Distribuzioni doppie

Si definiscono ponendo in corrispondenza eventi e una coppia di variabili casuali

ESPERIMENTO

Ripetizione, per 3 volte, del lancio di una moneta

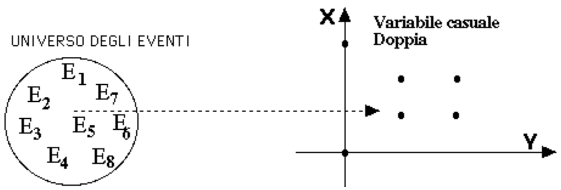
VARIABILI CASUALI

X: Numero di teste

Y: Numero di scambi nella sequenza

Eventi	Lanci	X	Y
E_1	TTT	3	0
E_2	TTC	2	1
E_3	TCT	2	2
E_4	TCC	1	1
E_5	CTT	2	1
E_6	CTC	1	2
E_7	CCT	1	1
E_8	CCC	0	0

Ad ogni evento dello spazio campionario è associata una coppia ordinata di valori.



La distribuzione congiunta (discrete)

Siano "X" e "Y" due v.c. discrete con "k" ed "h" valori. La distribuzione di probabilità congiunta è definita da

$$P(X = x_i,\ Y = y_j) = p_{ij} \quad \textbf{dove} \quad \boxed{\begin{array}{l} 1.\ p_{ij} \ge 0 \\ 2.\ \sum_{i=1}^{h}\sum_{j=1}^{k} p_{ij} = 1 \end{array}}$$

Per sintetizzare si utilizza la tabella a doppia entrata

In generale

Y/X	x_1	x_2	…	x_k	
y_1	p_{11}	p_{21}	:	p_{k1}	$p_{.1}$
y_2	p_{12}	p_{22}	:	p_{k2}	$p_{.2}$
…	…	…	:	:	:
y_h	p_{1h}	p_{2h}	:	p_{kh}	$p_{.h}$
	$p_{1.}$	$p_{2.}$	…	$p_{k.}$	1

Per l'esempio

Y/X	0	1	2	3	
0	$1/8$	0	0	$1/8$	$2/8$
1	0	$2/8$	$2/8$	0	$4/8$
2	0	$1/8$	$1/8$	0	$2/8$
	$1/8$	$3/8$	$3/8$	$1/8$	1

Notare che

somma di colonna: $p_{i.} = \sum_{j=1}^{h} p_{ij} \quad i = 1,2,\ldots,k$

somma di riga: $p_{.j} = \sum_{i=1}^{k} p_{ij} \quad j = 1,2,\ldots,h$

il punto indica l'indice rispetto a cui si è sommato

Esempio

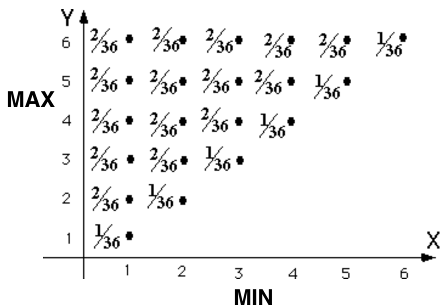
Si lanciano due dadi ben equilibrati e si indicano con D1 e D2 il risultato del primo ed il risultato del secondo. Per ognuno dei 36 risultati possibili si definiscono le due seguenti variabili casuali:

$$\begin{cases} X = \min\{D1, D2\} \\ Y = \max\{D1, D2\} \end{cases}$$

Ad esempio: se escono D1=6 e D2=4 si ha X=4 e Y=6; se esce D1=5 e D2=5 allora X=5 Y=5.

La funzione di distribuzione congiunta della v.c. doppia (X,Y) è

$$p(X,Y) = \begin{cases} \dfrac{1}{36} & \text{se } 1 \le X = Y \le 6 \\[2mm] \dfrac{2}{36} & \text{se } 1 \le X < Y \le 6 \end{cases}$$



Ad esempio:

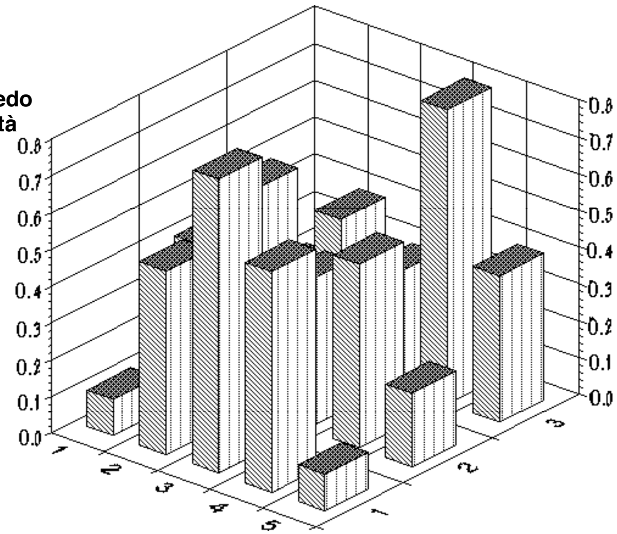
$$P(Y = 3,\ X = 2) = \frac{2}{36} \Rightarrow (2,3) \text{ e } (3,2)$$

Istogramma tridimensionale

Variabili casuali DISCRETE

il volume di ogni parallelepipedo è proporzionale alla probabilità della combinazione di X e Y

La somma dei volumi di tutti i parallelepipedi è pari ad uno (o altro valore prefissato)



Distribuzioni marginali

A partire dalla funzione di probabilità congiunta è possibile definire le funzioni di distribuzione per ciascuna delle v.c. ignorando l'altra

Funzione di probabilità marginale per la "X"

$$P(X=x_i)=P\left\{(X=x_i)\cap\left[\bigcup_{j=1}^{h}(Y=y_j)\right]\right\}=P\left[\bigcup_{j=1}^{h}(X=x_i)\cap(Y=y_j)\right]=\sum_{j=1}^{h}p_{ij}=p_{i.}$$

X	P(X = x)
0	1/8
1	3/8
2	3/8
3	1/8
	1

N.B. $A\cap I=A$

Funzione di probabilità marginale per la "Y"

$$P(Y=y_j)=P\left\{(Y=y_j)\cap\left[\bigcup_{i=1}^{k}(X=x_i)\right]\right\}=P\left[\bigcup_{i=1}^{k}(X=x_i)\cap(Y=y_j)\right]=\sum_{i=1}^{k}p_{ij}=p_{.j}$$

y	P(Y = y)
0	2/8
1	4/8
2	2/8
	1

Per ottenere la distribuzione di probabilità marginale si somma rispetto alla variabile che NON interessa

Esempio

Supponiamo che la funzione di distribuzione congiunta sia definita come:

$$P(X,Y)=\frac{X+Y}{21}\quad \text{per}\quad X=1,2,3;\quad Y=1,2$$

Marginale della "X"

$$\sum_{y=1}^{2}\frac{X+Y}{21}=\frac{X+1}{21}+\frac{X+2}{21}=\frac{2X+3}{21}\Rightarrow\quad P(X)=\frac{2X+3}{21}\quad\text{per}\quad x=1,2,3$$

Marginale della "Y"

$$\sum_{x=1}^{3}\frac{X+Y}{21}=\frac{1+Y}{21}+\frac{2+Y}{21}+\frac{3+Y}{21}=\frac{6+3Y}{21}\Rightarrow\quad P(Y)=\frac{2+Y}{7}\quad\text{per}\quad y=1,2$$

Da notare che sia P(X) che P(Y) verificano le condizioni per essere delle distribuzioni di probabilità: non negatività e somma unitaria

Esercizio

Sia data la seguente funzione di distribuzione congiunta

$$P(X=x,Y=y)=\frac{x^2-y^2}{82};\qquad x=1,2,3,4;\quad y=-1,0,1$$

Ricavare le funzioni di distribuzioni marginali:

$$P(X=x)=\sum_{y=-1}^{1}\frac{x^2-y^2}{82}=\frac{x^2-(-1)^2}{82}+\frac{x^2-(0)^2}{82}+\frac{x^2-(1)^2}{82}=\frac{3x^2-2}{82}$$

$$P(Y=y)=\sum_{x=1}^{4}\frac{x^2-y^2}{82}=\frac{1^2-y^2}{82}+\frac{2^2-y^2}{82}+\frac{3^2-y^2}{82}+\frac{4^2-y^2}{82}=\frac{30-4y^2}{82}$$

Da notare come la distribuzione marginale di una variabile NON dipenda più dall'altra variabile

Distribuzioni di probabilità condizionate

l'interesse per le distribuzioni doppie nasce dalla scoperta della dipendenza. Solo in questo caso il conoscere una delle variabili aiuta a conoscere l'altra.

Per studiare quale sia il comportamento della "Y" rispetto ai valori della "X" è possibile segmentare la distribuzione doppia in tante sottodistribuzioni

La distribuzione della Y CONDIZIONATA (PARZIALE) dal fatto che "X" è ad un dato livello è

$$P(Y=y|X=x_i)=\frac{P(Y=y,X=x_i)}{P(X=x_i)}$$

per cui si possono definire, "k" diverse distribuzioni parziali della "Y".

