

Lezione 2: Descrizione dei Dati

<http://www.mi.infn.it/~palombo/didattica/AnalisiStatistica/Lezioni/Lezione2.pdf>

Funzione Densità di Probabilità

- ❑ La variabile casuale assume valori diversi con diverse probabilità. Vogliamo ora vedere come possiamo caratterizzare la distribuzione delle probabilità.
- ❑ Per semplicità noi supponiamo che la variabile casuale sia una variabile continua. Il passaggio a variabile discreta è generalmente semplice (da integrali a sommatorie, ecc).

Funzione Densità di probabilità con una sola variabile casuale

- ❑ Faccio una misura di una variabile casuale continua i cui valori possano andare da $-\infty$ a $+\infty$. Sia $F(x_0)$ la probabilità che la mia misura x (variabile casuale) sia minore di x_0 . Quindi $F(-\infty) = 0$, $F(+\infty) = 1$ e la probabilità che x sia compreso tra x e $x + dx$ è :

$$P(x \in [x, x + dx]) = F(x + dx) - F(x)$$

Funzione Densità di Probabilità

- Si definisce **funzione densità di probabilità** (p.d.f.) la funzione :
$$f(x) = dF/dx$$

Quindi $P(x \in [x, x + dx]) = F(x + dx) - F(x) = f(x)dx$

e

$$\int_{-\infty}^{+\infty} f(x)dx = 1 \quad (\text{integrale di normalizzazione})$$

- Se la variabile assume valori discreti , l'integrale va sostituito con una sommatoria, ecc

p.d.f. in due dimensioni

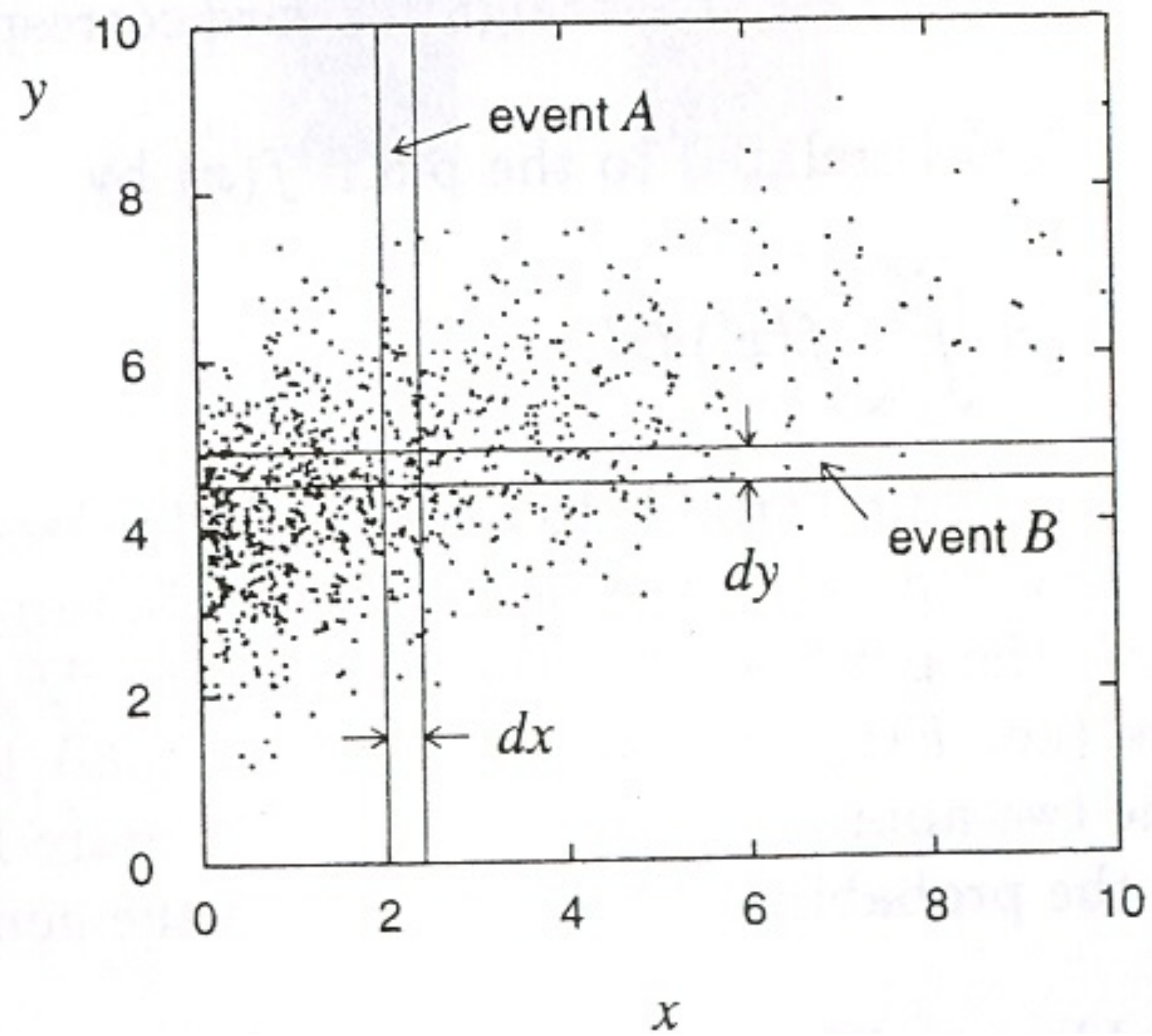
- Se la misura è caratterizzata da due variabili casuali x e y , allora la probabilità che la variabile X assuma un valore tra x e $x + dx$ e la variabile Y assuma un valore tra y e $y + dy$ è data da :

