20/5/05 | STATISTICA

VARIABILI ALEATORIE DISCRETE m-DIMENSIONALI

Def è una funzione $X : \Omega \to \mathbb{R}^m$

$$X(\omega) = (X_1(\omega), X_2(\omega), \dots, X_m(\omega))$$

dove ognuna delle $X_i : \Omega \to \mathbb{R}$ è una v.a. discreta (scalare)

Fissato $x = (x_1, \dots, x_m) \in \mathbb{R}^m$

$$X = x \Leftrightarrow X_1 = x_1, X_2 = x_2, \dots, X_m = x_m$$

Quindi l'evento

$$\{X = x\} = \{X_1 = x_1\} \cap \dots \cap \{X_m = x_m\}$$

Inoltre

Prodotto cartesiano

$$\text{Range } X \subseteq (\text{Range } X_1) \times (\text{Range } X_2) \times \dots \times (\text{Range } X_m)$$

sono insiemi al più numerabili perché le X_i sono v.a. discrete

$\Rightarrow$ anche Range X è al più numerabile

Si definisce anche in questo caso la densità di $X = (X_1, \dots, X_m)$:

$$p(x) = P\{X = x\} \qquad x \in \mathbb{R}^m$$

Analogamente al caso scalare:

i) $p(x) \in [0,1]$

ii) $p(x) > 0$ solo per al più un'infinità numerabile di valori

$$x^{(1)}, x^{(2)}, x^{(3)}, \dots, x^{(k)}, \dots$$

$$x^{(k)} = (x_1^{(k)}, x_2^{(k)}, \dots, x_m^{(k)})$$

iii) $\sum_k p(x^{(k)}) = 1$

Si prende $A \subset \mathbb{R}^m$

$$P\{X \in A\} = P\left(\bigcup_{x \in A} \{X = x\} \right) =$$

$$= \sum_{x \in A} P\{X = x\} = \sum_{x \in A} p(x)$$

Definizione: Sia $X = (X_1, \dots, X_m)$ v.a. discreta la densità $p(x)$ di X si dice DENSITÀ CONGIUNTA delle v.a. $X_1, X_2, \dots, X_m$.

Le densità $p_1, \dots, p_m$ delle v.a. $X_1, \dots, X_m$ vengono chiamate DENSITÀ MARGINALI DI X

Conoscendo la densità di $X = (X_1, \dots, X_m)$ si può risalire alle densità marginali delle componenti

CASO $m=2$ $X = (X_1, X_2)$

$$p(x) > 0 \iff x \in \{x^{(1)}, x^{(2)}, \dots, x^{(k)}, \dots\}$$

$$x^{(k)} = (x_1^{(k)}, x_2^{(k)})$$

$$p_1(z) = P\{X_1 = z\} = P\left(\bigcup_{k=1}^{\infty} \{X_1 = z, X_2 = x_2^{(k)}\} \right) =$$

$$= \sum_{k=1}^{\infty} P\left(X_1 = z, X_2 = x_2^{(k)} \right) =$$

$$= \sum_{k=1}^{\infty} P\left(X = (z, x_2^{(k)}) \right) = \sum_k p(z, x_2^{(k)})$$

In modo simmetrico:

$$p_2(z) = P(X_2 = z) = \sum_k p(x_1^{(k)}, z)$$

Invece, in generale, senza ulteriori ipotesi su X_1, X_2, dalla conoscenza di p_1 e p_2 NON si risale alla densità congiunta di (X_1, X_2).

