

ANALISI MATEMATICA

Ottavio Caligaris - Pietro Oliva

CAPITOLO 1

UN PO' DI LOGICA

Diciamo proposizione una affermazione di cui siamo in grado di stabilire se è vera o è falsa.

Indichiamo con lettere maiuscole le proposizioni e scriviamo $c : P$, leggendo c tale che P è vera, se c è un elemento in corrispondenza del quale la proposizione P è vera.

Assegnata una proposizione P si può costruire una nuova proposizione, che definiamo negazione di P ed indichiamo con $\text{not } P$, come la proposizione che è vera se P è falsa ed è falsa se P è vera.

Si può identificare la proposizione $\text{not } P$ anche mediante una tabella, detta tabella di verità, che elenca in corrispondenza dei due casi possibili la verità o la falsità della proposizione in questione:

P	not P
1	0
0	1

TABELLA 1.1

È inoltre necessario definire nuove proposizioni che dipendono da una o più proposizioni note.

Assegnate due proposizioni P e Q ,

- $(P \text{ and } Q)$ è vera se P e Q sono entrambe vere
- $(P \text{ or } Q)$ è vera se almeno una tra P e Q è vera
- $(P \text{ xor } Q)$ è vera se una ed una sola tra P e Q è vera.

Le corrispondenti tabelle di verità possono essere raggruppate nella seguente:

P	Q	$\text{not } P$	$\text{not } Q$	$P \text{ and } Q$	$P \text{ or } Q$	$P \text{ xor } Q$
1	1	0	0	1	1	0
0	1	1	0	0	1	1
1	0	0	1	0	1	1
0	0	1	1	0	0	0

TABELLA 1.2

È immediato verificare che proposizione $P \text{ xor } Q$ è vera o falsa a seconda che sia vera o falsa la proposizione

$$(P \text{ and } (\text{not } Q)) \text{ or } (Q \text{ and } (\text{not } P))$$

come si può verificare dalla tabella 1.2.

Possiamo anche verificare come not interagisce con and e or mediante la tabella 1.3

P	Q	not P	not Q	P and Q	P or Q	not(P and Q)	not (P or Q)
1	1	0	0	1	1	0	0
0	1	1	0	0	1	1	0
1	0	0	1	0	1	1	0
0	0	1	1	0	0	1	1

TABELLA 1.3

Dalla tabella 1.3 possiamo verificare che

- $(\text{not}(P \text{ and } Q))$ è vera tutte e sole le volte che è vera $((\text{not } P) \text{ or } (\text{not } Q))$
- $(\text{not}(P \text{ or } Q))$ è vera tutte e sole le volte che è vera $((\text{not } P) \text{ and } (\text{not } Q))$

Si può inoltre affermare che le seguenti affermazioni sono sempre vere

- $(P \text{ or } (\text{not } P))$ (legge del terzo escluso)
- $(\text{not}(P \text{ and } (\text{not } P)))$ (legge di non contraddizione)

Assegnate due proposizioni P e Q si possono inoltre costruire le seguenti proposizioni

$$(P \Rightarrow Q) , (P \Leftarrow Q) , (P \Leftrightarrow Q)$$

che leggiamo, rispettivamente ' P implica Q ', ' P è implicato da Q ', ' P è equivalente a Q ' e che sono identificate come segue

- $(P \Rightarrow Q)$ significa che Q è vera ogni volta che P è vera;
- $(P \Leftarrow Q)$ significa che P è vera ogni volta che Q è vera;
- $(P \Leftrightarrow Q)$ significa che P è vera tutte e sole le volte in cui Q è vera.

In altre parole $(P \Rightarrow Q)$ significa che o non è vera P oppure, se P è vera, allora è vera anche Q ; in simboli:

$$(1.1) \quad (P \Rightarrow Q) \Leftrightarrow ((\text{not } P) \text{ or } Q)$$

Possiamo verificare dalla tabella 1.4 che due proposizioni sono equivalenti se assumono gli stessi valori nella loro tabella di verità, cioè se sono entrambe vere o entrambe false.

P	Q	$P \Rightarrow Q$	$P \Leftarrow Q$	$P \Leftrightarrow Q$
1	1	1	1	1
0	1	1	0	0
1	0	0	1	0
0	0	1	1	1

TABELLA 1.4

Per convincerci che la definizione di implicazione corrisponde a criteri di senso comune, è opportuno mettere in evidenza la negazione della proposizione $(P \Rightarrow Q)$; avremo che

$$(1.2)$$

$$\text{not}(P \Rightarrow Q) \quad \text{se e solo se} \quad \text{not}((\text{not } P) \text{ or } Q) \quad \text{se e solo se} \quad (P \text{ or } (\text{not } Q))$$

Infatti è chiaro che $\text{not}(P \Rightarrow Q)$ se accade che P è vera e Q è falsa.

La seguente tabella permette di verificare che le due proposizioni contenute in 1.2 hanno la stessa tabella di verità; cioè sono equivalenti.

P	Q	$\text{not } Q$	$P \Rightarrow Q$	$(P \text{ and } (\text{not } Q))$	$\text{not } (P \Rightarrow Q)$
1	1	0	1	0	0
0	1	0	1	0	0
1	0	1	0	1	1
0	0	1	1	0	0

TABELLA 1.5

Osserviamo anche

$$(1.3) \quad (P \Rightarrow Q) \Leftrightarrow ((\text{not } Q) \Rightarrow (\text{not } P))$$

Per cui possiamo aggiungere una colonna alla tabella 1.5

P	Q	$\text{not } Q$	$\text{not } P$	$P \Rightarrow Q$	$(\text{not } Q) \Rightarrow (\text{not } P)$
1	1	0	0	1	1
0	1	0	1	1	1
1	0	1	0	0	0
0	0	1	1	1	1

TABELLA 1.6

$$(P \Leftrightarrow Q) \Leftrightarrow ((P \Rightarrow Q) \text{ and } (P \Leftarrow Q))$$

$$(P \Rightarrow Q) \Leftrightarrow ((\text{not } Q) \Rightarrow (\text{not } P)) \Leftrightarrow (\text{not}(P \text{ and } (\text{not } Q)))$$

Quest'ultima relazione è nota come **principio di dimostrazione per assurdo**.

Ricordiamo che si suppone noto il concetto di *insieme*.

Usualmente gli insiemi sono identificati da una lettera maiuscola, mentre le lettere minuscole, di solito, designano gli elementi di un insieme.

Ricordiamo anche che

- $a \in A$ significa che a è un elemento di A , a appartiene ad A ;
- $b \notin A$ significa che b non è un elemento di A , b non appartiene ad A .
- se A è un insieme, $a \in A$ e P_a è una proprietà che dipende da a , tale che cioè sia vera per certi valori di a e falsa per altri valori di a , scriviamo

$$\{a \in A : P_a\} \quad \text{oppure} \quad \{a \in A : P_a \text{ è vera}\}$$

per indicare l'insieme degli elementi di A tali che P_a è vera.

Occorre infine ricordare che si dice data una relazione binaria su un insieme A se dati due elementi $a, b \in A$ è possibile stabilire se è vera o falsa la proposizione ' **a è in relazione con b** '.

Scriveremo $a\mathcal{R}b$ e $a\mathcal{R}b$ per significare che la proposizione in oggetto è rispettivamente vera o falsa.

Una relazione binaria si dice *di relazione di equivalenza* se sono verificate le seguenti condizioni

- $(a\mathcal{R}b) \Rightarrow (b\mathcal{R}a)$ (simmetrità);
- $(a\mathcal{R}a)$ (riflessività);
- $((a\mathcal{R}b) \text{ and } (b\mathcal{R}c)) \Rightarrow (a\mathcal{R}c)$ (transitività).

Una *relazione binaria* si dice *relazione d'ordine* o *ordinamento* se sono verificate le seguenti condizioni

- $(a\mathcal{R}a)$ (antiriflessività);
- $((a\mathcal{R}b) \text{ and } (b\mathcal{R}c)) \Rightarrow (a\mathcal{R}c)$ (transitività).

Siano A, B due insiemi, diciamo che

$$(1.4) \quad A \subset B \quad (B \supset A) \quad \text{se} \quad (a \in A) \Rightarrow (a \in B)$$

Diciamo che

$$(1.5) \quad A = B \quad \text{se} \quad (a \in A) \Leftrightarrow (a \in B)$$

Definiamo

$$A \setminus B = \{a \in A \text{ and } a \notin B\}$$

Nel caso in cui $B \subset A$ l'insieme $A \setminus B$ si dice anche *complementare di B in A* e si indica con B^c essendo omessa l'indicazione che il complementare è fatto rispetto ad A , in quanto sarà sempre chiara, quando si userà tale simbolo, l'identità di A .

Definiamo inoltre

- $A \cup B = \{a \in A \text{ or } a \in B\}$ (A unione B)
- $A \cap B = \{a \in A \text{ and } a \in B\}$ (A intersezione B)
- $A \times B = \{(a, b) : a \in A \text{ and } b \in B\}$ (prodotto cartesiano).

Si possono provare facilmente proprietà del tipo

- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- $(A \cup B)^c = A^c \cap B^c$
- $(A \cap B)^c = A^c \cup B^c$

Le ultime due uguaglianze sono note come formule di De-Morgan.

Indichiamo con \emptyset l'insieme vuoto, cioè l'insieme privo di elementi

Se P è una proposizione ed A è un insieme possiamo considerare le seguenti proposizioni

- ogni elemento di A soddisfa P ;
- qualche elemento di A soddisfa P ;
- uno ed un solo elemento di A soddisfa P .

Le tre affermazioni di cui sopra si scrivono in simboli

- $\forall x \in A \quad : \quad P_x$

- $\exists x \in A \quad : \quad P_x$
- $\exists! x \in A \quad : \quad P_x$

Osserviamo che le negazioni delle prime due precedenti proposizioni sono

- $\exists x \in A \quad \text{not } P_x$
- $\forall x \in A \quad \text{not } P_x$

CAPITOLO 2

I NUMERI REALI

Introduciamo l'insieme \mathbb{R} dei numeri reali per via assiomatica; elenchiamo cioè le proprietà cui deve soddisfare l'insieme dei numeri reali prescindendo dalla verifica dell'esistenza di un modello di \mathbb{R} e dalla costruzione di tale modello.

A tale proposito ci limitiamo a ricordare che la retta euclidea su cui siano stati fissati due punti (0 ed 1), sia stato definito il verso positivo e siano state definite la somma ed il prodotto per via geometrica, costituisce un buon modello dei numeri reali.

Diciamo che sono assegnati i numeri reali, che indicheremo con \mathbb{R} , se:

- è assegnato un insieme \mathbb{R}
- sono assegnate due leggi, che chiamiamo somma o addizione e prodotto o moltiplicazione e che indichiamo con $+$ e \cdot rispettivamente, ciascuna delle quali associa ad ogni coppia $(x, y) \in \mathbb{R} \times \mathbb{R}$ un elemento di \mathbb{R} che indicheremo con $x+y$ ed $x \cdot y$ rispettivamente (in realtà useremo sempre xy in luogo di $x \cdot y$)
- è assegnata in \mathbb{R} una relazione di equivalenza che indicheremo con il simbolo $=$ (rispetto alla quale esistono in \mathbb{R} almeno due elementi distinti)
- è assegnata in \mathbb{R} una relazione d'ordine che indicheremo con $<$

valgono le seguenti proprietà per ogni $x, y, z \in \mathbb{R}$:

- (1) $x + y = y + x$
(proprietà commutativa dell'addizione)
- (2) $(x + y) + z = x + (y + z)$
(proprietà associativa dell'addizione)
- (3) esiste $\theta \in \mathbb{R}$ tale che $x + \theta = x$, per ogni $x \in \mathbb{R}$
(esistenza di un elemento neutro rispetto all'addizione)
 - *l'elemento neutro rispetto alla somma è unico in \mathbb{R}*
infatti se z, z' sono due elementi neutri rispetto alla somma si ha

$$z = z + z' = z' + z = z'$$

- **sarà indicato d'ora innanzi con 0**

- (4) $xy = yx$
(proprietà commutativa della moltiplicazione)
- (5) $(xy)z = x(yz)$
(proprietà associativa della moltiplicazione)
- (6) esiste $\zeta \in \mathbb{R}$ tale che $x\zeta = x$ per ogni $x \in \mathbb{R}$

(esistenza di un elemento neutro rispetto alla moltiplicazione)