$$P(x \in [x, x + dx], y \in [y, y + dy]) = f(x, y) dx dy$$

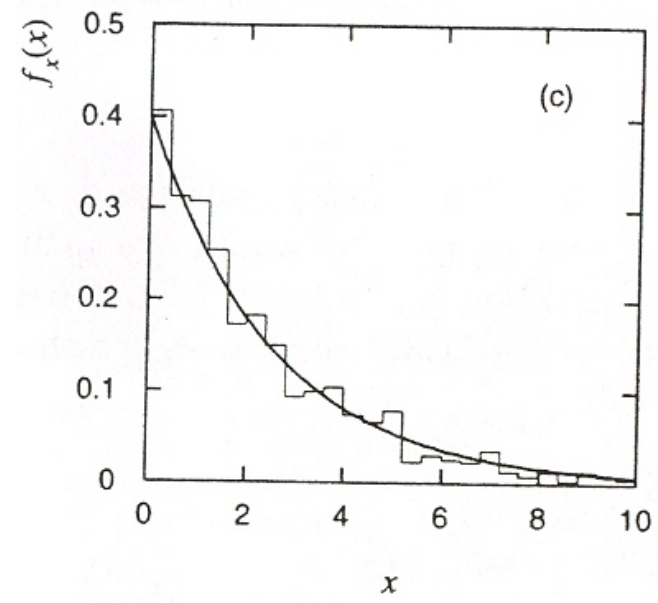
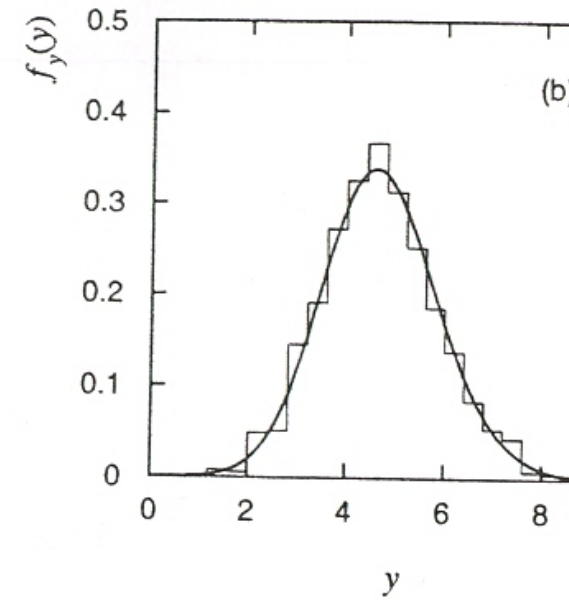
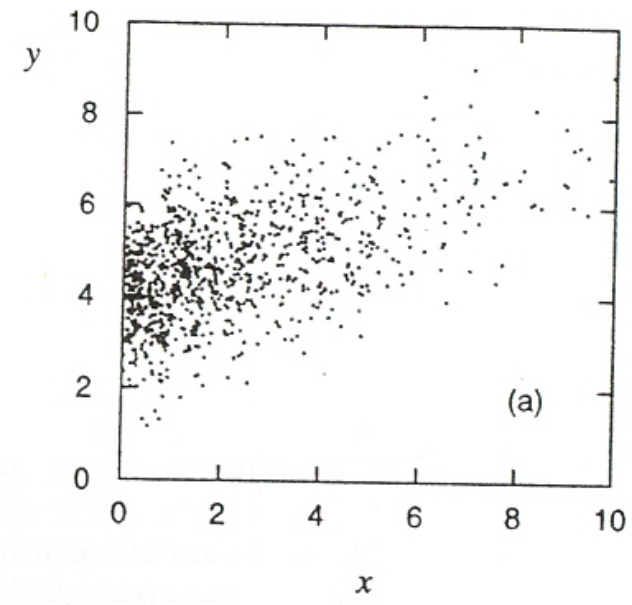
- La funzione $f(x, y)$ è detta **funzione densità di probabilità congiunta**. L'integrale di questa funzione esteso a tutti i possibili valori di x e y vale 1 (Condizione di normalizzazione).
- Nota la p.d.f. congiunta possiamo essere interessati ai valori che assume la x indipendentemente da quelli assunti da y (integriamo sulla variabile y):

$$f_x(x) = \int_{-\infty}^{+\infty} f(x, y) dy$$

- La funzione $f_x(x)$ è **detta p.d.f. marginale per x** . Analogamente per la p.d.f. marginale per y ($f_y(y)$). Sono proiezioni sugli assi (**scatter plot**)



Scatter Plot



p.d.f. Condizionale

- ❑ Calcoliamo la probabilità che y sia compreso tra y e $y + dy$ per qualunque valore di x (evento B) con la condizione che x sia compreso tra x e $x + dx$ con qualunque y (evento A) :

$$P(B | A) = \frac{P(A \cap B)}{P(A)} = \frac{\int f(x, y) dx dy}{\int f_x(x) dx}$$

- ❑ La p.d.f. condizionale $h(y | x)$ che si verifichi y dato x è così definita :

$$h(y | x) = \frac{f(x, y)}{f_x(x)} = \frac{f(x, y)}{\int f(x, y') dy'}$$

- ❑ In modo analogo è definita la p.d.f. condizionale $g(x | y)$ che si verifichi x dato y .

