,\lambda$.

3) a) Siano X e Z rispettivamente il risultato del primo e del secondo lancio. Allora $Y = X \wedge Z$ e

$$p_{X,Y}(x,y) = \mathrm{P}(X = x, Y = y) = \mathrm{P}(X = x, X \wedge Z = y) = \mathrm{P}(X = x, x \wedge Z = y).$$

Se $(x,y) \notin \{1,\dots,6\}^2$, ovviamente $p_{X,Y}(x,y) = 0$. Studiamo quindi il comportamento di $p_{X,Y}$ per $(x,y) \in \{1,\dots,6\}^2$. Si ha,

$$\begin{aligned} p_{X,Y}(x,y) &= \mathrm{P}(X = x, x \wedge Z = y, Z \le x) + \mathrm{P}(X = x, x \wedge Z = y, Z > x) \\ &= \mathrm{P}(X = x, x = y, Z \le x) + \mathrm{P}(X = x, Z = y, Z > x) \\ &= \begin{cases} 0 & \text{se } x > y \\ \mathrm{P}(X = x, Z \le x) & \text{se } x = y \\ \mathrm{P}(X = x, Z = y) & \text{se } x < y \end{cases} \end{aligned}$$

[3] $|\,\mathrm{Cov}(X,Y)\,| \le \sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}$, si veda l'esercizio 3 b) del Tutorato V.

Ricordando che X e Z rappresentano il risultato di due tiri (indipendenti) di un dado, possiamo scrivere, per $(x,y) \in \{1,\dots,6\}^2$,

$$p_{X,Y}(x,y) = \begin{cases} 0 & \text{se } x > y \\ \frac{x}{6^2} & \text{se } x = y \\ \frac{1}{6^2} & \text{se } x < y \end{cases}$$

Calcoliamo le distribuzioni marginali di X e Y: per $x, y \in \{1,\dots,6\}$,

$$p_X(x) = \sum_{y=1}^{6} p_{X,Y}(x,y) = \frac{x}{6^2} + \sum_{y=x+1}^{6} \frac{1}{6^2} = \frac{x}{6^2} + \frac{6-x}{6^2} = \frac{6}{6^2} = \frac{1}{6},$$

$$p_Y(y) = \sum_{x=1}^{6} p_{X,Y}(x,y) = \sum_{x=1}^{y-1} \frac{1}{6^2} + \frac{y}{6^2} = \frac{y-1}{6^2} + \frac{y}{6^2} = \frac{2y-1}{6^2}.$$

b) Per calcolare la covarianza tra X e Y occorrono $\mathbb{E}[X]$, $\mathbb{E}[Y]$ e $\mathbb{E}[X\,Y]$. Si ha[4]

$$\mathbb{E}[X] = \sum_{x=1}^{6} x\, p_X(x) = \sum_{x=1}^{6} x \frac{1}{6} = \frac{1}{6} \cdot \frac{6 \cdot 7}{2} = \frac{7}{2};$$

$$\mathbb{E}[Y] = \sum_{y=1}^{6} y\, p_Y(y) = \sum_{y=1}^{6} y \frac{2y-1}{6^2} = \frac{1}{6^2}\Big[2\sum_{y=1}^{6} y^2 - \sum_{y=1}^{6} y\Big] = \frac{1}{6^2}\Big[2 \cdot \frac{6\cdot 7\cdot 13}{6} - \frac{6\cdot 7}{2}\Big] = \frac{161}{36} (\simeq 4.72);$$

e

$$\mathbb{E}[X\,Y] = \sum_{x,y=1}^{6} x\,y\,p_{X,Y}(x,y) = \sum_{x=1}^{6}\sum_{y=1}^{6} x\,y\,p_{X,Y}(x,y)$$

$$= \sum_{x=1}^{6} x \sum_{y=x}^{6} y\,p_{X,Y}(x,y) = \sum_{x=1}^{6} x\Big[x \cdot \frac{x}{6^2} + \sum_{y=x+1}^{6} y \cdot \frac{1}{6^2}\Big] = \frac{1}{6^2}\sum_{x=1}^{6} x\Big[x^2 + \sum_{y=1}^{6} y - \sum_{y=1}^{x} y\Big]$$

$$= \frac{1}{6^2}\sum_{x=1}^{6} x\Big[x^2 + \frac{6\cdot 7}{2} - \frac{x(x+1)}{2}\Big] = \frac{1}{6^2}\sum_{x=1}^{6} x\,\frac{x^2 - x + 42}{2} = \frac{1}{2\cdot 6^2}\Big[\sum_{x=1}^{6} x^3 - \sum_{x=1}^{6} x^2 + 42\sum_{x=1}^{6} x\Big]$$

$$= \frac{1}{2\cdot 6^2}\Big[\Big(\frac{6\cdot 7}{2}\Big)^2 - \frac{6\cdot 7\cdot 13}{6} + 42 \cdot \frac{6\cdot 7}{2}\Big] = \frac{154}{9} (\simeq 17.11)$$

Quindi:

$$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] = \frac{35}{24}$$

e la dipendenza tra le due variabili è positiva: all'aumentare dell'una ci si aspetta un aumento anche dell'altra.

c) L'equazione della retta di regressione è:

$$y = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}\Big(x - \mathbb{E}[X]\Big) + \mathbb{E}[Y].$$

Occorre quindi calcolare $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}^2[X]$. Si ha

$$\mathbb{E}[X^2] = \sum_{x=1}^{6} x^2\, p_X(x) = \sum_{x=1}^{6} x^2 \cdot \frac{1}{6} = \frac{1}{6} \cdot \frac{6\cdot 7\cdot 13}{6} = \frac{91}{6}$$

