By Davide Micheli

Definizione assiomatica di probabilità

- La *probabilità* P(A) di un evento A è un numero non negativo assegnato a questo evento

$$P(A) \geq 0$$ **Equazione 1**

- La *probabilità* dell'evento certo è uguale a 1
- Se gli eventi A e B sono mutuamente esclusivi allora

$$P(A \cup B) = P(A) + P(B)$$ **Equazione 2**

Definizione frequenza della probabilità

La *probabilità* P(A) di un evento A è il limite

$$\boxed{P(A) = \lim_{n \to \infty} \frac{n_A}{n}}$$ **Equazione 3**

dove n_A è il numero delle occorrenze di A ed n è il numero delle prove

Definizione classica della probabilità

La *probabilità* P(A) di un evento A è determinata a priori dal rapporto

$$\boxed{P(A) = \frac{N_A}{N}}$$ **Equazione 4**

dove N è il numero delle possibili uscite (risultati) ed N_A è il numero delle uscite (risultati) favorevoli all'evento A.
per esempio nell'esperimento dei dadi le possibili uscite sono sei e le uscite favorevoli all'evento pari sono tre dunque la probabilità dell'evento pari è P(pari)=3/6=0.5

In questa definizione in ogni caso è importante notare che il significato dei numeri N e N_A non è sempre chiaro.

Una versione migliorata di tale definizione è la seguente:

la probabilità di un evento è uguale al rapporto tra le uscite favorevoli e il numero totale delle uscite, ammesso che tutte le uscite siano <u>egualmente probabili</u>.

Teoria del set

Un *set* A è una collezione di oggetti chiamati elementi, questi possono essere indicati da lettere greche:

$$A = \{\xi_1, \ldots\ldots\ldots, \xi_n\}$$

Un *subset* B di A è un altro *set* i cui elementi sono anche elementi di A.
Tutti i *set* considerati saranno un *subset* del *set* S che è chiamato *spazio*

Spazio di probabilità

Nella teoria della probabilità è usata la seguente terminologia:

- Lo *spazio* S o Ω è chiamato *evento certo,* nell'esperimento con il dado, l'evento certo è lo spazio è costituito dalle sei facce $f_1, f_2, f_3, f_4, f_5, f_6$
- Gli *elementi* ξ_i dello spazio S o Ω sono le uscite (risultati) dell'esperimento considerato
- I *subset* di tali *elementi* (risultati degli esperimenti) sono gli *eventi*
  Il set vuoto $set\{\ \}$ è l'evento impossibile,
  l'evento $event\{\xi_i\}$ di un solo *elemento* è un *evento elementare*
- Una singola performance di un esperimento è chiamata *Trial(prova)*
- Per ogni *trial* si osserva una singola uscita (risultato) ξ_i
- L'evento A avviene durante questo *trial* se esso contiene l'elemento ξ_i

Probabilità condizionata

La *probabilità condizionata* di un evento A assumendo un altro evento M, denotata con $P(A/M)$, è per definizione il rapporto:

$$P(A/M) = \frac{P(AM)}{P(M)}$$

dove *P(M)* deve essere diverso da zero.

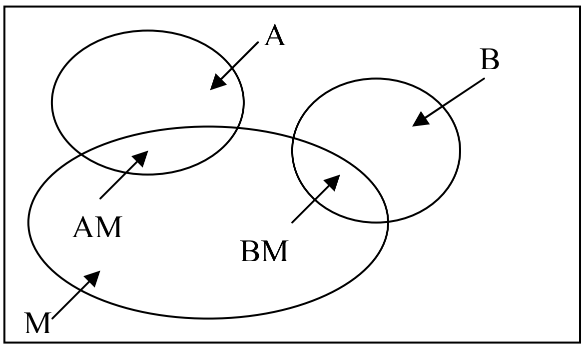


Figura 1

Se:

$$M \subset A \quad \Rightarrow \quad P(A/M) = 1 \quad perchè \quad AM = M$$

$$A \subset M \quad \Rightarrow \quad P(A/M) = \frac{P(A)}{P(M)} \geq P(A)$$

Teorema di Bayes

Se U=[A_1……A_n] è una partizione di S e B è un evento allora:

$$\boxed{\begin{aligned} P(B) &= P(B \mid A_1)P(A_1) + \ldots\ldots\ldots\ldots + P(B \mid A_n)P(A_n) \\ P(A_i \mid B) &= \frac{P(B \mid A_i)P(A_i)}{P(B \mid A_1)P(A_1) + \ldots\ldots\ldots\ldots + P(B \mid A_n)P(A_n)} \end{aligned}}$$

Equazione 5

dove la probabilità **a priori** è usata per **P(A $_i$)** e quella a **posteriori** è usata per **P(A $_i$ | B)**

Indipendenza

Due eventi A e B sono *statisticamente indipendenti* se

$$\boxed{P(AB) = P(A)P(B)}$$

Equazione 6

il concetto di indipendenza è fondamentale. Infatti è questo concetto che giustifica lo sviluppo matematico della probabilità, il significato della indipendenza può in particolare essere apprezzato nel contesto di *trial ripetuti*

Variabili casuali (aleatorie o random)

Una variabile casuale è un numero $X(\xi)$ assegnato ad ogni uscita ξ di un esperimento; per esempio in un esperimento si assegna il numero 1 ad ogni uscita pari ed il numero 0 ad ogni uscita dispari.

Significato di funzione

Una variabile casuale è una *funzione* il cui dominio è il *set S* di tutte le uscite degli esperimento.
La definizione non ambigua di funzione è la seguente:
Dati due set di oggetti S_α e S_β consistenti degli elementi α e β rispettivamente. Diciamo che β è una funzione di α se a ogni elemento del set S_α facciamo corrispondere un elemento β del set S_β. *Il set* S_α è il dominio della funzione e il set S_β *è il range (*codominio*).*

Variabile casuale

Dato un esperimento specificato per mezzo dello *spazio S,* i subset degli elementi ξ di S sono gli *eventi* di S, una data probabilità sia assegnata a questi eventi.

Ad ogni uscita (risultato) ξ dell'esperimento sia assegnato un numero $X(\xi)$, in tal modo è stata creata la funzione X con dominio il set S di elementi ξ e range un set di numeri $X(\xi)$. Questa funzione è chiamata variabile casuale.

Eventi generati da variabili casuali

Nello studio delle variabili casuali, sorge la seguente domanda:

- qual'è la probabilità che una variabile casuale X sia minore di un dato numero x ?
- qual'è la probabilità che una variabile casuale X sia compresa tra i numeri x_1 e x_2 ?

Dal momento che le *probabilità sono assegnate solo agli eventi* occorre esprimere le varie condizioni imposte su X come eventi.

Si può iniziare con la notazione:

$$\{X \le x\}$$

Questa notazione rappresenta un *subset* di S consistente delle uscite ξ tali che $X(\xi) \le x$.
$\{X \le x\}$ non è un set di numeri ma un *set di uscite sperimentali.*
Per esempio si può immaginare una variabile random X specificata su una tabella. Sulla colonna sinistra si possono listare gli elementi ξ_i di S e sulla colonna destra i corrispondenti valori (numeri) $X(\xi_i)$. Dato un arbitrario numero x si cercano tutti i numeri $X(\xi_i)$ che non eccedono x.
I corrispondenti numeri sulla colonna sinistra sono il *set* $\{X \le x\}$.

Funzione distribuzione probabilità

La *probabilità* $P\{X \leq x\}$ dell'evento $\{X \leq x\}$ è un numero che dipende dal valore *x*. Questo numero è denotato con $F_x(x)$ ed è chiamato *funzione distribuzione (cumulativa)* della variabile casuale X.

$$\boxed{F_x(x) = P\{X \leq x\}}$$ **Equazione 7**

definita per ogni x nell'intervallo .
Per esempio la notazione $F_x(w) = P\{X \leq w\}$ è ancora la funzione distribuzione della variabile casuale X.