Allo stesso modo si definiscono "h" distribuzioni condizionate della "X"

$$P(X=x|Y=y_j)=\frac{P(X=x,Y=y_j)}{P(Y=y_j)}$$

Esempio

Consideriamo la seguente funzione di distribuzione congiunta

$$\mathbf{P}(x,y)=\begin{cases}\dfrac{x^2+y}{32} & \text{per } \ x=0,1,2,3;\ Y=0,1\\ 0 & \text{altrove}\end{cases}$$

a) Sviluppare la tabella a doppia entrata della funzione di distribuzione doppia
b) Ottenere le distribuzioni marginali; c) Ottenere le condizionate;
d) Definire la condizionata della X per y=1 e) Definire la condizionata della Y per X=3
f) Verificare se X e Y sono indipendenti

Per risolvere la parte "a)" basta sostituire i valori nella funzione di distribuzione

Y/X	0	1	2	3	
0	0	1/32	4/32	9/32	14/32
1	1/32	2/32	5/32	10/32	18/32
	1/32	3/32	9/32	19/32	1

Per ricavare le distribuzioni marginali di una variabile occorre SOMMARE la congiunta rispetto ai valori dell'altra in modo da eliminarne l'influenza

$$\mathbf{P}(x)=\sum_{j=1}^{2}\mathbf{P}(x,y_j)=\sum_{j=1}^{2}\frac{x^2+y_j}{32}=\frac{x^2+0}{32}+\frac{x^2+1}{32}=\frac{2x^2+1}{32}$$

$$\mathbf{P}(y)=\sum_{i=1}^{4}\mathbf{P}(x_i,y)=\sum_{i=1}^{4}\frac{x_i^2+y}{32}=\frac{0+y}{32}+\frac{1+y}{32}+\frac{4+y}{32}+\frac{9+y}{32}=\frac{14+4y}{32}$$

Esercizio

La distribuzione congiunta di due variabili casuali discrete è la seguente:

$$P(X,Y)=\frac{(X-Y)^2}{19} \quad \text{per } \ X=1,2;\quad Y=-1,0,1;$$

Determinare la distribuzione condizionata di YIX e di XIY

Cerchiamo prima le distribuzioni marginali:

$$P(X)=\sum_{y=-1}^{1}\frac{(X-Y)^2}{19}=\frac{3X^2+2}{19};\qquad P(Y)=\sum_{x=1}^{2}\frac{(X-Y)^2}{19}=\frac{2Y^2-6Y+5}{19};$$

$$P(Y|X)=\frac{P(X,Y)}{P(X)}=\frac{\dfrac{(X-Y)^2}{19}}{\dfrac{3X^2+2}{19}}=\frac{(X-Y)^2}{3X^2+2}$$

$$P(X|Y)=\frac{P(X,Y)}{P(Y)}=\frac{\dfrac{(X-Y)^2}{19}}{\dfrac{2Y^2-6Y+5}{19}}=\frac{(X-Y)^2}{2Y^2-6Y+5}$$

Per ottenere una particolare distribuzione parziale basta fissare il valore di quella in base alla quale si vuole studiare l'altra

Continuazione esempio

Le condizionate possono essere ottenute in forma analitica

$$f(x|y)=\frac{f(x,y)}{f(y)}=\frac{\dfrac{x^2+y}{32}}{\dfrac{14+4y}{32}}=\frac{x^2+y}{14+4y}\qquad f(y|x)=\frac{f(x,y)}{f(x)}=\frac{\dfrac{x^2+y}{32}}{\dfrac{2x^2+1}{32}}=\frac{x^2+y}{2x^2+1}$$

$$f(x|y=1)=\frac{x^2+1}{14+4}=\frac{x^2+1}{18};\qquad f(y|x=3)=\frac{9+y}{18+1}=\frac{y+9}{19};$$

da queste si ottengono le particolari fissando i valori della condizionante

Per verificare l'indipendenza basta controllare se il prodotto delle marginali coincide o no con la distribuzione congiunta

$$f(x)*f(y)=\left(\frac{2x^2+1}{32}\right)*\left(\frac{14+4y}{32}\right)=\frac{28x^2+8x^2y+4y+14}{1024}$$

che è chiaramente diversa dalla congiunta: $f(x,y)=\dfrac{x^2+y}{32}$

Esercizio

Date la distribuzione doppia

Y/X	−2	0	2	
0	0	1/6	1/12	3/12
2	1/4	0	1/2	9/12
	3/12	2/12	7/12	1

a) Individuare le distribuzioni marginali;
b) Definire la condizionata della X per y=2
c) Definire la condizionata della Y per X=2
d) Verificare se X e Y sono indipendenti

Indipendenza

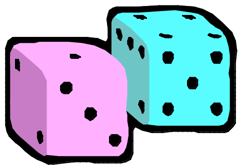
Perché abbia senso lo studio CONGIUNTO esso deve essere più informativo dello studio SEPARATO delle due componenti

Se la "X" assume valori in relazione ad eventi indipendenti da quelli che generano i valori della "Y" non può esistere legame probabilistico tra di esse

ESEMPIO

Lancio di due dadi di diverso colore

X: punteggio del dado rosso;
Y: punteggio del dado blù;



Sapere che lanciando i due dadi, X= 4 e, contemporaneamente, Y= 3 è come sapere che X=4 (ignorando "Y") e che Y=3 (ignorando "X")

$$P(X = 4 \cap Y = 3) = P(X = 4) * P(Y = 3)$$

I due eventi sono infatti INDIPENDENTI

Indipendenza/2

Due variabili casuali discrete "X" e "Y" si dicono indipendenti quando tali sono gli eventi (X=x) e (Y=y) per ogni possibile coppia di valori

$$P\left(X = x_i \cap Y = y_j\right) = P(X = x_i) * P\left(Y = y_j\right) \quad \forall\ i\ e\ j$$

Questo significa che la funzione di distribuzione congiunta è data dal prodotto delle due distribuzioni marginali

$$p_{ij} = p_{i.} * p_{.j} \quad \forall\ i\ e\ j$$

La condizione di indipendenza è di verifica immediata

Y/X	10	11	12	
1	0.2	0.1	0.0	0.3
2	0.1	0.2	0.1	0.4
3	0.0	0.1	0.2	0.3
	0.3	0.4	0.3	1

$$P(X = 10, Y = 1) = 0.2 \neq P(X = 10) * P(Y = 1) = 0.3 * 0.3 = 0.09$$

E' sufficiente che in una cella non si abbia l'uguaglianza tra congiunta e prodotto delle marginali perché sussista DIPENDENZA IN SENSO STATISTICO tra le due v.c.

Esempio

Supponiamo che la distribuzione congiunta della v.c. doppia (X,Y) sia

$$P(X,Y) = \frac{XY^2}{30} \qquad x = 1,2,3; \quad Y = 1,2.$$

Ricaviamo le due d.p. marginali

$$P(X) = \sum_{y=1}^{2} \frac{XY^2}{30} = \frac{X}{30} + \frac{X4}{30} = \frac{5X}{30} = \frac{X}{6} \qquad \text{per} \quad X = 1,2,3$$

$$P(Y) = \sum_{x=1}^{3} \frac{XY^2}{30} = \frac{Y^2}{30} + \frac{2Y^2}{30} + \frac{3Y^2}{30} = \frac{6Y^2}{30} = \frac{Y^2}{5} \quad \text{per} \quad Y = 1,2.$$

Poiché: $P(X) * P(Y) = \frac{X}{6} * \frac{Y^2}{5} = \frac{XY^2}{30} = P(X,Y)$ **Le due variabili sono indipendenti**

In questo caso, lo studio congiunto non aggiunge alcuna informazione rispetto allo studio separato delle due variabili casuali

Esercizio

Le variabili casuali X e Y assumono valori: -1,0,1 con distribuzione di probabilità

$$P(X,Y) = \frac{(X+Y)^2}{13} \qquad X, Y = -1,0,1;$$

Verificare che si tratti di variabili casuali DIPENDENTI:

$$P(X) = \sum_{Y=-1}^{1} \frac{(X+Y)^2}{13} = \frac{3X^2+2}{13}; \qquad P(Y) = \sum_{X=-1}^{1} \frac{(X+Y)^2}{13} = \frac{3Y^2+2}{13}$$

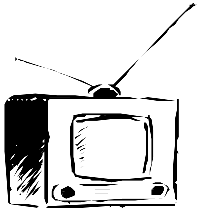
$$P(X) * P(Y) = \left[\frac{3X^2+2}{13}\right] * \left[\frac{3Y^2+2}{13}\right] \neq \frac{X^2+2XY+Y^2}{13} = P(X,Y)$$

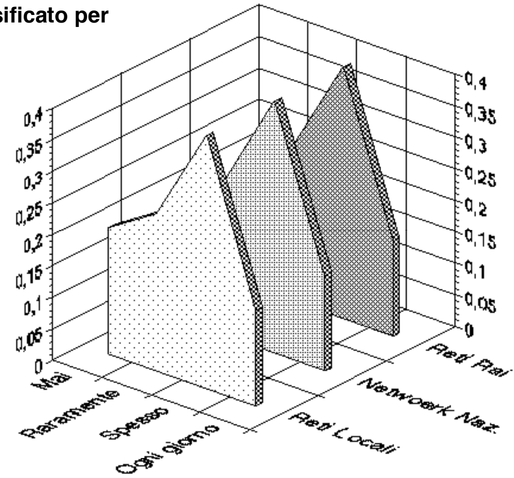
Poiché la congiunta non è ricavabile dal prodotto delle due marginale, le due v.c. sono da considerarsi dipendenti.

Lo studio congiunto è più informativo di quello separato a causa del legame di dipendenza tra le due variabili

Esempio di indipendenza

Un ipotetico campione di famiglie classificato per l'attenzione ai programmi televisivi





Indipendenza tra attenzione e canali significa che si tendono a guardare con la stessa probabilità tutti i network ovvero la probabilità con cui si tende a guardare la TV prescinde dal network

AT93

Definizione equivalente di indipendenza

Due v.c. sono indipendenti se la distribuzione condizionata YIX non varia al variare di X

$$f(Y_j \mid X_i)=f(Y_j) \qquad \text{per} \qquad i=1,2,\ldots,r \quad \text{e per ogni "j"}$$

ovvero le probabilità con cui compaiono le modalità di Y rimangono costanti al variare della X

Che sia una definizione equivalente lo si evince dal fatto che $f(y|x)=\dfrac{f(y,x)}{f(x)}$

e che, in caso di indipendenza $f(y|x)=\dfrac{f(y)*f(x)}{f(x)}=f(y)$

l'indipendenza statistica è BILATERALE: se X è indipendente da Y è vero anche il contrario

$$f(x|y)=\frac{f(y)*f(x)}{f(y)}=f(x)$$

AT93

Esempio

Y/X	1	3	5	7	
−1	0.1	0.0	0.2	0.0	0.3
0	0.1	0.1	0.1	0.2	0.5
1	0.0	0.1	0.1	0.0	0.2
	0.2	0.2	0.4	0.2	1.0

Si voglia determinare la distribuzione di "X" dato che Y=-1
Questa risulta

$X \mid Y=-1$	$P(X \mid Y=-1)=\dfrac{P(X=x \cap Y=-1)}{P(Y=-1)}$
1	$0.1/0.3=0.33$
3	$0.0/0.3=0.00$
5	$0.2/0.3=0.67$
7	$0.0/0.3=0.00$
	1.00

Determiniamo ora la distribuzione della Y dato che la X=5

$Y \mid X=5$	$P(Y \mid X=5)=\dfrac{P(Y=y \cap X=5)}{P(X=5)}$
−1	$0.2/0.4=0.50$
0	$0.1/0.4=0.25$
1	$0.1/0.4=0.25$
	1.00

AT93

Distribuzioni di probabilità condizionate

l'interesse per le distribuzioni doppie nasce dalla scoperta della dipendenza. Solo in questo caso il conoscere una delle variabili aiuta a conoscere l'altra.