- *l'elemento neutro rispetto al prodotto è unico in \mathbb{R} . Se u, u' sono due elementi neutri rispetto al prodotto si ha*

$$u = uu' = u'$$

- **sarà indicato d'ora innanzi con 1**

(7) $x(y + z) = xy + xz$

(proprietà distributiva della moltiplicazione rispetto all'addizione)

(8) è vera una ed una sola delle seguenti affermazioni

$$x < y \quad , \quad x = y \quad , \quad y < x$$

(legge di tricotomia)

(9) se $x < y$ allora $x + z < y + z$

(invarianza dell'ordine rispetto all'addizione)

(10) se $x < y$ e $0 < z$ allora $xz < yz$

(invarianza dell'ordine rispetto alla moltiplicazione per elementi positivi)

(11) Per ogni $x \in \mathbb{R}$ esiste $x' \in \mathbb{R}$ tale che $x' + x = 0$

(esistenza dell'inverso rispetto all'addizione)

- *per ogni $x \in \mathbb{R}$ l'inverso di x rispetto alla somma è unico, infatti siano x', x'' tali che $x' + x = x'' + x = 0$ allora si ha $x' + x + x' = x'' + x + x'$ e ne segue che $x' = x''$*

- **verrà indicato solitamente con $-x$**

(12) per ogni $x \in \mathbb{R} \setminus \{0\}$ esiste $x'' \in \mathbb{R}$ tale che $x''x = 1$

(esistenza dell'inverso rispetto alla moltiplicazione)

- *per ogni $x \in \mathbb{R} \setminus \{0\}$ l'inverso di x rispetto al prodotto è unico. Siano x', x'' tali che $x'x = x''x = 1$ si ha $x'xx' = x''xx'$ e ne segue $x' = x''$*

- **verrà indicato solitamente con $1/x$ o con x^{-1}**

(13) Per ogni $A, B \subset \mathbb{R}$, $A, B \neq \emptyset$ tali che

$$a \leq b \quad \forall a \in A, \quad \forall b \in B$$

esiste $c \in \mathbb{R}$ tale che

(2.1) $a \leq c \leq b \quad \forall a \in A, \quad \forall b \in B$

(esistenza di un elemento separatore).

Se $x, y \in \mathbb{R}$ scriveremo $x > y$ in luogo di $y < x$ e converremo di usare il simbolo $x \leq y$ se $(x = y)$ or $(x < y)$.

Ricordiamo inoltre che in caso di più operazioni in sequenza, se non vi sono parentesi, per convenzione il prodotto ha priorità sulla somma.

Passiamo ora a provare alcune fondamentali proprietà dei numeri reali.

TEOREMA 2.1. - *proprietà dei numeri reali - Valgono i seguenti fatti:*

- (1) $\forall x, y, z \in \mathbb{R} \quad x + z = y + z \Leftrightarrow x = y$ (legge di cancellazione rispetto alla somma)

- (2) $\forall x, y, z \in \mathbb{R}, z \neq 0, xz = yz \Leftrightarrow x = y$ (*legge di cancellazione rispetto al prodotto*)
- (3) $x0 = 0, \quad \forall x \in \mathbb{R}$
- (4) $-(-x) = x, \quad \forall x \in \mathbb{R}$
- (5) $(-x) + (-y) = -(x + y), \quad \forall x, y \in \mathbb{R}$
- (6) $(-x)y = -xy, \quad \forall x, y \in \mathbb{R}$
- (7) $(-x)(-y) = xy, \quad \forall x, y \in \mathbb{R}$
- (8) $xy = 0$ se e solo se $(x = 0)$ or $(y = 0)$
- (9) $(xy)^{-1} = x^{-1}y^{-1}, \quad \forall x, y \in \mathbb{R} \setminus \{0\}$
- (10) $x > 0$ implica $-x < 0$
- (11) $xx > 0, \quad \forall x \in \mathbb{R} \setminus \{0\}$
- (12) $1 \neq 0$
- (13) $1 > 0$
- (14) $x > 0$ implica $x^{-1} > 0, \quad \forall x \in \mathbb{R}$

DIMOSTRAZIONE.

- (1) Sia z' tale che $z + z' = 0$ allora

$$x = (x + z) + z' = (y + z) + z' = y$$

- (2) Sia z' tale che $zz' = 1$ allora

$$x = (xz)z' = (yz)z' = y$$

- (3) $x0 = x(0 + 0) = x0 + x0$ da cui $x0 = 0$.
- (4) $(-x) + x = 0$ da cui $-(-x) = x$.
- (5) $x + y + (-x) + (-y) = 0$ da cui $-(x + y) = (-x) + (-y)$.
- (6) $(-x)y + xy = (-x + x)y = 0$ onde $(-x)y = -xy$.
- (7) $(-x)(-y) + (-xy) = (-x)(-y) + (-x)y = (-x)(-y + y) = 0$.
- (8) Sia $xy = 0$, se fosse $y \neq 0$ si avrebbe $x = xyy^{-1} = 0$.
- (9) $(xy)x^{-1}y^{-1} = 1$.
- (10) Se $x > 0$ allora $0 = x - x > 0 - x = -x$.
- (11) Se $x > 0$ allora $xx > 0$ mentre se $x < 0$ si ha $xx = (-x)(-x) > 0$.
- (12) Se fosse $1 = 0$ si avrebbe, per ogni $x \in \mathbb{R}, x = x1 = x0 = 0$.
- (13) $1 = 1 \cdot 1 > 0$.
- (14) Se $x > 0$ e $x^{-1} < 0$ allora $1 = xx^{-1} < 0$.

□

Possiamo ora costruire un modello di \mathbb{R} identificando gli elementi di \mathbb{R} con i punti di una retta euclidea su cui è fissato un punto 0 ed un punto 1.

Si dice positivo il verso di percorrenza da 0 ad 1 e si dice altresì positivo un punto (elemento di \mathbb{R}) che sta dalla stessa parte di 1 rispetto a 0 e negativo in caso contrario.

Si definisce somma di due elementi x ed y l'elemento $x + y$ individuato dal secondo estremo del segmento composto affiancando i segmenti di estremi 0 ed x e 0 ed y come si vede in figura 2.

Si definisce il prodotto di due elementi x ed y mediante la costruzione indicata in figura 2.

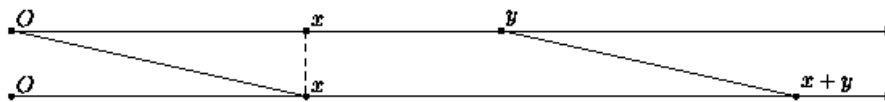


FIGURA 2.1. Costruzione della somma di due numeri reali

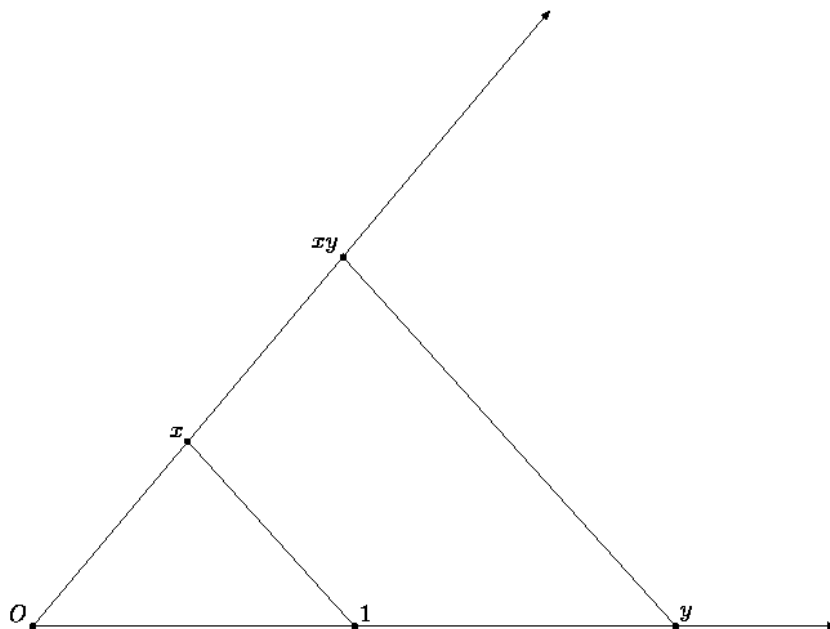


FIGURA 2.2. Costruzione del prodotto di due numeri reali

Con le operazioni di somma e di prodotto e la relazione d'ordine introdotte si può dimostrare che le proprietà richieste sono verificate e permettono di identificare nella retta euclidea un buon modello dei numeri reali.

Occorre ora identificare in \mathbb{R} l'insieme \mathbb{N} dei numeri naturali, l'insieme \mathbb{Z} dei numeri interi e l'insieme \mathbb{Q} dei numeri razionali.

A questo scopo definiamo il concetto di sottoinsieme induttivo in \mathbb{R}

DEFINIZIONE 2.1. Sia $E \subset \mathbb{R}$, diciamo che E è un insieme induttivo se soddisfa le due seguenti proprietà:

- (1) $1 \in E$
- (2) $x \in E \Rightarrow x + 1 \in E$

Osserviamo che esistono certamente insiemi induttivi in quanto, ad esempio, \mathbb{R} stesso è un insieme induttivo; è pure utile osservare che anche

$$\{x \in \mathbb{R} : x \geq 1\}$$

è un insieme induttivo.

DEFINIZIONE 2.2. Sia \mathcal{E} l'insieme degli insiemi induttivi di \mathbb{R} ; definiamo

$$\mathbb{N} = \bigcap_{E \in \mathcal{E}} E$$

La definizione assicura che \mathbb{N} è il più piccolo sottoinsieme induttivo di \mathbb{R} .

TEOREMA 2.2. \mathbb{N} è non vuoto, $1 \in \mathbb{N}$, $1 \leq n \quad \forall n \in \mathbb{N}$ ed inoltre vale la seguente proprietà:

se $A \subset \mathbb{N}$ è un insieme induttivo allora $A = \mathbb{N}$

DIMOSTRAZIONE. Dal momento che 1 appartiene ad ogni sottoinsieme induttivo e dal momento che $\{x \in \mathbb{R} : x \geq 1\}$ è un insieme induttivo si ha che $1 \in \mathbb{N}$ ed inoltre $n \geq 1$ se $n \in \mathbb{N}$.

Per provare la seconda affermazione possiamo osservare che se A è un insieme di \mathbb{N} induttivo, allora evidentemente $A \in \mathcal{E}$ e pertanto $A \supset \mathbb{N}$ onde $A = \mathbb{N}$ \square

L'ultima affermazione del teorema 2.2 è nota come principio di induzione.

Applicando il principio di induzione all'insieme

$$B = A \cup \{n \in \mathbb{N} : n < n_0\}$$

possiamo dedurre il seguente corollario:

COROLLARIO 2.1. Sia $n_0 \in \mathbb{N}$, $n_0 > 1$, e supponiamo che sia $A \subset \mathbb{N}$ tale che

$$(1) \quad n_0 \in A$$

$$(2) \quad n \in A \Rightarrow n + 1 \in A$$

Allora

$$A \supset \{n \in \mathbb{N} : n \geq n_0\}.$$

DEFINIZIONE 2.3. Siano $m, n \in \mathbb{N}$, $m \geq n$, e sia $a_k \in \mathbb{R}$ per ogni $k \in \mathbb{N}$; definiamo

$$\sum_{k=n}^n a_k = a_n \quad , \quad \sum_{k=n}^{m+1} a_k = a_{m+1} + \sum_{k=n}^m a_k \quad .$$

$$\prod_{k=n}^n a_k = a_n \quad \prod_{k=n}^{m+1} a_k = a_{m+1} \cdot \prod_{k=n}^m a_k \quad .$$

DEFINIZIONE 2.4. Chiamiamo insieme dei numeri interi l'insieme

$$\mathbb{Z} = \{a \in \mathbb{R} : a = m - n, m, n \in \mathbb{N}\} ,$$

inoltre diciamo insieme dei numeri razionali l'insieme

$$\mathbb{Q} = \{q \in \mathbb{R} : q = mn^{-1} = m/n, m \in \mathbb{Z}, n \in \mathbb{N}\} .$$

Non entriamo nel dettaglio delle proprietà di \mathbb{Z} e \mathbb{Q} , ricordiamo solo che

$$\mathbb{Z} = (-\mathbb{N}) \cup \{0\} \cup \mathbb{N}$$

DEFINIZIONE 2.5. Sia $A \subset \mathbb{R}$, diciamo che $M \in \mathbb{R}$ è un maggiorante di A se

$$\forall a \in A, a \leq M$$

Diciamo che $m \in \mathbb{R}$ è un minorante di A se

$$\forall a \in A, a \geq m$$

Chiamiamo

$$M(A) = \{M \in \mathbb{R} : \forall a \in A, a \leq M\}$$

$$m(A) = \{m \in \mathbb{R} : \forall a \in A, a \geq m\}$$

In altre parole $M(A)$ è l'insieme dei maggioranti di A , mentre $m(A)$ è l'insieme dei minoranti di A .