Percentile

Il percentile *u* di una variabile casuale X è il *più piccolo* numero x_u tale che: $u = P\{X \leq x_u\} = F(x_u)$

- anziché fissare x_u e trovare la probabilità che $P\{X \leq x_u\}$, si desidera invece trovare x_u tale che soddisfi una data probabilità *P*.
- il fatto che debba essere il più piccolo numero x_u tale che: $u = P\{X \leq x_u\} = F(x_u)$, indica che per una funzione distribuzione di probabilità in cui a più valori di *x* corrisponda un unico valore di *P,* va scelto il numero *più piccolo* di x:

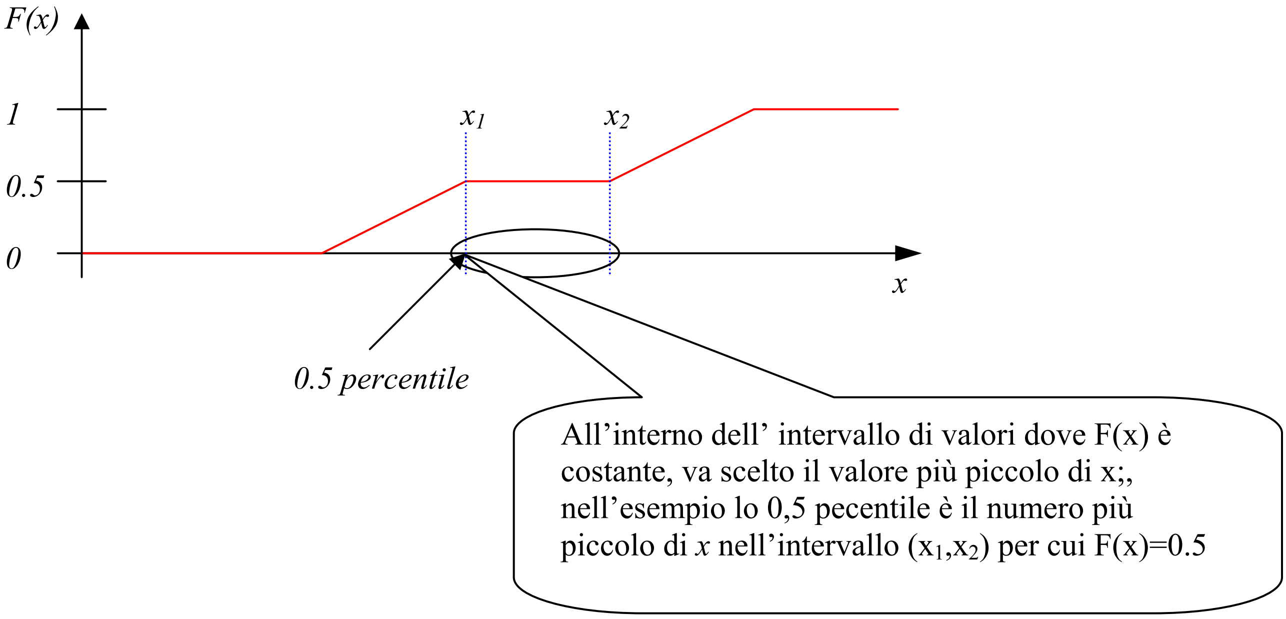


Figura 2

Funzione densità di probabilità

La derivata della funzione distribuzione di probabilità $F_x(x)$ è chiamata funzione densità di probabilità $f_x(x)$ della variabile casuale X.

$$\boxed{f_x(x) \equiv \frac{dF_x(x)}{dx}}$$

Equazione 8

$$f_x(x) \equiv \frac{dF_x(x)}{dx} = \lim_{\Delta x \to 0} \frac{F_x(x+\Delta x) - F_x(x)}{\Delta x} \geq 0$$

Equazione 9

dalla natura monotona non decrescente di $F_x(x)$, segue che $f_x(x) \geq 0$ per tutte le x. Se x è una variabile casuale continua allora anche $f_x(x)$ sarà una funzione continua.
In ogni caso se x è una variabile casuale discreta come nella figura sotto allora la sua p.d.f. ha la forma generale:

$$\boxed{f_x(x) = \sum_i p_i \delta(x - x_i)}$$

Equazione 10

dove x_i rappresenta i punti di salto-discontinuità in $F_x(x)$ ed è usata la delta di Dirac per rappresentare il valore che assume la funzione nei punti x_i.

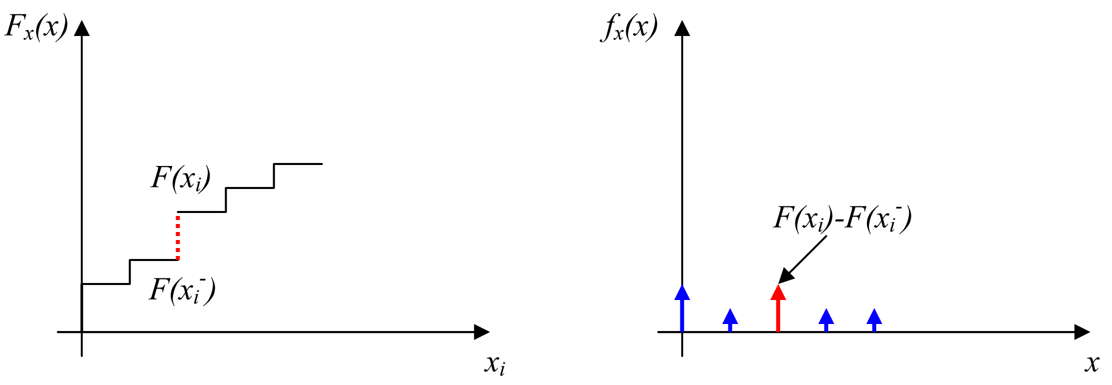


Figura 3

Come mostrato in figura $f_x(x)$ rappresenta una collezione di valori positivi discreti in casi discreti e essa è conosciuta come funzione massa di probabilità (p.m.f.)

Dalla definizione di $f_x(x)$ segue allora che integrando si riottienen $F_x(x)$:

$$\boxed{F_x(x) = \int_{-\infty}^{x} f_x(u)du}$$ **Equazione 11**

essendo $\boxed{F_x(+\infty) = 1}$ segue che

$$\boxed{F_x(x) = \int_{-\infty}^{+\infty} f_x(x)dx = 1}$$ **Equazione 12**

e questo giustifica il nome di funzione densità per la *f(x)*.

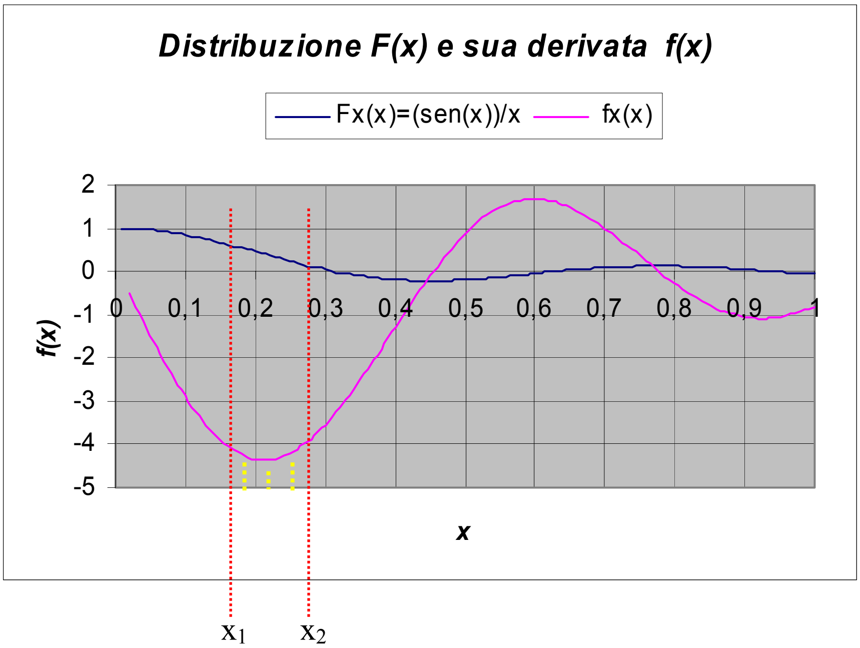


Figura 4

dalla figura sopra si osserva che ***l'area*** sotto $f_x(x)$ nell'intervallo (x_1,x_2) rappresenta la probabilità che la variabile casuale X giaccia nell'intervallo (x_1,x_2):

$$P\{x_1 < \xi < x_2\} = F_x(x_2) - F_x(x_1) = \int_{-x1}^{x2} f_x(x)dx$$

Distribuzione normale gaussiana

Si dice che X è una variabile casuale *normale o gaussiana* con media μ e varianza σ^2 se la sua funzione densità di probabilità è:

$$f_x(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Equazione 13

Poiché $f_x(x)$ dipende dai parametri μ e σ^2, allora viene usata la notazione $X \approx N(\mu, \sigma^2)$ per rappresentare la distribuzione normale.
Un esempio di densità gaussiana e graficato nella figura sottostante.