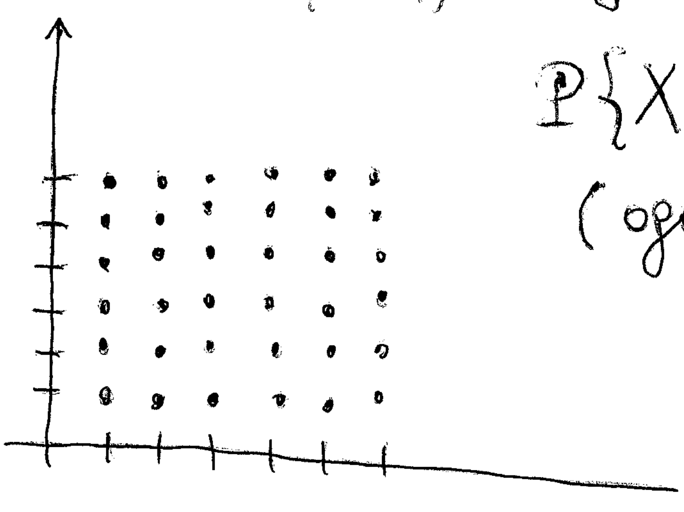
$\boxed{\text{Esempio}}$ Consideriamo un'urna con 6 palline numerate da $1, \dots, 6$

Estraiamone 2

a) con rimpiazzo: $X = (X_1, X_2)$

b) senza rimpiazzo: $Y = (Y_1, Y_2)$

$\boxed{\text{Caso a}}$ $X \in \{(i,j) : i,j = 1, 2, \dots, 6\}$



$$P\{X = (i,j)\} = \frac{1}{36} = p(i,j)$$

(ogni estrazione è ugualmente probabile)

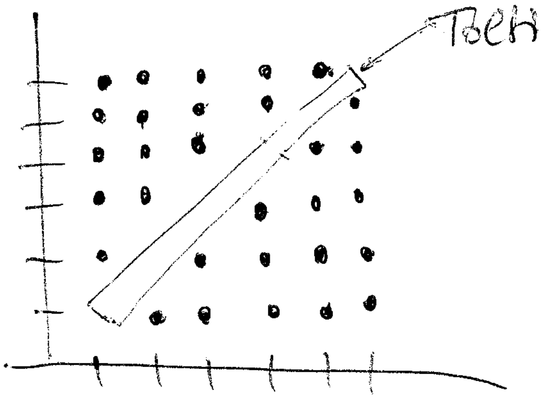
Densità marginale di X_1:

$$P_1(i) = \mathbb{P}(X_1 = i) = \sum_{j=1}^{6} p(i,j) = \sum_{j=1}^{6} \frac{1}{36} = \frac{1}{6}$$
$i = 1, 2, \dots, 6$

Similmente $$P_2(j) = \sum_{i=1}^{6} p(i,j) = \frac{1}{6}$$
$j = 1, 2, \dots, 6$

Sia X_1 che X_2 sono uniformemente distribuite su $\{1, 2, \dots, 6\}$.

$\boxed{\text{Caso b}}$ $Y \in \{(i,j) : i,j = 1, \dots, 6,\ i \neq j\}$



$$p(i,j) = \mathbb{P}(Y = (i,j)) = \frac{1}{30} \qquad i,j = 1, \dots, 6,\ \boxed{i \neq j}$$

Densità marginale di Y_1

$$p_1(i) = \mathbb{P}(Y_1 = i) = \sum_{\substack{j=1 \\ j \neq i}}^{6} p(i,j) = \sum_{\substack{j=1 \\ j \neq i}}^{6} \frac{1}{30} = 5 \cdot \frac{1}{30} = \frac{1}{6}$$
$i = 1, \dots, 6$

$$P_2(j) = \mathbb{P}(Y_2 = j) = \sum_{\substack{i=1 \\ i \neq j}}^{6} \frac{1}{30} = 5 \cdot \frac{1}{30} = \frac{1}{6} \qquad j = 1, \dots, 6$$

Di nuovo sono venute densità uniformi su $\{1, \dots, 6\}$

Quindi ho ottenuto due v.a. X e Y <u>diverse</u> (con diverse densità) ma che hanno le stesse marginali!!!