Osserviamo che

$$M(\emptyset) = m(\emptyset) = \mathbb{R} .$$

DEFINIZIONE 2.6. Sia $A \subset \mathbb{R}$, diciamo che A è un insieme superiormente limitato se $M(A) \neq \emptyset$.

Diciamo che A è un insieme inferiormente limitato se $m(A) \neq \emptyset$.

Diciamo che A è limitato se A è sia superiormente che inferiormente limitato, cioè se tanto $M(A)$ quanto $m(A)$ sono non vuoti.

DEFINIZIONE 2.7. Diciamo che $A \subset \mathbb{R}$ ammette minimo (massimo) se esiste $m \in A$ ($M \in A$) tale che $m \leq a$ ($M \geq a$) per ogni $a \in A$.
Scriveremo in tal caso

$$m = \min A, \quad M = \max A$$

Osserviamo che non sempre è vero che un insieme di numeri reali ammette minimo o massimo: si consideri ad esempio

$$A = \mathbb{R} \quad \text{oppure} \quad A = \{x \in \mathbb{R} : 0 < x < 1\}$$

Osserviamo anche che

$$\min A = m(A) \cap A, \quad \max A = M(A) \cap A$$

e che tali insiemi, se non sono vuoti, contengono un solo elemento.

TEOREMA 2.3. *Sia $A \subset \mathbb{R}$, $A \neq \emptyset$ se A è inferiormente limitato allora $m(A)$ ammette massimo, mentre se A è superiormente limitato $M(A)$ ammette minimo.*

DIMOSTRAZIONE. Proviamo ad esempio che se A è inferiormente limitato allora $m(A)$ ammette massimo.

Si ha

$$m \leq a \quad \forall m \in m(A), \forall a \in A$$

e pertanto per la 2.1

$$\exists \alpha \in \mathbb{R} \quad \text{tale che} \quad m \leq \alpha \leq a \quad \forall m \in m(A), \forall a \in A$$

Pertanto si può affermare che

$$\alpha = \max m(A)$$

□

DEFINIZIONE 2.8. *Sia $A \subset \mathbb{R}$, $A \neq \emptyset$, A inferiormente limitato; definiamo estremo inferiore di A e lo indichiamo con $\inf A$, il massimo dei minoranti di A , definiamo cioè*

$$\inf A = \max m(A).$$

Analogamente se A è superiormente limitato definiamo estremo superiore di A e lo indichiamo con $\sup A$, il minimo dei maggioranti di A , definiamo cioè

$$\sup A = \min M(A).$$

Definiamo inoltre

- $\inf A = -\infty$, se A non è inferiormente limitato
- $\sup A = +\infty$, se A non è superiormente limitato
- $\inf \emptyset = +\infty$
- $\sup \emptyset = -\infty$

È importante trovare una caratterizzazione dell'estremo inferiore e dell'estremo superiore di un insieme.

TEOREMA 2.4. *Sia $A \subset \mathbb{R}$, $A \neq \emptyset$, A inferiormente limitato; allora $\lambda = \inf A$ se e solo se valgono le seguenti condizioni:*

- (1) $\lambda \leq a \quad \forall a \in A$
- (2) $\forall \varepsilon > 0, \exists a_\varepsilon \in A \text{ tale che } a_\varepsilon < \lambda + \varepsilon$

DIMOSTRAZIONE. $\lambda = \inf A \Leftrightarrow \lambda = \max m(A) \Leftrightarrow \lambda \in m(A)$

e

$\forall \varepsilon > 0, \lambda + \varepsilon \notin m(A).$

D'altra parte

$\lambda \in m(A) \Leftrightarrow \lambda \leq a \quad \forall a \in A \Leftrightarrow (1);$

mentre

$\forall \varepsilon > 0 \quad \lambda + \varepsilon \notin m(A) \Leftrightarrow \forall \varepsilon > 0 \exists a_\varepsilon \in A : a_\varepsilon < \lambda + \varepsilon \Leftrightarrow (2) \quad \square$

In maniera analoga si può dimostrare

TEOREMA 2.5. *Sia $A \subset \mathbb{R}$, $A \neq \emptyset$, A superiormente limitato; allora $\mu = \sup A$ se e solo se valgono le seguenti condizioni:*

(1) $\mu \geq a \quad \forall a \in A$

(2) $\forall \varepsilon > 0 \exists a_\varepsilon \in A : a_\varepsilon > \mu - \varepsilon$

Dal momento che $\inf A = -\infty$ e $\sup A = +\infty$ se, rispettivamente, A non è inferiormente, o superiormente limitato, si ha che

TEOREMA 2.6. *Sia $A \subset \mathbb{R}$, $A \neq \emptyset$ allora*

(1) $\inf A = -\infty \Leftrightarrow \forall k \in \mathbb{R} \exists a_k \in A$ tale che $a_k < k$

(2) $\sup A = +\infty \Leftrightarrow \forall k \in \mathbb{R} \exists a_k \in A$ tale che $a_k > k$

Proviamo a questo punto che l'insieme dei numeri interi non è superiormente limitato; proviamo cioè che

TEOREMA 2.7. - *Principio di Archimede* - $\forall x \in \mathbb{R} \exists n \in \mathbb{Z} : n \geq x$

DIMOSTRAZIONE. Se per assurdo esistesse $y \in \mathbb{R}$ tale che

$$y > n \quad \forall n \in \mathbb{Z}$$

allora $y \in M(\mathbb{Z})$ e

$$\lambda = \sup \mathbb{Z} \in \mathbb{R}$$

Pertanto, da

$$\lambda \geq n \quad \forall n \in \mathbb{Z}$$

possiamo dedurre che

$$\lambda \geq n + 1 \quad \forall n \in \mathbb{Z}$$

e

$$\lambda - 1 \geq n \quad \forall n \in \mathbb{Z}$$

ma ciò contraddice il teorema 2.5

Infatti per $\varepsilon = 1$ non esiste alcun $n \in \mathbb{Z}$ tale che $\lambda - 1 < n$. \square

LEMMA 2.1. *Per ogni $x \in \mathbb{R}$ esiste $n_x \in \mathbb{Z}$ tale che*

$$n_x \leq x < n_x + 1.$$

Inoltre si ha $n_x = \max \{n \in \mathbb{Z} : n \leq x\}$.

DIMOSTRAZIONE. Definiamo

$$A = \{n \in \mathbb{Z} : n \leq x\} .$$

Evidentemente A è superiormente limitato e non vuoto in quanto, per il teorema 2.7, esiste $-n_0 \in \mathbb{Z}$, $-n_0 \geq -x$; si può pertanto affermare che $n_0 \in A$

$$\lambda = \sup A \in \mathbb{R}$$

Se per assurdo si avesse che

$$\forall n \in A \text{ si abbia } n + 1 \in A$$

avremmo allora che $A \supset \{n \in \mathbb{Z} : n \geq n_1\}$ e quindi A non potrebbe risultare limitato.

Quindi è lecito affermare che

$$\text{esiste } n_x \in A, \text{ tale che } n_x + 1 \notin A;$$

da cui, essendo n_x e di conseguenza $n_x + 1$ interi, si ha

$$n_x \leq x < n_x + 1 .$$

Osserviamo inoltre che, se esistesse $n \in A$, $n > n_x$, si avrebbe $n \geq n_x + 1 > x$ ed $n \notin A$. \square

E' pertanto lecito porre:

DEFINIZIONE 2.9. Sia $x \in \mathbb{R}$, definiamo $E(x)$, parte intera di x , come

$$E(x) = \max\{n \in \mathbb{Z} : n \leq x\}$$

Osserviamo che $E(x)$ è il più grande intero più piccolo di x .

DEFINIZIONE 2.10. Sia $x \in \mathbb{R}$, definiamo $|x|$, modulo, o valore assoluto, o norma di x :

$$(2.2) \quad |x| = \begin{cases} x & \text{se } x > 0 \\ 0 & \text{se } x = 0 \\ -x & \text{se } x < 0 \end{cases}$$

TEOREMA 2.8. Sono verificati i seguenti fatti, $\forall a, x, y \in \mathbb{R}$:

- (1) $|x| \geq 0$
- (2) $|x| = 0 \Leftrightarrow x = 0$
- (3) $|xy| = |x| |y|$
- (4) $|x| \leq a \Leftrightarrow -a \leq x \leq a$
- (5) $|x + y| \leq |x| + |y|$

$$(6) ||x| - |y|| \leq |x - y|$$

$$(7) |x| < \varepsilon \quad \forall \varepsilon > 0 \quad \Leftrightarrow \quad x = 0.$$

DIMOSTRAZIONE. (1), (2) e (3) seguono immediatamente dalla definizione di modulo; per quel che riguarda la (4) si noti che

$$|x| \leq a \quad \Leftrightarrow \quad 0 \leq x \leq a \quad \text{oppure} \quad -a \leq x \leq 0.$$

Proviamo ora la (5): per (4) si ha

$$-|x| \leq x \leq |x| \quad \text{e} \quad -|y| \leq y \leq |y|.$$

Pertanto, sommando membro a membro,

$$-(|x| + |y|) \leq x + y \leq |x| + |y|$$

e la tesi, riutilizzando la (4).

Per quel che riguarda (6), si ha

$$|x| = |x - y + y| \leq |x - y| + |y|$$

$$|y| = |y - x + x| \leq |x - y| + |x|;$$

perciò

$$-|x - y| \leq |x| - |y| \leq |x - y|$$

e, da (4), la tesi.

Infine, per quel che riguarda la (7), se fosse $x \neq 0$, si potrebbe scegliere ε tale che $0 < \varepsilon < |x|$. \square

DEFINIZIONE 2.11. Sia $x \in \mathbb{R}$, definiamo per ogni $n \in \mathbb{N}$:

$$x^0 = 1, \quad x^n = xx^{n-1}$$

x^n si dice potenza ennesima di base x .

Definiamo inoltre, se $x \neq 0$,

$$x^{-n} = 1/x^n.$$

TEOREMA 2.9. Siano $x, y \in \mathbb{R}$; per ogni $n, m \in \mathbb{N}$ si ha

$$(2.3) \quad x^{n+m} = x^n x^m$$

$$(2.4) \quad (x^n)^m = x^{nm}$$

$$(2.5) \quad (xy)^n = x^n y^n$$

$$(2.6)$$

Fin qui abbiamo definito cosa intendiamo per numero reale, naturale, intero e razionale ma non abbiamo introdotto un simbolismo adeguato.

Abbiamo fino ad ora identificato un numero utilizzando un simbolo, ma è chiaro che in tal modo possiamo utilizzare contemporaneamente solo pochi numeri dato che, per chiarezza, è necessario servirsi solo di un piccolo numero di segni (simboli o cifre) diversi; è quindi utile introdurre un sistema di rappresentazione che utilizzi solo un numero piccolo di cifre e sia in grado di fornire una adeguata rappresentazione dei numeri, anche molto grandi, che ci interessano.

Tale tipo di rappresentazione fu introdotta in Europa da Leonardo Pisano, detto Fibonacci, cioè figlio di Bonaccio attorno al 1400, ma era impiegata dagli arabi già da molto tempo.