Per studiare quale sia il comportamento della "Y" rispetto ai valori della "X" è possibile segmentare la distribuzione doppia in tante sottodistribuzioni

La distribuzione della Y CONDIZIONATA (PARZIALE) dal fatto che "X" è ad un dato livello è

$$P(Y=y \mid X=x_i)=\frac{P(Y=y, X=x_i)}{P(X=x_i)}$$

per cui si possono definire, "k" diverse distribuzioni parziali della "Y".

Allo stesso modo si definiscono "h" distribuzioni condizionate della "X"

$$P(X=x \mid Y=y_j)=\frac{P(X=x, Y=y_j)}{P(Y=y_j)}$$

In queste definizioni entra in modo determinante il 5° postulato

AT93

Esempio

Consideriamo la seguente funzione di distribuzione congiunta

$$\mathbf{P}(x,y)=\begin{cases}\dfrac{x^2+y}{32} & \text{per } \; x=0,1,2,3;\; Y=0,1\\ 0 & \text{altrove}\end{cases}$$

a) Sviluppare la tabella a doppia entrata della funzione di distribuzione doppia
b) Ottenere le distribuzioni marginali; c) Ottenere le condizionate;
d) Definire la condizionata della X per y=1 e) Definire la condizionata della Y per X=3
f) Verificare se X e Y sono indipendenti

Per risolvere la parte "a)" basta sostituire i valori nella funzione di distribuzione

Y/X	0	1	2	3	
0	0	$1/32$	$4/32$	$9/32$	$14/32$
1	$1/32$	$2/32$	$5/32$	$10/32$	$18/32$
	$1/32$	$3/32$	$9/32$	$19/32$	1

Per ricavare le distribuzioni marginali di una variabile occorre SOMMARE la congiunta rispetto ai valori dell'altra in modo da eliminarne l'influenza

$$\mathbf{P}(x)=\sum_{j=1}^{2}\mathbf{P}(x,y_j)=\sum_{j=1}^{2}\frac{x^2+y_j}{32}=\frac{x^2+0}{32}+\frac{x^2+1}{32}=\frac{2x^2+1}{32}$$

$$\mathbf{P}(y)=\sum_{i=1}^{4}\mathbf{P}(x_i,y)=\sum_{i=1}^{4}\frac{x_i^2+y}{32}=\frac{0+y}{32}+\frac{1+y}{32}+\frac{4+y}{32}+\frac{9+y}{32}=\frac{14+4y}{32}$$

Esempio

La distribuzione congiunta di due variabili casuali discrete è la seguente:

$$P(X,Y)=\frac{(X-Y)^2}{19} \quad \text{per } \; X=1,2;\quad Y=-1,0,1;$$

Determinare la distribuzione condizionata di YIX e di XIY

Cerchiamo prima le distribuzioni marginali:

$$P(X)=\sum_{y=-1}^{1}\frac{(X-Y)^2}{19}=\frac{3X^2+2}{19};\qquad P(Y)=\sum_{x=1}^{2}\frac{(X-Y)^2}{19}=\frac{2Y^2-6Y+5}{19};$$

$$P(Y|X)=\frac{P(X,Y)}{P(X)}=\frac{\dfrac{(X-Y)^2}{19}}{\dfrac{3X^2+2}{19}}=\frac{(X-Y)^2}{3X^2+2}$$

$$P(X|Y)=\frac{P(X,Y)}{P(Y)}=\frac{\dfrac{(X-Y)^2}{19}}{\dfrac{2Y^2-6Y+5}{19}}=\frac{(X-Y)^2}{2Y^2-6Y+5}$$

Per ottenere una particolare distribuzione parziale basta fissare il valore della X o della Y in base alla quale si vuole studiare l'altra

Continuazione esempio

Le condizionate possono essere ottenute in forma analitica

$$f(x|y)=\frac{f(x,y)}{f(y)}=\frac{\dfrac{x^2+y}{32}}{\dfrac{14+4y}{32}}=\frac{x^2+y}{14+4y}\qquad f(y|x)=\frac{f(x,y)}{f(x)}=\frac{\dfrac{x^2+y}{32}}{\dfrac{2x^2+1}{32}}=\frac{x^2+y}{2x^2+1}$$

$$f(x|y=1)=\frac{x^2+1}{14+4}=\frac{x^2+1}{18};\qquad f(y|x=3)=\frac{9+y}{18+1}=\frac{y+9}{19};$$

da queste si ottengono le particolari fissando i valori della condizionante

Per verificare l'indipendenza basta controllare se il prodotto delle marginali coincide o no con la distribuzione congiunta

$$f(x)*f(y)=\left(\frac{2x^2+1}{32}\right)*\left(\frac{14+4y}{32}\right)=\frac{28x^2+8x^2y+4y+14}{1024}$$

che è chiaramente diversa dalla congiunta: $f(x,y)=\dfrac{x^2+y}{32}$

La uniforme discreta bivariata

Se tutte le coppie di valori delle due variabili casuali sono equiprobabili allora la funzione di distribuzione congiunta è data da

$$P(X=x;Y=y)=\frac{1}{k*h};\quad X=X_1,X_2,\ldots,X_k;\quad Y=Y_1,Y_2,\ldots,Y_h$$

Poiché ci sono k*h combinazioni equiprobabili dei valori di X ed Y

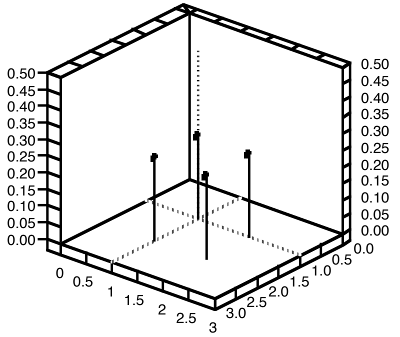
ESEMPIO

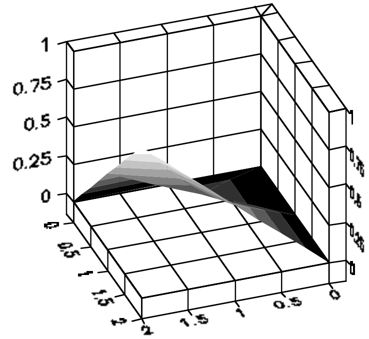
y/x	2	4	6	
1	$1/12$	$1/12$	$1/12$	$3/12$
2	$1/12$	$1/12$	$1/12$	$3/12$
3	$1/12$	$1/12$	$1/12$	$3/12$
4	$1/12$	$1/12$	$1/12$	$3/12$
	$4/12$	$4/12$	$4/12$	1

Naturalmente le due variabili nella discreta bivariata sono indipendenti dato che

$$\frac{3}{12}*\frac{4}{12}=\frac{12}{144}=\frac{1}{12}$$

Uniforme discreta bivariata/2





Funzione di distribuzione

	$y=0$	$y=1$
$X=0$	$1/4$	$1/4$
$X=1$	$1/4$	$1/4$

Funzione di ripartizione

	$y<0$	$0\leq y\leq 1$	$y\geq 1$
$x<0$	0	0	0
$0\leq x\leq 1$	0	$1/4$	$2/4$
$x\geq 1$	0	$2/4$	$4/4$

Esercizio

L'esperimento consiste nel lanciare 10 volte due dadi non truccati e su questo si definiscono due variabili casuali:

X = Numero di volte che esce il "6"

Y = Numero di volte che esce il "5"

a) Si costruisca e rappresenti la funzione di distribuzione congiunta (X,Y) nonché la la funzione di ripartizione.

b) Si calcolino le seguenti probabilità:

a) $P(X+Y=6)$; b) $P(x^2+y^2\leq 20)$;

c) $P(X+Y>9)$; d) $P(XY<10)$

I momenti delle v.c. doppie

il concetto di VALORE ATTESO si applica anche alle v.c. doppie.

Sia g(x,y) una qualche loro funzione

CASO DISCRETO $E[g(x,y)]=\sum_{i=1}^{k}\sum_{j=1}^{h}g(x_i,y_j)p_{ij}$

ponendo $g(x,y)=x^r y^s$ **si definiscono i momenti misti** $\mu_{rs}=\sum_{i=1}^{k}\sum_{j=1}^{h}x_i^r y_j^s p_{ij}$

se $g(x,y)=(x-\mu_x)^r(y-\mu_y)^s$ **si ottengono i momenti misti delle v.c. scarto**

$$\bar{\mu}_{rs}=\sum_{i=1}^{k}\sum_{j=1}^{h}(x_i-\mu_x)^r(y_j-\mu_y)^s p_{ij}$$

Casi particolari

$\mu_{00}=1;$
$\mu_{r0}=\mu_r(\text{momenti della singola } X);$
$\mu_{0s}=\mu_s(\text{momenti della singola } Y);$
$\bar{\mu}_{20}=Var(x);$
$\bar{\mu}_{02}=Var(y);$

$\bar{\mu}_{11}=Cov(X,Y)$

Esempio

Un'urna contiene 8 dischetti: 3 hanno lo zero su entrambe le facce: (0,0) due hanno un uno sul dritto ed uno zero sul rovescio: (1,0), altri due hanno (0,1) ed uno ha (1,1).