V.A. INDIPENDENTI

<u>Def</u> Siano $X_1, \dots, X_m$ v.a. discrete. Si dicono INDIPENDENTI se

$$\mathbb{P}\{X_1 \in A_1, X_2 \in A_2, \dots, X_m \in A_m\} = \mathbb{P}\{X_1 \in A_1\} \cdot \dots \cdot \mathbb{P}\{X_m \in A_m\}$$

per ogni possibile scelta di $A_1, \dots, A_m \subseteq \mathbb{R}$

[Equivale a chiedere che gli eventi $\{X_1 \in A_1\}, \dots, \{X_m \in A_m\}$ siano indipendenti]

<u>NOTA</u>- Similmente, infinite v.a. $\{X_n\}_{n \in \mathbb{N}}$ si dicono <u>indipendenti</u> se $X_1, \dots, X_m$ sono indipendenti $\forall m$.

- Se $X = (X_1, \dots, X_m)$, e $X_1, \dots, X_m$ sono indipendenti $x \in \mathbb{R}^m$, $x = (x_1, \dots, x_m)$, allora:

$$p(x) = \mathbb{P}\{X = x\} = \mathbb{P}\{X_1 = x_1, X_2 = x_2, \dots, X_m = x_m\} =$$

$$= \mathbb{P}\{X_1 = x_1\} \cdot \mathbb{P}\{X_2 = x_2\} \cdot \dots \cdot \mathbb{P}\{X_m = x_m\} =$$

per l'indipendenza delle v.a.

$$= p_1(x_1) \cdot \dots \cdot p_m(x_m)$$

densità marginali

Prop: $X_1, \dots, X_m$ sono indipendenti SE E SOLO SE la loro densità congiunta è il prodotto delle loro densità marginali $p_1, \dots, p_m$.

Esempio dell'urna con e senza rimpiazzo

$X = (X_1, X_2)$ (con rimpiazzo)

$$\frac{1}{36} = p(i,j) = p_1(i)\, p_2(j) = \frac{1}{6} \cdot \frac{1}{6}$$

$\Rightarrow$ X_1 e X_2 sono indipendenti

$Y = (Y_1, Y_2)$ (senza rimpiazzo)

$$p(i,j) = \frac{1}{30} \neq \frac{1}{6} \cdot \frac{1}{6} = p_1(i) \cdot p_2(j)$$

e Y_1 e Y_2 non sono indipendenti

Prop1. Se X e Y sono indipendenti e prendo due funzioni $\phi, \psi : \mathbb{R} \to \mathbb{R}$, allora le v.a. $\phi(X)$ e $\psi(Y)$ sono indipendenti

Prop2 Se $X_1, \dots, X_m, Y_1, \dots, Y_k$ v.a. indipendenti e $\phi : \mathbb{R}^m \to \mathbb{R}$, $\psi : \mathbb{R}^k \to \mathbb{R}$, allora le v.a. $\phi(X_1, \dots, X_m)$ e $\psi(Y_1, \dots, Y_k)$ sono indipend.

Esercizio - Calcolare la probabilità che, da un'urna contenente 6 palline numerate da 1 a 6, se ne estraggano <u>senza rimpiazzo</u> due i cui corrispondenti numeri differiscano al più di 2.

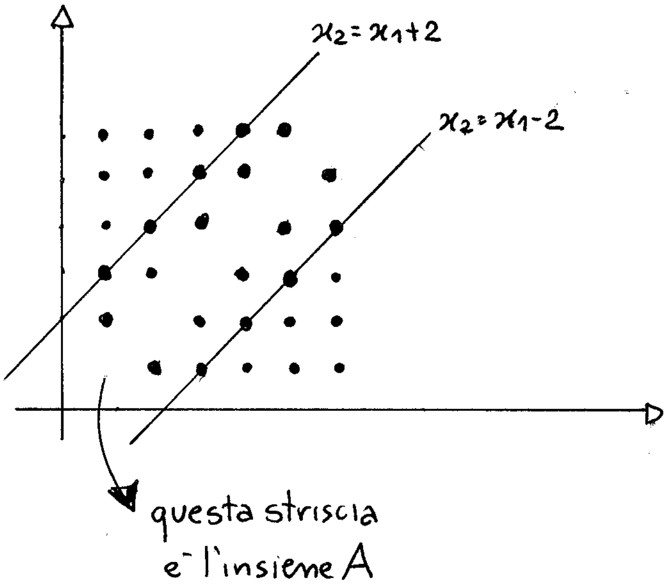
V.a. $Y=(Y_1, Y_2)$ distribuita come nel caso b) precedente