Essa prende il nome di notazione posizionale e si fonda sul seguente semplice fatto

LEMMA 2.2. *Per ogni $a \in \mathbb{N} \cup \{0\}$, e per ogni $b \in \mathbb{N}$, esistono e sono unici $q, r \in \mathbb{N} \cup \{0\}$ tali che*

$$(2.7) \quad a = bq + r, \quad r < b$$

DIMOSTRAZIONE. Posto $q = E(a/b)$ si ha

$$q \leq a/b < q + 1 \quad \text{e} \quad bq \leq a < b(q + 1)$$

Pertanto, posto $r = a - bq$ si ha $r \in \mathbb{N} \cup \{0\}$ e

$$bq + r = a < bq + b \quad \text{da cui} \quad r < b$$

□

1. Rappresentazione dei numeri naturali in base b

Siano $a_0, b \in \mathbb{N}$, $a_0 \geq b > 1$; e definiamo il seguente algoritmo per ogni $n \in \mathbb{N} \cup \{0\}$ indichiamo con a_n e c_n gli unici elementi di $\mathbb{N} \cup \{0\}$ (vedi il lemma 2.2) tali che

$$(2.8) \quad a_n = a_{n+1}b + c_n, \quad c_n < b$$

I numeri a_n così generati soddisfano interessanti proprietà:

- (1) $a_n \neq 0 \Rightarrow a_{n+1} < a_n$ infatti
- Dal momento che $a_{n+1} = E(a_n/b)$ e poiché $b > 1$

$$a_{n+1} \leq a_n/b < a_n$$

- (2) Esiste $n_0 \in \mathbb{N}$ tale che $a_{n_0} \neq 0$ ed $a_{n_0+1} = 0$

- Se $a_n \neq 0$ implicasse $a_{n+1} \neq 0$ avremmo, per il principio di induzione che $a_n \neq 0$ per ogni $n \in \mathbb{N}$ e quindi si avrebbe allora per il (lemma 2.2),

$$a_1 \leq a_0 - 1 \quad \text{essendo} \quad a_1 < a_0$$

ed inoltre si potrebbe affermare che

$$a_n \leq a_0 - n \Rightarrow a_{n+1} < a_n \leq a_0 - n \Rightarrow a_{n+1} \leq a_0 - (n + 1)$$

Ne verrebbe pertanto, per il principio di induzione, che

$$a_n \leq a_0 - n \quad \forall n \in \mathbb{N}$$

e ciò non è possibile in quanto, per $n > a_0$, si avrebbe $a_n < 0$

- (3) Risulta:

$$a_0 = \sum_{k=0}^{n_0} c_k b^k.$$

- Si ha,

$$\begin{aligned}a_0 - a_1b &= c_0 \\a_1 - a_2b &= c_1 \\a_2 - a_3b &= c_2 \\&\dots\dots\dots = \dots\dots\dots \\a_{n_0} &= c_{n_0}\end{aligned}$$

da cui moltiplicando la seconda uguaglianza per b , la terza uguaglianza per b^2 e così via fino a moltiplicare l'ultima per b^{n_0} si ottiene

$$\begin{aligned}a_0 - a_1b &= c_0 \\a_1b - a_2b^2 &= c_1b \\a_2b^2 - a_3b^3 &= c_2b^2 \\&\dots\dots\dots = \dots\dots\dots \\a_{n_0}b^{n_0} &= c_{n_0}b^{n_0}\end{aligned}$$

e sommando membro a membro si ottiene

$$\begin{aligned}(2.9) \quad a_0 - a_1b + a_1b - a_2b^2 + a_2b^2 - a_3b^3 + \dots\dots\dots - a_{n_0}b^{n_0} &= \\&= c_0 + c_1b + c_2b^2 + \dots\dots\dots + c_{n_0}b^{n_0}\end{aligned}$$

e cioè

$$a_0 = \sum_{k=0}^{n_0} c_k b^k$$

È evidente a questo punto che possiamo identificare in maniera univoca il numero a_0 mediante la sequenza dei numeri c_k , che risultano interi positivi o nulli, minori di b .

In altre parole conveniamo di rappresentare in base b il numero a_0 mediante l'allineamento ordinato dei numeri c_k trovati seguendo il procedimento descritto; definiamo cioè

$$r(a_0) = c_{n_0}c_{n_0-1}c_{n_0-2}\dots\dots\dots c_2c_1c_0.$$

Osserviamo esplicitamente che

$$0 \leq c_k < b$$

e che

$$r(a_0) = a_0 \quad \Leftrightarrow \quad a_0 < b$$

Possiamo verificare che, per come è stata costruita

- (1) La rappresentazione in base b di un numero naturale è unica;

(2) Ogni allineamento finito di cifre in base b

$$\alpha_k, \quad 0 \leq \alpha_k < b$$

, $k = 0, 1, \dots, n_0$ con $\alpha_{n_0} \neq 0$, identifica un numero $a \in \mathbb{N}$ mediante la

$$a = \sum_{k=0}^{n_0} \alpha_k b^k$$

Si può pertanto concludere che ogni numero naturale può essere individuato non appena si disponga di b simboli diversi che chiameremo cifre.

Usualmente si adopera per questo scopo un numero di simboli o cifre che è pari al numero delle dita delle mani di un uomo; tali simboli sono:

$$0, 1, 2, 3, 4, 5, 6, 7, 8, 9$$

I primi due sono usati per identificare rispettivamente zero (l'elemento neutro rispetto alla somma) ed uno (l'elemento neutro rispetto al prodotto), mentre i successivi servono ad indicare i numeri naturali da due a nove (secondo la terminologia in uso nella lingua italiana). I numeri da dieci in poi si indicano invece facendo ricorso a più di una cifra.

Naturalmente la scelta della base $b = 10$ non è l'unica possibile né è la sola usata frequentemente.

Oltre alla notazione decimale infatti si fa spesso ricorso alla notazione binaria, che corrisponde alla scelta $b = 2$ e che fa uso delle sole cifre

$$0, 1$$

alla notazione ottale, che corrisponde alla scelta $b = 8$ e usa le cifre

$$0, 1, 2, 3, 4, 5, 6, 7$$

e alla notazione esadecimale che corrisponde alla scelta $b = 16$ (decimale) e che fa uso delle cifre

$$0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F$$

Osservazione. Il ruolo della base $b = 2$ è diventato basilare in seguito allo sviluppo degli elaboratori; infatti la memoria di un elaboratore è in grado di registrare in ogni singola posizione di memoria due stati: attivo e non attivo, vero e falso, 1 e 0. Può pertanto in maniera semplice memorizzare un numero come una sequenza di stati binari.

Le basi $b = 8$ e $b = 16$ sono di conseguenza importanti in quanto $8 = 2^3$ e $16 = 2^4$ e la conversione di base tra numeri binari ottali o esadecimali risulta molto semplice. A titolo di esempio osserviamo che le seguenti rappresentazioni in base 2, 8 e 16 corrispondono al valore decimale 255

$$\begin{array}{ccc} 11 & 111 & 111 \\ 3 & 7 & 7 \end{array} \qquad \begin{array}{cc} 1111 & 1111 \\ F & F \end{array}$$

TABELLA 2.1

È anche utile ricordare che la numerazione in base 2 ha il vantaggio di usare poche cifre e quindi di necessitare di semplicissime tabelline di addizione e di moltiplicazione, mentre ha lo svantaggio di dover usare molte cifre anche per numeri piccoli.

Al contrario la numerazione in base 16 ha tabelline di addizione e di moltiplicazione complicate ma è in grado di rappresentare grandi numeri con poche cifre.

□

Dal momento che si ha

$$\mathbb{Z} = \mathbb{N} \cup \{0\} \cup (-\mathbb{N})$$

possiamo ottenere anche la rappresentazione in base b di ogni numero intero.

Per quanto concerne i numeri reali non interi non sarà in generale possibile identificarli mediante un allineamento finito di cifre, possiamo però provare che ogni numero reale si può approssimare con arbitraria precisione mediante allineamenti finiti di cifre.

2. Approssimazione dei numeri reali in base b

Anche in questo caso possiamo definire un algoritmo che è in grado di generare una successione, che può essere infinita, di cifre mediante la quale ogni numero reale può essere approssimato con arbitraria precisione.

Sia $x \in \mathbb{R}$, $x > 0$ e sia $b \in \mathbb{N}$, $b > 1$ definiamo

$$\begin{aligned} c_0 &= E(x) \\ c_1 &= E((x - c_0)b) \\ c_2 &= E((x - c_0 - \frac{c_1}{b})b^2) \\ \dots &= \dots\dots\dots \\ c_n &= E((x - \sum_{k=0}^{n-1} c_k b^{-k})b^n) \end{aligned}$$

Definiamo inoltre

$$(2.10) \quad x_n = \sum_{k=0}^n c_k b^{-k}$$

x_n si chiama approssimazione in base b di ordine n del numero reale x .

Usualmente si scrive

$$x_n = c_0, c_1 c_2 c_3 \dots c_n$$

oppure

$$x_n = c_0.c_1 c_2 c_3 \dots c_n$$

Se $x \in \mathbb{R}$, $x < 0$ si definisce

$$x_n = -(-x)_n$$

Nel seguito tuttavia faremo sempre riferimento al caso in cui $x > 0$

$$(1) \quad 0 \leq x - x_n < b^{-n}, \quad 0 \leq c_{n+1} < b$$

• Infatti osservando che

$$c_n = E((x - x_{n-1})b^n) \quad \text{e che} \quad x_n = x_{n-1} + c_nb^{-n}$$

si ha

$$c_n \leq (x - x_{n-1})b^n < c_n + 1$$

ed anche

$$c_nb^{-n} \leq x - x_{n-1} < c_nb^{-n} + b^{-n}$$

da cui

$$0 \leq x - x_n < b^{-n}$$

Inoltre moltiplicando la precedente disuguaglianza per b^{n+1} si ottiene

$$0 \leq (x - x_n)b^{n+1} < b$$

e

$$c_{n+1} = E((x - x_n)b^{n+1}) < b$$

(2) Si deduce quindi subito che

$$x \geq x_n$$

Ciò si esprime dicendo che x_n è una approssimazione per difetto del numero reale x . Si vede altresì che l'approssimazione di x può essere fatta con precisione arbitraria pur di aumentarne l'ordine. Infatti

(3) Per ogni $x \in \mathbb{R}$ si ha

$$x = \sup\{x_n : n \in \mathbb{N}\}$$

• Abbiamo già osservato che

$$x \geq x_n$$

D'altro canto si ha

$$x - x_n < 1/b^n$$

e possiamo anche affermare che

Se $b \in \mathbb{N}$, $b > 1$ si ha $b^n > n$, $\forall n \in \mathbb{N}$

- Infatti per $n = 1$ si ha $b > 1$
- inoltre se $b^n > n$ allora $b^{n+1} > n + 1$ in quanto

$$b^{n+1} = b^n b > bn \geq 2n = n + n \geq n + 1.$$

Poichè per ogni $\varepsilon > 0$, esiste $n \in \mathbb{N}$ tale che $\varepsilon > 1/n$ possiamo allora concludere che

$$\varepsilon > 1/n_\varepsilon > 1/b^{n_\varepsilon}.$$

Si possono altresì provare i seguenti risultati.

(1) Per ogni $x \in \mathbb{R}$ e per ogni $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$, esiste $q \in \mathbb{Q}$: $|x - q| < \varepsilon$.

- Scelto $q = x_{n_\varepsilon}$, con $b^{n_\varepsilon} > 1/\varepsilon$, è immediato verificare che $q \in \mathbb{Q}$ e che $|x - q| < \varepsilon$.
 - (2) Per ogni $x, y \in \mathbb{R}$, $x < y$, esiste $q \in \mathbb{Q}$ tale che $x < q < y$
 - (3) Per ogni $x, y \in \mathbb{Q}$, $x < y$, esiste $z \in \mathbb{R} \setminus \mathbb{Q}$ tale che $x < z < y$
- Infatti detto $z = (x + y)/2$ e scelto $q = z_{n_\varepsilon}$ tale che $b^{n_\varepsilon} > 3/(y - x)$, è immediato verificare che $q \in \mathbb{Q}$ e $x < q < y$.

La seconda affermazione segue dall'esistenza di almeno un irrazionale; se infatti α è un numero irrazionale compreso in $(0, 1)$, ad esempio $\alpha = \sqrt{2} - 1$, avremo che

$$(2.11) \quad x + \alpha(y - x)$$

non è razionale (se lo fosse, poichè $x, y \in \mathbb{Q}$ si avrebbe anche $\alpha \in \mathbb{Q}$) ed è compreso in (x, y)

Osserviamo che quindi anche tra due reali $x < y$ esiste un irrazionale, infatti si possono trovare $x' < y'$ con $x < x' < y' < y$

Questo risultato si esprime usualmente dicendo che \mathbb{Q} è **denso in** \mathbb{R} .

In realtà il risultato provato è più preciso in quanto assicura che il sottoinsieme dei numeri razionali che si possono scrivere nella forma usata in 2.10 è denso in \mathbb{R} .

Nel caso in cui $b = 10$ i numeri che si possono scrivere in tale forma si chiamano **numeri decimali finiti**.

Osservazione. È d'uso, lavorando con i numeri reali, adoperare la retta euclidea come modello dei numeri reali.

Infatti, assumendo i postulati della geometria euclidea ed il postulato di continuità di Dedekind, si possono definire sulla retta le operazioni di addizione e moltiplicazione, una relazione di equivalenza ed una d'ordine, in modo che siano verificati gli assiomi che identificano i numeri reali. \square

È inoltre utile costruire una rappresentazione geometrica del prodotto cartesiano $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$.

