DISTRIBUZIONI CONGIUNTE

Nei casi trattati, gli esiti di un esperimento erano considerati realizzazioni di una singola variabile casuale. Tuttavia, in alcune situazioni può essere necessario o desiderabile ottenere esiti simultanei da più variabili casuali.

Bisogna perciò estendere il concetto di v. casuale, di funzione di ripartizione, di funzione di densità di una v. casuale al caso n-dimensionale.

DEFINIZIONE: Chiamiamo variabile casuale n-dimensionale una funzione $X=(X_1,\dots,X_n):\Omega\to\mathbb{R}^n$ tale che $\forall\,(x_1,\dots,x_n)\in\mathbb{R}^n$

$$\{\omega\in\Omega : X_1(\omega)\le x_1,\dots, X_n(\omega)\le x_n\} \text{ è un evento.}$$

Quindi una v.c. n-dim è un' n-pla di v. casuali, le quali associano un numero ad un risultato.

DEFINIZIONE: Chiamiamo funzione di ripartizione congiunta di $X_1,\dots,X_n$, la funzione

$$F_{X_1,\dots,X_n}:\mathbb{R}^n\to[0,1]$$

tale che $\forall\; x=(x_1,\dots,x_n)$

$$F_{X_1,\dots,X_n}(x_1,\dots,x_n)=P\left[X_1\le x_1,\dots, X_n\le x_n\right]$$

Come nel caso 1-D, per definire la funzione di densità bisogna distinguere il caso discreto dal caso continuo. Noi ci occuperemo solo di v. casuali m-dim discrete.

DEFINIZIONE: La v. casuale $X=(X_1,\dots,X_m)$ è detta **v.c. discreta m-dimensionale** se può assumere valori solo in un insieme numerabile $(x_1,\dots,x_m)$ di punti di $\mathbb{R}^n$.

DEFINIZIONE: Chiamiamo **funzione di densità discreta congiunta** di $(X_1,\dots,X_m)$, v.c. discreta n-dim, la funzione

$$f_{X_1,\dots,X_m}(x_1,\dots,x_m)=P[X_1=x_1,\dots X_m=x_m]$$

- $f_{X_1,\dots,X_m}(x_1,\dots,x_m)\geq 0 \quad \forall (x_1,\dots,x_m)$
- $\sum_{x_1}\sum_{x_2}\dots\sum_{x_m} f_{X_1,\dots,X_m}(x_1,\dots,x_m)=1$
- $\forall$ regione $A\subset\mathbb{R}^m$ $P[(X_1,\dots,X_m)\in A]=\sum_A f_{X_1,\dots,X_m}(x_1,\dots,x_n)$

Nel caso $m=2$ elenchiamo le proprietà della funzione di ripartizione congiunta relative alla variabile casuale discreta bidimensionale (X,Y).

PROPRIETÀ DI $F_{X,Y}(x,y)$

- $\lim_{x\to-\infty} F(x,y)=0 \ \forall y$; $\lim_{y\to-\infty} F(x,y)=0 \ \forall x$
- $\lim_{\substack{x\to+\infty \\ y\to+\infty}} F(x,y)=1$
- se $x_1<x_2$ e $y_1<y_2$ allora
$$P[x_1\leq X\leq x_2 \,;\, y_1\leq Y\leq y_2]=F(x_2,y_2)-F(x_2,y_1)-F(x_1,y_2)+F(x_1,y_1)\geq 0$$
- $\lim_{h\to0^+} F(x+h,y)=\lim_{h\to0^+} F(x,y+h)=F(x,y)$ continuità da destra

DEFINIZIONE: Date X, Y v. casuali congiunte discrete

$$F_{X,Y}(\bar{x}, \bar{y}) = \sum_{x_i \leq \bar{x}}' \sum_{y_i \leq \bar{y}} f_{X,Y}(x_i, y_i)$$

dove x_i, y_i sono i punti massa per le v.c. X e Y.

Data la funzione di densità congiunta $f_{X,Y}(x,y)$ delle v.c. discrete congiunte X, Y è possibile ricavare le funzioni di densità $f_X(\cdot)$ di X e $f_Y(\cdot)$ di Y, dette f. di densità **marginali**.