$x_1, x_2 \in \mathbb{R}$ differiscono al più di 2 se e solo se

$$|x_1 - x_2| \leq 2 \iff x_1 - 2 \leq x_2 \leq x_1 + 2$$

L'evento di cui cerco la probabilità si può scrivere come:

$$\{Y \in A\} \quad \text{dove} \quad A = \left\{ (x_1, x_2) \in \mathbb{R}^2 : \ x_1 - 2 \leq x_2 \leq x_1 + 2 \right\}$$



$$P\{Y \in A\} = \sum_{(x_1, x_2) \in A} P\{Y = (x_1, x_2)\} = 18 \cdot \frac{1}{30} = \frac{3}{5}$$

Esercizio 2 - Si hanno una moneta e un dado non truccati. Calcolare la probabilità che esca testa prima che esca "6" (NOTA - si suppone di lanciare insieme ripetutamente la moneta e il dado)

Consideriamo le due v.a:

$$\left.\begin{array}{l} X = \text{primo lancio della moneta in cui è uscita testa} \\ Y = \text{primo lancio del dado in cui è uscito "6"} \end{array}\right\} \text{SONO INDIPENDENTI!}$$

Si osservi che X segue una legge <u>geometrica</u> di parametro $\frac{1}{2}$, mentre Y segue una legge <u>geometrica</u> di parametro $\frac{1}{6}$. Dunque:

$$p_X(k) = \begin{cases} \frac{1}{2}\left(1-\frac{1}{2}\right)^{k-1} = \left(\frac{1}{2}\right)^k & \text{se } k = 1,2,3,\dots \\ 0 & \text{altrimenti} \end{cases}$$

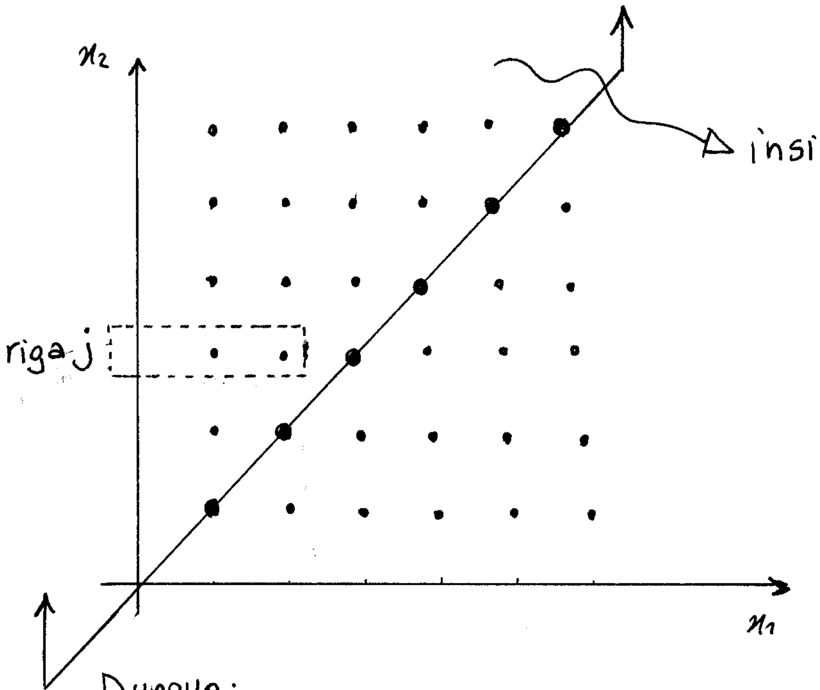
$$p_Y(k) = \begin{cases} \frac{1}{6}\left(1-\frac{1}{6}\right)^{k-1} = \frac{1}{6}\left(\frac{5}{6}\right)^{k-1} & \text{se } k = 1,2,3,\dots \\ 0 & \text{altrimenti} \end{cases}$$