Ciò può essere ottenuto identificando \mathbb{R}^2 con un piano.

Consideriamo pertanto un piano α e fissiamo su di esso due rette r_1 ed r_2 , dette **assi cartesiani**, che si intersecano nel punto O , detto **origine**.

Usualmente adopereremo le lettere x ed y per indicare i punti di r_1 ed r_2 rispettivamente; per tale ragione diremo che r_1 è l'asse x e che r_2 è l'asse y .

Ognuna delle rette può essere interpretata come \mathbb{R} ed è chiaro che procedendo come in figura 2 si può identificare ogni coppia di numeri reali con un punto del piano α e viceversa.

Qualora r_1 ed r_2 siano tali che ruotando r_1 in senso antiorario, di un angolo inferiore ad un angolo piatto, fino a sovrapporla ad r_2 , i punti che rappresentano le unità sulle due rette stanno dalla stessa parte rispetto al punto O , la rappresentazione si chiama **destrorsa**; in caso contrario si dice **sinistrorsa**.

Qualora le rette r_1 ed r_2 siano perpendicolari, la rappresentazione si chiama **ortogonale**.

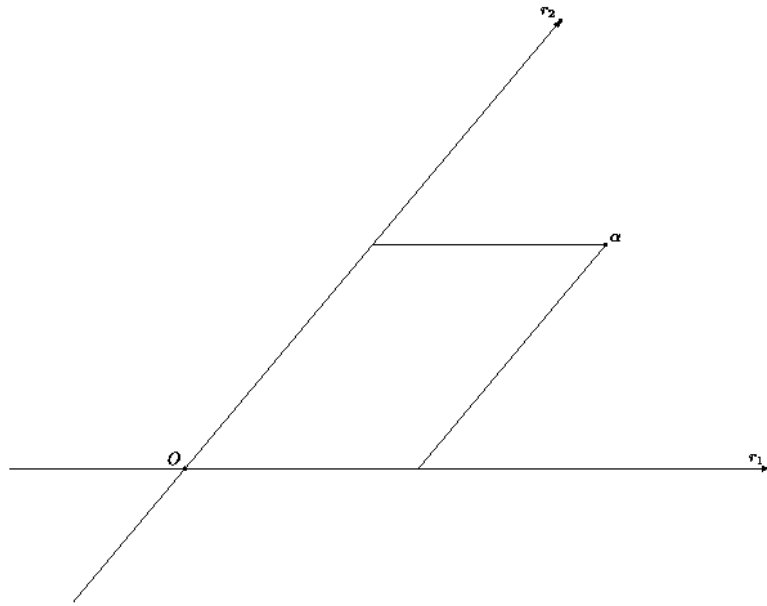


FIGURA 2.3. Sistema di riferimento Cartesiano

Qualora si scelgano in r_1 ed r_2 segmenti unitari uguali, la rappresentazione si dice **monometrica**.

Usualmente adopereremo una rappresentazione *destrorsa, ortogonale e monometrica*, che chiamiamo **sistema cartesiano**.

Per concludere ricordiamo alcune notazioni:

DEFINIZIONE 2.12. Siano $a, b \in \mathbb{R}$, definiamo

$$(a, b) = \{x : x \in \mathbb{R}, a < x < b\}$$

$$[a, b] = \{x : x \in \mathbb{R}, a \leq x \leq b\}$$

$$[a, b) = \{x : x \in \mathbb{R}, a \leq x < b\}$$

$$(a, b] = \{x : x \in \mathbb{R}, a < x \leq b\}$$

$$(a, +\infty) = \{x : x \in \mathbb{R}, x > a\}$$

$$a, +\infty) = \{x : x \in \mathbb{R}, x \geq a\}$$

$$(-\infty, a) = \{x : x \in \mathbb{R}, x < a\}$$

$$(-\infty, a] = \{x : x \in \mathbb{R}, x \leq a\}$$

Definiamo inoltre

$$\mathbb{R}_+ = \{x : x \in \mathbb{R}, x > 0\}$$

$$\overline{\mathbb{R}}_+ = \{x : x \in \mathbb{R}, x \geq 0\}$$

$$\mathbb{R}_- = \{x : x \in \mathbb{R}, x < 0\}$$

$$\overline{\mathbb{R}}_- = \{x : x \in \mathbb{R}, x \leq 0\}$$

DEFINIZIONE 2.13. Sia $A \subset \mathbb{R}$, diciamo che A è aperto se

$$\forall x \in A \ \exists r > 0 \quad \text{tale che} \quad (x - r, x + r) \subset A$$

Diciamo che A è chiuso se A^c è aperto.

Diciamo che A è compatto se è chiuso e limitato.

CAPITOLO 3

FUNZIONI REALI DI UNA VARIABILE REALE

Il concetto di funzione è di fondamentale importanza;

DEFINIZIONE 3.1. Diciamo che è data una funzione reale di una variabile reale se sono assegnati un sottoinsieme $D \subset \mathbb{R}$ ed una corrispondenza f che ad ogni elemento $x \in D$ associa uno ed un solo elemento $y \in \mathbb{R}$.

DEFINIZIONE 3.2. Chiamiamo D dominio della funzione e denotiamo con $f(x)$ (si legge f di x) il corrispondente di x secondo la legge assegnata f ; spesso useremo il termine valore di f in x oppure f calcolata in x per indicare $f(x)$ e chiamiamo x argomento di $f(x)$; per indicare la corrispondenza f scriviamo anche $x \mapsto f(x)$ oppure $x \mapsto y = f(x)$.

Chiamiamo rango di f l'insieme

$$R(f) = \{y \in \mathbb{R} : \exists x \in D, y = f(x)\}.$$

Per indicare una funzione scriviamo $f : D \longrightarrow \mathbb{R}$, specificando prima della freccia il dominio D di f , ma non curandoci di precisare, dopo la freccia, il suo rango.

Osservazione. Distinguiamo fin d'ora due notazioni che saranno usate con significati completamente diversi. Useremo f per indicare la legge di corrispondenza di una funzione ed $f(x)$ per indicare il valore di f in x , (quindi $f(x)$ è un numero reale). \square

Spesso, nell'assegnare una funzione, daremo soltanto la legge di corrispondenza f ; in tal caso sottointendiamo sempre che D è il più grande sottoinsieme di \mathbb{R} i cui elementi possono essere usati come argomenti di f .

Ad ogni funzione è possibile associare un sottoinsieme del prodotto cartesiano $\mathbb{R} \times \mathbb{R}$ (che indicheremo spesso con \mathbb{R}^2) che la caratterizza in maniera completa e dalla quale è completamente caratterizzato.

DEFINIZIONE 3.3. Sia $f : D \longrightarrow \mathbb{R}$, definiamo grafico di f

$$G(f) = \{(x, y) \in \mathbb{R}^2 : x \in D, y = f(x)\}$$

Mediante la definizione 3.2 è possibile associare in maniera univoca un sottoinsieme $G (= G(f))$ ad ogni funzione $f : D \longrightarrow \mathbb{R}$.

Non è altrettanto vero che ad ogni sottoinsieme $G \subset \mathbb{R}^2$ è possibile associare una funzione $f : D \longrightarrow \mathbb{R}$. Ciò accade solo nel caso in cui G soddisfi una particolare proprietà.

DEFINIZIONE 3.4. Diciamo che $G \subset \mathbb{R}^2$ soddisfa la proprietà (g) se

$$(g) \quad (x, y_1), (x, y_2) \in G \Rightarrow y_1 = y_2$$

TEOREMA 3.1. *Per ogni $G \subset \mathbb{R}^2$, G soddisfacente la proprietà (g), esistono unici $D \subset \mathbb{R}$ ed $f : D \rightarrow \mathbb{R}$ tali che $G = \text{gph} f$.*

DIMOSTRAZIONE. Definiamo $D = \{x \in \mathbb{R} : \exists y \in \mathbb{R}, (x, y) \in G\}$ e sia $f(x) = y$ dove y è l'unico elemento di \mathbb{R} tale che $(x, y) \in G$.

Dal momento che G soddisfa la proprietà (g), D ed f verificano le proprietà richieste. \square

DEFINIZIONE 3.5. *Sia $f : D \rightarrow \mathbb{R}$, sia $A \subset D$, definiamo restrizione di f ad A la funzione $g : A \rightarrow \mathbb{R}$ tale che $g(x) = f(x) \forall x \in A$.*

Indichiamo con $f|_A$ la restrizione di f ad A .

Siano invece $B \supset D$ e $g : B \rightarrow \mathbb{R}$; diciamo che g è un prolungamento di f a B se $g|_D = f$.

DEFINIZIONE 3.6. *Sia $f : D \rightarrow \mathbb{R}$, diciamo che*

(1) *f è iniettiva se*

$$\forall x_1, x_2 \in D \quad f(x_1) = f(x_2) \Rightarrow x_1 = x_2$$

In altre parole f è iniettiva se ogni retta parallela all'asse delle x interseca $\text{gph} f$ in un solo punto.

(2) *Sia $A \subset \mathbb{R}$, diciamo che f è surgettiva su A se $R(f) = A$, cioè se*

$$\forall y \in A \quad \exists x \in D \quad : \quad y = f(x).$$

Per esprimere che f è surgettiva su A diremo anche che $f : D \rightarrow A$ è surgettiva. (Si osservi che in questo caso abbiamo specificato dopo la freccia il rango di f).

Sia $f : D \rightarrow A$, diciamo che f è bigettiva se è iniettiva e surgettiva.

Osservazione. Ogni funzione è surgettiva sul suo rango. \square

DEFINIZIONE 3.7. *Siano $f, g : D \rightarrow \mathbb{R}$, definiamo*

(1) *$f + g : D \rightarrow \mathbb{R}$ come $(f + g)(x) = f(x) + g(x) \quad \forall x \in D$*

(2) *$f \cdot g : D \rightarrow \mathbb{R}$ come $(f \cdot g)(x) = f(x) \cdot g(x) \quad \forall x \in D$*

(3) *$\frac{1}{g} : D_1 \rightarrow \mathbb{R}$ come $\frac{1}{g}(x) = \frac{1}{g(x)} \quad \forall x \in D_1 = \{x \in D : g(x) \neq 0\}$*

DEFINIZIONE 3.8. *Siano $f : D \rightarrow \mathbb{R}$ e $g : B \rightarrow \mathbb{R}$ e supponiamo che $R(g) \subset D$.*

Definiamo funzione composta di g ed f la funzione che ad ogni $x \in B$ associa $f(g(x)) \in \mathbb{R}$.

DEFINIZIONE 3.9. *Diciamo che $f : D \rightarrow A$ è invertibile se esiste $g : A \rightarrow D$ tale che*

$$(3.1) \quad f(g(y)) = y \quad \forall y \in A$$

$$(3.2) \quad g(f(x)) = x \quad \forall x \in D$$

Per indicare g usiamo il simbolo f^{-1} cosicché

$$f^{-1} : A \rightarrow D$$

è l'inversa di f .

TEOREMA 3.2. *Sia $f : D \longrightarrow A$, f è invertibile se e solo se f è bigettiva.*

DIMOSTRAZIONE. Supponiamo f invertibile; allora

$$\forall y \in A \quad f(f^{-1}(y)) = y$$

e ciò prova che f è surgettiva su A .

Siano poi $x_1, x_2 \in D$ e sia $f(x_1) = f(x_2)$, allora

$$x_1 = f^{-1}(f(x_1)) = f^{-1}(f(x_2)) = x_2$$

e ciò prova l'iniettività di f .

Supponiamo viceversa che f sia bigettiva e definiamo, per ogni $y \in A$, $f^{-1}(y) = x$ dove x è l'unico elemento di D tale che $f(x) = y$.