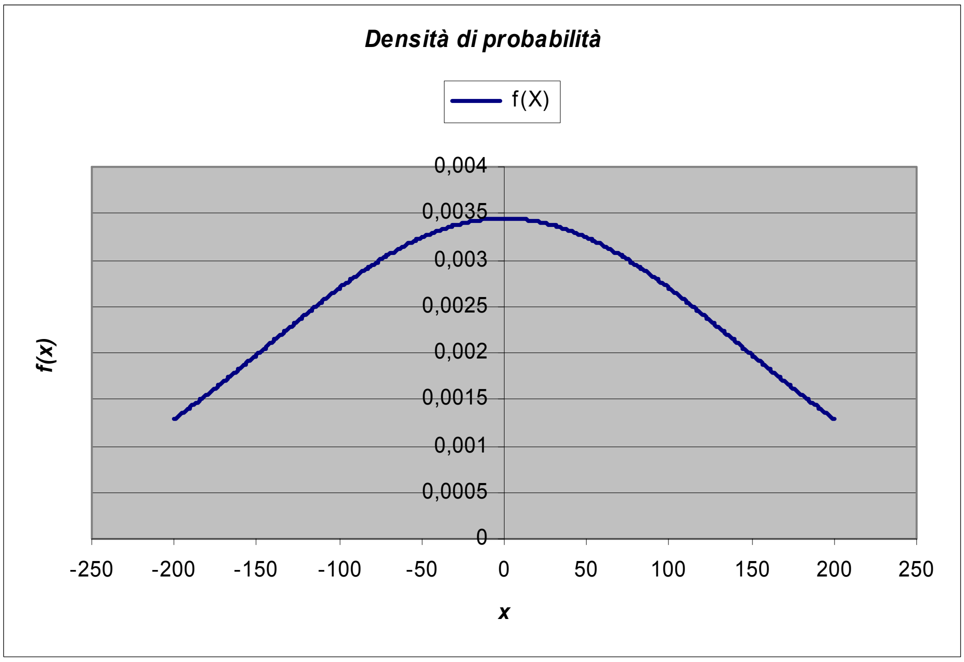


Figura 5

la curva è simmetrica intorno al parametro $\mu \neq 0$ e la relativa *funzione distribuzione di probabilità* è:

$$F_x(x) = \int_{-\infty}^{x} f_x(y)dy = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y-\mu)^2}{2\sigma^2}} dy$$

Equazione 14

usando la sostituzione: $z = \frac{y-\mu}{\sigma}$ di conseguenza $y = \sigma z + \mu \Rightarrow dy = \sigma dz$ pertanto

$$F(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(y-\mu)^2}{2\sigma^2}}dy=\int_{-\infty}^{z}\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-z^2}{2}}\sigma dz=\int_{-\infty}^{z}\frac{1}{\sqrt{2\pi}}e^{\frac{-z^2}{2}}dz=G(z)=G\left(\frac{y-\mu}{\sigma}\right)$$ **Equazione 15**

la funzione che ne risulta è quella maggiormente tabulata

$$\boxed{G(z)=\int_{-\infty}^{z}\frac{1}{\sqrt{2\pi}}e^{\frac{-z^2}{2}}dz}$$ **Equazione 16**

Un esempio di distribuzione cumulativa *F(X)* relativa ad una densità di tipo gaussiano è graficata nella figura sottostante

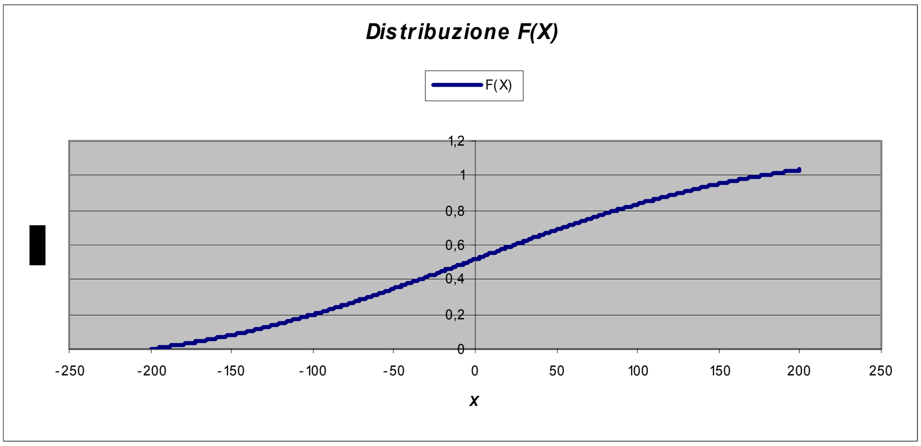


Figura 6

la costante $\sqrt{2\pi\sigma^2}$ è una *costante di normalizzazione* che mantiene l'area sotto $f_x(x)$ *unitaria* nell'intervallo da meno infinto a più infinito, pertanto se l'intervallo di integrazione della $f_x(x)$ va da meno infinito a più infinito allora l'area rappresentata da F(x) è unitaria cioè F(x)=1

questo si dimostra ponendo:

$$Q=\int_{-\infty}^{+\infty}e^{-\frac{\xi^2}{2}}d\xi=\int_{-\infty}^{+\infty}e^{-\frac{x^2}{2\sigma^2}}dx$$ **Equazione 17**

allora si può scrivere:

$$Q^2=\int_{-\infty}^{+\infty}e^{-\frac{x^2}{2\sigma^2}}dx\cdot\int_{-\infty}^{+\infty}e^{-\frac{y^2}{2\sigma^2}}dy=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}e^{-\frac{x^2+y^2}{2\sigma^2}}dxdy$$ **Equazione 18**

usando la trasformazione[1]:

[1] Per ulteriori approfondimenti sulla sostituzione di variabili in un integrale multiplo si veda l'appendice

$x = r\cos\vartheta \qquad y = r\sin\vartheta$

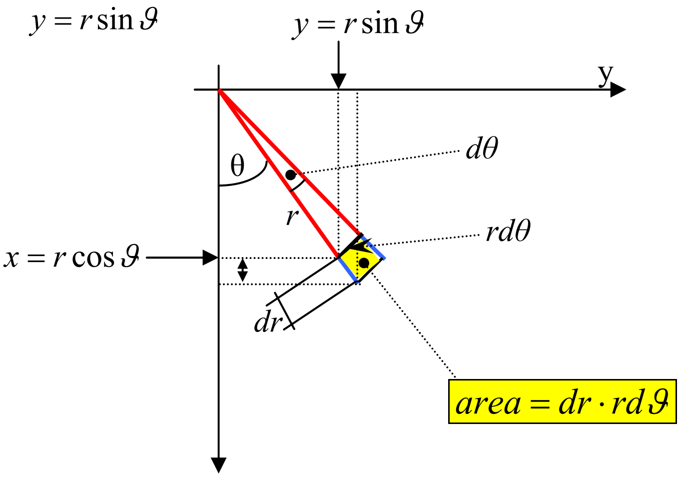


Figura 7

si ottiene:

$$Q^2 = \int_0^{2\pi}\int_0^{+\infty} e^{-\frac{r^2\cos^2\vartheta + r^2\sin^2\vartheta}{2\sigma^2}}\, r\,dr\,d\theta = \int_0^{2\pi}\int_0^{+\infty} e^{-\frac{r^2}{2\sigma^2}}\, r\,dr\,d\theta = 2\pi\int_0^{+\infty} e^{-\frac{r^2}{2\sigma^2}}\, r\,dr = -2\pi\sigma^2\int_0^{+\infty} -\frac{re^{-\frac{r^2}{2\sigma^2}}}{\sigma^2}\,dr \quad \textbf{Equazione 19}$$