Ci interessa calcolare $P\{X<Y\} = P\{(X,Y)\in A\}$

dove A è l'insieme: $A = \{(x_1,x_2)\in\mathbb{R}^2 : x_1 < x_2\}$

Ricaviamo la densità congiunta di X e Y. Essendo indipendenti:

$$p(i,j) = p_X(i)\,p_Y(j) = \begin{cases} \frac{1}{6}\left(\frac{1}{2}\right)^{i}\left(\frac{5}{6}\right)^{j-1} & \text{se } i,j = 1,2,3,\dots \\ 0 & \text{altrimenti} \end{cases}$$



$$P\{(X,Y)\in A\} = \sum_{i<j} p(i,j) = \sum_{j=2}^{\infty}\sum_{i=1}^{j-1} p(i,j)$$

Scompongo l'insieme delle coppie $(i,j)\in A$ per righe orizzontali

Dunque:

$$P\{X<Y\} = \frac{1}{6}\sum_{j=2}^{\infty}\left(\frac{5}{6}\right)^{j-1}\underbrace{\sum_{i=1}^{j-1}\left(\frac{1}{2}\right)^{i}} = \frac{1}{6}\sum_{j=2}^{\infty}\left(\frac{5}{6}\right)^{j-1}\left[1-\left(\frac{1}{2}\right)^{j-1}\right] = \dots$$

$$\sum_{k=0}^{j-2}\left(\frac{1}{2}\right)^{k+1} = \frac{1}{2}\underbrace{\sum_{k=0}^{j-2}\left(\frac{1}{2}\right)^{k}} = \frac{1}{2}\,\frac{1-\left(\frac{1}{2}\right)^{j-1}}{1-\frac{1}{2}} = 1-\left(\frac{1}{2}\right)^{j-1}$$

si ricordi che: $\sum_{k=0}^{N} x^k = \frac{1-x^{N+1}}{1-x}$

$$\dots = \frac{1}{6}\left[\sum_{j=2}^{\infty}\left(\frac{5}{6}\right)^{j-1} - \sum_{j=2}^{\infty}\left(\frac{5}{12}\right)^{j-1}\right] = \frac{1}{6}\left(\frac{\frac{5}{6}}{1-\frac{5}{6}} - \frac{\frac{5}{12}}{1-\frac{5}{12}}\right) = \dots$$

<u>NOTA</u> - $\displaystyle\sum_{j=2}^{\infty} x^{j-1} = \sum_{k=0}^{\infty} x^{k+1} = x \underbrace{\sum_{k=0}^{\infty} x^k}_{\text{serie geometrica}} = x\frac{1}{1-x} = \frac{x}{1-x}$

$$\dots = \frac{1}{6}\left(5 - \frac{5}{7}\right) = \frac{5}{\cancel{6}}\frac{\cancel{6}}{7} = \frac{5}{7}$$

SPERANZA MATEMATICA (VALORE ATTESO, MEDIA)

Def Sia X una v.a. discreta che assume i valori $x_1, x_2, \dots, x_k, \dots$ e sia $p(x)$ la sua densità discreta di probabilità. Allora si dice che X ha <u>speranza matematica</u> finita se

$$\sum_k |x_k|\, p(x_k) < +\infty$$

(cioè, in termini di Analisi I, $\sum_k x_k\, p(x_k)$ converge assolutamente).
In tal caso si pone:

$$E[X] \triangleq \sum_k x_k\, p(x_k)$$

<u>Valore atteso di X</u> (in inglese, "expected value")

<u>Esempio</u> - X = risultato del lancio di un dado a 6 facce non truccato

$p(k) = \frac{1}{6}$, $k = 1, 2, \dots, 6$

$$E[X] = 1\cdot\frac{1}{6} + 2\cdot\frac{1}{6} + 3\cdot\frac{1}{6} + 4\cdot\frac{1}{6} + 5\cdot\frac{1}{6} + 6\cdot\frac{1}{6}$$

$$= \frac{1+2+3+4+5+6}{6}$$

coincide con la media aritmetica dei valori assunti da X!