La loro funzione di distribuzione di probabilità congiunta è

$$P(X_1,X_2)=\frac{3-X_1-X_2}{8} \quad \text{per} \quad X_1,X_2=0,1$$

Calcoliamo il valore atteso della loro somma e del loro prodotto:

$$E(X_1+X_2)=\sum_{X_1=0}^{1}\sum_{X_2=0}^{1}(X_1+X_2)P(X_1,X_2)=(0)\frac{3}{8}+(1)\frac{2}{8}+(1)\frac{2}{8}+(2)\frac{1}{8}=\frac{3}{4}$$

$$E(X_1*X_2)=\sum_{X_1=0}^{1}\sum_{X_2=0}^{1}(X_1*X_2)P(X_1,X_2)=(0*0)\frac{3}{8}+(0*1)\frac{2}{8}+(1*0)\frac{2}{8}+(1*1)\frac{1}{8}=\frac{1}{8}$$

Esercizio

Consideriamo la seguente distribuzione doppia

Y/X	10	11	12	
1	0.2	0.1	0.0	0.3
2	0.1	0.2	0.1	0.4
3	0.1	0.1	0.1	0.3
	0.4	0.4	0.2	1.0

Calcolo delle medie marginali

$$E(X) = 10*0.4 + 11*0.4 + 12*0.2 = 10.8$$

$$E(Y) = 1*0.3 + 2*0.4 + 3*0.3 = 2$$

Calcolo del momenti misto scarto di ordine (1,1)

$$E\left[(x-\mu_x)(y-\mu_y)\right] = (10-10.8)*(1-2)*0.2 + (10-10.8)*(2-2)*0.1 + (10-10.8)*(3-2)*0.1 +$$

$$+ (11-10.8)*(1-2)*0.1 + (11-10.8)*(2-2)*0.2 + (11-10.8)*(3-2)*0.1 +$$

$$+ (12-10.8)*(1-2)*0.0 + (12-10.8)*(2-2)*0.1 + (12-10.8)*(3-2)*0.1 =$$

=0.2

Distribuzioni doppie (caso continuo)

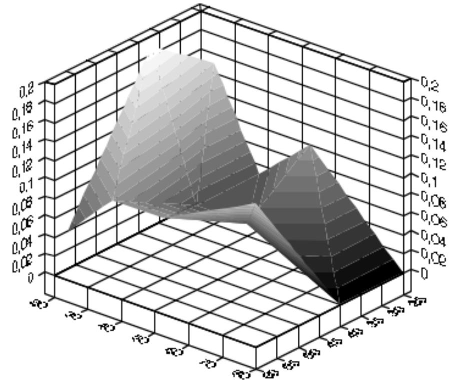
Consideriamo una popolazione di redditieri di cui si rilevano il reddito (Y) e l'età (X) divise in "k" ed "h" classi

IN GENERALE

Y/X	$L_1^x - U_1^x$	$L_2^x - U_2^x$	...	$L_k^x - U_k^x$	
$L_1^y - U_1^y$	p_{11}	p_{21}	...	p_{k1}	$p_{.1}$
$L_2^y - U_2^y$	p_{12}	p_{22}	...	p_{k2}	$p_{.2}$
:	:	:	:	:	:
$L_h^y - U_h^y$	p_{1h}	p_{2h}	...	p_{kh}	$p_{.h}$
	$p_{1.}$	$p_{2.}$	...	$p_{k.}$	1

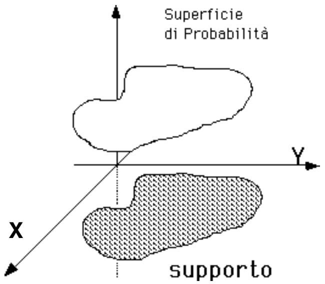
AD ESEMPIO

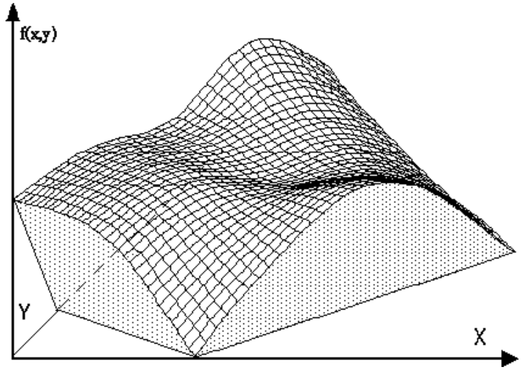
Y/X	20 – 30	30 – 50	50 – 70	
18 – 30	0.15	0.20	0.05	0.40
30 – 45	0.05	0.05	0.10	0.20
45 – 60	0.10	0.05	0.10	0.25
60 – 100	0.00	0.00	0.15	0.15
	0.30	0.30	0.40	1.00



Nello scegliere classi sempre più piccole, le probabilità tendono a descrivere una superficie continua. Questa è la funzione di densità bivariata

Superficie di Probabilità





il SUPPORTO è ora una porzione del piano e non un intervallo della retta.

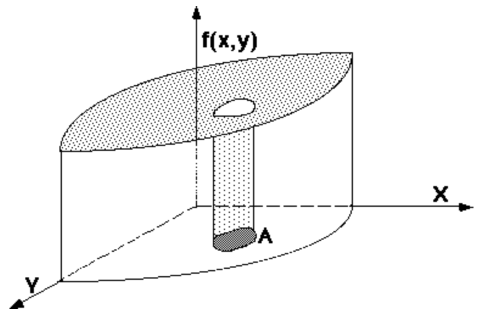
La funzione di densità descrive una superfice. Il volume coperto da questa è pari ad uno;

Definizione della densità bivariata

La variabile casuale doppia (X,Y) è di tipo continuo e con funzione di densità f(X,Y) se, per ogni evento "A" identificabile con una porzione del piano, la probabilità di tale evento è data da:

$$P\left[(x,y)\in A\right] = \iint_A f(x,y)\,dx\,dy$$

Integrale doppio su "A"

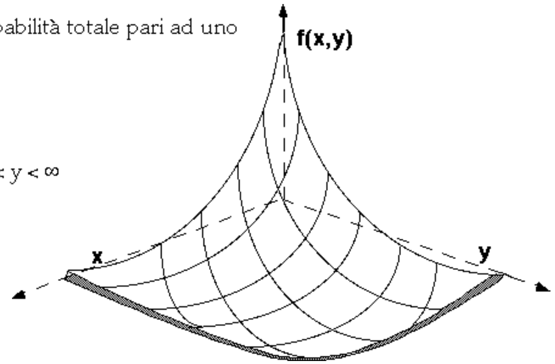


Che rappresenta il volume del solido che ha base definita da "A" coperto dalla corrispondente porzione di superficie descritta da f(x,y)

Proprietà formali

Perché la f(x,y) sia una funzione di densità, deve verificare le seguenti condizioni

1. $f(x;y) \geq 0 \quad \text{per } (x,y) \in A$ Non negatività nel supporto
2. $f(x;y) = 0 \quad \text{per } (x,y) \notin A$ Estensione a tutto R^2
3. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$ Probabilità totale pari ad uno



Esempio:

$$f(x,y) = \begin{cases} e^{-x-y} & \text{per } 0 < x < \infty;\ 0 < y < \infty \\ 0 & \text{altrove} \end{cases}$$

La funzione di ripartizione congiunta

Definizione: $F(x,y) = P(X \leq x,\ Y \leq y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(s,t)\,ds\,dt$

Ha proprietà simili al caso univariato anche se ora si parla di volume piuttosto che di area

1. $F(x,y)$ é non decrescente per entrambi gli argomenti
2. $F(x,y)$ è continua a destra per ciascun argomento
3. $F(-\infty,y) = F(x,-\infty) = 0$ (la superficie si appiattisce ai valori estremi)
4. $F(\infty,\infty) = 1$ La probabilità dell'evento certo è pari ad uno
5. Per ogni $x_1 < x_2$ e $y_1 < y_2$ vale la disuguaglianza:

$$F(x_1,y_1) + F(x_2,y_2) \geq F(x_2,y_1) + F(x_1,y_2)$$

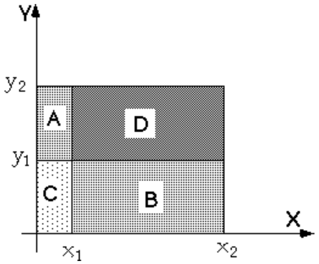
Funzione di ripartizione/2

La disuguaglianza

$$F(x_1,y_1) + F(x_2,y_2) \geq F(x_2,y_1) + F(x_1,y_2)$$

[C]+[A+B+C+D] ≥[C+B] + [A+C]

Indica semplicemente che la probabilità in una regione rettangolare di tipo "D" è non negativa.



A+B+C+D **C+B** **C** **A+C**

$$P(X \leq x_2, Y \leq y_2) - P(X \leq x_2, Y \leq y_1) + P(X \leq x_1, Y \leq y_1) - P(X \leq x_1, Y \leq y_2) \geq 0$$

$$P(X \leq x_2, y_1 \leq Y \leq y_2) - P(X \leq x_1, y_1 \leq Y \leq y_2) \geq 0$$

$$P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) \geq 0$$

Se questo non si verifica la "f" non potrà essere una funzione di densità congiunta.

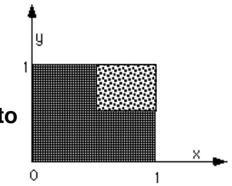
Esempio

Consideriamo la seguente funzione di densità congiunta:

$$f(x,y) = 1 \qquad 0 \leq x \leq 1;\quad 0 \leq y \leq 1$$

Quanto vale $P\left(X > \frac{1}{2}, Y > \frac{1}{2}\right)$ **?**

Supporto

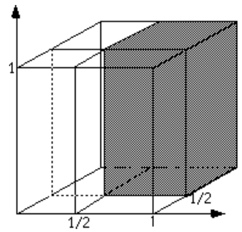


In questo caso il calcolo può essere condotto in soli termini geometrici.