essendo

$$\frac{d}{dr}e^{-\frac{r^2}{2\sigma^2}} = -\frac{2r}{2\sigma^2}e^{-\frac{r^2}{2\sigma^2}} = -\frac{r}{\sigma^2}e^{-\frac{r^2}{2\sigma^2}} \quad \textbf{Equazione 20}$$

allora l'integrale diviene

$$Q^2 = -2\pi\sigma^2\int_0^{+\infty} -\frac{re^{-\frac{r^2}{2\sigma^2}}}{\sigma^2}\,dr = -2\pi\sigma^2\int_0^{+\infty}\frac{d}{dr}e^{-\frac{r^2}{2\sigma^2}} = -2\pi\sigma^2\left[e^{-\frac{r^2}{2\sigma^2}}\right]_0^{+\infty} = -2\pi\sigma^2(0-1) = 2\pi\sigma^2 \quad \textbf{Equazione 21}$$

$$Q = \sqrt{2\pi\sigma^2}$$

Media di una variabile casuale

Il valore atteso o media di una variabile casuale (aleatoria) X è per definizione l'integrale:

$$\boxed{E(X)=\int_{-\infty}^{\infty} x\,f(x)dx} \qquad \textbf{Equazione 22}$$

dove $f(x)$ è la funzione densità di probabilità.
Questo numero può anche essere denotato per mezzo di η_x o η e μ_x o μ

Se X è uniforme nell'intervallo (x_1,x_2) allora $f(x)=\frac{1}{x_2-x_1}$ in tale intervallo, dunque si ha:

$$E(X)=\int_{-\infty}^{\infty} x\,f(x)dx=\int_{x_1}^{x_2} x\frac{1}{x_2-x_1}dx=\frac{1}{x_2-x_1}\int_{x_1}^{x_2} x\,dx=\frac{1}{x_2-x_1}\left[\frac{x^2}{2}\right]_{x_1}^{x_2}=\frac{1}{x_2-x_1}\left[\frac{x_2^2-x_1^2}{2}\right]=$$

Equazione 23

$$=\frac{1}{x_2-x_1}\left[\frac{(x_2+x_1)(x_2-x_1)}{2}\right]=\frac{(x_2+x_1)}{2}$$

se la linea verticale x=a è un asse di simmetria di $f(x)$ allora E(X)=a; in particolare se

$f(-x)=f(x)$ allora E(X)=0

Media di una variabile casuale discreta

Per una variabile casuale discreta l'integrale del valore medio può essere scritto come una somma. Si supponga che X possa assumere i valori x_i con probabilità p_i in tal caso si dimostra che la funzione di densità di probabilità $f(x)$ può essere scritta come:

$$\boxed{f(x)=\sum_i p_i\delta(x-x_i)} \qquad \textbf{Equazione 24}$$

dove x_i rappresenta i punti di salto di discontinuità in F(x), usando la definizione di valore atteso si ha allora:

$$E(X)=\int_{-\infty}^{\infty} x\,f(x)dx=\int_{-\infty}^{\infty} x\sum_i p_i\delta(x-x_i)dx=\int_{-\infty}^{\infty}\sum_i p_i x\delta(x-x_i)dx \qquad \textbf{Equazione 25}$$

usando le proprietà delle funzioni delta di Dirac per cui l'integrale di una funzione per una delta di Dirac su un punto restituisce il valore della funzione in quel punto allora si può utilizzare l'identità

$$\int_{-\infty}^{\infty} x\delta(x-x_i)dx = x_i$$ **Equazione 26**

di conseguenza dalla linearità si ha:

$$\boxed{\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x)dx = \int_{-\infty}^{\infty} x\sum_i p_i\delta(x-x_i)dx = \int_{-\infty}^{\infty}\sum_i p_i x\delta(x-x_i)dx = \\ &= \sum_i \int_{-\infty}^{\infty} p_i x\delta(x-x_i)dx = \sum_i p_i \int_{-\infty}^{\infty} x\delta(x-x_i)dx = \sum_i p_i x_i \end{aligned}}$$ **Equazione 27**

dove

$p_i = P(X=x_i)$

Interpretazione frequenza

La media aritmetica dei valori x_i di X tende al valore E(X) per $n \to \infty$

$$\bar{x} = \frac{x_1 + x_2 + \dots\dots\dots + x_n}{n} \to E(X)$$ **Equazione 28**

Media di una funzione g(X)

Data una variabile causale X e una funzione $g(x)$, può essere formata la variabile casuale Y=g(X), la media di questa variabile è data da:

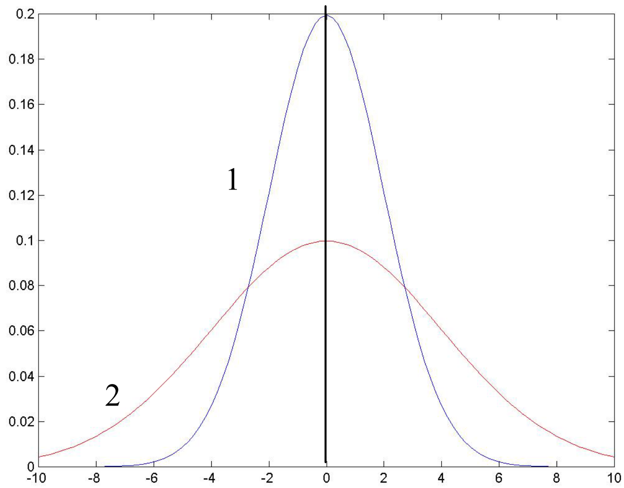
$$E(Y) = \int_{-\infty}^{\infty} y f_y(y)dy$$ **Equazione 29**

dove $f_y(y)$ è la funzione densità di probabilità. Tale funzione in ogni caso non è necessaria giacché la media si può esprimere direttamente in funzione di g(x) e di $f_x(x)$ di X:

$$E\{g(x)\} = \int_{-\infty}^{\infty} g(x) f_x(x)dx$$ **Equazione 30**

Varianza di una variabile casuale

La *Media* da sola non è in grado di rappresentare completamente la *funzione densità di probabilità P.d.f.* di una variabile casuale. Osservando la figura sotto si vede che:



le distribuzioni sono ambedue gaussiane con **media nulla**: il comportamento, totalmente differente, può essere rappresentato proprio da quanto mediamente ci si allontana dal valor medio 0: nel caso blu questa distanza è mediamente sqrt(2), nel caso rosso è mediamente 2.
In MATLAB :
x=(-10:.2:10);
y2= normpdf(x,0,2) ;
y4= normpdf(x,0,4) ;
plot(x,y2,x,y4,'r')

Figura 8

le distribuzioni hanno la stessa media (nulla), ma sono piuttosto differenti, la curva blu (1) è maggiormente concentrata attorno al valore medio, mentre la curva rossa (2) è più piatta ed allargata (wider spread).
Chiaramente si ha la necessità di un addizionale parametro per misurare questa dispersione attorno al valore medio.

Per una variabile casuale X con media μ, la quantità $(X-\mu)$ rappresenta la deviazione della variabile casuale dalla sua media. *Poiché tale deviazione può essere positiva o negativa si considera la quantità* $(X-\mu)^2$.

Il valore medio $E\left[(X-\mu)^2\right]$ rappresenta la *deviazione quadratica media* della variabile casuale X intorno alla sua media μ. Si definisce

$$\boxed{\sigma_x^2 \equiv E\left[(X-\mu)^2\right] > 0}$$ **Equazione 31**

ponendo :

$g(x) = (X-\mu)^2$

e usando :

$$E\{g(x)\} = \int_{-\infty}^{\infty} g(x) f_x(x) dx$$ **Equazione 32**

si ottiene:

$$\sigma_x^2 = E\{g(x)\} = \int_{-\infty}^{\infty} g(x) f_x(x) dx = \int_{-\infty}^{\infty} (X-\mu)^2 f_x(x) dx > 0$$ **Equazione 33**

la costante positiva σ_x^2 è conosciuta come *varianza della variabile casuale* X e la sua radice quadrata:

$$\sigma_x = \sqrt{E\{g(x)\}} = \sqrt{E\{g(X-\mu)^2\}}$$ **Equazione 34**

è conosciuta come *deviazione standard* di X e rappresenta il *valore quadratico medio* della *deviazione* della variabile casuale X intorno alla sua media μ.

