LEZIONE (8) 11/05/09

CONVERGENZA DI SUCCESSIONI DI V.C.

__Definizione__: una __successione di v.c.__ è un esperimento casuale il cui esito è una successione di numeri reali (impropriamente considerato un vettore di v.c. di dimensione infinita).

I numeri saranno indicati con $X_1, X_2, \dots, X_n, \dots$ e la successione con

$$\{X_n\}$$

Esempio: lanciando n volte un dado si ottengono i seguenti risultati:

X_1	X_2	X_3	...	X_n
1	3	6		5

Si definisca:

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Si ottiene una successione di v.c. $\bar{X}_1, \bar{X}_2, \bar{X}_3, \dots, \bar{X}_n$, con i seguenti risultati:

$n = 1$ $\quad 1$

$n = 2$ $\quad (3+1)/2 = 2$

$n = 3$ $\quad (1+3+6)/3 = 3{,}33$

In pratica si hanno due successioni: $\{X_n\}$ e $\{\bar{X}_n\}$.
La prima di esse non converge ad alcun risultato, mentre ci si può ragionevolmente aspettare che la seconda converge.

Per le v.c. esistono diverse nozioni di convergenza

1) __Convergenza in distribuzione__

Si dice che una successione $\{X_n\}$ converge in distribuzione ad una variabile X se

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

$F_{X_n}(x)$ è la f.d.d. dell'ennesimo termine della successione.

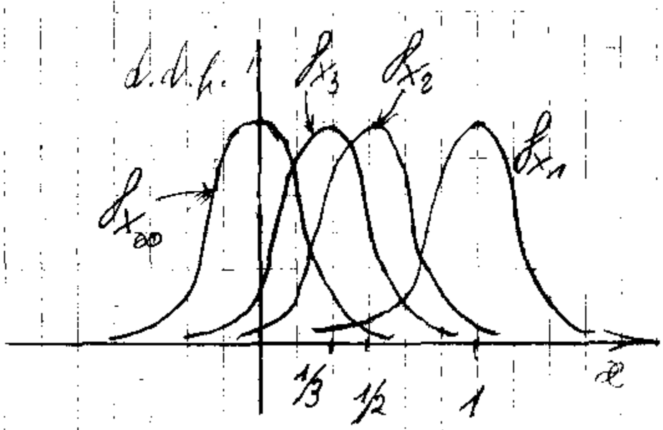
Ogni singolo elemento di una successione di v.c. è a sua volta una v.c. e possiede una f.d.d.

Esempio

Immaginiamo che le v.c. X_n siano delle Gaussiane con

$$E[X_n] = \frac{1}{n}, \quad \sigma^2_{X_n} = 1$$

$$X_n \sim N\left(\frac{1}{n}, 1\right)$$



Otteniamo:

$$X_1 \sim N(1, 1)$$
$$X_2 \sim N(1/2, 1)$$
$$X_3 \sim N(1/3, 1)$$

Le v.c. X_n hanno d.d.f. che al crescere di n, hanno la stessa varianza, ma una media che decresce con legge $1/n$.

Quando $n \to \infty$, si ha $X_\infty \sim N(0, 1)$

Al procedere della successione la d.d.f. si modifica e tende ad f_{X_∞}.

2) Convergenza in probabilità

$\{X_n\}$ converge ad X in probabilità se, $\forall \varepsilon > 0$,

$$\lim_{n \to \infty} P(|X_n - X| \le \varepsilon) = 1$$

Proprietà: la convergenza in probabilità implica la convergenza in distribuzione, ma non vale il viceversa.