DEFINIZIONE: Le funzioni di densità **marginali** di X e di Y sono:

$$f_X(x_K) = \sum_J' f_{X,Y}(x_K, y_J)$$

$$f_Y(y_K) = \sum_J' f_{X,Y}(x_J, y_K)$$

N.B. Dalla f. di densità congiunta è sempre possibile ricavare le f. di densità marginali, ma non vale il viceversa.

Ricordando la definizione di probabilità condizionata

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \text{se } P(B) > 0$$

se ora A è l'evento definito da $X = x$ e B è l'evento definito da $Y = x$, allora

$$P[X = x \mid Y = y] = \frac{P[X = x, Y = y]}{P[Y = y]}, \quad P[Y = y] > 0$$

allora è possibile dare la definizione formale di funzione di densità condizionata.

DEFINIZIONE: Date X, Y v.c. discrete congiunte con funzione di densità congiunta $f_{X,Y}$, chiamiamo funzione di densità condizionata di X, dato $Y = y$, la funzione:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \quad se\ f_Y(y) > 0$$

analogamente si definisce la funzione di densità condizionata di Y, dato $X = x$, la funzione

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}, \quad se\ f_X(x) > 0$$

Conseguentemente è possibile definire le funzioni di ripartizione condizionate di Y, dato $X=x$ e di X, dato $Y=y$.

$$F_{Y|X}(y|x) = P[Y \leq y \mid X = x] \quad per\ f_X(x) > 0$$

$$= \sum_{\{j:\ y_j \leq y\}} f_{Y|X}(y_j|x)$$

e

$$F_{X|Y}(x,y) = P[X \leq x \mid Y = y] \quad per\ f_Y(y) > 0$$

$$= \sum_{\{j:\ x_j \leq x\}} f_{X,Y}(x_j|y)$$

N.B. $\begin{cases} f_{X,Y}(x,y) = f_{X|Y}(x|y) \cdot f_Y(y) \\ f_{X,Y}(x,y) = f_{Y|X}(y|x) \cdot f_X(x) \end{cases}$

Esempio: Lancio di due tetraedri (poliedri regolari a 4 facce) aventi facce numerate da 1 a 4.

X v.c. che indica il n° sulla faccia rivolta verso il basso del 1° tetraedro.

Y v.c. che indica il n° più grande fra quelli indicati sulle facce rivolte verso il basso.

I valori congiunti di X e di Y sono:

(1,1) (1,2) (1,3) (1,4)
(2,2) (2,3) (2,4)
(3,3) (3,4)
(4,4)

Vogliamo calcolare la funzione di densità congiunta.

1°T	2°T	
1	1	⇒ (1,1)
1	2	⇒ (1,2)
1	3	⇒ (1,3)
1	4	⇒ (1,4)
2	1	⇒ (2,2)
2	2	⇒ (2,2)
2	3	⇒ (2,3)
2	4	⇒ (2,4)
3	1	⇒ (3,3)
3	2	⇒ (3,3)
3	3	⇒ (3,3)
3	4	⇒ (3,4)

1°T	2°T	
4	1	(4,4)
4	2	(4,4)
4	3	(4,4)
4	4	(4,4)

Ω spazio campione è formato da 16 elementi (matrice 4×4).

in quanti modi ottengo il risultato $(2,2)$? $\frac{2}{16}$

in quanti modi ottengo il risultato $(3,3)$? $\frac{3}{16}$

in quanti modi ottengo il risultato $(4,4)$? $\frac{4}{16}$

Posso riassumere tutti i risultati in una tabella:

(x,y)	$(1,1)$	$(1,2)$	$(1,3)$	$(1,4)$	$(2,2)$	$(2,3)$	$(2,4)$	$(3,3)$	$(3,4)$	$(4,4)$
$f_{X,Y}(x,y)$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{4}{16}$

$$\sum f_{X,Y}(x,y) = 1$$

Dalla definizione di funzione di ripartizione congiunta si ha:

$$\begin{cases} F_{X,Y}(1,1) = f_{X,Y}(1,1) = \frac{1}{16} \\ F_{X,Y}(1,2) = f_{X,Y}(1,1) + f_{X,Y}(1,2) = \frac{1}{16} + \frac{1}{16} = \frac{2}{16} \\ F_{X,Y}(1,3) = f_{X,Y}(1,1) + f_{X,Y}(1,2) + f_{X,Y}(1,3) = \frac{1}{16} + \frac{1}{16} + \frac{1}{16} = \frac{3}{16} \\ F_{X,Y}(1,4) = f_{X,Y}(1,1) + f_{X,Y}(1,2) + f_{X,Y}(1,3) + f_{X,Y}(1,4) = \frac{4}{16} \end{cases}$$