Dalla definizione segue che la varianza σ_x^2 è la media della variabile $g(x) = (X-\mu)^2$ che si può scrivere dunque:

$$\begin{aligned} Var\{X\} = \sigma^2 &= E\{(X-\mu)^2\} = E\{X^2 - 2\mu X + \mu^2\} = E\{X^2\} - 2\mu E\{X\} + \mu^2 = \\ &= E\{X^2\} - 2\mu \cdot \mu + \mu^2 = E\{X^2\} - 2\mu^2 + \mu^2 = E\{X^2\} - \mu^2 = E\{X^2\} - E\{X\}^2 \end{aligned}$$

Equazione 35

Varianza di una variabile casuale di tipo discreto

Se la variabile casuale X è di tipo discreto tale che $f(x) = \sum_i p_i \delta(x - x_i)$ allora

$$\sigma^2 = \sum_i P_i (x_i - \eta)^2$$ **Equazione 36**

con $P_i = P(X = x_i)$

Momenti

Momento di una variabile casuale X

Il momento di ordine n per una variabile casuale X è definito come:

$$m_n = E(X^n) = \int_{-\infty}^{\infty} x^n f(x)dx \qquad \textbf{Equazione 37}$$

Momento centrale

$$\boxed{\mu_n = E\{(X-\mu)^n\} = \int_{-\infty}^{\infty} (X-\mu)^n f(x)dx} \qquad \textbf{Equazione 38}$$

$$\mu_n = E\{(X-\mu)^n\} = E\left\{\sum_{k=0}^{n}\binom{n}{k}X^k(-n)^{n-k}\right\} \qquad \textbf{Equazione 39}$$

per linearità essendo $m_k = E(X^k) = \int_{-\infty}^{\infty} x^k f(x)dx$ allora la precedente si può riscrivere come:

$$\mu_n = E\{(X-\mu)^n\} = \sum_{k=0}^{n}\binom{n}{k}m_k(-n)^{n-k} \qquad \textbf{Equazione 40}$$

La ***varianza*** alla luce di quanto detto è il *momento centrale* del *secondo ordine* definita come:

$$\boxed{\sigma^2 = \mu_2 = E\{(X-\mu)^2\} = \int_{-\infty}^{\infty} (X-\mu)^2 f(x)dx} \qquad \textbf{Equazione 41}$$

Si nota che il *momento centrale di ordine 1 è nullo poiché risulta* ***(μ-μ)***, e d'altra parte questo è insito nel concetto di media che rappresenta la ricerca del baricentro della distribuzione.

$$\boxed{\mu_1 = E\{(X-\mu)\} = \int_{-\infty}^{\infty} (X-\mu) f(x)dx = 0} \qquad \textbf{Equazione 42}$$

nella figura sotto è riportato un esempio di calcolo di momenti fino al 4 ordine:

variabile casuale X	momento centrale di ordine 1	momento centrale di ordine 2	momento centrale di ordine 3	momento centrale di ordine 4
2	-3	9	-27	81
4	-1	1	-1	1
6	1	1	1	1
5	0	0	0	0
8	3	9	27	81
9	4	16	64	256
1	-4	16	-64	256
5	0	7,428571429	0	96,57142857
media(x)	***media[(x-media(x))]***	***media[(x-media(x))^2]***	***media[(x-media(x))^3]***	***media[(x-media(x))^4]***

Figura 9

Momento assoluto

$$E\left\{|X|^n\right\} \qquad E\left\{|X-\mu|^n\right\}$$

Equazione 43

Matrice di Covarianza

Si considerino s esperimenti casuali definiti su un certo spazio campionario.
Se **X** è un **vettore colonna** *n x 1* di variabili casuali a valori reali X_i, appartenente ad $\mathbf{R}^n$ e **Y** è un **vettore <u>colonna</u>** *m x 1* di variabili casuali a valori reali Y_i appartenente ad $\mathbf{R}^m$ allora la ***matrice di covarianza*** *n x m* è in definita come

$$COV(\mathbf{X},\mathbf{Y}) = E\left\{[\mathbf{X}-E(\mathbf{X})]\cdot[\mathbf{Y}-E(\mathbf{Y})]^T\right\}$$

$$= E\left\{\begin{bmatrix} X_1-E(X_1) \\ \vdots \\ X_n-E(X_n)\end{bmatrix}\cdot\begin{bmatrix} Y_1-E(Y_1) \\ \vdots \\ Y_m-E(Y_m)\end{bmatrix}^T\right\} =$$

Equazione 44

$$= E\left\{\begin{bmatrix} X_1-E(X_1) \\ \vdots \\ X_n-E(X_n)\end{bmatrix}\cdot\begin{bmatrix} Y_1-E(Y_1) & \cdots & \cdots & Y_m-E(Y_m)\end{bmatrix}\right\}$$

dove $E(X_i)$ e $E(Y_j)$ sono i valori attesi (le medie) delle variabili casuali X_i e Y_j sull'insieme degli s esperimenti considerati.

Per esempio per una variabile casuale discreta a valori egualmente probabili con probabilità p_k unitaria, la media si può scrivere:

Valore assunto dalla variabile aleatoria X_i nel k-esimo esperimento

$$E(X_i) = \frac{\sum_{k=1}^{s} (x_i)_k}{s}$$ con i=1,2…..,n e s= numero esperimenti

$$E(Y_j) = \frac{\sum_{k=1}^{s} (y_j)_k}{s}$$ con j=1,2,…,m e s= numero esperimenti

Il prodotto righe per colonne fornisce la matrice di covarianza di *n x m* variabili casuali a valori reali, equivalentemente la matrice di covarianza è una matrice casuale :

$$COV(\mathbf{X,Y}) = E\left\{\begin{matrix} (X_1 - E(X_1))(Y_1 - E(Y_1)) & \cdots & \cdots & (X_1 - E(X_1))(Y_m - E(Y_m)) \\ \vdots & \cdots & \cdots & \vdots \\ & \cdots & \cdots & \\ (X_n - E(X_n))(Y_1 - E(Y_1)) & \cdots & \cdots & (X_n - E(X_n))(Y_m - E(Y_m)) \end{matrix}\right\}$$

Equazione 45

il valore atteso (la media) di tale matrice è equivalente alla matrice dei valori attesi di singoli elementi della matrice:

$$COV(\mathbf{X,Y}) = \left\{\begin{matrix} E[(X_1 - E(X_1))(Y_1 - E(Y_1))] & \cdots & \cdots & E[(X_1 - E(X_1))(Y_m - E(Y_m))] \\ \vdots & \cdots & \cdots & \vdots \\ & \cdots & \cdots & \\ E[(X_n - E(X_n))(Y_1 - E(Y_1))] & \cdots & \cdots & E[(X_n - E(X_n))(Y_m - E(Y_m))] \end{matrix}\right\}$$

$$= \left\{\begin{matrix} C_{X_1 Y_1} & \cdots & \cdots & C_{X_1 Y_m} \\ \vdots & \cdots & \cdots & \vdots \\ & \cdots & \cdots & \\ C_{X_n Y_1} & \cdots & \cdots & C_{X_n Y_m} \end{matrix}\right\}$$

Equazione 46

dove il generico termine $c_{i,j}$ rappresenta la covarianza $C_{X_I Y_J}$ delle variabili casuali X_i e Y_j :

$$c_{i,j} = E\left[\left(X_i - E(X_i)\right)\left(Y_j - E(Y_j)\right)\right] = C_{X_I Y_J}$$ **Equazione 47**

e dove il valore atteso (la media) di ogni singolo elemento della matrice di covarianza potrebbe essere scritto sotto opportune condizioni come:

$$E\left[\left(X_i - E(X_i)\right)\cdot\left(Y_j - E(Y_j)\right)\right] = \frac{\sum_{k=1}^{s}\left[\left((x_i)_k - E(X_i)\right)\cdot\left((y_j)_k - E(Y_j)\right)\right]}{s}$$