La convergenza in probabilità è più forte di quella in distribuzione

3) Convergenza in media quadratica

$\{X_n\}$ converge in media quadratica se

$$\lim_{n \to \infty} E[(X_n - X)^2] = 0$$

con $(X_n - X)^2$ = errore quadratico medio

La proprietà si indica anche con

$$\underset{n \to \infty}{l.i.m.}\ X_n = X$$

con

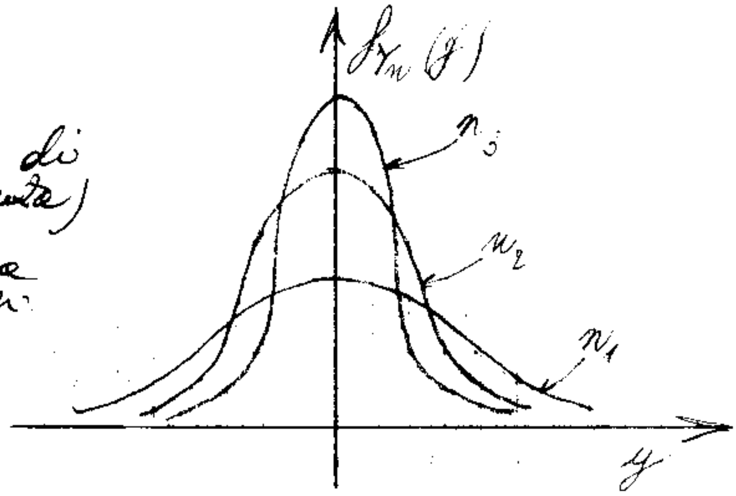
$$l.i.m. = \text{limite in media}$$

<u>Interpretazione grafica</u>

Definiamo

$Y_n := X_n - X$ (errore di convergenza)

Attendersi che X_n converga ad X significa attendersi che Y_n converga a zero.



L'errore è piccolo se

$$E\left[Y_n^2\right]$$

è piccolo. $E[Y_n^2]$ misura la dispersione attorno all'origine. Chiedere che $E[Y_n^2]$ tenda a zero equivale a chiedere che la d.d.p. di Y_n tenda ad una delta di Dirac centrata nell'origine.

<u>Proprietà</u>

La convergenza in media quadratica implica la convergenza in probabilità. Non vale il viceversa.

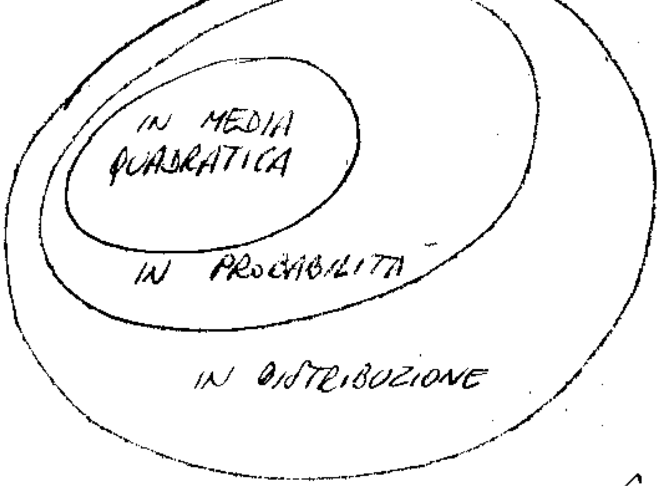
Questa proprietà è più forte delle due precedenti.

<u>Proprietà</u>: Se $X = a$, ~~dice che~~ con a = costante deterministica,

$$\left.\begin{aligned} &\lim_{n\to\infty} E[X_n] = a \\ &\lim_{n\to\infty} Var[X_n] = 0 \end{aligned}\right\} \Longleftrightarrow \underset{n \to \infty}{l.i.m.}\ X_n = a$$

Il simbolo $\Longleftrightarrow$ indica equivalenza nel senso di "se e solo se". Le due condizioni a sinistra devono valere entrambe.

Le relazioni tra le nozioni di convergenza si possono riassumere nel seguente grafico



Teorema: Legge dei Grandi Numeri

Data una successione $\{X_n\}$ di v.c. indipendenti e identicamente distribuite (d'ora in poi i.i.d.), si supponga che

$$Var[X_n] < \infty$$

e si definisca $m := E[X_n]$;

allora

$$\text{l.i.m.}\ \overline{X}_n = m$$

con $\overline{X}_n$, media campionaria dei primi n elementi

$$\overline{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i$$

Il teorema afferma che, al crescere di n le medie campionarie convergono ad m, valor atteso.

Nel caso del dado (pag. 64), al crescere del numero dei lanci, le medie campionarie convergono a 3,5.

Dimostrazione. Si sfrutta la 2ª proprietà enunciata per la convergenza in media quadratica.