- ❑ Si verifica facilmente che $g(x | y) = \frac{h(y | x) f_x(x)}{f_y(y)}$ (teorema di Bayes)

- ❑ Se le due variabili x e y sono statisticamente indipendenti (il valore che assume una non dipende dal valore assunto dall'altra), allora $f(x, y) = f(x) f(y)$ e la p.d.f. marginale in x è semplicemente $f(x)$

p.d.f. Marginali. Esempio

- Sia $f(x,y) = \frac{3}{2} \cdot (x^2 + y^2)$ per $x = 0 < x < 1$, $0 < y < 1$ e uguale a 0 altrove. Calcolare la probabilità che la variabile x assuma un valore compreso tra 0 e 0.5, supponendo che y sia compreso tra 0 e 0.5. Calcolare inoltre la probabilità che x sia nell'intervallo (0, 0.5) quando la variabile y assume il valore 0.5. Le due variabili x e y sono indipendenti ?

- Sia A l'evento x compreso tra 0 e 0.5 e sia B l'evento y compreso tra 0 e 0.5

Allora :

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- Il numeratore si calcola con doppia integrazione della $f(x,y)$ negli intervalli dati e vale 1/16. Per calcolare il denominatore prima calcolo la p.d.f. marginale per y :

$$f_y(y) = \frac{3}{2} \int_0^1 (x^2 + y^2) dx = \frac{1}{2} + \frac{3}{2} y^2$$

- La probabilità che si verifichi l'evento B è: $\int_0^{0.5} \left(\frac{1}{2} + \frac{3}{2} y^2 \right) dy = \frac{5}{16}$

p.d.f. Marginali. Esempio

- Perciò la probabilità che si realizzi A dato B è uguale a 1/5
- La p.d.f. condizionale che si verifichi x dato y è :

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

- Ponendo $y = 0.5$ si ha $f(x|y = 0.5) = \frac{3}{7} + \frac{12}{7}x^2$

- e quindi $P(A|B) = \int_0^{0.5} \left(\frac{3}{7} + \frac{12}{7}x^2 \right) dx = \frac{2}{7}$

- Le due variabili X e Y non possono essere indipendenti perché la p.d.f. congiunta non può esprimersi come prodotto di una funzione della variabile X ed una della variabile Y.

Funzione Distribuzione Cumulativa

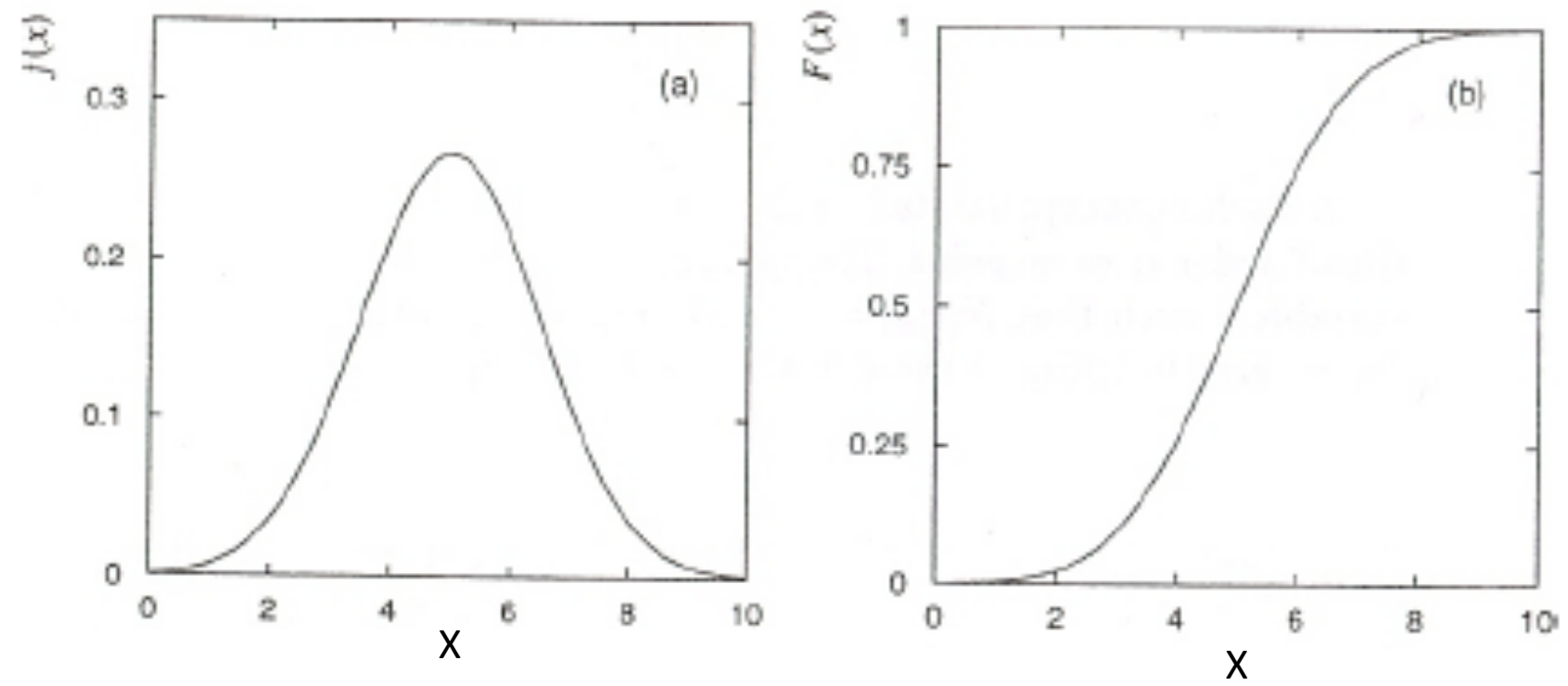
- ❑ La distribuzione di probabilità di una variabile casuale X può essere caratterizzata dalla sua **funzione di distribuzione cumulativa** (c.d.f.) $F(x)$ così definita:

$$F(x) = \int_{-\infty}^x f(u) du$$

dove $f(x)$ è la p.d.f. La c.d.f. $F(x)$ è la probabilità che la variabile casuale assuma un valore minore o uguale ad x