$\Rightarrow E[X] = 3.5$

Esempio - $X \sim B(1,p)$, cioè $X \in \{0,1\}$ e $P\{X=1\}=p$.

Allora $E[X] = 0 \cdot (1-p) + 1 \cdot p = p$.

Valore atteso di funzioni di variabili aleatorie

Prop: Sia $X = (X_1, \dots, X_m)$ una v.a. discreta che assume i valori $x^{(1)}, x^{(2)}, \dots, x^{(k)}, \dots$, e sia $\phi: \mathbb{R}^m \to \mathbb{R}$ una funzione.
Si consideri la variabile aleatoria $Z = \phi(X)$. Sia inoltre $p_X(x)$ la densità di probabilità congiunta di $X_1, \dots, X_m$. Allora:

1) Z ha speranza matematica finita se e solo se

$$\sum_k |\phi(x^{(k)})| \, p_X(x^{(k)}) < +\infty$$

2) in questo caso:

$$E[Z] = \sum_k \phi(x^{(k)}) \, p_X(x^{(k)})$$

NOTA - Si osservi che, per calcolare $E[Z]$ dalla definizione, cioè:

$$E[Z] = \sum_h z_h \, p_Z(z_h),$$

dove z_h sono i valori distinti assunti da Z, avremmo dovuto calcolare prima la densità di probabilità di Z, cioè $p_Z(z)$.

Esempio - $Z = |X|$. Quanto vale $E[Z]$?

$\rightsquigarrow \phi(x) = |x|$

$$\Rightarrow E[Z] = \sum_k |x_k| \, p_X(x_k)$$

Quindi X ha speranza matematica finita se e solo se $E[|X|] < +\infty$.

PROPRIETÀ DEL VALORE ATTESO

1) $E[X+Y] = E[X] + E[Y]$

dim. $\phi(x,y) = x+y$

densità congiunta di X e Y

$$E[X+Y] = \sum_{i,j} \phi(x_i, y_j)\, p(x_i, y_j) =$$

$$= \sum_{i,j} (x_i + y_j)\, p(x_i, y_j) =$$

$$= \sum_{i,j} x_i\, p(x_i, y_j) + \sum_{i,j} y_j\, p(x_i, y_j) =$$

$$= \sum_i x_i \underbrace{\sum_j p(x_i, y_j)}_{\substack{p_X(x_i) \\ \text{marginale di } X}} + \sum_j y_j \underbrace{\sum_i p(x_i, y_j)}_{\substack{p_Y(y_j) \\ \text{marginale di } Y}} = \dots$$

$$\dots = \sum_i x_i\, p_X(x_i) + \sum_j y_j\, p_Y(y_j) = E[X] + E[Y]$$

2) $E[cX] = c\,E[X]$, $c \in \mathbb{R}$

NOTA - 1) e 2) $\Rightarrow$ l'operatore $E[\cdot]$ è lineare.

Esempio - Una variabile $X \sim B(n,p)$ [numero di successi in n prove indipendenti, in ciascuna delle quali la probabilità di successo è p] può essere scritta come:

$$X = X_1 + \dots + X_n$$

dove $X_i \sim B(1,p)$, $i = 1, \dots, n$.

$$X_i = \begin{cases} 0 & \text{con probabilità } 1-p \\ 1 & \text{con probabilità } p \end{cases}$$ (risultato della i-esima prova)

Dunque: $E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p = np.$

OSSERVAZIONE - $E[X]$ dipende solo dalla densità $p(x)$ di X, cioè due v.a. con uguale densità hanno lo stesso valore atteso.

Esercizio - Provare a calcolare $E[X]$ per una v.a. X che segue una legge geometrica di parametro p:

$$p(k) = \begin{cases} p(1-p)^{k-1} & k = 1, 2, 3, \dots \\ 0 & \text{altrimenti} \end{cases}$$

SUGGERIMENTO: si ricordi che $\sum_{k=0}^{\infty} k x^k = \dfrac{x}{(1-x)^2}$.