Matrice di covarianza di un solo vettore di variabili casuali X

Un caso particolare del precedente si ha considerando un solo vettore di variabili casuali **X=[**$X_1, \ldots X_n$**].** In questo caso la matrice di covarianza si ottiene dalle precedenti sostituendo X_i ad Y_j come di seguito illustrato:

$$COV(\boldsymbol{X}) = E\left\{\left[\boldsymbol{X} - E(\boldsymbol{X})\right]\cdot\left[\boldsymbol{X} - E(\boldsymbol{X})\right]^T\right\}$$

$$= E\left\{\begin{bmatrix} X_1 - E(X_1) \\ \vdots \\ \vdots \\ X_n - E(X_n) \end{bmatrix}\cdot\begin{bmatrix} X_1 - E(X_1) \\ \vdots \\ \vdots \\ X_n - E(X_n) \end{bmatrix}^T\right\} =$$

Equazione 48

$$= E\left\{\begin{bmatrix} X_1 - E(X_1) \\ \vdots \\ \vdots \\ X_n - E(X_n) \end{bmatrix}\cdot\left[X_1 - E(X_1) \quad \cdots \quad \cdots \quad X_n - E(X_n)\right]\right\}$$

$$COV(\boldsymbol{X}) = \begin{Bmatrix} E[(X_1 - E(X_1))(X_1 - E(X_1))] & \cdots & \cdots & E[(X_1 - E(X_1))(X_n - E(X_n))] \\ \vdots & \cdots & \cdots & \vdots \\ & \cdots & \cdots & \\ E[(X_n - E(X_n))(X_1 - E(X_1))] & \cdots & \cdots & E[(X_n - E(X_n))(X_n - E(X_n))] \end{Bmatrix} =$$

$$= \begin{Bmatrix} \sigma^2_{X_1} & \cdots & \cdots & C_{X_1\,X_n} \\ \vdots & \cdots & \cdots & \vdots \\ & \cdots & \cdots & \\ C_{X_n\,X_1} & \cdots & \cdots & \sigma^2_{X_n} \end{Bmatrix}$$

Equazione 49

dove il generico termine $c_{i,i}$ della diagonale principale rappresenta la ***varianza*** $\sigma^2_{X_i}$ della variabile casuale X_i :

$$c_{i,i} = E[(X_i - E(X_i))(X_i - E(X_i))] = E\left[(X_i - E(X_i))^2\right] = \sigma^2_{X_i}$$ **Equazione 50**

mentre il generico termine $c_{i,j}$ fuori dalla diagonale principale rappresenta la ***covarianza*** $C_{X_I\,X_J}$ delle variabili casuali X_i e X_j :

$$c_{i,j} = E\left|(X_i - E(X_i))(X_j - E(X_j))\right| = C_{X_I\,X_J}$$ **Equazione 51**

Covarianza di due sole due variabili casuali X e Y

Un caso particolare rispetto al primo più generale è quando anziché due vettore di variabili casuali, si considerano solo due variabili casuali a valori reali X e Y relative ai *k* esperimenti.

La covarianza di due variabili casuali, misura la loro tendenza a variare insieme cioè a co-variare. Mentre la varianza è il valore quadratico medio della deviazione della variabile casuale dalla sua media, la covarianza è la media dei prodotti delle deviazioni delle variabili casuali dalla loro rispettiva media.

Siano *E(X)* ed *E(Y)* i valori attesi (le medie) di tali variabili, allora la covarianza di X e Y è definita come:

$$\boxed{COV(X,Y) = E\{[X - E(X)]\cdot[Y - E(Y)]\}}$$ **Equazione 52**

pertanto la covarianza si può esplicitare come:

$$\begin{aligned} COV(X,Y) &= E\{[X - E(X)]\cdot[Y - E(Y)]\} = E\{X\cdot Y - X\cdot E(Y) - Y\cdot E(X) + E(X)\cdot E(Y)\} = \\ &= E(X\cdot Y) - E(X)\cdot E(Y) - E(Y)\cdot E(X) + E(X)\cdot E(Y) = \\ &= E(X\cdot Y) - E(X)\cdot E(Y) \end{aligned}$$

Equazione 53

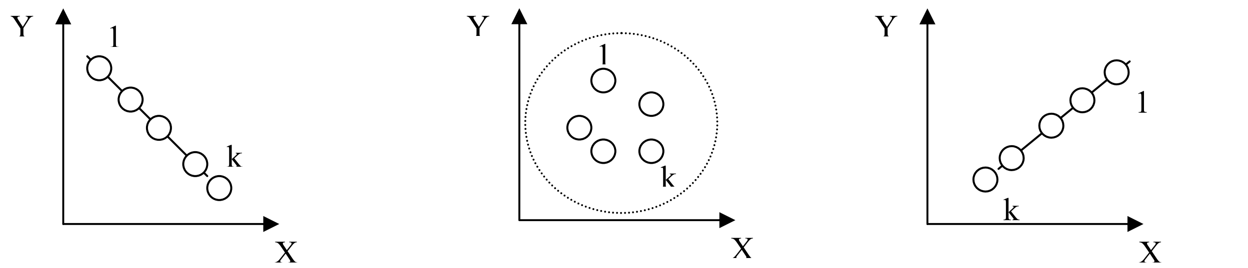
la covarianza ha le seguenti proprietà:

- Se X e Y hanno valori che si incrementano contemporaneamente allora COV(X,Y) > 0
- Se X si incrementa quando Y si decrementa allora COV(X,Y) < 0
- Se X e Y sono indipendenti allora COV(X,Y)=0
- $-sd(X)\cdot sd(Y) \le COV(X,Y) \le sd(X)\cdot sd(Y)$
- $COV(X,X) \le sd(X)^2 =$ varianza(X)

dove con *sd(X)* si è indicata la *deviazione standard* (Standard Deviation) della variabile casuale X dalla sua media.

La covarianza misura la dipendenza tra le variabili casuali X e Y, se due variabili sono indipendenti allora $E(X\cdot Y) = E(X)\cdot E(Y)$ *ne consegue che COV(X,Y)=0*

le tre figure sotto riportate mostrano tre casi diversi di dipendenza tra le variabili casuali X e Y nel caso dei valori assunti in ogni esperimento k:



$$COV(X,Y) = -sd(X)\cdot sd(Y) \qquad COV(X,Y) = 0 \qquad COV(X,Y) = sd(X)\cdot sd(Y)$$

Figura 10

Correlazione di due variabili casuali X e Y

Il coefficiente di correlazione di due variabili casuali X e Y è per definizione il rapporto:

$$cor(X,Y) = \frac{\mathrm{cov}(X,Y)}{sd(X)\cdot sd(Y)} = \frac{C_{XY}}{\sigma_x \cdot \sigma_y} = \rho_{XY}$$

Equazione 54

dove con σ_x e σ_y sono indicate le deviazioni standard per le variabili casuali X e Y.

La *correlazione* è quindi una versione modificata della covarianza in particolare può essere vista come una *covarianza normalizzata che può assumere valori compresi tra -1 e +1.*

Covarianza e correlazione hanno sempre lo stesso segno, quando il segno è *positivo* le variabili si dicono *positivamente correlate*; quando il segno è *negativo* le variabili si dicono *negativamente correlate*, quando è *zero* le variabili si dicono *incorrelate*.

In particolare se X e Y sono *indipendenti*, allora sono anche *incorrelate*, il contrario non è vero cioè se X e Y sono incorrelate non è detto che siano indipendenti.

Ortogonalità

Due variabili casuali X e Y sono *ortogonali* se il *valore atteso* è nullo:

$$\boxed{E(X \cdot Y) = 0} \qquad \textbf{Equazione 55}$$

se inoltre le due variabili sono anche *indipendenti* allora si ha

$$\boxed{E(X \cdot Y) = E(X) \cdot E(Y) = 0} \qquad \textbf{Equazione 56}$$

quindi consegue che la *covarianza* è *nulla se si verificano le condizioni:*

1. *le due variabili sono indipendenti*
2. *oltre ad essere indipendenti sono anche ortogonali*

cioè $COV(X,Y) = E(X \cdot Y) - E(X) \cdot E(Y) = 0$ **Equazione 57**

viceversa se la *covarianza* di due variabili casuali è *nulla* non possiamo affermare nulla sulla loro *ortogonalità* ma soltanto sulla loro indipendenza:

$COV(X,Y) = 0$ **Equazione 58**

implica solo

$E(X \cdot Y) = E(X) \cdot E(Y)$ **Equazione 59**

ma non che

$E(X \cdot Y) = E(X) \cdot E(Y) = 0$ **Equazione 60**

pertanto riassumendo:

condizione *necessaria e sufficiente* affinché due variabili casuali siano *statisticamente indipendenti* è che la loro *covarianza* sia nulla.