Poiché il volume complessivo deve essere pari ad uno e poiché l'evento

$$\left(X > \tfrac{1}{2}\right) \cap \left(Y > \tfrac{1}{2}\right)$$

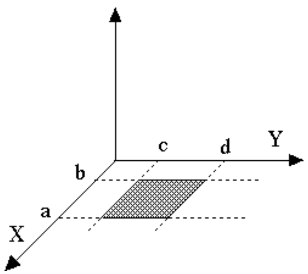
corrisponde ad una divisione in quattro cubetti di lato 1/2, allora il volume di pertinenza dell'evento è pari a 1/4.



Calcolo della funzione di ripartizione

Per calcolare i valori della funzione di ripartizione occorre calcolare un integrale doppio.

Se il supporto della v.c. doppia è una regione rettangolare, il calcolo è semplice



$$P(X \le x_0, Y \le y_0) = \int_a^{x_0}\int_c^{y_0} f(x,y)\,dx\,dy$$

Si integra prima rispetto alla "X" trattando la "Y" come una costante. Si ottiene una quantità che ora dipende solo dalla "Y"

$$\int_a^{x_0} f(x,y)\,dx = A(y) \quad \textbf{= f(y)}$$

A questo punto si integra tale funzione rispetto alla "Y"

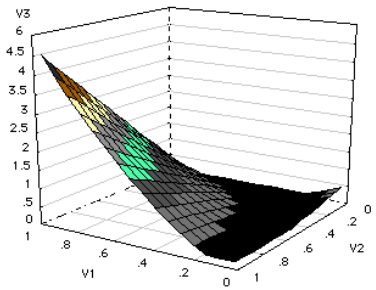
$$\int_c^{y_0} A(y)\,dy = F(x_0, y_0)$$

L'ordine di integrazione può essere scambiato rispetto alle due variabili

Esempio

Data la funzione di densità congiunta: $f(x,y) = 6x^2y \quad \text{per} \quad 0<x<1; \quad 0<y<1$

calcoliamo $P\left(0 < x < \frac{3}{4}, \frac{1}{3} < y < \frac{1}{2}\right)$



Fase_1: integrazione rispetto alla "X" tenendo la "Y" costante:

$$\int_0^{3/4} 6x^2y\,dx = 2yx^3\Big|_0^{3/4} = 2y\left(\frac{3}{4}\right)^3 = \frac{27}{32}y$$

Fase_2: integrazione rispetto alla "y"

$$\int_{1/3}^{1/2} \frac{27}{32}y\,dy = \frac{27}{64}y^2\Big|_{1/3}^{1/2} = \frac{27}{64}\left(\frac{1}{4}-\frac{1}{9}\right) = \frac{27}{64} * \frac{5}{36} = 0.0586$$

Identico risultato si sarebbe ottenuto cominciando l'integrazione con la "Y" e tenendo costante la "X".

Esercizio

Supponiamo che la funzione di densità congiunta di due v.c. sia

$$f(x,y) = \begin{cases} 3x^2y + x & \text{per } 0 \le x \le 1;\ 0 \le y \le 1 \\ 0 & \text{altrove} \end{cases}$$

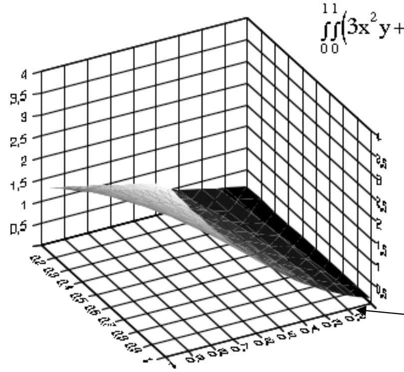
Verifichiamo che si tratti di una funzione di densità

L'integrale doppio si calcola agendo su di una variabile alla volta

$$\int_0^1\int_0^1 \left(3x^2y + x\right)dx\,dy = \int_0^1 \left(3x^2y + x\right)dx$$

$$= \left(x^3y + \frac{x^2}{2}\right)\Bigg|_0^1 = \left(1^3 * y + \frac{1^2}{2} - 0^3 * y - \frac{0^2}{2}\right) = y + \frac{1}{2}$$

$$= \int_0^1 \left(y + \frac{1}{2}\right)dy = \left(\frac{y^2}{2} + \frac{y}{2}\right)\Bigg|_0^1 = \frac{1}{2} + \frac{1}{2} - \frac{0}{2} - \frac{0}{2} = 1$$

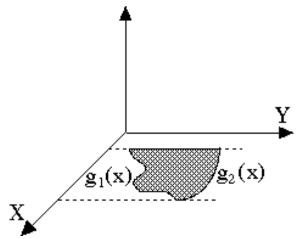


Il volume va calcolato limitatamente al supporto della densità.

Calcolo della funzione di ripartizione/2

Quando il supporto della funzione di densità ha forma diversa dalla rettangolare, il calcolo della F(x,y) richiede un accorgimento in più.

Supponiamo ad esempio che il dominio sia definito con le relazioni seguenti



$$a \le X \le b; \quad g_1(x) \le Y \le g_2(x)$$

In questo caso si integra prima rispetto alla "Y " adottando come estremi le due funzioni

$$A(x) = \int_{g_1(x)}^{g_2(x)} f(x,y)\,dy$$

Si ottiene una funzione della sola "X" che può essere integrata in (a,b)

$$\int_a^{x_0} A(x)\,dx = F(x_0, y_0)$$

La stessa impostazione andrà seguita nel caso di ruolo scambiato delle variabili

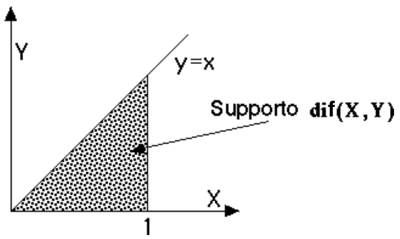
Esempio

Consideriamo la funzione di densità congiunta

$$f(x,y)=\begin{cases}2 & \text{per } 0\le y\le x\le 1\\ 0 & \text{altrove}\end{cases} \longleftrightarrow \begin{array}{l}1.\ 0\le x\le 1\\ 2.\ 0\le y\le x\end{array}$$

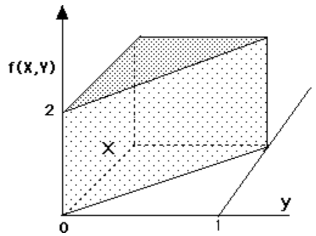
Le funzioni che delimitano il dominio della v.c. sono:

$$g_1(x)=0;\quad g_2(x)=x;\quad g_1(x)\le y\le g_2(x)$$



Verifichiamo se è una funzione di densità

$$\int_0^x 2y^0dy=2y\Big|_0^x=2x;\quad \int_0^1 2xdx=x^2\Big|_0^1=1$$



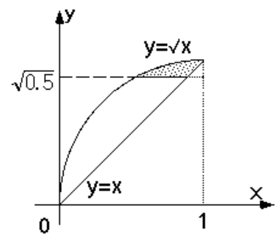
Esercizio

La v.c. doppia (X,Y) ha funzione di densità $f(x,y)=12xy \quad \text{per} \quad 0\le x\le 1;\quad x\le y\le \sqrt{x}$

Calcoliamo $P\left(0.5\le x\le 1,\ \sqrt{0.5}\le y\le 1\right)$

$$\int_x^{\sqrt{x}} 12xydy=12x\left.\frac{y^2}{2}\right|_x^{\sqrt{x}}=6x\left(x-x^2\right)=6\left(x^2-x^3\right)$$

$$\int_{1/2}^1 6\left(x^2-x^3\right)dx=6\left.\left[\frac{x^3}{3}-\frac{x^4}{4}\right]\right|_{1/2}^1=0.3437$$



Esercizio

I pneumatici anteriori di un certo tipo di automobile dovrebbero avere una pressione di 2.6 atmosfere.

Si supponga che la pressione reale dei due pneumatici anteriori sia una v.c. con funzione di densità

$$f(R,S)=\begin{cases}k\left(r^2+s^2\right) & \text{per } 2\le r\le 3;\ 2\le s\le 3\\ 0 & \text{altrove}\end{cases}$$

dove S è il pneumatico sul lato guida ed R quello a fianco

Quando deve essere la costante "k" perché la "f" sia una funzione di densità?

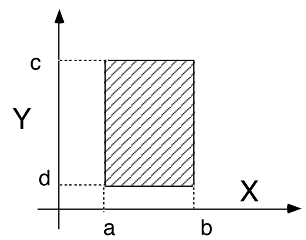
$$k\int_2^3\int_2^3\left(r^2+s^2\right)drds=1\Rightarrow \int_2^3\left[\int_2^3\left(r^2+s^2\right)ds\right]dr=\int_2^3\left[r^2s+\frac{r^3}{3}\Big|_2^3\right]dr=\int_2^3\left(r^2+\frac{19}{3}\right)dr=\frac{1}{k}$$

$$\left.\frac{r^3}{3}+\frac{19}{3}r\right|_2^3=\frac{38}{3}\Rightarrow k=\frac{3}{38}$$

Una semplificazione

Se il supporto è di tipo RETTANGOLARE

$a\le X\le b;\quad c\le Y\le d$



e se la funzione di densità è scomponibile in due fattori moltiplicativi

f(X,Y)= g(X)*h(Y)