Occorre mostrare che

1) $\lim_{n \to \infty} E[\overline{X}_n] = m$

2) $\lim_{n \to \infty} Var[\overline{X}_n] = 0$

Ciò automaticamente significherebbe che $\text{l.i.m.}\ \overline{X}_n = m$

1) Calcoliamo $E[\overline{X}_n]$

$$E[\overline{X}_n] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i]$$ per la linearità di $E[\cdot]$

Per ipotesi le X_i sono tutte identicamente distribuite. Di conseguenza i loro valori attesi sono tutti uguali ad m. Perciò:

$$E[\overline{X}_n] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{\cancel{n}}\cancel{n}\,m = m$$

Affermare che le X_i sono identicamente distribuite, significa affermare che

$$E[X_i] = m \qquad \forall i$$

non avendo truccato dado, qualunque sia il dado.

2) Calcolo di $Var[\overline{X}_n]$

$$Var[\overline{X}_n] = Var\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2} Var\left[\sum_{i=1}^{n} X_i\right]$$

In generale la ~~somma~~ varianza della somma di v.c. non è la somma delle varianze, in quanto occorre considerare tutti i doppi prodotti (covarianze).

Poiché le v.c sono anche indipendenti (i.i.d.) le X_i hanno tutte covarianze nulle. Perciò:

$$X_i \text{ indipendenti} \Rightarrow X_i \text{ incorrelate} \Rightarrow$$
$$\Rightarrow \sigma_{X_i X_j} = 0 \quad \forall i \neq j$$

$$\frac{1}{n^2} Var\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} Var[X_i]$$

Essendo le X_i identicamente distribuite, ne consegue

$$Var[X_i] = \sigma^2, \quad \forall i$$

Perciò:
$$\frac{1}{n^2}\sum_{i=1}^{n} Var[X_i] = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n} = Var[\overline{X}_n]$$

Ne consegue che: $\lim_{n\to\infty} Var[\overline{X}_n] = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0.$

TEOREMA CENTRALE DEL LIMITE e VARIABILI CASUALI CONGIUNTAMENTE GAUSSIANE

- Teorema centrale del limite

Data una successione $\{X_n\}$ di v.c. indipendenti e identicamente distribuite, con

$$m = E[X_n] \quad e \quad Var[X_n] < \infty \quad e$$
$$\sigma^2 := Var[X_n],$$

si definiscono:

$$\tilde{X}_n := \sum_{i=1}^{n} X_i \qquad S_n := \frac{\tilde{X}_n - n \cdot m}{\sqrt{n\sigma^2}} = \frac{\tilde{X}_n - E[\tilde{X}_n]}{\sqrt{Var[\tilde{X}_n]}}$$

La S_n rappresenta la $\tilde{X}_n$ standardizzata.
Il teorema afferma che in queste condizioni S_n converge in distribuzione ad una v.c. gaussiana standard, con

$$m = 0 \;, \; Var = 1$$

Osservazione: specificare che la varianza è $< \infty$, significa affermare che essa esiste.

Per calcolare media e varianza di una v.c. occorre calcolare un integrale esteso da $-\infty$ a $+\infty$, ma l'integrale potrebbe non convergere, come nel caso della variabile di Cauchy.

$\tilde{X}_n$ è la somma delle prime n X_i, mentre S_n standardizza la $\tilde{X}_n$.

In pratica ci stiamo chiedendo quale sia la d.d.p. della somma di infinite v.c. i.i.d.

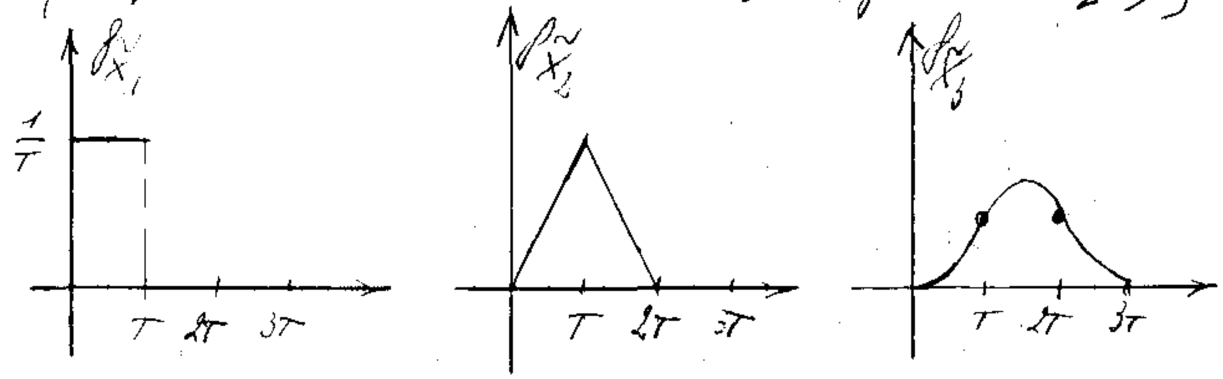
Apparentemente sembrerebbe un problema difficilmente risolvibile, dovendo ricorrere ad infinite convoluzioni (pag. 98)