$$\begin{cases} F_{X,Y}(2,1) = f_{X,Y}(1,1) = \frac{1}{16} \\ F_{X,Y}(2,2) = f_{X,Y}(1,1) + f_{X,Y}(1,2) + f_{X,Y}(2,2) = \frac{4}{16} \\ F_{X,Y}(2,3) = f_{X,Y}(1,1) + f_{X,Y}(1,2) + f_{X,Y}(1,3) + f_{X,Y}(2,2) \\ \qquad + f_{X,Y}(2,3) = \frac{6}{16} \\ F_{X,Y}(2,4) = \frac{8}{16} \end{cases}$$

$$\begin{cases} F_{X,Y}(3,1) = f_{X,Y}(1,1) = \frac{1}{16} \\ F_{X,Y}(3,2) = f_{X,Y}(1,1) + f_{X,Y}(1,2) + f_{X,Y}(2,2) = \frac{4}{16} \\ F_{X,Y}(3,3) = f_{X,Y}(1,1) + f_{X,Y}(1,2) + f_{X,Y}(1,3) + f_{X,Y}(2,2) \\ \qquad + f_{X,Y}(2,3) + f_{X,Y}(3,3) = \frac{9}{16} \\ F_{X,Y}(3,4) = \frac{12}{16} \end{cases}$$

$$\begin{cases} F_{X,Y}(4,1) = \frac{1}{16} \\ F_{X,Y}(4,2) = \frac{4}{16} \\ F_{X,Y}(4,3) = \frac{9}{16} \\ F_{X,Y}(4,4) = 1 \end{cases}$$

Ricorda che $F_{X,Y}(\bar{x},\bar{y}) = \sum_{\substack{x_i \leq \bar{x} \\ y_i \leq \bar{y}}} f_{X,Y}(x_i, y_i)$

Dalla definizione di densità marginale si ha:

$$\begin{cases} f_Y(1) = \sum_{x \leq 1} f_{X,Y}(x, 1) = 1/16 \\ f_Y(2) = \sum_{x \leq 2} f_{X,Y}(x, 2) = 3/16 \\ f_Y(3) = \sum_{x \leq 3} f_{X,Y}(x, 3) = 5/16 \\ f_Y(4) = \sum_{x \leq 4} f_{X,Y}(x, 4) = 7/16 \end{cases}$$

$f_X(1) = \sum_{1 \leq y \leq 4} f_{X,Y}(1, y) = 4/16$

$f_X(2) = \sum_{2 \leq y \leq 4} f_{X,Y}(2, y) = 4/16$

$f_X(3) = \sum_{3 \leq y \leq 4} f_{X,Y}(3, y) = 4/16$ $\qquad f_X(4) = f_{X,Y}(4,4) = 4/16$

Tabella dei valori di $F_{X,Y}$

$4 \leq y$	0	4/16	8/16	12/16	1
$3 \leq y < 4$	0	3/16	6/16	9/16	9/16
$2 \leq y < 3$	0	2/16	4/16	4/16	4/16
$1 \leq y < 2$	0	1/16	1/16	1/16	1/16
$y < 1$	0	0	0	0	0
	$x < 1$	$1 \leq x < 2$	$2 \leq x < 3$	$3 \leq x < 4$	$4 \leq x$

Tabella dei valori di $f_{X,Y}$

4	1/16	1/16	1/16	4/16		$\rightarrow f_Y(4)$
3	1/16	1/16	3/16			$\rightarrow f_Y(3)$
2	1/16	2/16				$\rightarrow f_Y(2)$
1	1/16					$\rightarrow f_Y(1)$
y/x	1	2	3	4	1	
	$\downarrow f_X(1)$	$\downarrow f_X(2)$	$\downarrow f_X(3)$	$\downarrow f_X(4)$		

OSSERVAZIONE: Le densità marginali si calcolano più semplicemente sommando per righe o per colonne i valori di $f_{X,Y}$ della tabella a doppia entrata.