Appendice

Cambiamento di variabile in un integrale multiplo

Caso particolare: cambiamento di variabile in un integrale doppio

Consideriamo un integrale doppio:

$$\iint_D f(x,y)\,dx\,dy$$

Equazione 61

Supponiamo di volere effettuare il seguente cambiamento di variabile:

$$\begin{cases} x = X(u,v) \\ y = Y(u,v) \end{cases}$$

Equazione 62

L' Equazione 63 definisce un'applicazione che manda un punto *(u,v)* del piano *uv* in un punto immagine *(x,y)* del piano *xy*. Indichiamo con T l'insieme del piano *uv* che viene mandato nell'insieme D del piano *xy* [2]. Ogni punto di T viene trasformato in un punto di D, come illustrato in Figura 11

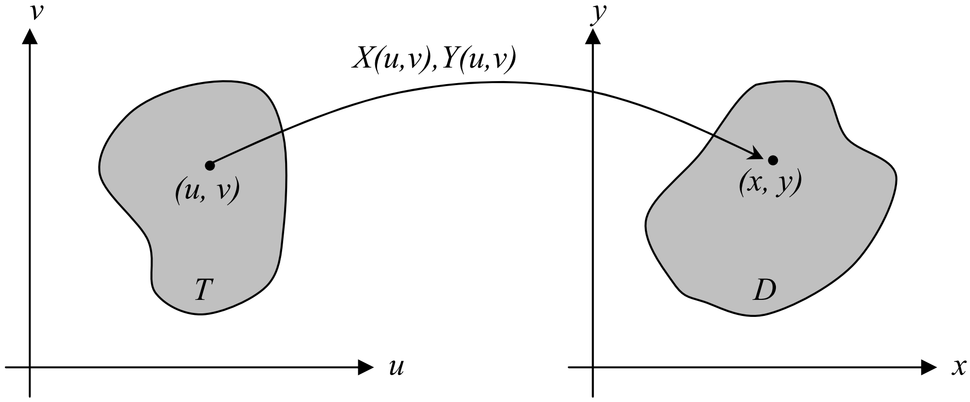


Figura 12

[2] Pertanto D e T sono sottoinsiemi dello spazio R^2.

Si suppone che siano valide le seguenti ipotesi:

1. Si suppone che, nell'insieme T, le funzioni X e Y siano continue e hanno derivate parziali continue $\frac{\partial X}{\partial u}$, $\frac{\partial X}{\partial v}$, $\frac{\partial Y}{\partial u}$, $\frac{\partial Y}{\partial v}$.
2. Si suppone che la trasformazione da T a D sia ***iniettiva***, cioè mandi punti distinti di T in punti distinti di D ([3]). Ciò consente (almeno in teoria) di "tornare" da D a T attraverso la trasformazione inversa (che, naturalmente, è ancora *iniettiva*).
3. Indichiamo con $J(u,v)$ la ***matrice Jacobiana***. Essa è definita nel modo seguente:

$$J(u,v) = \begin{bmatrix} \frac{\partial X}{\partial u} & \frac{\partial X}{\partial v} \\ \frac{\partial Y}{\partial u} & \frac{\partial Y}{\partial v} \end{bmatrix}$$

Equazione 64

Si suppone che il determinante della ***matrice Jacobiana*** (anche detto ***determinante jacobiano***) sia diverso da *0*.

Sotto le suddette ipotesi, l'integrale dell'Equazione 61 può essere riscritto nel modo seguente:

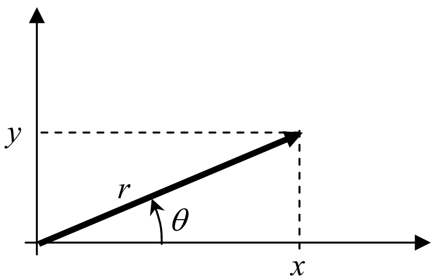
$$\iint_D f(x,y)\,\boxed{dx\,dy} = \iint_T f[X(u,v),Y(u,v)]\,\boxed{|\det[J(u,v)]|\,du\,dv}$$

Equazione 65

dove $|\det[J(u,v)]|$ è il modulo del ***determinante jacobiano***.

L' Equazione 66 è detta ***formula di trasformazione***.

Applicazione: passaggio da coordinate cartesiane a coordinate polari



In questo caso si ha

[3] In altri termini, nessuna coppia di punti distinti di T è mandata nello stesso punto di D.

$$\begin{cases} x = X(r,\theta) = r\cos(\theta) \\ y = Y(r,\theta) = r\,sen(\theta) \end{cases}$$

Equazione 67

da cui

$$J(u,v) = \begin{bmatrix} \dfrac{\partial X}{\partial r} & \dfrac{\partial X}{\partial \theta} \\ \dfrac{\partial Y}{\partial r} & \dfrac{\partial Y}{\partial \theta} \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -r\,sen(\theta) \\ sen(\theta) & r\cos(\theta) \end{bmatrix} \Rightarrow \left|\det[J(u,v)]\right| = \left|r\cos^2(\theta) + r\,sen^2(\theta)\right| = r$$

e quindi l'Equazione 65 diventa

$$\boxed{\iint_D f(x,y)\,dx\,dy = \iint_T f[r\cos(\theta), r\,sen(\theta)]\,r\,dr\,d\theta}$$

Equazione 68

Estensione al caso generale di un integrale su uno spazio a n dimensioni, con n ≥ 3

Consideriamo un integrale su uno spazio ad n dimensioni:

$$\int \cdots \int_D f(x_1, x_2, \cdots x_n)\,dx_1\,dx_2 \cdots dx_n$$

Equazione 69

Supponiamo di volere effettuare il seguente cambiamento di variabile:

$$\begin{cases} x_1 = X_1(u_1, u_2, \cdots u_n) \\ x_2 = X_2(u_1, u_2, \cdots u_n) \\ \quad\vdots \\ x_n = X_n(u_1, u_2, \cdots u_n) \end{cases}$$

Equazione 70

L' Equazione 71 definisce un'applicazione che manda un punto dello sapzio ad n-dimensioni $(u_1, u_2, \cdots u_n)$ in un punto immagine dello spazio ad n-dimensioni $(x_1, x_2, \cdots x_n)$. Indichiamo con T l'insieme dello spazio $(u_1, u_2, \cdots u_n)$ che viene mandato nell'insieme D dello spazio $(x_1, x_2, \cdots x_n)$ [4]. Ogni punto di T viene trasformato in un punto di D.

Sotto ipotesi analoghe a quelle riportate nel paragrafo 0, l'integrale dell' Equazione 72 può essere riscritto nel modo seguente:

[4] Pertanto D e T sono sottoinsiemi dello spazio R^n.

$$\boxed{\int_D \cdots \int f(x_1,\cdots x_n)\, dx_1\, dx_2 \cdots dx_n = \int_T \cdots \int f(X_1(u_1,\cdots n_n),\cdots X_n(u_1,\cdots n_n))\,|\det[J(u_1,\cdots u_n)]|\, du_1 \cdots du_n}$$

Equazione 73

dove $|\det[J(u_1,\cdots u_n)]|$ è il modulo del ***determinante jacobiano***. La ***matrice Jacobiana*** è definita nel modo seguente:

$$J(u_1,u_2,\cdots u_n)=\begin{bmatrix} \frac{\partial X_1}{\partial u_1} & \frac{\partial X_1}{\partial u_2} & \cdots & \frac{\partial X_1}{\partial u_n} \\ \frac{\partial X_2}{\partial u_1} & \frac{\partial X_2}{\partial u_2} & & \frac{\partial X_2}{\partial u_n} \\ \vdots & & & \\ \frac{\partial X_n}{\partial u_1} & \frac{\partial X_n}{\partial u_2} & & \frac{\partial X_n}{\partial u_n} \end{bmatrix}$$

Equazione 74

L'Equazione 73 è detta ***formula di trasformazione***.