Il teorema, invece, afferma che, per n sufficientemente grande, il risultato è una densità di tipo gaussiano.

Si osservi che non è stata fatta alcuna ipotesi sulla d.d.p. delle X_i, che potrebbe essere qualsiasi.

•) Osservazione: spesso l'approssimazione è già buona per piccoli valori di n

Esempio: si considerino v.c. X_i uniformi in $[0, T]$



La $f_{\tilde{X}_1} = f_{X_1}$

La $f_{\tilde{X}_2}$ è la d.d.f. della somma di X_1 e X_2, entrambe uniformi e fornisce una densità a triangolo (pag. 98 e segg.).

La $f_{\tilde{X}_3}$ risulta dalla convoluzione delle prime due v.c.

Il risultato fornisce tre archi di parabola tra loro raccordati e permette di intravedere un grafico che ricorda quello di una Gaussiana, che è riprodotta bene, già a partire da n = 5 o 6.

La convergenza enunciata dal teorema è abbastanza rapida.

•) Osservazione su un caso particolare (media campionaria di v.c. i.i.d.)

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{\tilde{X}_n}{n} \qquad \text{(pag. 137)}$$

Il teorema centrale del limite afferma che S_n diventa gaussiana. Tra S_n e $\tilde{X}_n$ c'è solo una trasformazione di standardizzazione, cioè una trasformazione lineare che non cambia la forma della d.d.f. che resta gaussiana.

Se $\tilde{X}_n$ diventa gaussiana, anche $\overline{X}_n$ lo diventa.

Nota: si sa che

$$E[\overline{X}_n] = \mu \quad e \quad Var[\overline{X}_n] = \frac{\sigma^2}{n} \qquad \text{(pag. 137)}$$

Di conseguenza si conosce la d.d.f. di $\overline{X}_n$ che è una

Mo distribuzione Normale : $N\left(m, \frac{\sigma^2}{n}\right)$

Si sa che una Gaussiana è nota se si conosce la media ed anche la varianza. Inoltre sappiamo che la media campionaria è pressoché gaussiana e la varianza è stata precedentemente calcolata.
Però per n grande la d.d.p. della media campionaria è completamente nota anche se non si conoscono le d.d.p. delle X_i originarie.

•) Altro caso particolare : <u>v.c. Binomiale</u>

Sia Y_n Binomiale di ordine n, che rappresenta il numero di successi in n prove di Bernoulli.

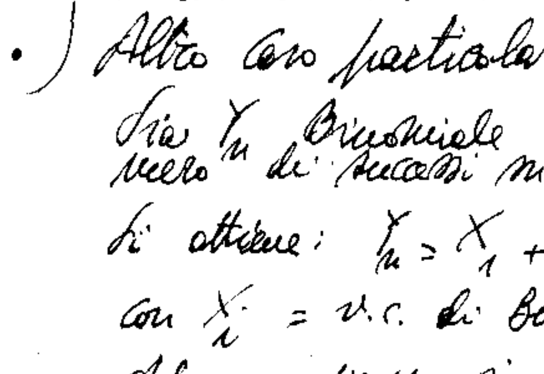
Si ottiene: $Y_n = X_1 + X_2 + \dots + X_n$

con X_i = v.c. di Bernoulli i.i.d.

Ad ogni prova si può associare una v.c. di Bernoulli che vale 0 oppure 1.

La Binomiale, che rappresenta il numero di successi, si ottiene sommando tutte le variabili binarie.

Usando il teorema centrale del limite, ne consegue che Y_n tende a diventare gaussiana, per n grande.