Vogliamo calcolare ora le funzioni di densità condizionate di Y dato $X=2$ e dato $X=3$

$$f_{Y|X}(y|2) = \frac{f_{X,Y}(2,y)}{f_X(2)}$$

se $X=2 \Rightarrow Y \geq 2$

$$f_{X,Y}(2|2) = \frac{f_{X,Y}(2,2)}{f_X(2)} = \frac{2/16}{4/16} = \frac{1}{2}$$

$$f_{X,Y}(3|2) = \frac{f_{X,Y}(2,3)}{f_X(2)} = \frac{1/16}{4/16} = \frac{1}{4}$$

$$f_{X,Y}(4|2) = \frac{f_{X,Y}(2,4)}{f_X(2)} = \frac{1/16}{4/16} = \frac{1}{4}$$

$\Sigma = 1$

se $X=3 \Rightarrow Y \geq 3$

$$f_{Y|X}(3|3) = \frac{f_{X,Y}(3,3)}{f_X(3)} = \frac{3/16}{4/16} = \frac{3}{4}$$

$$f_{Y|X}(4|3) = \frac{f_{X,Y}(3,4)}{f_X(3)} = \frac{1/16}{4/16} = \frac{1}{4}$$

$\Sigma = 1$

Per esercizio calcolare $f_{Y|X}$ dati $X=1$ e $X=4$.

Quanto vale $f_{X|Y}$ dati $Y = 1, 2, 3, 4$?

INDIPENDENZA

DEF: Data $X = (X_1, .., X_m)$ v.c. m-dim discreta (o continua) con funzione di densità congiunta $f_{X_1,..,X_m}$, diciamo che $X_1, ..., X_m$ sono v.c. INDIPENDENTI sse

$$f_{X_1,..,X_m}(x_1,..,x_m) = f_{X_1}(x_1) \cdots f_{X_m}(x_m)$$

per tutti gli $x_1, .., x_m$.

Cioè la f. di densità congiunta si scrive come PRODOTTO delle f. di densità marginali.

Nel caso 2-D si ha:

X, Y indipendenti sse $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$.

Dalla definizione di densità condizionata si ha:

$$f_{X,Y}(x,y) = f_{Y|X}(y|x)\, f_X(x)$$

perciò $f_{Y|X}(y|x) = f_Y(y)$

cioè la densità condizionata di Y dato $X = x$ è la densità non condizionata di Y.

Per mostrare che due v.c. non sono indipendenti basta mostrare che $f_{Y|X}(y|x)$ dipende da x.

OSSERVAZIONE

Si può provare che se $X_1, .., X_m$ sono v.c. indipendenti e $g_1, .., g_m$ sono m funzioni tali che $Y_k = g_k(X_k)$ $\forall\, k = 1, .., m$ siano v.c. allora $Y_1, .., Y_m$ sono indipendenti.

Nell' esempio del lancio dei due tetraedri, alla domanda "Le v.c. X, Y sono indipendenti?"

La risposta è ovviamente **NO!**

Basta trovare una coppia di valori (x, y) per i quali p. esempio:

$$f_{Y|X}(y|x) \neq f_Y(y)$$

se $X=3$, $Y=2$

$f_{Y|X}(2|3) = P[Y=2 | X=3] \equiv 0$ poichè $Y < X$.

e $f_Y(2) = P[Y=2] = \frac{3}{16}$

Esempio

Da un gruppo di 12 batterie (3 nuove, 4 usate, 5 difettose) ne vengono scelte 3 a caso. Indicato con

X = n° batterie nuove

Y = n° batterie usate

tra quelle scelte, determinare la $f_{X,Y}(x,y)$.

Possibili risultati:

NNN $\rightarrow f(3,0)$

NNU $\rightarrow f(2,1)$

NND $\rightarrow f(2,0)$

NUU $\rightarrow f(1,2)$

NDD $\rightarrow f(1,0)$

NUD $\rightarrow f(1,1)$

UUU $\rightarrow f(0,3)$

UUD $\rightarrow f(0,2)$

UDD $\rightarrow f(0,1)$

DDD $\rightarrow f(0,0)$

(non conta l'ordine).