In altre parole

$$f^{-1}(y) = x \quad \Leftrightarrow \quad y = f(x) \quad .$$

Si ha allora

$$f^{-1}(f(x)) = f^{-1}(y) = x \quad \forall x \in D$$

$$f(f^{-1}(y)) = f(x) = y \quad \forall y \in A.$$

□

DEFINIZIONE 3.10. *Sia $f : D \longrightarrow \mathbb{R}$ e supponiamo che D sia simmetrico rispetto all'origine (cioè $x \in D \Rightarrow -x \in D$); diciamo che f è una funzione pari se*

$$f(x) = f(-x) \quad \forall x \in D;$$

diciamo che f è una funzione dispari se

$$f(x) = -f(-x) \quad \forall x \in D.$$

DEFINIZIONE 3.11. *Sia $f : D \longrightarrow \mathbb{R}$, diciamo che f è crescente (strettamente crescente) se*

$$\forall x, y \in D, \quad x < y \quad \Rightarrow \quad f(x) \leq f(y) \quad (f(x) < f(y)) \quad .$$

Diciamo che f è decrescente (strettamente decrescente) se

$$\forall x, y \in D, \quad x < y \quad \Rightarrow \quad f(x) \geq f(y) \quad (f(x) > f(y))$$

DEFINIZIONE 3.12. *Sia $f : D \longrightarrow \mathbb{R}$, diciamo che f è monotona (strettamente monotona) se f è crescente oppure decrescente (strettamente crescente oppure strettamente decrescente).*

TEOREMA 3.3. *Sia $f : D \longrightarrow \mathbb{R}$, f è monotona (strettamente monotona) se e solo se*

$$\forall x, y, z \in D, \quad x < y < z \quad \Rightarrow \quad [f(y) - f(x)][f(z) - f(y)] \geq 0 \quad (> 0).$$

DIMOSTRAZIONE. La sufficienza è ovvia, proviamo la necessità.

Siano $a, b, x_1, x_2 \in D$, $a < b \leq x_1 \leq x_2$,

$$[f(a) - f(b)][f(b) - f(x_1)][f(b) - f(x_1)][f(x_1) - f(x_2)] \geq 0;$$

perciò $f(x_1) - f(x_2)$ ha lo stesso segno di $f(a) - f(b)$ ed f è monotona in $D \cap [b, +\infty)$.

In maniera analoga si prova che f è monotona su $D \cap (-\infty, b]$ e quindi su D in quanto se $a < b < c$

$$[f(a) - f(b)][f(b) - f(c)] \geq 0$$

Nel caso in cui f sia strettamente monotona tutte le disuguaglianze sono da intendersi in senso stretto. \square

TEOREMA 3.4. Sia $f : D \longrightarrow A$ surgettiva e strettamente monotona, allora f è invertibile ed $f^{-1} : A \longrightarrow D$ è strettamente monotona.

Più precisamente se f è strettamente crescente (decrecente), f^{-1} è strettamente crescente (decrecente).

DIMOSTRAZIONE. Supponiamo che f sia strettamente crescente e vediamo che f è iniettiva.

Siano $x, y \in D$ tali che $f(x) = f(y)$; se si avesse, per assurdo, $x < y$, si potrebbe dedurre che $f(x) < f(y)$ e ciò è in contrasto con l'ipotesi che $f(x) = f(y)$. Pertanto $x = y$.

Si ha quindi che f è invertibile e $f^{-1}(y) = x$ se e solo se $y = f(x)$ per ogni $x \in D$ e per ogni $y \in A$.

Siano ora $y_1, y_2 \in A$, $y_1 < y_2$ e siano $x_1, x_2 \in D$ in modo che

$$y_1 = f(x_1) \quad \text{e} \quad y_2 = f(x_2)$$

se fosse $x_1 \geq x_2$ si avrebbe $f(x_1) \geq f(x_2)$ e $y_1 \geq y_2$ e ciò è assurdo.

Se ne deduce che $x_1 < x_2$ e la stretta crescita di f^{-1} . \square

DEFINIZIONE 3.13. Sia $f : D \longrightarrow \mathbb{R}$, diciamo che f è superiormente limitata se esiste $M \in \mathbb{R}$ tale che

$$f(x) \leq M \quad \forall x \in D$$

diciamo che f è inferiormente limitata se esiste $m \in \mathbb{R}$ tale che

$$f(x) \geq m \quad \forall x \in D$$

diciamo che f è limitata se è sia superiormente che inferiormente limitata.

Osservazione. f è limitata (superiormente) [inferiormente] se e solo se $R(f)$ è limitato (superiormente) [inferiormente]. \square

DEFINIZIONE 3.14. Sia $f : D \longrightarrow \mathbb{R}$, diciamo che $x_0 \in D$ è un punto di minimo assoluto per f se

$$f(x_0) \leq f(x) \quad \forall x \in D$$

diciamo che $x_0 \in D$ è un punto di massimo assoluto per f se

$$f(x_0) \geq f(x) \quad \forall x \in D$$

TEOREMA 3.5. *Sia $f : D \rightarrow \mathbb{R}$, affinché $x_0 \in D$ sia un punto di minimo (massimo) assoluto per f è sufficiente che f sia decrescente (crescente) in $(-\infty, x_0] \cap D$ e crescente (decrescente) in $[x_0, +\infty) \cap D$.*

TEOREMA 3.6. *Siano $f : D \rightarrow \mathbb{R}$, $g : A \rightarrow \mathbb{R}$ tali che $R(g) \subset D$, allora*

- *f strettamente crescente, g strettamente crescente $\Rightarrow f(g(\cdot))$ strettamente crescente;*
- *f strettamente crescente, g strettamente decrescente $\Rightarrow f(g(\cdot))$ strettamente decrescente;*
- *f strettamente decrescente, g strettamente crescente $\Rightarrow f(g(\cdot))$ strettamente decrescente;*
- *f strettamente decrescente, g strettamente decrescente $\Rightarrow f(g(\cdot))$ strettamente crescente.*

Inoltre, le stesse asserzioni valgono abolendo ovunque la parola strettamente.

CAPITOLO 4

LE FUNZIONI ELEMENTARI

Per costruire modelli che coinvolgono funzioni occorre avere un certo numero di funzioni, che chiameremo elementari, di cui sono note le proprietà.

Usando tali funzioni si possono costruire la maggior parte delle funzioni necessarie per l'impostazione di modelli matematici.

È pertanto molto importante una buona conoscenza e della definizione delle funzioni elementari e delle loro principali proprietà.

Naturalmente la classe delle funzioni elementari, sebbene codificata e delimitata dalla letteratura e dalla tradizione matematica è in qualche modo aperta a nuovi ingressi che si rendano di uso frequente in applicazioni future.

1. Le funzioni Potenze

Cominciamo con il definire cosa si intende per potenza di esponente naturale;

DEFINIZIONE 4.1. Sia $n \in \mathbb{N}$, definiamo la funzione $p_n : \mathbb{R} \longrightarrow \mathbb{R}$ mediante la

$$p_n(x) = x^n;$$

p_n si dice potenza di esponente n e di base x .

TEOREMA 4.1. Sia $n \in \mathbb{N}$, p_n è strettamente crescente in $\overline{\mathbb{R}}_+$.

DIMOSTRAZIONE. Procediamo per induzione; è innanzi tutto ovvio che p_1 è strettamente crescente in $\overline{\mathbb{R}}_+$ ed inoltre se supponiamo p_n crescente in $\overline{\mathbb{R}}_+$, presi $x > y \geq 0$ si ha

$$p_{n+1}(x) = xp_n(x) > xp_n(y) \geq yp_n(y) = p_{n+1}(y)$$

e p_{n+1} è strettamente crescente su $\overline{\mathbb{R}}_+$ □

TEOREMA 4.2. Sia $n \in \mathbb{N}$, n pari, allora

- (1) $p_n(x) = p_n(-x)$ per ogni $x \in \mathbb{R}$
- (2) p_n è strettamente decrescente su $\overline{\mathbb{R}}_-$
- (3) $R(p_n) = \overline{\mathbb{R}}_+$

DIMOSTRAZIONE.

- (1) Poichè n è pari si ha $n = 2k$, $k \in \mathbb{N}$ e

$$x^n = x^{2k} = (x^2)^k = ((-x)^2)^k = (-x)^{2k} = (-x)^n$$

(2) Siano $x < y \leq 0$, allora $-x > -y \geq 0$ e dal teorema 4.2

$$p_n(x) = p_n(-x) > p_n(-y) = p_n(y)$$

(3) È evidente che $R(p_n) \subset \overline{\mathbb{R}}_+$ in quanto $x^n = x^{2k} \cdot x^2 \geq 0$, l'inclusione opposta dipende dal fatto che p_n è una funzione continua e dal teorema dei valori intermedi. Proveremo tale inclusione a suo tempo.

□

TEOREMA 4.3. *Sia $n \in \mathbb{N}$, n dispari, allora*

- (1) $p_n(x) = -p_n(-x)$ per ogni $x \in \mathbb{R}$
- (2) p_n è strettamente crescente su $\overline{\mathbb{R}}_-$
- (3) $R(p_n) = \mathbb{R}$.

DIMOSTRAZIONE.

(1) Poiché n è dispari si ha $n = 2k + 1$ e

$$p_n(x) = x^{2k+1} = x x^{2k} = -(-x)(-x)^{2k} = -(-x)^n = -p_n(-x)$$

(2) Siano $x < y \leq 0$, allora $-x > -y \geq 0$ e, per il teorema 4.2, $-p_n(x) = p_n(-x) > p_n(-y) = -p_n(y)$.

(3) Anche in questo caso rimandiamo la dimostrazione al seguito.

□

Abbiamo visto che, se $n \in \mathbb{N}$, n pari, allora $p_n : \overline{\mathbb{R}}_+ \longrightarrow \overline{\mathbb{R}}_+$ è strettamente crescente e surgettiva; pertanto p_n è invertibile ed è lecito definire

$$r_n : \overline{\mathbb{R}}_+ \longrightarrow \overline{\mathbb{R}}_+ \quad \text{come} \quad r_n = (p_n)^{-1}.$$

Se $x \in \overline{\mathbb{R}}_+$, $r_n(x)$ si dice radice n -esima di x .

Sia $n \in \mathbb{N}$, n dispari, abbiamo già visto che $p_n : \mathbb{R} \longrightarrow \mathbb{R}$ è strettamente crescente e surgettiva; pertanto p_n è invertibile ed è lecito definire

$$r_n : \mathbb{R} \longrightarrow \mathbb{R} \quad \text{come} \quad r_n = (p_n)^{-1}$$

Se $x \in \mathbb{R}$, $r_n(x)$ si dice radice n -esima di x .

TEOREMA 4.4. *Sia $n \in \mathbb{N}$*

- (1) *se n è pari, $r_n : \overline{\mathbb{R}}_+ \longrightarrow \overline{\mathbb{R}}_+$ è strettamente crescente;*
- (2) *se n è dispari, $r_n : \mathbb{R} \longrightarrow \mathbb{R}$ è strettamente crescente.*

Definiamo ancora le potenze ad esponente negativo.

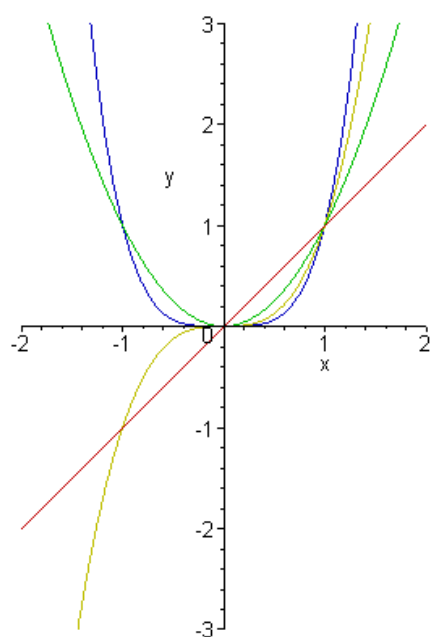


FIGURA 4.1. Potenze ad esponente naturale

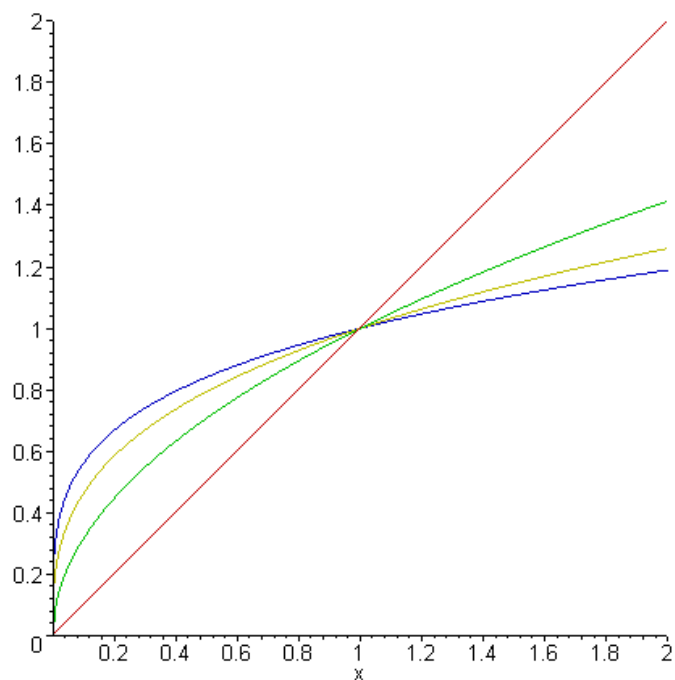


FIGURA 4.2. Radici ad esponente naturale

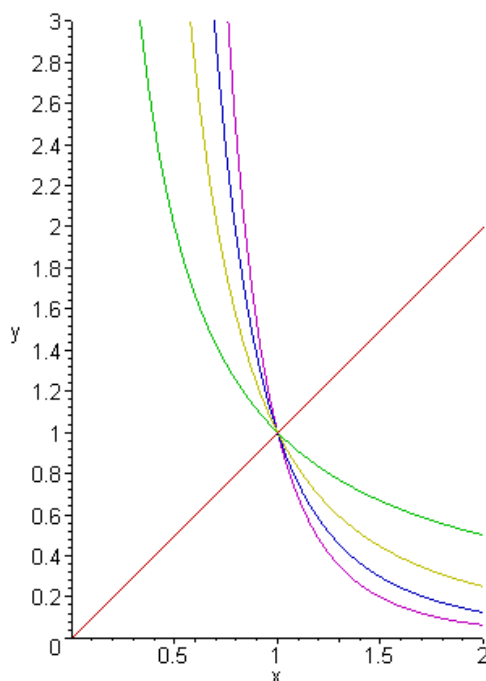


FIGURA 4.3. Potenze ad esponente negativo

DEFINIZIONE 4.2. Sia $n \in \mathbb{N}$, $p_{-n} : \mathbb{R} \setminus \{0\} \longrightarrow \mathbb{R}$ è definita come