- ❑ Supponendo che $F(x)$ sia una funzione strettamente crescente, allora c'è corrispondenza biunivoca tra il valore della variabile x e il valore assunto dalla c.d.f. $F(x)$. → La funzione $F(x)$ può essere invertita.
- ❑ Si **definisce punto α (o quantile di ordine α)** x_α il valore della variabile x per il quale si ha:
$$F(x_\alpha) = \alpha \quad \text{con } 0 \leq \alpha \leq 1.$$
- ❑ Il quantile è l'inverso della distribuzione cumulativa: **$x_\alpha = F^{-1}(\alpha)$**
- ❑ Il quantile più usato è quello di ordine 0.5 che è detto mediana. Questo quantile (la mediana) divide la distribuzione in due parti uguali.

p.d.f. e c.d.f.



p.d.f. e corrispondente c.d.f.

Funzione Caratteristica

- ❑ Un terzo modo per rappresentare la distribuzione di una variabile casuale fa uso della **funzione caratteristica** (ch.f.) ϕ
- ❑ La funzione ϕ è definita come la trasformata di Fourier della p.d.f.
- ❑ Con una sola variabile casuale continua, si ha $\phi(t) = \int_{-\infty}^{+\infty} e^{itz} f(x) dx$
- ❑ Si ha una corrispondenza biunivoca tra p.d.f. e ch.f : la conoscenza che si ha con una di queste è del tutto equivalente a quella che si ha con l'altra funzione.
- ❑ Data la funzione caratteristica ϕ , la p.d.f. corrispondente si ottiene mediante l'antitrasformata di Fourier : $f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-ixt} \phi(t) dt$
- ❑ In una trattazione teorica, dove si devono dimostrare teoremi per esempio, l'uso della funzione caratteristica è molto più comodo e molto più semplice. Noi non useremo questo tipo di rappresentazione.

Funzioni di Variabili Casuali

- ❑ Una funzione di variabile casuale è essa stessa una variabile casuale.
- ❑ Sia $Y = \phi(X)$ una funzione continua di x e sia $f(x)$ la p.d.f. della variabile X . Qual è la p.d.f. $g(y)$ della variabile Y ?
- ❑ Si ha: $P(Y < y) = P(\phi(X) < y)$
- ❑ Se ϕ è una funzione **crescente**, si può scrivere:
 $P(Y < y) = P[X < \phi^{-1}(y)]$ Si noti che $x = \phi^{-1}(y)$
- ❑ Chiamando $G(y)$ e $F(x)$ le c.d.f. delle variabili Y ed X , si ha:
 $G(y) = F[\phi^{-1}(y)]$
- ❑ Differenziando rispetto ad y si ha: $g(y) = f[\phi^{-1}(y)] d\phi^{-1}/dy$ [1]
- ❑ Se la funzione $\phi(Y)$ è una funzione **decrescente**, allora

Funzioni di Variabili Casuali

- ❑ $P(Y < y) = P[X > \phi^{-1}(y)]$
che possiamo riscrivere: $G(y) = 1 - F[\phi^{-1}(y)]$
- ❑ Differenziando rispetto ad y , si ha: $g(y) = -f[\phi^{-1}(y)] \frac{d\phi^{-1}}{dy}$ [2]
- ❑ Mettendo assieme la [1] e la [2], si ha:
 $g(y) = f[\phi^{-1}(y)] |d\phi^{-1}(y)/dy|$ cioè anche :

 $g(y) = f(x) |dx/dy|$ [si noti che $d\phi^{-1}/dy = (d\phi/dx)^{-1} = dx/dy$]
- ❑ Il termine preso in valore assoluto è detto **Jacobiano della trasformazione**

Vediamo un esempio:

Funzione di Variabile Casuale

- ❑ Sia $f(x) = 1$ per x compreso tra 0 e 1 e $f(x) = 0$ altrove per una variabile casuale X . Preso $Y = X^2$, determinare la p.d.f. $g(y)$ della variabile casuale Y .
- ❑ $dy/dx = 2x$; $g(y) = 1/2x$ ($0 < x < 1$)
essendo $x = \sqrt{y} \rightarrow g(y) = 1/(2\sqrt{y})$
- ❑ Si verifica facilmente la condizione di normalizzazione :

$$\int_0^1 \frac{1}{2\sqrt{y}} dy = 1$$

- ❑ I risultati ottenuti si generalizzano al caso di funzioni di più variabili casuali !

p.d.f. della Somma di Due Variabili Casuali

□ $Z = X + Y$ e siano $g(x)$ e $h(y)$ le p.d.f. di X e Y , variabili casuali indipendenti

□ p.d.f. congiunta: $f(x,y) = g(x) h(y)$

□ Sia $P(z)$ la probabilità di avere $Z = X + Y \leq z$. Nel piano XY la regione in cui $X + Y \leq z$ è la parte di piano a sinistra sotto la retta di equazione $X + Y = Z$