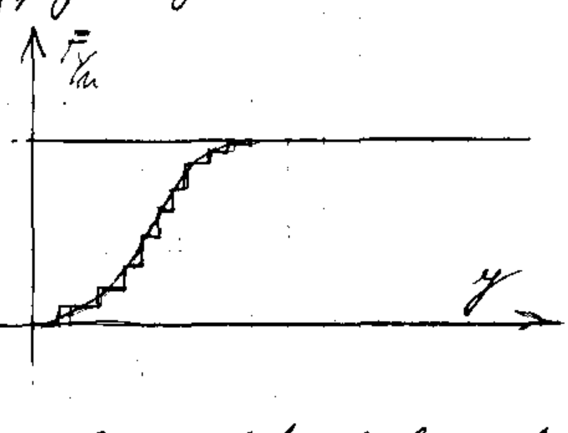


La v.c. Binomiale ha una d.d.p. a forma di pettine, essendo una v.c. discreta (il numero di successi è necessariamente un numero intero).

Questa funzione non può convergere ad una Gaussiana, essendo discreta (pag. 107).

La convergenza in distribuzione non è formulata sulla d.d.p., bensì sulle f.d.d. (pag. 133).

La f.d.d. è composta da scalini, che se diventano piccoli, possono convergere alla curva che rappresenta la f.d.d. di una Gaussiana.



<u>Considerazioni finali sul teorema centrale del limite.</u>

La somma di tante v.c. i.i.d., indipendentemente della loro distribuzione, ha una f.d.d. gaussiana.

Teoremi simili a questo valgono anche se le v.c. non sono i.i.d.

In natura molti fenomeni naturali sono gaussiani.

— Variabili casuali congiuntamente gaussiane

Consideriamo la d.d.f. di due v.c. X_1 e X_2 indipendenti e le cui marginali f_{X_1} e f_{X_2} sono gaussiane.

Definiamo:

$$\mu := \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} E[X_1] \\ E[X_2] \end{bmatrix} \quad \text{vettore dei valori attesi}$$

$$\sigma_i^2 := Var[X_i] \qquad i = 1,2$$

nell'ipotesi che $\sigma_i^2 > 0 \qquad i = 1,2$

Sfruttando l'indipendenza la d.d.f. congiunta del vettore

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

e'

$$f_X(x) = f_{X_1}(x_1)\, f_{X_2}(x_2) =$$

$$= \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}} \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}} =$$

$$= \frac{1}{\sqrt{(2\pi)^2\sigma_1^2\sigma_2^2}} e^{-\frac{1}{2}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]}$$

Sappiamo che

$$E[X] = \begin{bmatrix} E[X_1] \\ E[X_2] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \qquad \text{(pag. 112)}$$

$$\hat{Var}[X] = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \qquad \text{(pag. 113)}$$

142 Essendo indipendenti, X_1 e X_2 sono anche incorrelate e le covarianze sono nulle e la matrice varianza è diagonale.

Definiamo la matrice V

$$V = Var[X]$$

Scriviamo la $f_X(x)$ in modo compatto

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^2 \det V}} \, e^{-\frac{1}{2}(x-\mu)^T V^{-1} (x-\mu)}$$

Infatti:

$$\det V = \det \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} = \sigma_1^2 \sigma_2^2$$

Inoltre:

$$(x-\mu)^T V^{-1} (x-\mu) =$$

$$= \begin{bmatrix} (x_1-\mu_1) & (x_2-\mu_2) \end{bmatrix} \frac{1}{\sigma_1^2 \sigma_2^2} \begin{bmatrix} \sigma_2^2 & 0 \\ 0 & \sigma_1^2 \end{bmatrix} \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix} =$$

$$= \begin{bmatrix} (x_1-\mu_1) & (x_2-\mu_2) \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} \end{bmatrix} \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix} =$$

$$= \begin{bmatrix} \frac{x_1-\mu_1}{\sigma_1^2} & \frac{x_2-\mu_2}{\sigma_2^2} \end{bmatrix} \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix} = \frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}$$

L'espressione della $f_X(x)$ non cambia anche se n ha un valore qualsiasi, con

$$X_1, X_2, \dots, X_n$$

indipendenti tra loro e gaussiane, con

$$\mu = E[X] = \begin{bmatrix} E[X_1] \\ E[X_2] \\ \dots \\ E[X_n] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \dots \\ \mu_n \end{bmatrix}$$

$$V = Var[X] = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_n^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & & O \\ & \sigma_2^2 & \\ O & & \ddots \ \sigma_n^2 \end{bmatrix}$$

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Se V non è una matrice diagonale l'espressione di $f_X(x)$ è ancora una d.d.p., purché V sia simmetrica e definita positiva.