In cui ogni fattore dipende da una sola variabile, allora

$$F(X,Y)=F(x)*F(y)=\left[\int_a^x g(t)dt\right]*\left[\int_a^y h(t)dt\right]$$

ESEMPIO $\quad f(x,y)=\dfrac{x*y}{16};\quad 0\le x\le 2;\ 0\le y\le 4$

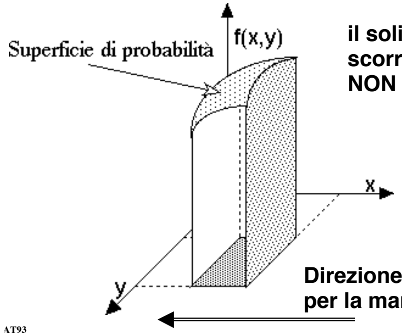
$$F4(4,2)=F(4)*F(2)=\left[\int_0^4\frac{t}{4}dt\right]*\left[\int_0^2\frac{t}{4}dt\right]=\frac{4}{8}*\frac{16}{8}=1$$

Distribuzioni marginali

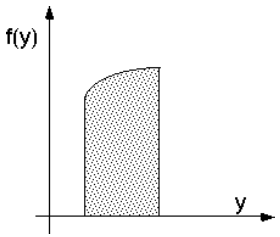
Le distribuzioni marginali e condizionali delle doppie continue hanno struttura identica alle v.c. doppie discrete

Per le marginali si ha $f(x) = \int_A f(x,y)\,dy; \qquad f(y) = \int_A f(x,y)\,dx;$

Cioè si "integra" sul supporto di definizione rispetto alla variabile che NON interessa



il solido è schiacciato con uno scorrimento lungo l'asse che NON interessa

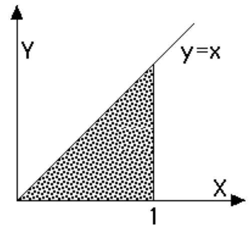


Direzione dello schiacciamento per la marginale della "y"

Esempio

Consideriamo la seguente funzione di densità doppia

$$f(x,y) = \begin{cases} 8xy & \text{per } 0 \le x \le 1;\ 0 \le y \le x \\ 0 & \text{altrove} \end{cases}$$



Costruiamo le marginali della "X" e della "Y"

$$f(x) = \int_0^x 8xy\,dy = 8x \left.\frac{y^2}{2}\right|_0^x = 4x^3 \quad \text{per } 0 \le x \le 1$$

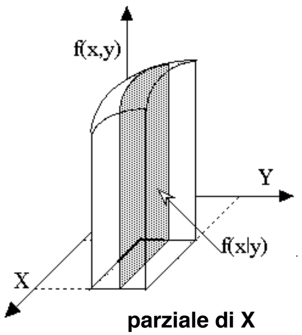
$$f(y) = \int_y^1 8xy\,dx = 8y \left.\frac{x^2}{2}\right|_y^1 = 4y\left(1-y^2\right) \quad \text{per } 0 \le y \le 1$$

occore tenere conto che i valori della "x" sono limitati inferiormente da quelli della "y".

Distribuzioni condizionali

La definizione parte sempre dal rapporto tra la congiunta e la marginale della variabile che condiziona

$$f(x|y) = \frac{f(x,y)}{f(y)}; \qquad f(y|x) = \frac{f(x,y)}{f(x)};$$



parziale di X dato Y

è come se si fosse "affettato" del pane

Esempio: $f(x,y) = 3x \quad \text{per } 0 \le x \le 1; \quad 0 \le y \le x$

Determiniamo f(y|x).

$$f(x) = \int_0^x 3x\,dy = 3x\ y\Big|_0^x = 3x^2 \quad \text{per } 0 \le x \le 1$$

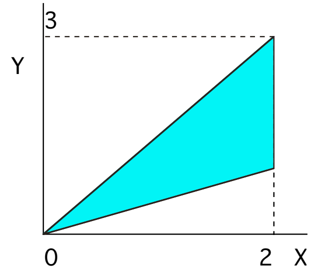
$$f(y|x) = \frac{3x}{3x^2} = \frac{1}{x} \quad \text{per } 0 < y \le x;$$

$$f(y|x = 0.5) = \frac{1}{0.5} = 2 \quad \text{per } 0 < y \le 0.5$$

Esercizio

Per la funzione di densità doppia

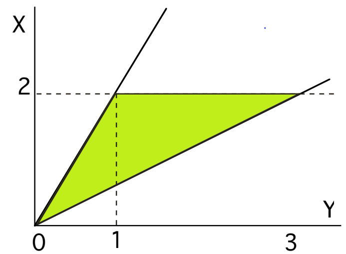
$$f(x,y) = 3(x-y)^2; \quad 0 \le x \le 2; \quad \frac{x}{2} \le y \le \frac{3x}{2}$$



definire le marginali della "X" e della "Y"

$$f(x) = 3\int_{\frac{x}{2}}^{\frac{3}{2}x} \left(x^2 + y^2 - 2xy\right)dy; = 3\left[yx^2 + \frac{y^3}{3} - xy^2\right]_{\frac{x}{2}}^{\frac{3}{2}x} = \frac{x^3}{4} \quad 0 \le x \le 2$$

$$f(y) = \begin{cases} 8 + 6y^2 - 12y - 2y^3 & per \quad 1 \le y \le 3 \\ \frac{28}{27}y^3 & per \quad 0 \le y \le 1 \end{cases}$$



La rotazione degli assi modifica -apparentemente- l'aspetto del dominio

Esercizio

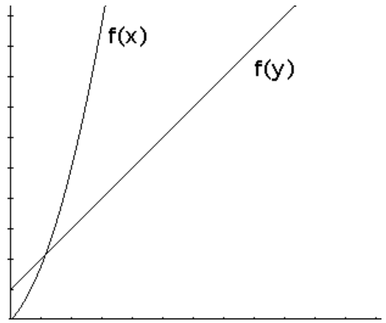
Consideriamo ancora la densità congiunta:

$$f(x,y) = 3x^2y + x \quad \text{per} \quad 0 \le x \le 1; \quad 0 \le y \le 1$$

Densità marginali

$$f(x) = \int_0^1 (3x^2y + x)dy = \left(3x^2\frac{y^2}{2} + xy\right)\Bigg|_0^1 = \frac{3x^2}{2} + x$$

$$f(y) = \int_0^1 (3x^2y + x)dx = \left(x^3y + \frac{x^2}{2}\right)\Bigg|_0^1 = y + \frac{1}{2}$$



Densità condizionate

$$f(y|x) = \frac{f(x,y)}{f(x)} = \frac{3x^2y + x}{\frac{3x^2}{2} + x} = \frac{6xy + 2}{3x + 2}$$

$$f(x|y) = \frac{f(x,y)}{f(y)} = \frac{3x^2y + x}{\frac{1}{2} + y} = \frac{6x^2y + 2x}{2x + 1}$$

Per ciascuno degli infiniti valori della x o della y è possibile definire una densità condizionata

Richiami sull'indipendenza

La dipendenza di una v.c. X da un'altra Y significa che è possibile formarsi delle opinioni di probabilità sulla prima anche partendo dalla conoscenza della seconda

Lo stesso vale ovviamente per le v.c. doppie continue: si dicono indipendenti se ...

1. Definizione: $f(x|y) = f(x)$ oppure $f(y|x) = f(y)$

2. Definizione: $f(x,y) = f(x) * f(y)$

Due v.c. sono indipendenti se la funzione di densità congiunta rispetta le due condizioni seguenti:

1. può essere scritta come il prodotto di due fattori in cui compaiono separatamente le due variabili;
2. il campo di variazione di una variabile non dipende dall'altra

Esempio di indipendenza

$$f(x,y) = \lambda_1\lambda_2 e^{-\lambda_1 x - \lambda_2 y} \qquad x > 0, \; y > 0$$

$$f(x,y) = \left(\lambda_1 e^{-\lambda_1 x}\right) * \left(\lambda_2 e^{-\lambda_2 y}\right) = f(x)f(y)$$

Indipendenza nelle doppie continue

In generale, abbiamo: $f(X,Y) = g(X) * h(Y)$

Esempio: $f(X,Y) = \dfrac{x^2(y+1)}{28} \quad 2 \le x \le 4; \quad 1 \le y \le 2 \qquad \Leftrightarrow g(X) = \dfrac{X^2}{28}; \quad h(Y) = (Y+1)$

Da notare che in questo caso si può sfruttare la regola di integrazione

$$\int_a^b\int_c^d g(x)h(y)dxdy = \left[\int_a^b g(x)dx\right] * \left[\int_c^d h(y)dy\right]$$

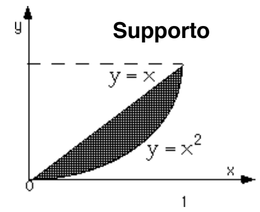
e la funzione di ripartizione doppia diventa: $F(x,y) = \left[\int_a^x g(t)dt\right] * \left[\int_c^y h(t)dt\right]$

Indipendenza nelle doppie continue/2

Se gli estremi di variazione sono interconnessi le due variabili sono dipendenti anche se la distribuzione può essere fattorizzata.

$$f(x,y) = 15y \qquad 0 \le x \le 1; \quad x^2 \le y \le x$$

Supporto



In questo caso le marginali sono:

$$f(x) = \int_{x^2}^{x} 15y dy = 15\frac{y^2}{2}\Bigg|_{x^2}^{x} = \frac{15}{2}(x^2 - x^4) \quad \text{per} \quad 0 \le x \le 1$$

$$f(y) = \int_y^{\sqrt{y}} 15y dx = 15yx\Big|_y^{\sqrt{y}} = 15y(\sqrt{y} - y^2) \quad \text{per} \quad 0 \le y \le 1$$

E' evidente che:

$$f(y) * f(x) = 15y(\sqrt{y} - y^2) * \frac{15}{2}(x^2 - x^4) \ne f(x,y)$$

Esercizio

Supponiamo che due funzioni di densità doppia abbiano le stesse marginali. Se ne può concludere che pure esse siano uguali?

