Corso di Laurea in Ingegneria Informatica
Anno Accademico 2016/2017
Calcolo delle Probabilità e Statistica Matematica

Nome ....

N. Matricola .... Ancona, 12 luglio 2017

1. (10 punti) Un campione di 100 studenti viene analizzato per vedere la correlazione tra il numero di ore di studio giornaliere e i voti riportati in un certo esame. I dati vengono raccolti nella seguente tabella:

	Ore di studio			
Voti	2	4	6	8
insuff	4	3	2	1
18-21	3	4	2	1
22-25	1	18	16	5
26-28	0	16	15	4
29-30	0	0	2	3

Si chiede di:

- calcolare le densità marginali delle variabili che rappresentano le ore di studio e il voto;
- dire se le ore di studio e i voti sono indipendenti;
- calcolare il voto medio e il numero medio di ore di studio;
- calcolare il coefficiente di correlazione tra numero di ore di studio e voti.

Soluzione. Sia X la variabile che descrive i voti e Y la variabile che descrive le ore di studio. I valori nella tabellina, divisi per 100, non sono altro che i valori della densità congiunta $p_{XY}(i, j)$. Prendendo il valore centrale degli intervalli dei voti e attribuendo zero al voto insufficiente, abbiamo che $X \in \{0, 19.5, 23.5, 27, 29.5\}$ mentre $Y \in \{2, 4, 6, 8\}$. Sommando righe e colonne otteniamo la nuova tabellina:

	Ore di studio				
Voti	2	4	6	8	
insuff	4	3	2	1	10
18-21	3	4	2	1	10
22-25	1	18	16	5	40
26-28	0	16	15	4	35
29-30	0	0	2	3	5
	8	41	37	14	100

Le marginali si ottengono dividendo per 100 i numeri nell'ultima colonna e nell'ultima riga:

$$p_X(0) = 0.1 \qquad p_X(19.5) = 0.1 \qquad p_X(23.5) = 0.4 \qquad p_X(27) = 0.35 \qquad p_X(29.5) = 0.05$$
$$p_Y(2) = 0.08 \qquad p_Y(4) = 0.41 \qquad p_Y(6) = 0.37 \qquad p_Y(8) = 0.14$$

Ovviamente le due variabili non sono indipendenti. Le medie sono date da:

$$E[X] = 0.1 \times 0 + 0.1 \times 19.5 + 0.4 \times 23.5 + 0.35 \times 27 + 0.05 \times 29.5 = 22.27$$
$$E[Y] = 0.08 \times 2 + 0.41 \times 4 + 0.37 \times 6 + 0.14 \times 8 = 5.14$$

Per il calcolo delle varianze, calcoliamo dapprima $E[X^2]$ e $E[Y^2]$:

$$E[X^2] = 0.1 \times 0 + 0.1 \times 19.5^2 + 0.4 \times 23.5^2 + 0.35 \times 27^2 + 0.05 \times 29.5^2 = 557.59$$
$$E[Y^2] = 0.08 \times 2^2 + 0.41 \times 4^2 + 0.37 \times 6^2 + 0.14 \times 8^2 = 29.16$$

da cui

$$Var(X) = E[X^2] - E[X]^2 = 557.59 - 496.18 = 61.41$$
$$Var(Y) = E[Y^2] - E[Y]^2 = 29.12 - 26.21 = 2.74$$

Per il calcolo del coefficiente di correlazione ci serve ancora $E[XY]$. Abbiamo

$$E[XY] = \sum_{i=1,5} \sum_{j=1,4} x_i \, y_j \, p_{XY}(i,j) = ... = 118.38$$

La covarianza è quindi

$$Cov(X,Y) = E[XY] - E[X]\,E[Y] = 118.38 - 29.5 \times 5.14 = 3.89$$

e finalmente

$$\rho = \frac{Cov(X,Y)}{Var(X)\,Var(Y)} = 0.3.$$

2. (10 punti) Il numero di pazienti che visita uno studio medico segue una legge di Poisson di diverso parametro λ nei diversi giorni della settimana, dal lunedì al venerdì. Si ha, in decine di pazienti al giorno (associando "1=lunedì", "2=martedì", etc.):

$$\lambda_1 = 2, \qquad \lambda_2 = 0.8, \qquad \lambda_3 = 1.5, \qquad \lambda_4 = 1.5, \qquad \lambda_5 = 0.2.$$