<u>Definizione</u> Le v.c. $X_1, X_2, \dots, X_n$ si dicono congiuntamente gaussiane se, essendo

$$X = [X_1 \; X_2 \; \dots \; X_n]^T$$

la densità congiunta del vettore X si può scrivere come indicato a pag. 162, con

$$m \in \mathbb{R}^{n \times 1} \quad , \quad V = V^T > 0 \quad , \quad V \in \mathbb{R}^{n \times n}$$

Nel caso particolare che V sia anche diagonale, le v.c. X_i sono delle Gaussiane indipendenti tra loro.

Il vettore X di v.c. congiuntamente gaussiane gode di alcune <u>proprietà</u>, qui elencate

•) Si verifica che

$$E[X] = m \quad , \quad Var[X] = V$$

•) la d.d.p. è completamente nota, se sono note m e V

<u>Osservazione</u>: quanti parametri servono per descrivere n variabili congiuntamente gaussiane?

Se la variabile è una, bastano due numeri: media e varianza.

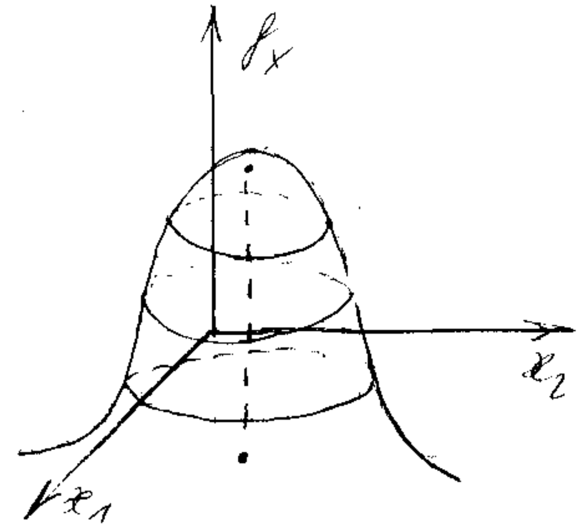
Se le variabili congiuntamente gaussiane sono due, servono i due valori medi ed una matrice 2×2. Essendo la matrice simmetrica, i parametri necessari sono in tutto 5.

•) Se X è un vettore di v.c. congiuntamente gaussiane, le d.d.p. marginali

$$f_{X_i}(x_i)$$

sono gaussiane con $E[X_i] = m_i$ e $Var[X_i] = V_{ii}$

 essendo V_{ii} l'elemento iesimo sulla diagonale principale



- per $n = 2$ la f_x ha la forma di una campana
- se X è un vettore di v.c. congiuntamente gaussiane
$$Y = A X$$
è ancora un vettore di v.c. congiuntamente gaussiane, con
$$E[Y] = A\, E[X]$$
$$Var[Y] = A\, Var[X]\, A^T$$
Ciò significa che la combinazione lineare di v.c. gaussiane è ancora una v.c. gaussiana.
X e Y potrebbero avere dimensioni diverse.

Corollario: la combinazione lineare
$$a_1 X_1 + a_2 X_2 + \dots + a_n X_n$$
di v.c. X_i congiuntamente gaussiane è gaussiana.

- Teorema: se $X_1, X_2, \dots, X_n$ congiuntamente gaussiane sono incorrelate, allora sono anche indipendenti.

Due variabili X_i, X_j sono incorrelate se e solo se
$$Cov[X_i, X_j] = 0 \qquad i \neq j \quad (\text{pag. } 108)$$
e se la matrice $Var[X]$ è diagonale.
Pertanto, per v.c. congiuntamente gaussiane, incorrelazione è equivalente ad indipendenza
$$\text{incorrelazione} \iff \text{indipendenza}$$
Ciò non vale in generale (pag. 108), cioè per v.c. congiunte di tipo qualsiasi.
Per v.c. gaussiane, verificare se c'è o meno indipendenza si riduce a verificare l'incorrelazione, cioè

a verificare che il coefficiente di correlazione (vedi pag 145) è uguale a zero.

