8 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 variabile casuale, di funzione di ripartizione, di funzione di densità di una variabile casuale al caso $n-$dimensionale.

DEFINIZIONE: Chiamiamo variabile casuale $n-$dimensionale una funzione $X = (X_1, \ldots, X_n)$: $\Omega \to \mathbb{R}^n$ tale che $\forall (r_1, \ldots, r_n) \in \mathbb{R}^n$

$$\{w \in \Omega : X_1(w) \leq r_1, \ldots, X_n(w) \leq r_n\} \text{ è un evento}$$

Quindi una variabile casuale $n-$dimensionale è una $n-$upla di variabili casuali, le quali associano un numero ad un risultato.

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

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

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

$$F_{X_1,...,X_n}(x_1, ..., x_n) = P[X_1 \leq x_1, ..., X_n \leq x_n]$$

Come nel caso 1-D, per definire la funzione di densità bisogna distinguere il caso discreto dal caso continuo. Noi ci occuperemo solo di variabili casuali $n-$dimensionali discrete.

DEFINIZIONE: La variabile casuale $X = (X_1, ..., X_n)$ è detta variabile casuale discreta $n-$dimensionale se può assumere valori solo in un insieme numerabile $(x_1, ..., x_n)$ di punti di $\mathbb{R}^n$.

DEFINIZIONE: Chiamiamo funzione di densità discreta congiunta di $X = (X_1, ..., X_n)$, variabile casuale discreta $n-$dimensionale, la funzione:

$$f_{X_1,...,X_n}(x_1, ..., x_n) = P[X_1 = x_1, ..., X_n = x_n]$$

Proprietà

- $f_{X_1,...,X_n}(x_1, ..., x_n) \geq 0 \quad \forall(x_1, ..., x_n)$
- $\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} f_{X_1,...,X_n}(x_1, ..., x_n) = 1$
- $\forall A \subset \mathbb{R}^n \; P[(x_1, ..., x_n) \in A] = \sum_{A} f_{X_1,...,X_n}(x_1, ..., x_n)$

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

Proprietà di $F_{X,Y}(x, y)$

- $\lim\limits_{x\to-\infty} F_{X,Y}(x, y) = 0 \quad \forall y$
- $\lim\limits_{y\to-\infty} F_{X,Y}(x, y) = 0 \quad \forall x$
- $\lim\limits_{\substack{x\to+\infty \\ y\to+\infty}} F_{X,Y}(x, y) = 1$
- se $x_1 < x_2$ e $y_1 < y_2$ allora

  $P[x_1 < X \leq x_2, y_1 < Y \leq y_2] = F_{X,Y}(x_2, y_2)+$

  $+F_{X,Y}(x_1, y_1) - F_{X,Y}(x_1, y_2) - F_{X,Y}(x_2, y_1) \geq 0$
- $\lim\limits_{h\to0^+} F_{X,Y}(x + h, y) = \lim\limits_{h\to0^+} F_{X,Y}(x, y + h) =$
  $= F_{X,Y}(x, y)$ (continuità a destra)

DEFINIZIONE: Date X, Y variabili casuali congiunte discrete

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

dove x_i, y_j sono i punti massa delle variabili casuali congiunte X, Y.
Dalla funzione di densità congiunta $f_{X,Y}(x, y)$ delle variabili casuali discrete congiunte X, Y è possibile ricavare le funzioni di densità $f_X(\cdot)$ di X e $f_Y(\cdot)$ di Y, dette funzioni di densità marginali.

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

$$f_X(x_k) = \sum_j f_{X,Y}(x_k, y_j)$$
$$f_Y(y_k) = \sum_j f_{X,Y}(x_j, y_k)$$

Nota bene: dalla funzione di densità congiunta è possibile sempre ricavare le funzioni di densità marginali, ma non vale il viceversa.