$$\boxed{f(x,y) = x + y; \quad 0 \le x \le 1; \quad 0 \le y \le 1;}$$

$$f(x) = \int_0^1 (x+y)\,dy = \left(xy + \frac{y^2}{2}\right)\Bigg|_0^1 = \left(x + \frac{1}{2}\right)$$

$$f(y) = \int_0^1 (x+y)\,dx = \left(\frac{x^2}{2} + yx\right)\Bigg|_0^1 = \left(\frac{1}{2} + y\right)$$

$$\boxed{f(x,y) = \left(x + \frac{1}{2}\right)\left(y + \frac{1}{2}\right); \quad 0 \le x \le 1; \quad 0 \le y \le 1;}$$

$$f(x) = \int_0^1 \left(x + \frac{1}{2}\right)\left(y + \frac{1}{2}\right)dy = \left(x + \frac{1}{2}\right)\left(\frac{y^2}{2} + \frac{y}{2}\right)\Bigg|_0^1 = \left(x + \frac{1}{2}\right)$$

$$f(y) = \int_0^1 \left(x + \frac{1}{2}\right)\left(y + \frac{1}{2}\right)dx = \left(y + \frac{1}{2}\right)\left(\frac{x^2}{2} + \frac{x}{2}\right)\Bigg|_0^1 = \left(y + \frac{1}{2}\right)$$

La risposta è no: marginali uguali non implicano congiunte uguali. Ciò succede solo in caso di indipendenza

Ancora sull'indipendenza

Se X ed Y sono indipendenti lo sono anche delle loro funzioni.

$$\text{Se } f(x,y) = f(x) * f(y) \Rightarrow f[g(x), g(y)] = f[g(x)] * f[g(y)]$$

dato che le trasformazioni possono essere applicate separatamente alle marginali

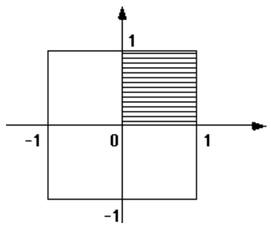
il contrario non è necessariamente vero:

$$\text{Se } f[g(x), g(y)] = f[g(x)] * f[g(y)] \not\Rightarrow f(x,y) = f(x) * f(y)$$

Esempio

Supponiamo che X ed Y abbiano la densità congiunta seguente

$$f(x,y) = \frac{1 + xy}{4} \quad \text{per } |x| < 1; \; |y| < 1$$



La struttura della "f" non suggerisce la fattorizzazione e infatti le due varibili sono dipendenti:

$$f(x) = \frac{1}{4}\int_{-1}^{1}(1 + xy)\,dy = \frac{1}{4}\left(y + x\frac{y^2}{2}\right)\Bigg|_{-1}^{1} = \frac{1}{2}; \qquad f(y) = \frac{1}{4}\int_{-1}^{1}(1 + xy)\,dx = \frac{1}{4}\left(x + y\frac{x^2}{2}\right)\Bigg|_{-1}^{1} = \frac{1}{2};$$

Ne consegue che f(x,y) ≠ f(x)*f(y)

Tuttavia, nella zona 0<x<1; 0<y<1 i quadrati delle variabili sono indipendenti

$$P(X^2 \le x, Y^2 \le y) = P(|X| \le x, |Y| \le y) = \frac{1}{4}\int_{-\sqrt{x}}^{\sqrt{x}}\int_{-\sqrt{y}}^{\sqrt{y}}(1 + s * t)\,ds\,dt = \frac{1}{4}\int_{\sqrt{x}}^{\sqrt{x}}\left[s + \frac{s^2}{2}t\Bigg|_{-\sqrt{y}}^{\sqrt{y}}\right]dt$$

$$= \frac{1}{4}\int_{-\sqrt{x}}^{\sqrt{x}}(2\sqrt{y})\,dt = \frac{1}{2} * 2\int_0^{\sqrt{x}}(\sqrt{y})\,dt = \sqrt{x}\sqrt{y} = P(X^2 \le x) * P(Y^2 \le y)$$

I momenti delle v.c. doppie

La struttura dei momenti è tale da non richiedere particolari modifiche quando dalle doppie discrete si passa alle doppie continue

Per una funzione generica g(x,y) si ha

$$E[g(x,y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f(x,y)\,dx\,dy$$

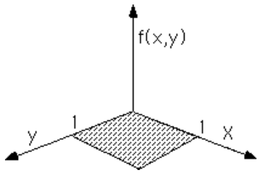
Momenti centrali **Momenti scarto**

$$\mu_{rs} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^r y^s f(x,y)\,dx\,dy \qquad \bar{\mu}_{rs} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_i - \mu_x)^r (y_j - \mu_y)^s f(x,y)\,dx\,dy$$

Valori attesi condizionati $E(X|y) = \int_{-\infty}^{\infty} x f(x|y)\,dx = \mu_x(y); \qquad E(Y|x) = \int_{-\infty}^{\infty} y f(y|x)\,dy = \mu_y(x)$

A parte la sostituzione delle sommatorie con gli integrali, nulla è cambiato rispetto alle v.c. discrete

Esempio



Consideriamo la seguente funzione di densità

$$f(x,y)=\begin{cases}2*(1-x) & \text{per } \; 0\le x\le 1;\;\; 0\le y\le 1\\ 0 & \text{altrove}\end{cases}$$

Per il calcolo delle medie marginali occorre la densità marginale

$$f(X)=\int_{-\infty}^{\infty} f(x,y)\,dy=\int_0^1 2*(1-x)\,dy=2*(1-x)*y\Big|_0^1=2*(1-x)$$

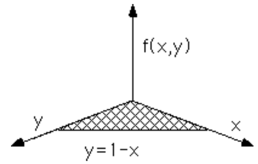
$$E(X)=\int_{-\infty}^{\infty} x f(x)\,dx=\int_0^1 x*2*(1-x)\,dx=2\int_0^1 (x-x^2)\,dx=\left[2*\left(\frac{x^2}{2}-\frac{x^3}{3}\right)\right]\Bigg|_0^1=\frac{1}{3}$$

$$f(y)=\int_0^1 2*(1-x)\,dx=1\Rightarrow E(y)=\int_0^1 y*1\,dy=\left(\frac{y^2}{2}\right)\Bigg|_0^1=\frac{1}{2}$$

Da notare che le due variabili sono indipendenti:

$$f(y,x)=2*(1-x)=[1]*[2*(1-x)]=f(y)*f(x)$$

Esercizio



Si supponga che la v.c. doppia abbia funzione di densità:

$$f(x,y)=24xy \quad \text{per} \quad 0\le x\le 1;\quad 0\le x+y\le 1$$

y≤1-x

Calcolare il valore atteso di g(x,y)=0.5+0.5X+Y

$$E[g(x,y)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)\,dx\,dy=\int_0^1\int_0^{1-x}(0.5+0.5x+y)24xy\;\;dy\,dx$$

$$=24\int_0^1\int_0^{1-x} 0.5xy+0.5x^2y+xy^2\;\;dy\,dx=24\int_0^1\left[0.25xy^2+0.25x^2y^2+\frac{xy^3}{3}\Bigg|_0^{1-x}\right]dx$$

$$=24\int_0^1 \frac{(x+x^2)(1-x)^2}{4}+\frac{x(1-x)^3}{3}\,dx=1.1$$

Calcolo della Covarianza

$$f(x,y)=\begin{cases}8xy & \text{per } \; 0\le x\le 1;\; 0\le y\le x\\ 0 & \text{altrove}\end{cases}$$

$$\mathrm{Cov}(x,y)=\int_0^1\int_0^x xy f(x,y)\,dy\,dx-E(Y)*E(Y)$$

$$=\int_0^1\int_0^x xy8xy\,dy\,dx-\frac{4}{5}*\frac{1}{3}$$

$$=8\int_0^1 x^2\left(\int_0^x y^2\,dy\right)dx-\frac{4}{15}$$

$$=8\int_0^1 x^2*\frac{x^3}{3}\,dx-\frac{4}{15}\quad=\frac{8}{18}-\frac{4}{15}=\frac{8}{45}$$

Densità marginali

$$f(x)=\int_0^x 8xy\,dy=8x\left(\frac{y^2}{2}\right)\Bigg|_0^x=8x\frac{X^2}{2}=4x^3$$

$$E(x)=\int_0^1 x*(4x^3)\,dx=4\int_0^1 x^4\,dx=4\left(\frac{x^5}{5}\right)\Bigg|_0^1=\frac{4}{5}$$

$$f(y)=\int_y^1 8xy\,dx=8y\left(\frac{x^2}{2}\right)\Bigg|_y^1=4y(1-y^2)$$ **0≤y≤1**

$$E(y)=\int_0^1 y*[4y(1-y^2)]\,dy=4\int_0^1 y^2-y^3\,dy=4\left(\frac{y^3}{3}-\frac{y^4}{4}\right)\Bigg|_0^1=\frac{1}{3}$$

Esercizio

Per la seguente funzione di densità

$$f(x,y)=\begin{cases}3y & \text{per } \; 0\le x\le y\le 1\\ 0 & \text{altrove}\end{cases}$$

N.B.

$$f(x)=\frac{3}{2}*(1-x^2) \qquad 0<x\le 1;$$

$$f(y)=3*y^2 \qquad 0\le y\le 1$$

Calcolare il coefficiente di correlazione

$$E(x)=\int_0^1 x\frac{3}{2}(1-x^2)\,dx=\frac{3}{8};\quad \sigma^2(x)=\int_0^1 x^2\frac{3}{2}(1-x^2)\,dx-\left(\frac{3}{8}\right)^2=\frac{19}{320}$$

$$E(y)=\int_0^1 y3y^2\,dy=\frac{3}{4};\quad \sigma^2(y)=\int_0^1 y^2 3y^2\,dy-\frac{9}{16}=\frac{3}{5}-\frac{9}{16}=\frac{3}{80}$$

$$E(xy)=\int_0^1\int_x^1 xy3y\,dx\,dy=\int_0^1\int_x^1 3xy^2\,dx\,dy$$

$$=\int_0^1\left(3x\int_x^1 y^2\,dy\right)dx=\int_0^1 x(1-x^3)\,dx=\frac{3}{10}$$

$$r(x,y)=\frac{E(xy)-E(x)E(y)}{\sigma(x)\sigma(y)}=\frac{\frac{3}{10}-\frac{3}{8}*\frac{3}{4}}{\sqrt{\frac{57}{25600}}}=0.3974$$