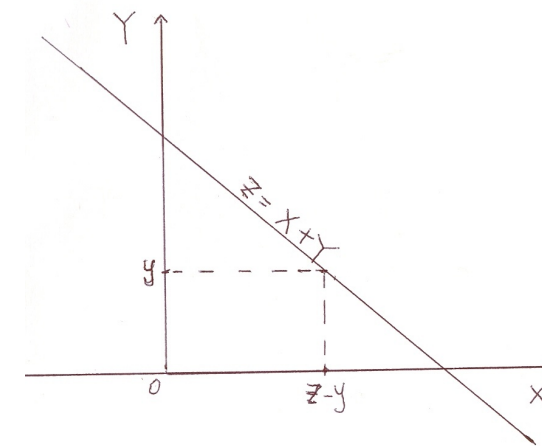
□ Per un dato y la probabilità $G(z-y)$ di avere $X \leq z-y$

$$G(z - y) = \int_{-\infty}^{z-y} g(u) du$$

□ La probabilità $P(z)$ si ottiene moltiplicando questa $G(z-y)$ per la probabilità di avere y e integrando su tutti i possibili y

$$P(z) = \int_{-\infty}^{+\infty} G(z - y)h(y)dy$$

□



p.d.f. della Somma di Due Variabili Casuali

- Differenziando rispetto a z , si ha $f(z) = \int_{-\infty}^{+\infty} g(z-y)h(y)dy$

Questa relazione si scrive così: $f = g \otimes h$ ed è detta **convoluzione (di Fourier)** di g e h . Si può anche riscrivere così:

$$f(z) = \int_{-\infty}^{+\infty} g(x)h(z-x)dx$$

- Esempio: la variabile casuale X abbia una distribuzione uniforme nell'intervallo $[0,1]$ e Y una distribuzione triangolare simmetrica nell'intervallo $[0,2]$. La variabile $Z = X + Y$ ha valore solo nell'intervallo $[0,3]$. Vogliamo determinare la p.d.f. della variabile Z .
- La p.d.f. di X è: $g(x) = 1$ per $0 < x < 1$;
La p.d.f. di Y è: $h(y) = y$ per $0 \leq y \leq 1$ e $h(y) = 2 - y$ per $1 \leq y \leq 2$
- La funzione $g(z-y) = 1$ per y tra $z-1$ e z mentre è 0 altrove.

p.d.f. della Somma di Due Variabili Casuali

□ Quindi $f(z) = \int_{z-1}^z h(y)dy$

□ $f(z)$ ha espressione diversa negli intervalli di z $[0,1]$, $[1,2]$, $[2,3]$:

$$f(z) = \int_0^z ydy = \frac{z^2}{2}, \text{ per } 0 < z < 1$$

$$f(z) = \int_{z-1}^1 ydy + \int_1^z (2-y)dy = 3(z-1/2) - z^2, 1 < z < 2$$

$$f(z) = \int_{z-1}^2 (2-y)dy = \frac{1}{2}(3-z)^2, 2 < z < 3$$

□ $f(z)$ è data da parti di tre parabole ed è simmetrica rispetto a $z=0.5$

p.d.f. di un Prodotto di Variabili Casuali

□ Sia $Z = XY$ con $g(x)$ e $h(y)$ le p.d.f. di X e Y , variabili casuali indipendenti

□ Si può far vedere che la p.d.f. $f(z)$ di Z è data da $f(z) = \int_{-\infty}^{+\infty} g(x)h\left(\frac{z}{x}\right)\frac{dx}{|x|}$

□ Questa relazione si può scrivere anche così

$$f(z) = \int_{-\infty}^{+\infty} g\left(\frac{z}{y}\right)h(y)\frac{dy}{|y|}$$

□ Questa relazione è spesso scritta così: $\mathbf{f = g \otimes h}$ ed è detta **convoluzione (di Mellin)** di g e h

Valori di Aspettazione

- **Media (aritmetica)** di una variabile casuale

Sia $f(x)$ la p.d.f. della variabile casuale X . Il valore di aspettazione $E[x]$ di questa variabile è definito da

$$E[x] = \int_{-\infty}^{+\infty} x f(x) dx = \mu$$

μ è detto anche media della popolazione (o semplicemente media). È la media aritmetica (quella più usata).

- **Media (aritmetica)** di una funzione di variabile casuale

Sia $a(x)$ una funzione della variabile casuale X distribuita secondo una p.d.f. $f(x)$ → il valore di aspettazione della variabile a è :

$$E[a] = \int_{-\infty}^{+\infty} a(x) f(x) dx$$

Valori di Aspettazione

- ❑ **Momento n-esimo** (momento di ordine n)

Il momento n-esimo di X è definito da: $E[x^n] = \int_{-\infty}^{+\infty} x^n f(x) dx = \mu'_n$
[$\mu'_1 = \mu$ (media aritmetica)]

- ❑ **Momento Centrale n-esimo** (o di ordine n)

È definito dalla quantità:

$$E[(x - E[x])^n] = \int_{-\infty}^{+\infty} (x - \mu)^n f(x) dx$$

- ❑ Il momento centrale secondo (o di ordine 2) è **la varianza** di x ed è indicato col simbolo σ^2 oppure con $V[x]$:

$$E[(x - E[x])^2] = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx = \sigma^2 = V[x]$$

- ❑ Si noti che $V[x] = E[x^2] - \mu^2$ Infatti:

Varianza e Deviazione Standard

- Per semplicità si abbiano N misure. Allora :

$$V[x] = \frac{1}{N} \sum_{i=1}^N (x_i)^2 - \frac{1}{N} \sum_{i=1}^N 2x_i\mu + \frac{1}{N} \sum_{i=1}^N \mu^2$$

$$V[x] = \frac{1}{N} \sum_{i=1}^N (x_i)^2 - 2\mu \frac{1}{N} \sum_{i=1}^N x_i + \mu^2 \frac{1}{N} \sum_{i=1}^N 1$$

$$V[x] = \bar{x}^2 - 2\mu^2 + \mu^2 = \bar{x}^2 - \mu^2$$

- La varianza è data dalla **media dei quadrati meno il quadrato della media**.
La varianza misura quanto è larga la distribuzione attorno al valore medio
- Deviazione Standard
La deviazione standard σ è definita come la radice quadrata della varianza