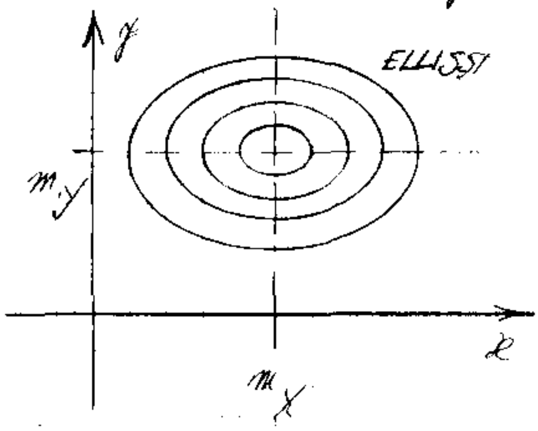
Interpretazione grafica ($n = 2$)

•) caso $r_{XY} = 0$

Ciò significa che tra le v.c. c'è un rapporto di incorrelazione e indipendenza.

Ci interessa studiare come è fatta, dal punto di vista grafico, la d.d.p. congiunta.

Essa rappresenta una superficie nello spazio, essendo una funzione di due variabili. Occorre studiare le forme delle linee di livello, che hanno equazione



$$f_{XY}(x,y) = \text{costante}$$

cui consegue

$$\frac{(x-m_X)^2}{\sigma_X^2} + \frac{(y-m_Y)^2}{\sigma_Y^2} = k$$

(vedi pag. 141)

Si osservi che nella formula di pag. 141 x e y ~~compaiono~~ appaiono solo nell'esponente.

L'equazione appena scritta rappresenta una famiglia di ellissi con centro in (m_X, m_Y).

Nel caso di v.c. standardizzate si ha:

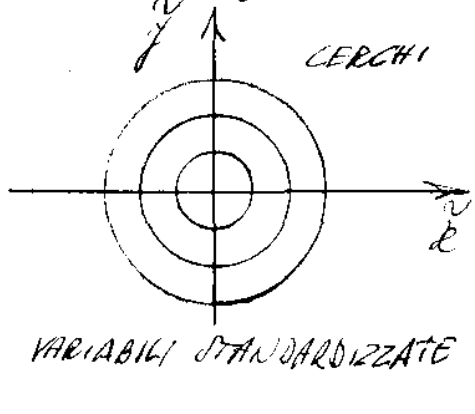
$$\tilde{X} := \frac{X - m_X}{\sigma_X}$$

$$\tilde{Y} := \frac{Y - m_Y}{\sigma_Y}$$

ottenendo

$$\tilde{X}^2 + \tilde{Y}^2 = k$$

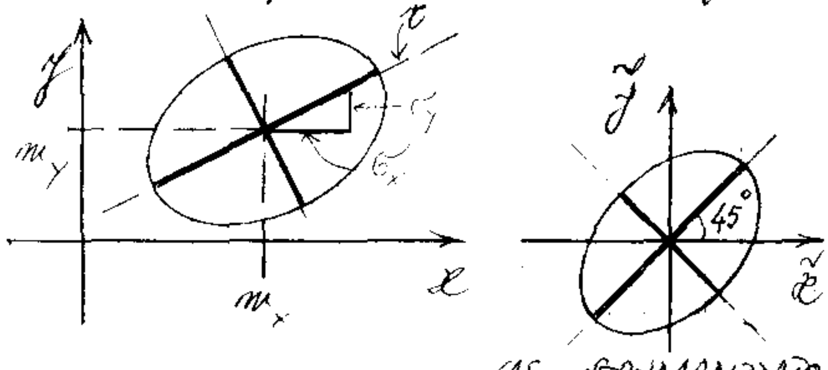
che rappresenta l'equazione di un cerchio.



VARIABILI STANDARDIZZATE

•) caso $r_{XY} \neq 0$

$f_{XY}(x,y) = \text{costante}$ rappresenta ancora degli ellissi, poco inclinati. La retta r è detta retta delle deviazioni standard.