Esercizio

La variabile casuale doppia (X,Y) ha densità $f(x,y)=1 \quad 0\le x\le 1; \quad x^2\le y\le \sqrt{x}$

calcolare il coefficiente di correlazione

$$f(x)=\int_{x^2}^{\sqrt{x}} 3dy = 3y\Big|_{x^2}^{\sqrt{x}} = 3\left(\sqrt{x}-x^2\right) \qquad 0\le x\le 1;$$

$$E(x)=\int_0^1 3x\left(\sqrt{x}-x^2\right)dx=\frac{9}{20}$$

$$E\left(x^2\right)=\int_0^1 3x^2\left(\sqrt{x}-x^2\right)dx=\frac{9}{35}$$

$$\sigma^2(x)=\frac{9}{35}-\left(\frac{9}{20}\right)^2=0.0546$$

$$f(y)=\int_{y^2}^{\sqrt{y}} 3dx = 3x\Big|_{y^2}^{\sqrt{y}} = 3\left(\sqrt{y}-y^2\right) \qquad 0\le y\le 1;$$

$$E(y)=\int_0^1 3y\left(\sqrt{y}-y^2\right)dy=\frac{9}{20}$$

$$E\left(y^2\right)=\int_0^1 3y^2\left(\sqrt{y}-y^2\right)dx=\frac{9}{35}$$

$$\sigma^2(y)=\frac{9}{35}-\left(\frac{9}{20}\right)^2=0.0546$$

$$E(xy)=3\int_0^1\int_{x^2}^{\sqrt{x}} xy\,dy\,dx=\frac{1}{4}\Rightarrow Cov(x,y)=\frac{1}{4}-\left(\frac{9}{20}\right)^2=0.0475\Rightarrow r(x,y)=\frac{0.0475}{\sqrt{0.0546*0.0546}}=0.87$$

Esercizio

Il tempo che Caterina Ruffolo impiega da casa all'università è una v.c. X_1 con valore atteso 25' e deviazione standard 5'.

il ritorno, è una v.c. X_2, indipendente dalla prima, con valore atteso di 20' e deviazione standard di 4'.

Che comportamento ha la differenza X_1-X_2 ***?***

Anche questa è una v.c. con valore atteso: $E(X_1-X_2)=E(X_1)-E(X_2)=25-20=4$

e con varianza: $\sigma^2(X_1-X_2)=\sigma^2(X_1)+\sigma^2(X_2)=25+16=41$

Che comportamento ha il tempo medio $\frac{X_1+X_2}{2}$ ***?***

$$E\left(\frac{X_1+X_2}{2}\right)=\left(\frac{1}{2}\right)E(X_1)+\left(\frac{1}{2}\right)E(X_2)=\left(\frac{25}{2}\right)+\left(\frac{20}{2}\right)=22.5$$

$$\sigma^2\left(\frac{X_1+X_2}{2}\right)=\left(\frac{1}{2}\right)^2\sigma^2(X_1)+\left(\frac{1}{2}\right)^2\sigma^2(X_2)=\frac{25}{4}+\frac{16}{4}=10.25$$

Esercizio

Per la funzione di densità:

$$f(X,Y)=\begin{cases}\frac{3}{2} & \text{per } \; 0\le x\le 1 \;\text{ e }\; x^2\le y\le 1\\ 0 & \text{altrove}\end{cases}$$

a) Verificare che si tratti in effetti di una funzione di densità;

b) Determinare le marginali f(X) e f(Y)

c) Determinare le condizionate f(XIY) e f(YIX)

d) Calcolare $E\left(X^2Y^3\right)$

e) Calcolare r(x,y)

La distribuzione uniforme doppia

Si tratta di una mera estensione della uniforme semplice

$$f(x,y)=k \qquad \text{per } (x,y)\in A$$

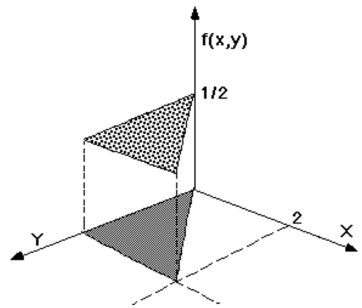
dove "A" è il supporto su cui la densità è positiva e k è tale che il volume coperto sia pari ad uno.

ESEMPIO:

$$f(x,y)=\frac{1}{2} \qquad \text{per } \; 0<x<2; \;\; 0<y<x$$

$$\int_0^2\int_0^x \frac{1}{2}\,dydx=\int_0^2\left[\int_0^x \frac{1}{2}dy\right]dx=\int_0^2\left[\frac{1}{2}y\Big|_0^x\right]dx$$

$$=\int_0^2 \frac{x}{2}dx=\frac{x^2}{4}\Big|_0^2=1$$



Ancora sulla uniforme doppia

Ricavare le marginali e le condizionali nella uniforme doppia è facile data la struttura molto semplice della funzione di densità:

Rispetto all'esempio si ha

$$f(y) = \int_y^2 \frac{1}{2}\,dx = \frac{1}{2}x\Big|_y^2 = 1 - \frac{y}{2} \quad \text{per} \quad 0 < y < 2; \qquad f(x) = \int_0^x \frac{1}{2}\,dy = \frac{1}{2}y\Big|_0^x = \frac{x}{2} \quad \text{per} \quad 0 < x < 2$$

Le marginali non sono necessariame delle uniformi a causa della dipendenza tra gli estremi del campo di variazione della X e della Y.

Le condizionate sono invece, necessariamente, delle uniformi

$$f(y|x) = \frac{\frac{1}{2}}{\frac{x}{2}} = \frac{1}{2} \quad \text{per } 0 < y < 2; \qquad f(x|y) = \frac{\frac{1}{2}}{\frac{2-y}{2}} = \frac{1}{2-y} \quad \text{per } 0 < x < 2;$$

La Normale Doppia

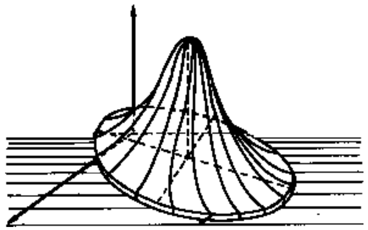
Tra le variabili continue bivariate ha rilevanza la distribuzione normale doppia

$$f(X,Y) = \frac{e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \quad \text{per} \quad -\infty < x, y < \infty$$

La funzione di densità dipende da 5 parametri: medie e varianze le conosciamo. Il parametro "ρ" esprime il coefficiente di correlazione tra la "X" e la "Y"

r(x,y)=ρ

che studieremo più avanti. Per ora ci basta sapere che $-1 \le \rho \le 1$



Ancora sulla normale doppia

1. Le marginali sono pure delle normali **Riproducibilità**

$$f(X) = \frac{e^{-\frac{1}{2}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right]}}{\sigma_x\sqrt{2\pi}} \quad \text{per} \quad -\infty < x < \infty; \quad f(y) = \frac{e^{-\frac{1}{2}\left[\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}}{\sigma_y\sqrt{2\pi}} \quad \text{per} \quad -\infty < y < \infty$$

2. Le condizionali sono pure normali

$$X|Y \sim N\left(\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y-\mu_y); \sigma_x^2 * (1-\rho^2)\right); \quad Y|X \sim N\left(\mu_y + \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x); \sigma_y^2 * (1-\rho^2)\right)$$

Con medie che sono funzioni lineari della variabile condizionante.

3. Se la X e la Y sono incorrelate allora sono anche indipendenti

$$\rho = 0 \Rightarrow f(x,y) = \frac{e^{-\frac{1}{2}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}}{2\pi\sigma_x\sigma_y} = \frac{e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}}{\sigma_x\sqrt{2\pi}} * \frac{e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2}}{\sigma_y\sqrt{2\pi}} = f(x)*f(y)$$

N.B. In generale questo non succede

Esempio

Supponiamo che $\mu_x = 10; \quad \sigma_x = 3; \quad \mu_y = 15; \quad \sigma_y = 4; \quad \rho = 0.8$

La distribuzione condizionale di Y dato X è

$$N\left[15 + 0.8\frac{4}{3}(x-10), 16(1-0.8^2)\right]$$

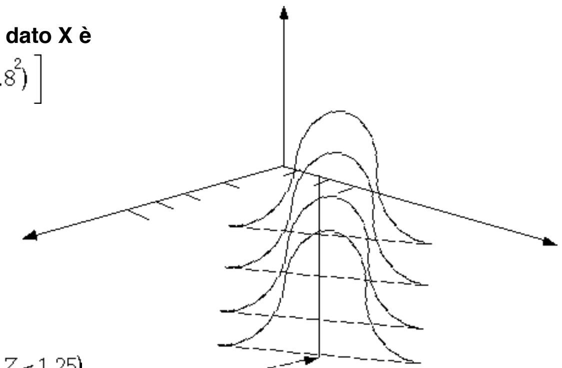
Calcoliamo $P(16 \le y \le 18|x = 10)$

La condizionata è ora:

$$N\left(\frac{13}{3} + \frac{32\mathbf{x}}{3}, 16(1-0.8^2)\right) = N(15, 5.76)$$

Quindi:

$$P(16 \le y \le 18|x = 10) = P(0.42 \le Z \le 1.25) = 0.2316$$



$$E(Y|x) = 15 + 0.8\frac{4}{3}(X-10) = \frac{13}{3} + \frac{3.2}{3}x$$

Esercizio

Si indichi con "X" il voto di diploma ed "Y" il voto di laurea. Siano dati i parametri

$$\mu_x = 48; \quad \mu_y = 101; \quad \sigma_x = 6; \quad \sigma_y = 9; \quad \rho = 0.8$$

a) Calcolare: $P(92 < Y < 105)$

a') Calcolare $P(50 < X < 58)$

c) Calcolare E(Y|X) e E(X|Y)

d) Calcolare σ(Y|X) e σ(X|Y)

e) Calcolare: $P(92 < Y < 105 | X = 60)$

f) Calcolare $P(50 < X < 58 | Y = 100)$