Calcolo della Deviazione Standard

- Facciamo un esempio di calcolo della deviazione standard. Particelle elementari siano distribuite in modo uniforme in un quadrato di lato a e con i lati paralleli agli assi. Si calcoli il valore medio lungo l'asse x e la deviazione standard σ_x

La probabilità lungo l'asse X è costante (k). k viene determinata dalla condizione di normalizzazione :

$$\int_0^a k dx = 1 \quad \text{da cui } k = 1/a$$

Allora si ottiene: $\bar{x} = \int_0^a \frac{1}{a} x dx = \frac{a}{2}$

$$\overline{x^2} = \int_0^a \frac{1}{a} x^2 dx = \frac{a^2}{3}$$

e di conseguenza :

$$\sigma_x^2 = \frac{a^2}{3} - \left(\frac{a}{2}\right)^2 = \frac{a^2}{12} \quad \text{e} \quad \sigma_x = a/\sqrt{12}$$

Matrice di Covarianza

- ❑ Consideriamo eventi che coinvolgono più variabili casuali. Iniziamo con due, X e Y
- ❑ È necessario tenere conto delle eventuali correlazioni che legano le due variabili. Questa correlazione è detta **covarianza** V_{xy} [**cov(x,y)**], definita da:

$$V_{xy} = E[(x - \mu_x)(y - \mu_y)] = E[xy] - \mu_x\mu_y$$

$$V_{xy} = \int_{-\infty}^{+\infty} xyf(x,y)dxdy - \mu_x\mu_y$$

con $\mu_x = E[x]$, $\mu_y = E[y]$

- ❑ La matrice di covarianza, detta anche **matrice degli errori**, è una matrice simmetrica ($V_{xy} = V_{yx}$)
- ❑ Gli elementi diagonali della matrice sono le varianze delle variabili ($V_{xx} = \sigma_x^2$, $V_{yy} = \sigma_y^2$)

Coefficiente di Correlazione

- ❑ Il **coefficiente di correlazione** tra le due variabili X e Y è definito da :

$$\rho_{xy} = \frac{V_{xy}}{\sigma_x \sigma_y}$$

È una quantità che assume valori compresi tra -1 e +1.

- ❑ È una grandezza **adimensionale** (ma gli elementi della matrice di covarianza hanno dimensione!!).
- ❑ Misura il grado di correlazione **lineare** tra le due variabili.
- ❑ Due variabili indipendenti hanno grado di correlazione lineare uguale zero.
- ❑ Ma due variabili che hanno grado di correlazione lineare uguale a zero non necessariamente sono indipendenti. Possono infatti avere correlazioni di ordine superiore (non lineari !).

Coefficiente di Correlazione

- ❑ Un correlazione positiva vuol dire che al crescere di una variabile cresce anche l'altra mentre la correlazione è negativa se al crescere di una variabile l'altra decresce.

- ❑ La varianza della somma di due variabili casuali X e Y è data da:

$$V_{[x+y]} = V_x + V_y + 2V_{xy}$$

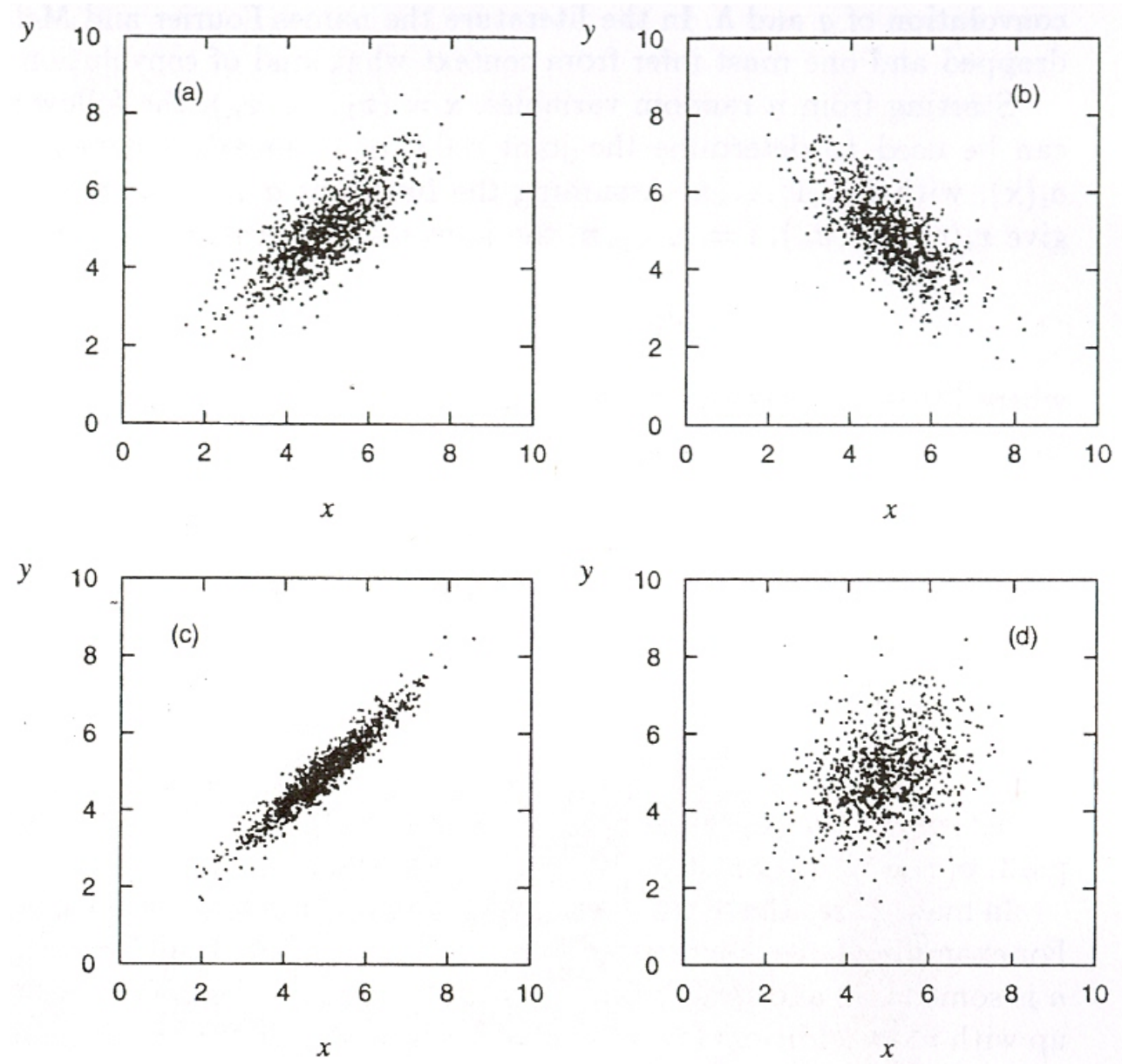
Se due variabili sono indipendenti ($V_{xy} = 0$) la varianza della somma è la somma delle varianze.