Ricordando la definizione di probabilità condizionata

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

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

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

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

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

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

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

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

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

$$F_{Y|X}(y|x) = P[Y \leq y|X = x] =$$
$$= \sum_{\{j: y_j \leq y\}} f_{Y|X}(y_j|x), \quad f_X(x) > 0$$

e

$$F_{X|Y}(x|y) = P[X \le x | Y = y] =$$
$$= \sum_{\{j: x_j \le x\}} f_{X|Y}(x_j|y), \quad f_Y(y) > 0$$

Nota bene

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

Esempio. Lancio di due tetraedri (poliedri regolari a 4 facce) aventi le facce numerate da 1 a 4.
Chiamiamo X la variabile casuale che indica il numero sulla faccia rivolta verso il basso del 1° tetraedro, Y la variabile casuale che indica il numero più grande fra quelli indicati sulle facce rivolte verso il basso dei due tetraedri.
I valori congiunti di X e Y sono:

$$\begin{array}{llll} (1,1) & (1,2) & (1,3) & (1,4) \\ (2,2) & (2,3) & (2,4) & \\ (3,3) & (3,4) & & \\ (4,4) & & & \end{array}$$

Tabella 1: I valori di (X, Y) nel lancio dei due tetraedri

1° T	**2° T**	(X, Y)
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)
4	1	(4,4)
4	2	(4,4)
4	3	(4,4)
4	4	(4,4)

Lo spazio campione Ω è formato da 16 elementi.
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 possibili risultati in una tabella.

Tabella 2: I valori della funzione di densità congiunta

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

Si noti che $\sum_{i,j} f_{X,Y}(x_i, y_j) = 1$.
Dalla definizione di funzione di ripartizione congiunta si ha (ricorda che $F_{X,Y}(\bar{x}, \bar{y}) = \sum_{x_i \leq \bar{x}, y_j \leq \bar{y}} f_{X,Y}(x_i, y_j)$):

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

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

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

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

Le seguenti figure riassumono i valori di $F_{X,Y}(x,y)$ ed i valori di $f_{X,Y}(x,y)$ con le relative funzioni di densità marginali.

$4 \leq y$	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
	$x < 1$	$1 \leq x < 2$	$2 \leq x < 3$	$3 \leq x < 4$	$4 \leq x$

Figura 1: Tabella dei valori di $F_{X,Y}(x,y)$

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

Figura 2: Tabella dei valori di $f_{X,Y}(x,y)$ con le densità marginali

OSSERVAZIONE. Le densità marginali si calcolano più facilmente sommando per righe o per colonne i valori di $f_{X,Y}$ della tabella a doppia entrata.
Vogliamo ora calcolare 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 $X = 2 \Rightarrow Y \geq 2$

$$\left.\begin{aligned} f_{X,Y}(2|2) &= \frac{f_{X,Y}(2,2)}{f_X(2)} = \frac{2/16}{4/16} = \frac{1}{2} \\ f_{X,Y}(3|2) &= \frac{f_{X,Y}(2,3)}{f_X(2)} = \frac{1/16}{4/16} = \frac{1}{4} \\ f_{X,Y}(4|2) &= \frac{f_{X,Y}(2,4)}{f_X(2)} = \frac{1/16}{4/16} = \frac{1}{4} \end{aligned}\right\} \sum = 1$$

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

$$\left.\begin{aligned} f_{X,Y}(3|3) &= \frac{f_{X,Y}(3,3)}{f_X(3)} = \frac{3/16}{4/16} = \frac{3}{4} \\ f_{X,Y}(4|3) &= \frac{f_{X,Y}(3,4)}{f_X(3)} = \frac{1/16}{4/16} = \frac{1}{4} \end{aligned}\right\} \sum = 1$$

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

INDIPENDENZA

DEFINIZIONE: Data $X = (X_1, ..., X_n)$ variabile casuale $n-$dimensionale (discreta o continua) con funzione di densità congiunta $f_{X_1,...,X_n}$, diciamo che $X_1, ..., X_n$ sono variabili casuali INDIPENDENTI se e solo se

$$f_{X_1,...,X_n}(x_1, ..., x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n),$$

cioè la funzione di densità congiunta si scrive come PRODOTTO delle funzioni di densità marginali.