[4] Nel seguito, useremo le uguaglianze:

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}, \quad \sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}, \quad \sum_{k=1}^{n} k^3 = \Big(\frac{n(n+1)}{2}\Big)^2.$$

da cui segue che

$$\mathrm{Var}(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}.$$

Quindi, la retta di regressione ha equazione

$$y = \frac{1}{2}\left(x - \frac{7}{2}\right) + \frac{161}{36}.$$

d) Denotiamo con X_i il risultato dell'$i-$esimo lancio e $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Allora $3.5 = \mathbb{E}[\bar{X}_n]$ e $\mathrm{Var}(\bar{X}_n) = \frac{1}{n}\mathrm{Var}(X_1)$, dove $\mathrm{Var}(X_1) = \frac{35}{12}$. La disuguaglianza di Chebycev assicura che

$$\mathrm{P}(|\bar{X}_n - 3.5| > 0.1) \leq \frac{\mathrm{Var}(\bar{X}_n)}{0.1^2} = 100 \cdot \frac{\mathrm{Var}(X_1)}{n}.$$

Perché si abbia $\mathrm{P}(|\bar{X}_n - 3.5| > 0.1) \leq 0.2$ è sufficiente che

$$100 \cdot \frac{\mathrm{Var}(X_1)}{n} \leq 0.2$$

da cui segue che $n \geq 2000 \cdot \mathrm{Var}(X_1) = 5833,\bar{3}$: basterà scegliere $n = 5834$.

4) a) Ovviamente, indipendenza implica covarianza nulla. Supponiamo ora che $\mathrm{Cov}(X,Y) = 0$:

$$0 = \mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathrm{P}(X=1,Y=1) - \mathrm{P}(X=1)\mathrm{P}(Y=1),$$

cioè $\mathrm{P}(X=1,Y=1) = \mathrm{P}(X=1)\mathrm{P}(Y=1)$. Allora

$$\begin{aligned}
\mathrm{P}(X=0,Y=1) &= \mathrm{P}(Y=1) - \mathrm{P}(X=1,Y=1) = \mathrm{P}(Y=1) - \mathrm{P}(X=1)\mathrm{P}(Y=1)\\
&= \mathrm{P}(Y=1)\Big(1 - \mathrm{P}(X=1)\Big) = \mathrm{P}(X=0)\mathrm{P}(Y=1),\\
\mathrm{P}(X=0,Y=0) &= \mathrm{P}(X=0) - \mathrm{P}(X=0,Y=1) = \mathrm{P}(X=0) - \mathrm{P}(X=0)\mathrm{P}(Y=1)\\
&= \mathrm{P}(X=0)\Big(1 - \mathrm{P}(Y=1)\Big) = \mathrm{P}(X=0)\mathrm{P}(Y=0),\\
\mathrm{P}(X=1,Y=0) &= \mathrm{P}(X=1) - \mathrm{P}(X=1,Y=1) = \mathrm{P}(X=1) - \mathrm{P}(X=1)\mathrm{P}(Y=1)\\
&= \mathrm{P}(X=1)\Big(1 - \mathrm{P}(Y=1)\Big) = \mathrm{P}(X=1)\mathrm{P}(Y=0),
\end{aligned}$$

da cui segue che X e Y sono indipendenti.

Supponiamo ora che X e Y possano assumere solo due valori $x_1 \neq x_2$ e $y_1 \neq y_2$ rispettivamente e che $\mathrm{Cov}(X,Y) = 0$. Consideriamo le v.a.

$$U = \frac{X - x_1}{x_2 - x_1} \quad \text{e} \quad V = \frac{Y - y_1}{y_2 - y_1}.$$

U e V sono bernoulliane di parametri, rispettivamente, $p = p_X(x_2)$ e $\hat{p} = p_Y(y_2)$. Inoltre,

$$\mathrm{Cov}(U,V) = \mathrm{Cov}\left(\frac{X - x_1}{x_2 - x_1}, \frac{Y - y_1}{y_2 - y_1}\right) = \frac{1}{(x_2 - x_1)(y_2 - y_1)}\mathrm{Cov}(X,Y) = 0,$$

quindi, per **a)**, U e V sono indipendenti. Infine, poiché

$$X = x_1 + (x_2 - x_1)U \quad \text{e} \quad Y = y_1 + (y_2 - y_1)V,$$

segue che anche X e Y sono indipendenti.

b) Si ha:

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{x \in E_X} x p_X(x) = -2 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} = 0\\
\mathbb{E}(Y) &= \sum_{y \in E_Y} y p_Y(x) = -1 \cdot \frac{1}{3} + 2 \cdot \frac{2}{3} = 0\\
\mathbb{E}(XY) &= \sum_{x \in E_X}\sum_{y \in E_Y} xy p_{X,Y}(x,y) = 2 \cdot \frac{1}{6} - 1 \cdot \frac{1}{6} - 4 \cdot \frac{1}{12} + 2 \cdot \frac{1}{12} = 0
\end{aligned}$$

da cui segue che $\mathrm{Cov}(X,Y)=0$. Inoltre, ad esempio, $\mathrm{P}(X=-2,Y=-1)=\frac{1}{6}\neq\frac{1}{9}=\mathrm{P}(X=-2)\mathrm{P}(Y=-1)$, dunque X e Y non sono indipendenti. Abbiamo quindi trovato un controesempio alla (evidentemente falsa, in generale) implicazione

$$\text{covarianza nulla} \Longrightarrow \text{indipendenza.}$$