DISTRIBUZIONI di PROBABILITA' IN 2 (o più) VARIABILI

e probabilità di una SOMMA

Supponiamo di avere 2 VARIABILI ALEATORIE, che per semplicità supporremo assumere un SET FINITO di VALORI

$X \in \{x_1, x_2, \dots, x_M\}$
$Y \in \{y_1, y_2, \dots, y_M\}$ — N.B. Notare che i 2 set sono stati assunti della stessa cardinalità (M): questo non è necessario...

Posso considerare l'evento $A_i \leftrightarrow X = x_i$, ovvero X assume il valore x_i
e l'evento $B_j \leftrightarrow Y = y_j$, ovvero Y assume il valore y_j

Posso ora chiedermi la PROBABILITA' dell'EVENTO PRODOTTO $P(A_i B_j)$

Essa sarà un NUMERO $p_{ij} = P(X = x_i \text{ et } Y = y_j)$

La collezione dei numeri p_{ij} è la DISTRIBUZIONE di PROBABILITA' (a due variabili) per le X e Y

NORMALIZZAZIONE $\longleftrightarrow \sum_{ij} p_{ij} = 1$

Notare che posso definire PROBABILITA' ad 1 variabile sommando su una delle due:

$$p_i^X = \sum_j p_{ij} \quad \text{(probabilità di avere } X = x_i \text{, indipendent. dal valore di } Y)$$
$$p_j^Y = \sum_i p_{ij} \quad \text{(probabilità di avere } Y = y_j \text{, indipendent. dal valore di } X)$$

(ovviamente $\sum_i p_i^X = 1 = \sum_j p_j^Y$) Queste distribuzioni sono dette MARGINALI

Consideriamo la SOMMA $X + Y$ (si tratta di una nuova variabile aleatoria, funzione di X e Y)

Per la MEDIA vale che
$$M[X+Y] = \sum_{ij} (x_i + y_j) p_{ij} = \sum_{ij} x_i p_{ij} + \sum_{ij} y_j p_{ij}$$
$$= \sum_i p_i^X x_i + \sum_j p_j^Y y_j = M[X] + M[Y]$$

In generale il calcolo della VARIANZA è più complicato...

(Cogliamo l'occasione per richiamare 2 casi banali:

COSTANTE c $\qquad M[c]=c$

$$D[c] = M[c^2] - M[c]^2 = c^2 - c^2 = 0$$

cX dove c costante e X variabile aleatoria

$$M[cX] = \sum_i c x_i p_i = c \sum_i x_i p_i = c M[X]$$

$$\begin{aligned} D[cX] &= \sum_i c^2 x_i^2 p_i - \left(\sum_i c x_i p_i\right)^2 \\ &= c^2 M[X^2] - c^2 M[X]^2 = c^2 \left(M[X^2] - M[X]^2\right) \\ &= c^2 D[X] \end{aligned}$$

... fine della digressione: anche maneggiando una sola variabile aleatoria e costanti il calcolo della varianza merita un minimo di attenzione in più...)

Per calcolare $D[X+Y]$ usiamo la definizione equivalente di varianza come 2° momento centrato

$$D[Z] = M\left[(Z-m_Z)^2\right] \quad \text{dove } m_Z = M[Z]$$

nel caso in esame $Z = X+Y$ $\quad m_Z = m_X + m_Y$ (dove ovviamente $m_X = M[X]$ e $m_Y = M[Y]$)

ora $Z - m_Z = X + Y - m_X - m_Y = (X-m_X) + (Y-m_Y)$

e dunque $(Z-m_Z)^2 = (X-m_X)^2 + (Y-m_Y)^2 + 2(X-m_X)(Y-m_Y)$

per cui $D[X+Y] = M[(Z-m_Z)^2] = M[(X-m_X)^2] + M[(Y-m_Y)^2] + 2M[(X-m_X)(Y-m_Y)]$

$$\equiv D[X] + D[Y] + 2K_{XY}$$

dove abbiamo definito la COVARIANZA

$$\begin{aligned} K_{XY} &= M[(X-m_X)(Y-m_Y)] \\ &= \sum_{ij} (x_i - m_X)(y_j - m_Y) p_{ij} \end{aligned}$$

Notare che nel CASO di VARIABILI X e Y INDIPENDENTI la DISTRIBUZIONE a due variabili è BANALE!

NB Notate come cambiamo notazione rispetto alle distribuzioni marginali, mettendo ad apice delle parentesi: non è la stessa cosa, perché ora le distribuzioni nella X e nella Y esistono "a priori"...

Per def. di INDIP. $p_{ij} = p_i^{(X)} p_j^{(Y)} \Leftrightarrow$ X e Y INDIPENDENTI

Questo comporta che NON C'E' COVARIANZA... $K_{XY} = \sum_{ij} (x_i - x_m)(y_j - m_y)\, p_{ij}$

$$= \sum_{ij} (x_i - x_m)(y_j - m_y)\, p_i^{(X)} p_j^{(Y)}$$

(il primo momento centrato è sempre nullo...) $= \sum_i (x_i - x_m)\, p_i^{(X)} \sum_j (y_j - m_y)\, p_j^{(Y)}$

$$= 0$$

N.B. X e Y si dicono NON CORRELATE SE $p_{ij} \neq p_i^{(X)} p_j^{(Y)}$ MA UGUALMENTE $K_{XY} = 0$.

N.B. 2 In generale possiamo pensare a distribuzioni a più variabili $\rightarrow p_{ijk\ldots l}$

Notare che sono definite non solo prob. come $p_i^{X_1} = \sum_{jk\ldots l} p_{ijk\ldots l}$

Spero sia chiara la notazione: chiamiamo $\{X_1 X_2 \ldots X_n\}$ le variabili per cui scriviamo la distribuzione di probabilità.

ma anche $p_{ij}^{X_1 X_2} = \sum_{k\ldots l} p_{ijk\ldots l}$

Le distribuzioni marginali possibili sono ora tante...

IN GENERALE LA VARIANZA ha SEMPRE CONTRIBUTI di COVARIANZE

$$D\Big[\sum_i X_i\Big] = \sum_i D[X_i] + 2 \sum_{i<j} K_{X_i X_j}$$

IMPORTANTE! Un caso fondamentale dell'annullamento di effetti di COVARIANZA è quello della osservabile MEDIA ARITMETICA di VARIABILI INDIPENDENTI!