- ❑ Tutto ciò si generalizza al caso di eventi con n variabili casuali X_1, X_2, \dots, X_n
L'elemento V_{ij} della matrice degli errori tra x_i e x_j è dato da :

$$V_{ij} = E[(x_i - \mu_i)(x_j - \mu_j)]$$

- ❑ Il corrispondente coefficiente di correlazione lineare è $\rho_{ij} = \frac{V_{ij}}{\sigma_i \sigma_j}$

Coefficiente di Correlazione



Propagazione degli Errori

- ❑ Caso di funzione f lineare di variabile X : $f(x) = a x + b$ (a, b costanti).
Dalle misure calcolo la media x_0 e la varianza $V[x]$ della variabile X . Per la funzione f otteniamo: $V[f] = E[(f - E[f])^2] = a^2 V[x]$

e $\sigma_f = |a| \sigma_x$

- ❑ Caso di funzione f non lineare. Sviluppo di $f(x)$ in serie di Taylor attorno al valore medio x_0 :

$$f(x) = f(x_0) + (x - x_0) \left(\frac{df}{dx} \right) \Big|_{x=x_0} \quad \text{relazione valida per errori piccoli}$$

- ❑ Se la funzione dipende da due variabili X_1 e X_2 : $y = f(x_1, x_2)$ e abbiamo già calcolato valori medi e varianze delle due variabili, allora si ha:

Propagazione degli Errori

$$V[y] = \left(\frac{\partial y}{\partial x_1}\right)^2 V[x_1] + \left(\frac{\partial y}{\partial x_2}\right)^2 V[x_2] + 2 \left(\frac{\partial y}{\partial x_1}\right) \left(\frac{\partial y}{\partial x_2}\right) V_{x_1 x_2}$$
$$\sigma_y^2 = \left(\frac{\partial y}{\partial x_1}\right)^2 \sigma_{x_1}^2 + \left(\frac{\partial y}{\partial x_2}\right)^2 \sigma_{x_2}^2 + 2 \left(\frac{\partial y}{\partial x_1}\right) \left(\frac{\partial y}{\partial x_2}\right) \rho \sigma_{x_1} \sigma_{x_2}$$

dove i differenziali sono calcolati nei valori medi delle variabili.

□ Nel caso di funzione di n variabili, si ha:
$$\sigma_y^2 = \sum_{i,j=1}^n \left[\frac{\partial y}{\partial x_i} \frac{\partial y}{\partial x_j} \right] V_{x_i x_j}$$

□ Se si hanno m variabili y_1, y_2, \dots, y_m funzioni delle n variabili x_1, \dots, x_n la matrice degli errori diventa:

$$U_{kl} = V_{y_k y_l} = \sum_{i,j=1}^n \left[\frac{\partial y_k}{\partial x_i} \frac{\partial y_l}{\partial x_j} \right] V_{x_i x_j}$$

□ In notazione matriciale: $\mathbf{U} = \mathbf{A} \mathbf{V} \mathbf{A}^T$. La matrice delle derivate è definita da

$$A_{ij} = \left[\frac{\partial y_i}{\partial x_j} \right] \quad \mathbf{A}^T \text{ è la matrice trasposta di } \mathbf{A}$$

Propagazione degli Errori

❑ Le formule viste permettono di propagare gli errori dalle variabili x_1, \dots, x_n alle variabili y_1, \dots, y_m . Nel caso di variabili non correlate i termini covarianti sono nulli.