$$f(3,0) = \frac{\binom{3}{3}}{\binom{12}{3}} = \dots = \frac{1}{220}$$

$$f(2,1) = \frac{\binom{3}{2}\binom{4}{1}}{\binom{12}{3}} = \dots = \frac{12}{220}$$

$$f(2,0) = \frac{\binom{3}{2}\binom{5}{1}}{\binom{12}{3}} = \dots = \frac{15}{220}$$

$$f(1,2) = \frac{\binom{3}{1}\binom{4}{2}}{\binom{12}{3}} = \dots = \frac{18}{220}$$

$$f(1,0) = \frac{\binom{3}{1}\binom{5}{2}}{\binom{12}{3}} = \dots = \frac{30}{220}$$

$$f(1,1) = \frac{\binom{3}{1}\binom{4}{1}\binom{5}{1}}{\binom{12}{3}} = \dots = \frac{60}{220}$$

$$f(0,3) = \frac{\binom{4}{3}}{\binom{12}{3}} = \frac{4}{220}$$

$$f(0,2) = \frac{\binom{4}{2}\binom{5}{1}}{\binom{12}{3}} = \frac{30}{220}$$

$$f(0,1) = \frac{\binom{4}{1}\binom{5}{2}}{\binom{12}{3}} = \frac{40}{220}$$

$$f(0,0) = \frac{\binom{5}{3}}{\binom{12}{3}} = \frac{10}{220}$$

Tabella di $f_{X,Y}$.

	$Y=0$	$Y=1$	$Y=2$	$Y=3$	$f_X(x)$
$X=0$	$1/22$	$2/11$	$3/22$	$1/55$	$42/110$
$X=1$	$3/22$	$3/11$	$9/110$	0	$27/55$
$X=2$	$15/220$	$3/55$	0	0	$27/220$
$X=3$	$1/220$	0	0	0	$1/220$
$f_Y(y)$	$14/55$	$28/55$	$12/55$	$1/55$	1

MODELLO DI V.A. N-DIM (caso discreto)

DISTRIBUZIONE MULTINOMIALE

Tale distribuzione è associata a prove ripetute e indipendenti, che generalizzano il caso delle prove di Bernoulli a 2 esiti a quello con più di due esiti.

Supponiamo che esistano $k+1$ esiti possibili distinti di un tentativo. Siano $s_1, \dots s_{k+1}$ tali esiti.

Sia $p_i = P[s_i]$ $i = 1, \dots, k+1$ con

$$\sum_{i=1}^{k+1} p_i = 1 \Rightarrow p_{k+1} = 1 - (p_1 + \dots + p_k)$$

Ripetiamo la prova n volte.

Sia X_i il n° di volte che si ottiene s_i sugli n tentativi, per $i = 1, \dots, k+1$.

Se le prove sono ripetute e indipendenti si ha:

$$f_{X_1, \dots, X_k}(x_1, \dots, x_k) = \frac{n!}{x_1! \cdots x_{k+1}!} p_1^{x_1} \cdots p_{k+1}^{x_{k+1}}$$

dove $\sum_{i=1}^{k+1} x_i = n$ e $X_{k+1} = n - \sum_{i=1}^{k} X_i$.

Pertanto cerco la probabilità che da n prove si abbiano esattamente x_1 esiti del tipo s_1, x_2 esiti del tipo s_2, ... x_{k+1} esiti del tipo s_{k+1}.

Un particolare ordinamento ha probabilità

$$p_1^{x_1} \cdots p_{k+1}^{x_{k+1}}$$

e di ordinamenti ne esistono:

$$\frac{n!}{x_1! \cdots x_{k+1}!}$$

Esempio

Su un quantitativo di merce, il 10% viene pagato in ritardo, il 30% viene restituito.

Vengono effettuati 20 ordini.

Calcolare la probabilità che 3 ordini vengano pagati in ritardo e 5 vengano restituiti (su un totale di 20).

IPOTESI: indipendenza

$n = 20$

$x_1 = 3$, $x_2 = 5$ $x_3 = n - x_1 - x_2 = 20 - 3 - 5 = 12$

$\downarrow$ $\downarrow$ $\downarrow$

$p_1 = 0.1$ $p_2 = 0.3$ $p_3 = 0.6 = 1 - p_1 - p_2$

il valore cercato è:

$$P = \frac{20!}{3!\,5!\,12!} (0.1)^3 \cdot (0.3)^5 \cdot (0.6)^{12} \approx 0.037.$$