$$p_{-n}(x) = 1/p_n(x)$$

inoltre, $p_0 : \mathbb{R} \longrightarrow \mathbb{R}$ è definita da

$$p_0(x) = 1 \quad \forall x \in \mathbb{R}$$

Per studiare le proprietà di crescenza, decrescenza e invertibilità di p_{-n} sarà sufficiente fare ricorso al seguente risultato.

TEOREMA 4.5. Sia $f : D \longrightarrow \mathbb{R}$ tale che $f(x) > 0$ per ogni $x \in D$; allora

- (1) f è (strettamente) crescente $\Leftrightarrow 1/f$ è (strettamente) decrescente;
- (2) f è (strettamente) decrescente $\Leftrightarrow 1/f$ è (strettamente) crescente.

DEFINIZIONE 4.3. Sia $s \in \mathbb{Q}$, $s = m/n$, $m \in \mathbb{Z}$, $n \in \mathbb{N}$; definiamo la funzione $f_s : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ mediante la

$$f_s(x) = p_m(r_n(x)) = r_n(p_m(x))$$

Se $x \in \mathbb{R}_+$, $f_s(x)$ si dice potenza ad esponente frazionario di esponente s di base x .

Osservazione. Se $x \in \mathbb{R}_+$ la definizione 4.3 è indipendente dalla rappresentazione di s in forma frazionaria e dall'ordine in cui viene fatta la composizione tra potenza e radice.

Se invece $x \in \mathbb{R}_-$ può accadere che $r_n(p_m(x))$ sia definito anche quando $p_m(r_n(x))$ non lo è.

Inoltre, se m/n , m'/n' sono due diverse rappresentazioni frazionarie dello stesso numero razionale s , può accadere che $r_n(p_m(x))$ sia definito mentre $r_{n'}(p_{m'}(x))$ non lo è.

(Si consideri ad esempio $m = 2$ ed $n = 4$; allora se $x < 0$ si ha che $r_4(p_2(x))$ è definito mentre $p_2(r_4(x))$ no.

Inoltre se $m' = 1$ e $n' = 2$ $r_4(p_2(x))$ è definito mentre $r_2(p_1(x))$ no.)

□

Pertanto per valori di $x \in \mathbb{R}_-$ consideriamo la composizione di potenze e radici ove essa ha senso, ma non definiamo in alcun modo la funzione potenza ad esponente razionale.

Ribadiamo ancora che ciò è dovuto al fatto che non è agevole, e talvolta non è possibile, definire in modo univoco la potenza ad esponente razionale in \mathbb{R}_- .

Ciò non significa comunque rinunciare a considerare la composizione di una potenza di esponente n e di una radice di indice m , qualora essa abbia senso, anche per valori dell'argomento negativi.

Ad esempio è chiaro che $p_2(r_3(-2))$ risulta perfettamente ed univocamente individuato.

In casi simili tuttavia, pur trattando la funzione composta

$$p_m(r_n(\cdot))$$

non parleremo di potenza ad esponente razionale $\frac{n}{m}$ e non pretenderemo di applicare a $p_m(r_n(\cdot))$ le proprietà delle potenze in quanto, come visto, potrebbero risultare false.

TEOREMA 4.6. Sia $s \in \mathbb{Q}$, $s > 0$, e sia $f_s : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$, allora f_s è strettamente crescente.

DIMOSTRAZIONE. Infatti se $s = \frac{m}{n} > 0$ allora si ha $m, n \in \mathbb{N}$ e quindi

$$x^s = p_m(r_n(x)) = r_n(p_m(x))$$

è la composizione di due funzioni strettamente crescenti. □

E' inoltre possibile dimostrare che le usuali regole di calcolo delle potenze naturali continuano a valere anche per le potenze ad esponente razionale. In altre parole si può provare che

Se $s, r \in \mathbb{Q}$, e se $x, y \in \mathbb{R}_+$, allora si ha:

$$(1) \quad x^{s+r} = x^s x^r$$

$$(2) \quad (x^s)^r = x^{sr}$$

$$(3) \quad (xy)^s = x^s y^s$$

Si dimostra altresì che

TEOREMA 4.7. Se $x > 1$ la funzione $\mathbb{Q} \ni r \mapsto x^r$ è crescente su \mathbb{Q} . In altre parole per $s, r \in \mathbb{Q}$, $s < r$; si ha

$$x^s < x^r$$

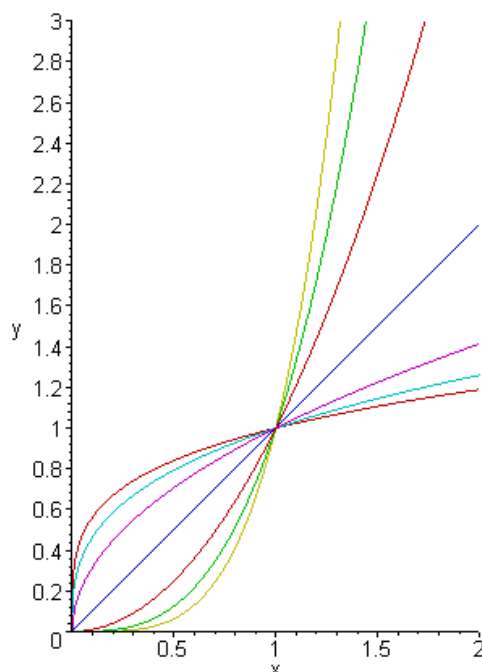


FIGURA 4.4. Potenze ad esponente positivo

DIMOSTRAZIONE. Si ha

$$x^r - x^s = x^s(x^{r-s} - 1) > 0$$

in quanto

$$x^s > 0 \quad \text{e} \quad x^{r-s} > 1$$

dal momento che essendo $r - s > 0$ la funzione $x \mapsto x^{r-s}$ è crescente ed $x > 1$. \square

Per poter definire la potenza anche per esponenti reali è necessario ricordare che piccole variazioni dell'esponente razionale corrispondono a piccole variazioni della potenza; più precisamente possiamo affermare che:

TEOREMA 4.8. *Sia $x > 1$ e sia $r \in \mathbb{Q}$, allora per ogni $\varepsilon > 0$ esiste $n_\varepsilon \in \mathbb{N}$ tale che*

$$0 < x^r - x^{r-1/n_\varepsilon} < \varepsilon$$

Sfruttando questa proprietà è facile capire come sia naturale definire x^a per ogni $x > 1$.

Nel caso in cui sia invece $0 < x < 1$ potremo considerare $\frac{1}{x} > 1$, calcolare $\left(\frac{1}{x}\right)^a$ e definire

$$x^a = \frac{1}{x^{-a}} = \left(\frac{1}{x}\right)^{-a}$$

Se $x = 1$, infine definiamo $1^a = 1$ per ogni a .

DEFINIZIONE 4.4. Sia $x > 1$ e sia $a \in \mathbb{R}$; definiamo

$$x^a = \sup\{x^r : r \in \mathbb{Q}, r \leq a\} = \sup\{x^r : r \in \mathbb{Q}, r < a\}$$

La precedente definizione afferma implicitamente che i due estremi superiori che vi figurano sono reali ed uguali; possiamo infatti verificare che

Siano $A = \{x^r : r \in \mathbb{Q}, r \leq a\}$ e $B = \{x^r : r \in \mathbb{Q}, r < a\}$ allora

$$\sup A = \sup B \in \mathbb{R}$$

DEFINIZIONE 4.5. Se $x = 1$ definiamo $1^a = 1$ per ogni $a \in \mathbb{R}$. Se $0 < x < 1$ definiamo $x^a = (1/x)^{-a}$.

A partire dalla definizione data possiamo verificare che la quantità x^a per $x > 0$ ed $a \in \mathbb{R}$ soddisfa le proprietà che siamo abituati ad usare quando maneggiamo potenze.

se $x, y > 0$ e se $a, b \in \mathbb{R}$, allora

- (1) $x^a > 0$
- (2) $x^{a+b} = x^a x^b$
- (3) $(x^a)^b = x^{ab}$
- (4) $(xy)^a = x^a y^a$
- (5) $1/(x^a) = x^{-a}$

DEFINIZIONE 4.6. Sia $a \in \mathbb{R}$, definiamo la funzione $p_a : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ mediante la

$$x \mapsto p_a(x) = x^a$$

potenza di esponente a

TEOREMA 4.9. Sia $a \in \mathbb{R}$, valgono i seguenti fatti:

- (1) Se $a > 0$ allora p_a è strettamente crescente
- (2) Se $a < 0$ allora p_a è strettamente decrescente
- (3) Se $a \neq 0$ allora $R(p_a) = \mathbb{R}_+$

DIMOSTRAZIONE.

- (1) Sia $0 < x < y$, occorre provare che $x^a < y^a$, cioè che $x^a - y^a < 0$; si ha

$$x^a - y^a = x^a(1 - (y/x)^a) \quad \text{e} \quad (y/x) > 1$$

pertanto $(y/x)^r > 1$ per ogni $r \in \mathbb{Q}, 0 < r \leq a$ e $(y/x)^a > 1$.

- (2) È immediata conseguenza di (1)
- (3) Il fatto che $R(p_a) \subset \mathbb{R}_+$ segue dalla definizione di potenza, mentre la dimostrazione dell'inclusione opposta si ottiene mediante il teorema dei valori intermedi.

□

TEOREMA 4.10. Sia $a \in \mathbb{R}, a \neq 0$, allora p_a è invertibile su \mathbb{R}_+ e $p_a^{-1} = p_{1/a}$.