Il numero di pazienti in un dato giorno è inoltre indipendente dal numero di pazienti negli altri giorni. Si chiede di:

- calcolare la media settimanale di pazienti;
- qual è la probabilità che in una settimana più di 70 pazienti visitino lo studio medico se nei primi tre giorni della settimana ci sono state complessivamente 50 visite?

Soluzione. La distribuzione di Poisson è riproducibile e pertanto il numero di pazienti in una settimana segue pure la legge di Poisson con $\lambda = \lambda_1 + \lambda_2 + ... + \lambda_5 = 6$, quindi la media di visite settimanali è di 60. Indicando con X_1, X_2, ..., X_5 le variabili che rappresentano le visite nelle varie giornate e con $X = X_1 + X_2 + ... + X_5$ la variabile che rappresenta le visite settimanali, la richiesta è di determinare $P(X \geq 7|X_1 + X_2 + X_3 = 5)$ (in decine di pazienti al giorno). Abbiamo:

$$
\begin{aligned}
P(X \geq 7|X_1 + X_2 + X_3 = 5) &= \frac{P(\{X \geq 7\} \cap \{X_1 + X_2 + X_3 = 5\})}{P(X_1 + X_2 + X_3 = 5)} = \\
&= \frac{P(\{X_4 + X_5 \geq 2\} \cap \{X_1 + X_2 + X_3 = 5\})}{P(X_1 + X_2 + X_3 = 5)} = \\
&= \frac{P(X_4 + X_5 \geq 2)\, P(X_1 + X_2 + X_3 = 5)}{P(X_1 + X_2 + X_3 = 5)} = P(X_4 + X_5 \geq 2) = \\
&= 1 - P(X_4 + X_5 < 2) = 1 - P(X_4 + X_5 = 0) - P(X_4 + X_5 = 1) = \\
&= 1 - e^{-1.7}\,(1 + 1.7) = 0.51
\end{aligned}
$$

dove negli ultimi passaggi abbiamo sfruttato la riproducibilità della distribuzione di Poisson e l'indipendenza delle variabili.

3. (10 punti) Si vuole determinare l'intervallo di confidenza per la densità dei portatori di carica in un campione di silicio. Si fanno due misurazioni che danno luogo ai campioni di rango $n = 10$: (in unità di 10^{10} cm^{-3}):

$$
\begin{aligned}
I &: 1.34, 1.47, 1.48, 1.41, 1.38, 1.35, 1.39, 1.47, 1.42, 1.48 \\
II &: 1.44, 1.48, 1.47, 1.50, 1.46, 1.54, 1.46, 1.48, 1.48, 1.43
\end{aligned}
$$

Determinare, per entrambi i campioni, gli intervalli di fiducia al 90%, 95% e 99%. In quali casi gli intervalli provenienti dai due campioni si sovrappongono?

Soluzione. Abbiamo

$$
\begin{aligned}
\overline{X}_n &= 1.42 \qquad S^2 = 0.0054 \text{ I campione} \\
\overline{X}_n &= 1.47 \qquad S^2 = 0.0031 \text{ I campione}
\end{aligned}
$$

mentre per il quantile di Student usiamo

$$
t_{0.05}(9) = 1.833 \qquad t_{0.025}(9) = 2.262 \qquad t_{0.005}(9) = 3.250
$$

Gli intervalli sono

$(1.39, 1.45)$	per il I campione	e	$(1.46, 1.50)$	per il II campione	al 90 %
$(1.38, 1.46)$	per il I campione	e	$(1.45, 1.50)$	per il II campione	al 95 %
$(1.36, 1.47)$	per il I campione	e	$(1.44, 1.51)$	per il II campione	al 99 %

Gli intervalli al 90% non si sovrappongono.