❑ Abbiamo già notato che se una funzione $y = x_1 + x_2$ e x_1, x_2 sono tra di loro scorrelate, si ha che $\sigma_y^2 = \sigma_{x_1}^2 + \sigma_{x_2}^2$ Gli errori si sommano in quadratura

❑ Se $y = x_1 x_2$ in generale si ha: $\frac{\sigma_y^2}{y^2} = \frac{\sigma_{x_1}^2}{x_1^2} + \frac{\sigma_{x_2}^2}{x_2^2} + 2 \frac{V_{x_1 x_2}}{x_1 x_2}$ Il rapporto tra la σ e il valore della variabile è detto **errore relativo**

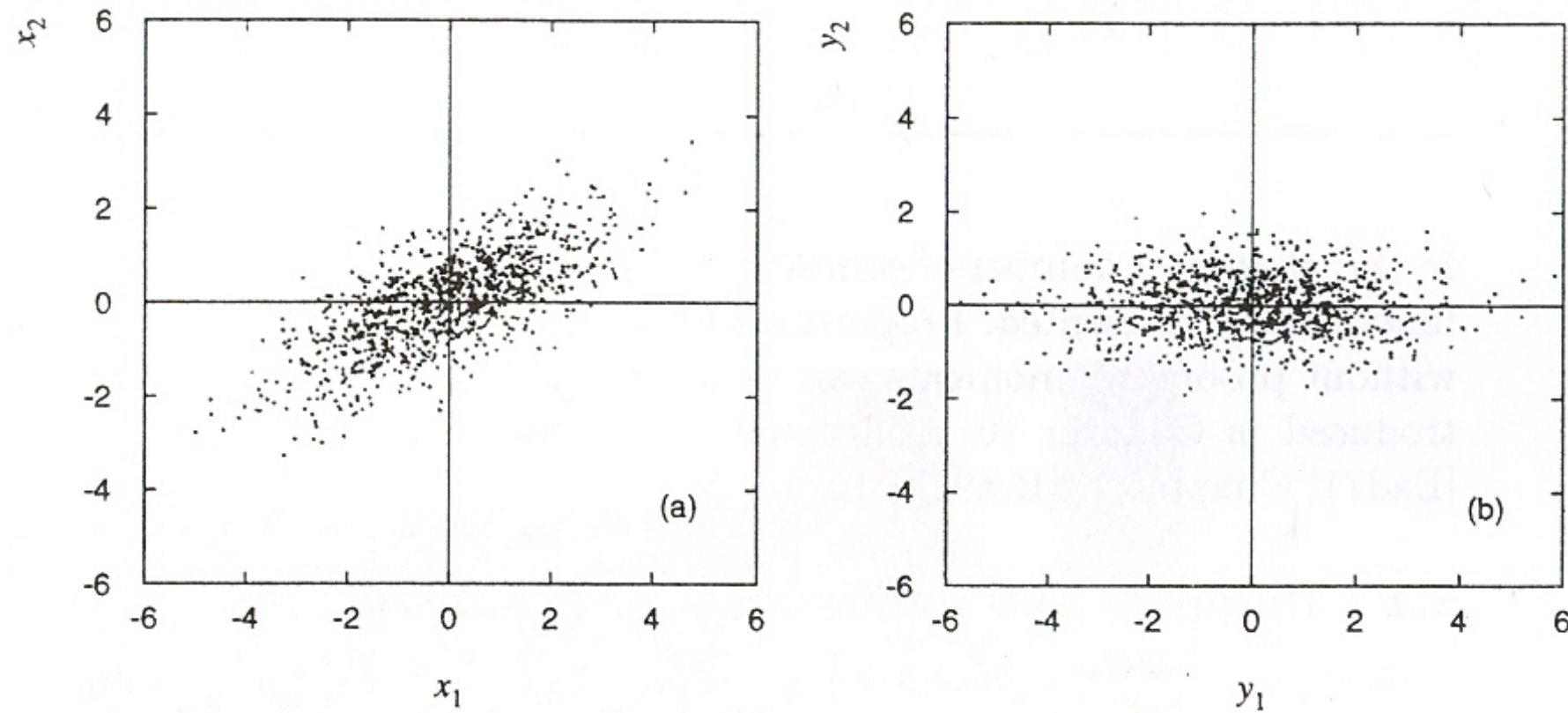
❑ Se le due variabili non sono correlate, allora si ha: $\frac{\sigma_y^2}{y^2} = \frac{\sigma_{x_1}^2}{x_1^2} + \frac{\sigma_{x_2}^2}{x_2^2}$
In questo caso gli errori relativi si sommano in quadratura

❑ Questo risultato vale anche per il rapporto $y = x_1/x_2$

Diagonalizzazione della Matrice di Covarianza

- ❑ Generalmente la matrice di covarianza non è diagonale a cause della presenza di correlazione tra le n variabili x_1, x_2, \dots, x_n .
- ❑ In molte circostanze è utile che le variabili non siano correlate. Per esempio se si vuole variare una variabile senza variare le altre (processi di minimizzazione, ecc)
- ❑ È possibile riscrivere la matrice di covarianza in termini di nuove variabili y_1, y_2, \dots, y_n che non siano tra di loro correlate e quindi con matrice di covarianza diagonale.
- ❑ Si può dimostrare che tramite una trasformazione lineare questo è sempre possibile:
$$y_i = \sum_{j=1}^n A_{ij} x_j$$
- ❑ La nuova matrice di covarianza U è
$$U_{ij} = \sum_{k,l=1}^n A_{ik} A_{jl} V_{x_k, x_l} = \sum_{k,l=1}^n A_{ik} V_{kl} A_{lj}^T$$
- ❑ In pratica si trova una matrice A tale che $U = AVA^T$ (diagonalizzazione di una matrice simmetrica e reale)

Diagonalizzazione della Matrice di Covarianza



Scatter plot di due variabili correlate (x_1, x_2) e delle trasformate (y_1, y_2) con matrice di covarianza diagonale