Università degli Studi di Firenze
Corso di Laurea in Ingegneria Informatica — Facsimile dell'Esame di Statistica — 7-4-2003

NOTA BENE: Nel rispondere alle domande teoriche' e' necessario definire TUTTI i termini introdotti e spiegare le notazioni. Per risolvere i problemi e' necessario spiegare i passaggi.

1. Supponete che lo 0.1% di una popolazione sia affetta da una malattia. Esiste un test per scoprire la malattia e la probabilita' che il test individui una persona infetta e' del 98%. La probabilita' che il test dichiari infetta una persona non infetta e' dello 0.2%. Si estrae a caso un individuo e il test lo giudica affetto dalla malattia. Qual e' la probabilita' che questa persona abbia la malattia?

2. Trovare la media e la varianza della variabile X con funzione di densita'

$$f(x) = \begin{cases} 1/6 & \text{per } -3 < x < 3 \\ 0 & \text{altrove} \end{cases}.$$

3. (a) Definire la covarianza tra due variabili aleatorie discrete X e Y. (b) Che cosa misura tale indice? (c) Che valori puo' assumere? Come si interpreta? (d) Qual e' l'interpretazione nelle situazioni estreme? (e) Data la variabile doppia

$X \setminus Y$	$y = 10$	$y = 20$
$x = 0$	0.1	0.4
$x = 1$	0.5	0

Calcolare la covarianza e il coefficiente di correlazione.

Soluzioni

1. Sia I l'evento 'e' infetto'. Quindi $P(I) = 0.001$ e $P(I^c) = 1 - 0.001 = 0.999$. Sia $+$ l'evento 'e positivo al test'. Sappiamo che $P(+ \mid I) = 0.98$ e che $P(+ \mid I^c) = 0.002$. Dobbiamo calcolare

$$P(I \mid +) = \frac{P(+ \mid I)P(I)}{P(+)}$$

per la formula di Bayes. Quindi,

$$P(+) = P(+ \mid I)P(I) + P(+ \mid I^c)P(I^c)$$

e dunque

$$P(I \mid +) = \frac{0.98 \times 0.001}{0.98 \times 0.001 + 0.002 \times 0.999} = 980/2978 = 0.329.$$

2. La distribuzione e' uniforme. La media e' pertanto $\mu = [(-3) + 3]/2 = 0$. La varianza e' dunque uguale a $E(X^2)$ cioe'

$$\int_{-3}^{3} x^2(1/6)dx = \frac{1}{6} \times [x^3/3]_{-3}^{3} = 3.$$

3. (a) La covarianza tra due variabili discrete Xe Y aventi funzione di probabilita' congiunta $p(x, y)$ e valori attesi $\mu_x = EX$ e $\mu_y = EY$ e'

$$\text{cov}(X, Y) = \sum_x \sum_y (x - \mu_x)(y - \mu_y)p(x, y) = E\{(Y - EX)(Y - EY)\}.$$

(b) L'indice misura il grado di associazione lineare tra X e Y. (c) Puo' assumere valori compresi nell'intervallo

$$-\sigma_x\sigma_y \leq \text{cov}(X, Y) \leq \sigma_x\sigma_y$$

dove σ_x e σ_y sono gli scarti quadratici medi (le radici quadrate delle varianze) di X e Y rispettivamente. Se e' positiva, c'e' concordanza tra X e Y (a scarti di un segno dalla media di X si accompagnano scarti dello stesso segno dalla media di Y). Se e' negativa, c'e' discordanza (a scarti di un segno dalla media di X si accompagnano scarti di segno opposto dalla media di Y). (d) La covarianza e' uguale al prodotto

$\sigma_x\sigma_y$ se e solo se $Y = a + bX$ con $b > 0$. La covarianza e' uguale a $-\sigma_x\sigma_y$ se e solo se $Y = a + bX$ con $b < 0$.

(e) $E(X) = \frac{1}{2}$, $E(Y) = 14$. $E(XY) = 10 \times \frac{1}{2} = 5$. Quindi $\sigma_{XY} = 5 - 14/2 = -2$. Inoltre, $\sigma_Y^2 = 100 \times 0.6 + 400 \times 0.4 - 14^2 = 24$ e $\sigma_X^2 = 0.5 \times (1 - 0.5) = 0.25$. Infine il coefficiente di correlazione e' $\rho_{XY} = -2/\sqrt{24/4} = -0.81$.