Applicazione: passaggio da coordinate cartesiane in R^3 a coordinate cilindriche

Si consideri la trasformazione dalle coordinate cartesiane (x,y,z) alle coordinate cilindriche (r,θ,z), come illustrato in Figura 13.

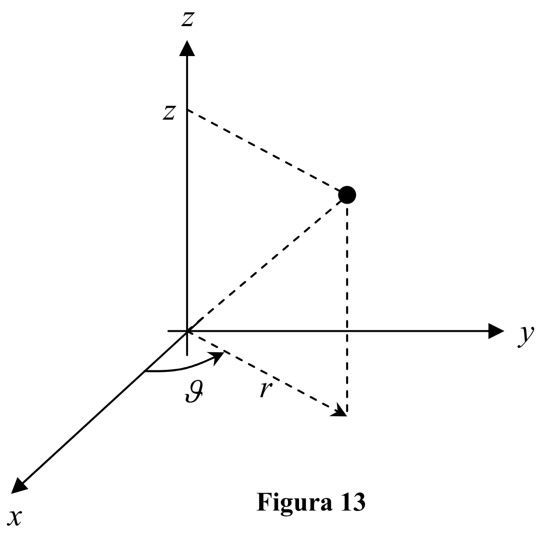


Figura 13

In questo caso la trasformazione di coordinate è data da

$$\begin{cases} x = X_1(r,\theta,z) = r\cos(\theta) \\ y = X_2(r,\theta,z) = r\,sen(\theta) \\ z = X_3(r,\theta,z) = z \end{cases}$$

Equazione 75

da cui

$$J(r,\theta,z) = \begin{bmatrix} \frac{\partial X_1}{\partial r} & \frac{\partial X_1}{\partial \theta} & \frac{\partial X_1}{\partial z} \\ \frac{\partial X_2}{\partial r} & \frac{\partial X_2}{\partial \theta} & \frac{\partial X_2}{\partial z} \\ \frac{\partial X_3}{\partial r} & \frac{\partial X_3}{\partial \theta} & \frac{\partial X_3}{\partial z} \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -r\,sen(\theta) & 0 \\ sen(\theta) & r\cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Equazione 76

$$\Downarrow$$

$$\left|\det[J(r,\theta,z)]\right| = \left|r\cos^2(\theta) + r\,sen^2(\theta)\right| = r$$

$$\Downarrow$$

$$\boxed{\begin{aligned} &\iiint_D f(x,y,z)\,dx\,dy\,dz = \iiint_T f[r\cos(\theta), r\,sen(\theta), z]\left|\det[J(r,\theta,z)]\right| dr\,d\theta\,dz = \\ &= \iiint_T f[r\cos(\theta), r\,sen(\theta), z]\,r\,dr\,d\theta\,dz \end{aligned}}$$

Equazione 77

Applicazione: passaggio da coordinate cartesiane in R^3 a coordinate sferiche

Si consideri la trasformazione dalle coordinate cartesiane (x,y,z) alle coordinate sferiche (ρ,θ,φ), come illustrato in Figura 13.

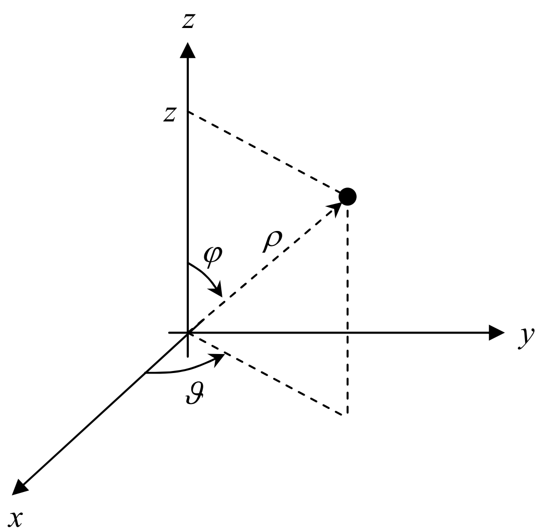


Figura 14

In questo caso la trasformazione di coordinate è data da

$$\begin{cases} x = X_1(\rho,\theta,\varphi) = \rho\; sen(\varphi)\cos(\theta) \\ y = X_2(\rho,\theta,\varphi) = \rho\; sen(\varphi)\, sen(\theta) \\ z = X_3(\rho,\theta,\varphi) = \rho\cos(\varphi) \end{cases}$$

Equazione 78

da cui

$$J(\rho,\theta,\varphi) = \begin{bmatrix} \dfrac{\partial X_1}{\partial \rho} & \dfrac{\partial X_1}{\partial \theta} & \dfrac{\partial X_1}{\partial \varphi} \\ \dfrac{\partial X_2}{\partial \rho} & \dfrac{\partial X_2}{\partial \theta} & \dfrac{\partial X_2}{\partial \varphi} \\ \dfrac{\partial X_3}{\partial \rho} & \dfrac{\partial X_3}{\partial \theta} & \dfrac{\partial X_3}{\partial \varphi} \end{bmatrix} = \begin{bmatrix} \cos(\theta)\, sen(\varphi) & -\rho\; sen(\theta)\, sen(\varphi) & \rho\cos(\theta)\cos(\varphi) \\ sen(\theta)\, sen(\varphi) & \rho\cos(\theta)\, sen(\varphi) & \rho\; sen(\theta)\cos(\varphi) \\ \cos(\varphi) & 0 & -\rho\; sen(\varphi) \end{bmatrix}$$

$$\Downarrow$$

$$\begin{aligned} &|\det[J(\rho,\theta,\varphi)]| = |\ \cos(\theta)sen(\varphi)\left[-\rho^2\cos(\theta)sen^2(\varphi)\right] + \\ &+\rho\, sen(\theta)sen(\varphi)\left[-\rho\, sen(\theta)sen^2(\varphi) - sen(\theta)\cos^2(\varphi)\right] + \rho\cos(\theta)\cos(\varphi)\left[-\rho\cos(\theta)\, sen(\varphi)\cos(\varphi)\right]\ | = \\ &= \left|\ \rho^2\ sen(\varphi)\left[-\cos^2(\theta)\, sen^2(\varphi) - sen^2(\theta) - \cos^2(\theta)\cos^2(\varphi)\right]\right| = \\ &= \left|\ -\rho^2\ sen(\varphi)\left\{\cos^2(\theta)\left[sen^2(\varphi)+\cos^2(\varphi)\right] + sen^2(\theta)\right\}\right| = \left|\ -\rho^2\ sen(\varphi)\right| = \rho^2\ sen(\varphi) \end{aligned}$$

Perché $sen(\varphi) \geq 0$ in quanto $0 \leq \varphi \leq \pi$.

$$\Downarrow$$

$$\begin{aligned} &\iiint_D f(x,y,z)\,dx\,dy\,dz = \\ &= \iiint_T f[\,\rho\cos(\theta)\,sen(\varphi),\ \rho\;sen(\theta)\,sen(\varphi),\ \rho\cos(\varphi)\,]\,|\det[J(\rho,\theta,\varphi)]|\ \rho^2\ sen(\varphi)\,d\rho\,d\theta\,d\varphi = \\ &= \iiint_T f[\,\rho\cos(\theta)\,sen(\varphi),\ \rho\;sen(\theta)\,sen(\varphi),\ \rho\cos(\varphi)\,]\ \rho^2\ sen(\varphi)\,d\rho\,d\theta\,d\varphi \end{aligned}$$

Equazione 79

Bibliografia:

[1] Probability, Random Variables and Stocasthastic Process (A.Papoulis)
[2] Volume terzo Analisi 2 (Tom M. Apostol)
[3] Covariance (www.engr.sjsu.edu/~knapp/HCIRODPR/PR_Mahal/cov.htm)
[4] covarianza e correlazione (www.eco.uninsubria.it/VL/VL_IT/expect/expect3.html)