CASO STANDARDIZZATO

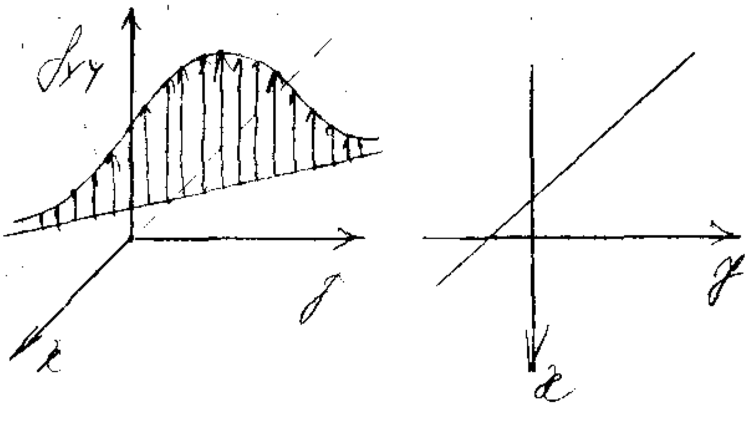
La sua inclinazione è definita da un triangolo che ha come cateti σ_X e σ_Y.

Nel caso standardizzato la retta è inclinata di 45°.

Avendo standardizzato le due deviazioni standard sono entrambe uguali ad 1.

•) Caso $r_{XY} = 1$

La d.d. f è degenere e diventa una muraglia di impulsi, detta di Dirac teoricamente alta infinito.



La base sta diventando sempre più stretta. Poiché il volume deve essere unitario, l'altezza ~~deve~~ diventa sempre più grande.

-) Densità e momenti condizionati

•) Teorema: se X e Y sono vettori di v.c. congiuntamente gaussiane, nel senso che

$$\begin{bmatrix} X \\ Y \end{bmatrix} \text{ è gaussiano,}$$

allora anche

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

è gaussiana.

Se X e Y fossero due v.c. gaussiane (ad es. peso ed altezza), uno desidererebbe che la d.d. delle altezze, dato il peso, è ancora gaussiana.

Inoltre, essendo

$$Var\left[\begin{bmatrix} X \\ Y \end{bmatrix}\right] = \begin{bmatrix} Var[X] & Cov[X,Y] \\ Cov[Y,X] & Var[Y] \end{bmatrix} =$$

$$= \begin{bmatrix} V_{XX} & V_{XY} \\ V_{XY}^T & V_{YY} \end{bmatrix} \quad \text{con } V_{XY}^T = V_{YX}$$

si ottiene:

$$E[Y/X=x] = m_Y + V_{YX} V_{XX}^{-1} (x - m_X)$$

$$Var[Y/X] = V_{YY} - V_{YX} V_{XX}^{-1} V_{XY}$$

Per comprendere il significato del teorema si immagini di avere due V.C. (peso e altezza) congiuntamente gaussiane.

Di una persona si conosca il peso e si voglia determinare l'altezza.

L'informazione è ricavabile dalla funzione d.d.p. che è ancora gaussiana. Devo conoscere media e varianza, calcolandole come indica il teorema.

Tutto può essere calcolato, conoscendo il vettore delle medie e la matrice varianza, cioè varianze e covarianze.

•) caso $n=2$ con X e Y scalari

$$E[Y/X=x] = m_Y + \varrho_{XY} \frac{\sigma_Y}{\sigma_X} (x - m_X)$$

In questo caso non servono le matrici inverse.

Inoltre

$$Var[Y/X=x] = \sigma_Y^2 (1 - \varrho_{XY}^2)$$

Da ricordare anche che

$$\frac{E[Y/X=x] - m_Y}{\sigma_Y} = \varrho_{XY} \left(\frac{x - m_X}{\sigma_X}\right)$$

L'espressione $\frac{x - m_X}{\sigma_X}$ rappresenta la X standardizzata

zata.

$E[Y|X=x]$ è la previsione della Y.

Spesso per fare previsioni si usa il valor atteso.

Il primo membro rappresenta la previsione standardizzata, che risulta uguale al prodotto del coefficiente di correlazione per la misura standardizzata.

Se $r_{XY}=0$, la previsione standardizzata è nulla.

In questo caso conoscere la X non serve a nulla e il valor atteso della Y è quello di partenza, m_Y.

Se $|r_{XY}|$ è quasi uno (correlazione forte) la previsione standardizzata dipende fortemente dalla misura.

Inoltre, se $|r_{XY}|=1$, $r_{XY}^2=1$ e

$$Var[Y|X=x]=0$$

Se $|r_{XY}|=1$ tra X e Y c'è una relazione deterministica e non rimane più alcuna incertezza sulla Y.

Se invece $r_{XY}=0$ permane l'incertezza di partenza