Nel caso 2-D si ha quindi che X, Y sono indipendenti se e solo se $f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$.

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 variabili casuali non sono indipendenti basta mostrare che $f_{Y|X}(y|x)$ dipende da x.

OSSERVAZIONE
Si può provare che se $X_1, ..., X_n$ sono variabili casuali indipendenti e se $g_1, ..., g_n$ sono n funzioni tali che $Y_k = g_k(X_k)$, $k = 1, .., n$, sono variabili casuali, allora $Y_1, ..., Y_n$ sono indipendenti.

Nell'esempio del lancio dei due tetraedri, alla domanda "Le variabili casuali X, Y sono indipendenti?", la risposta è NO!

Basta infatti trovare una coppia di valori (X, Y) per i quali:

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

Per esempio se $X = 3$, $Y = 2$
$f_{Y|X}(2|3) = P[Y = 2|X = 3] = 0$ poichè $Y < X$
ma
$f_Y(2) = P[Y = 2] = \frac{3}{16}$.

<u>Esempio</u>
Da un gruppo di 12 batterie (3 nuove, 4 usate, 5 difettose) ne vengono scelte 3 a caso. Indicato con
X = numero di batterie nuove
Y = numero di batterie usate

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

$$f(3,0) = \frac{\binom{3}{3}}{\binom{12}{3}} = \frac{1}{220} \qquad f(2,1) = \frac{\binom{3}{2}\binom{4}{1}}{\binom{12}{3}} = \frac{21}{220}$$

$$f(2,0) = \frac{\binom{3}{2}\binom{5}{1}}{\binom{12}{3}} = \frac{15}{220} \qquad f(1,2) = \frac{\binom{3}{1}\binom{4}{2}}{\binom{12}{3}} = \frac{18}{220}$$

$$f(1,0) = \frac{\binom{3}{1}\binom{5}{2}}{\binom{12}{3}} = \frac{30}{220} \qquad f(1,1) = \frac{\binom{3}{1}\binom{4}{1}\binom{5}{1}}{\binom{12}{3}} = \frac{60}{220}$$

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

$$f(0,1) = \frac{\binom{4}{1}\binom{5}{2}}{\binom{12}{3}} = \frac{40}{220} \qquad f(0,0) = \frac{\binom{5}{3}}{\binom{12}{3}} = \frac{10}{220}$$

La seguente figura riassume i valori di $f_{X,Y}(x,y)$.

	$Y=0$	$Y=1$	$Y=2$	$Y=3$	$f_X(x)$
$X=0$	$1/22$	$2/11$	$3/22$	$1/55$	$42/110$
$X=1$	$3/22$	$3/11$	$9/110$	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

Figura 3: Tabella dei valori di $f_{X,Y}(x, y)$ con le marginali

MODELLO DI VARIABILE CASUALE N DIMENSIONALE DISCRETA

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 $s_1, ..., s_{k+1}$ tali esiti.
Sia $p_i = P[s_i]$, $i = 1, ..., k+1$ con
$\sum_{i=1}^{k+1} p_i = 1, \Rightarrow p_{k+1} = 1 - (p_1 + ... + p_k)$
Ripetiamo le prove n volte.
Se le prove sono ripetute e indipendenti si ha:

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

dove

$$\sum_{i=1}^{k+1} x_i = n, \Rightarrow x_{k+1} = n - \sum_{i=1}^{k} x_i.$$

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 e ... 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:

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

<u>Esempio</u>

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).
<u>Ipotesi</u>: indipendenza
$n = 20$
$x_1 = 3, \qquad p_1 = 0.1,$
$x_2 = 5, \qquad p_2 = 0.3,$
$x_3 = n - x_1 - x_2 = 20 - 3 - 5 = 12,$
$p_3 = 1 - p_1 - p_2 = 1 - 0.1 - 0.3 = 0.6.$

Il valore cercato è

$$p = \frac{20!}{3!5!12!}\,(0.1)^3 \cdot (0.3)^5 \cdot (0.6)^{12} \;\sim\; 0.037$$