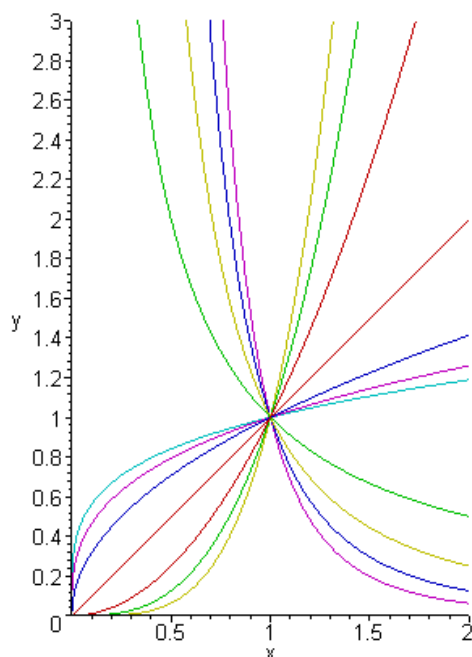


FIGURA 4.5. Potenze ad esponente reale

DIMOSTRAZIONE. Sarà sufficiente dimostrare che

$$p_a(p_{1/a}(y)) = y \quad \forall y \in \mathbb{R}_+$$

e

$$p_{1/a}(p_a(x)) = x \quad \forall x \in \mathbb{R}_+.$$

Si ha infatti

$$p_a(p_{1/a}(y)) = (y^{1/a})^a = y^{a/a} = y$$

e

$$p_{1/a}(p_a(x)) = (x^a)^{1/a} = x^{a/a} = x.$$

□

2. La funzione esponenziale

DEFINIZIONE 4.7. Sia $a \in \mathbb{R}_+$, definiamo la funzione $\exp_a : \mathbb{R} \longrightarrow \mathbb{R}_+$ mediante la

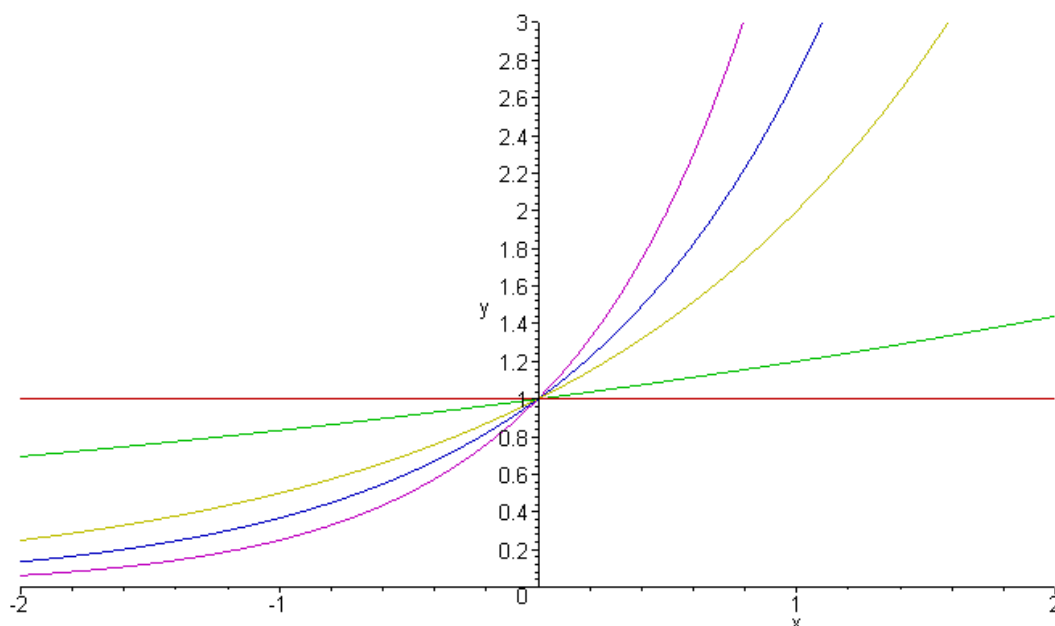
$$x \mapsto \exp_a(x) = a^x$$

esponenziale di base a

TEOREMA 4.11. Sia $a > 0$, valgono i seguenti fatti

- (1) Se $a > 1$ allora \exp_a è strettamente crescente
- (2) Se $0 < a < 1$ allora \exp_a è strettamente decrescente
- (3) Se $a \neq 1$ allora $R(\exp_a) = \mathbb{R}_+$.

DIMOSTRAZIONE.

FIGURA 4.6. Esponenziali di base $a > 1$

- (1) Siano $x, y \in \mathbb{R}$, $x < y$, allora esistono $r, s \in \mathbb{Q}$ tali che $x < r < s < y$ e perciò

$$a^x \leq a^r < a^s \leq a^y$$

dove la prima e la terza disuguaglianza discendono dalla definizione 4.4, mentre la seconda segue dal teorema 4.7.

- (2) è conseguenza di (1)
 (3) Il fatto che $R(\exp_a) \subset \mathbb{R}_+$ è conseguenza della definizione di potenza; l'inclusione opposta segue ancora dal teorema dei valori intermedi.

□

3. La funzione logaritmo

Definiamo ora la funzione inversa dell'esponenziale.

Sia $a \in \mathbb{R}_+$, $a \neq 1$, allora la funzione $\exp_a : \mathbb{R} \longrightarrow \mathbb{R}_+$ è strettamente monotona e surgettiva, pertanto essa è invertibile e si può considerare la funzione $\log_a : \mathbb{R}_+ \longrightarrow \mathbb{R}$ definita da

$$\log_a = \exp_a^{-1}$$

Se $x \in \mathbb{R}_+$, $\log_a(x)$ si dice logaritmo in base a di x .

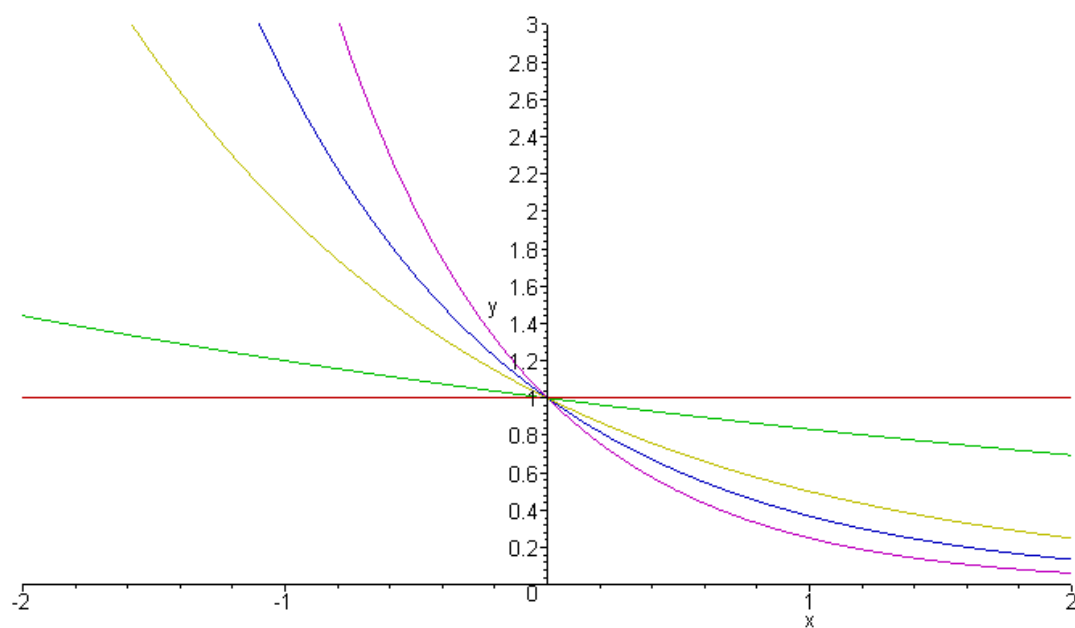
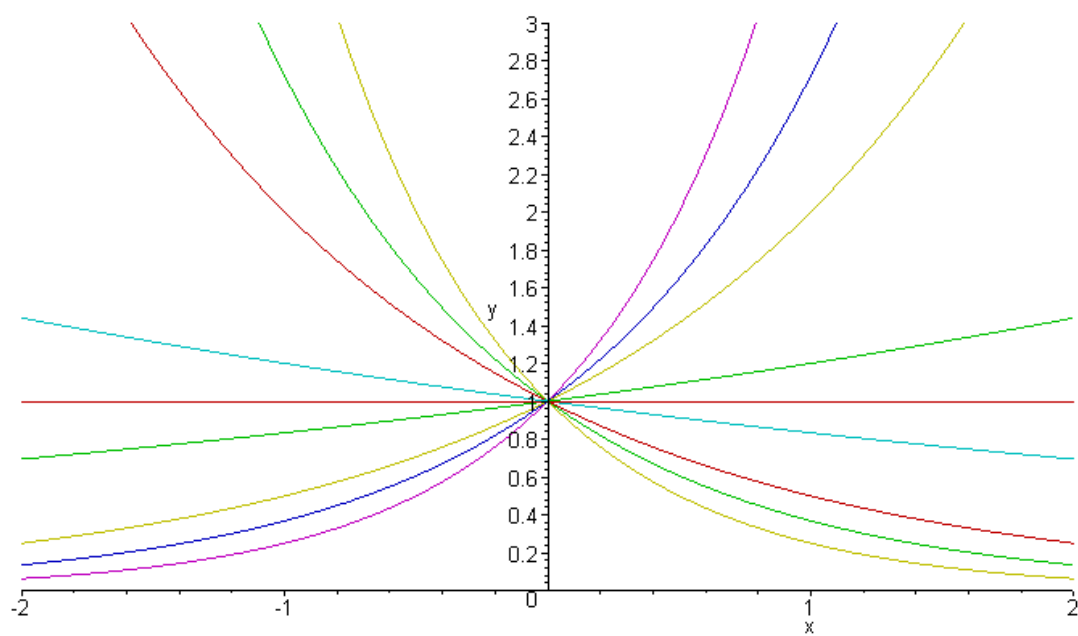
FIGURA 4.7. Esponenziali di base a , $0 < a < 1$ 

FIGURA 4.8. Esponenziali

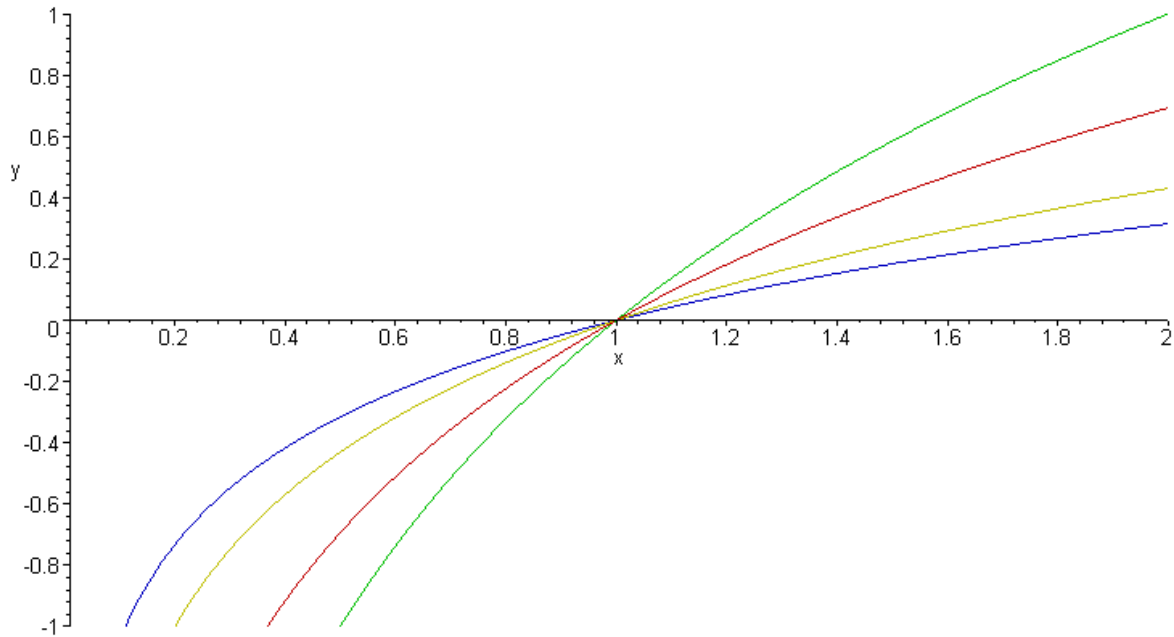


FIGURA 4.9. Logaritmi in base maggiore di 1

TEOREMA 4.12. *Sia $a \in \mathbb{R}$, allora*

- (1) *Se $a > 1$ allora \log_a è strettamente crescente*
- (2) *Se $0 < a < 1$ allora \log_a è strettamente decrescente*

TEOREMA 4.13. *Siano $a, b \in \mathbb{R}_+$, $a, b \neq 1$, $x, y \in \mathbb{R}_+$, $z \in \mathbb{R}$, allora*

- (1) $\log_a(xy) = \log_a(x) + \log_a(y)$
- (2) $\log_a(x^z) = z \log_a(x)$
- (3) $\log_a(x) = \log_a(b) \log_b(x)$.

DIMOSTRAZIONE.

- (1) Siano $u = \log_a(x)$ e $v = \log_a(y)$, allora $x = \exp_a(u)$ e $y = \exp_a(v)$, da cui

$$xy = \exp_a(u) \exp_a(v) = \exp_a(u + v)$$

e

$$u + v = \log_a(xy)$$

- (2) Sia $u = \log_a(x)$, allora $x = \exp_a(u)$ e $x^z = (\exp_a(u))^z = \exp_a(zu)$ per cui

$$zu = \log_a(x^z)$$

- (3)

$$\exp_a(\log_a(b) \log_b(x)) = (\exp_a(\log_a(b)))^{\log_b(x)} = \exp_b(\log_b(x)) = x$$

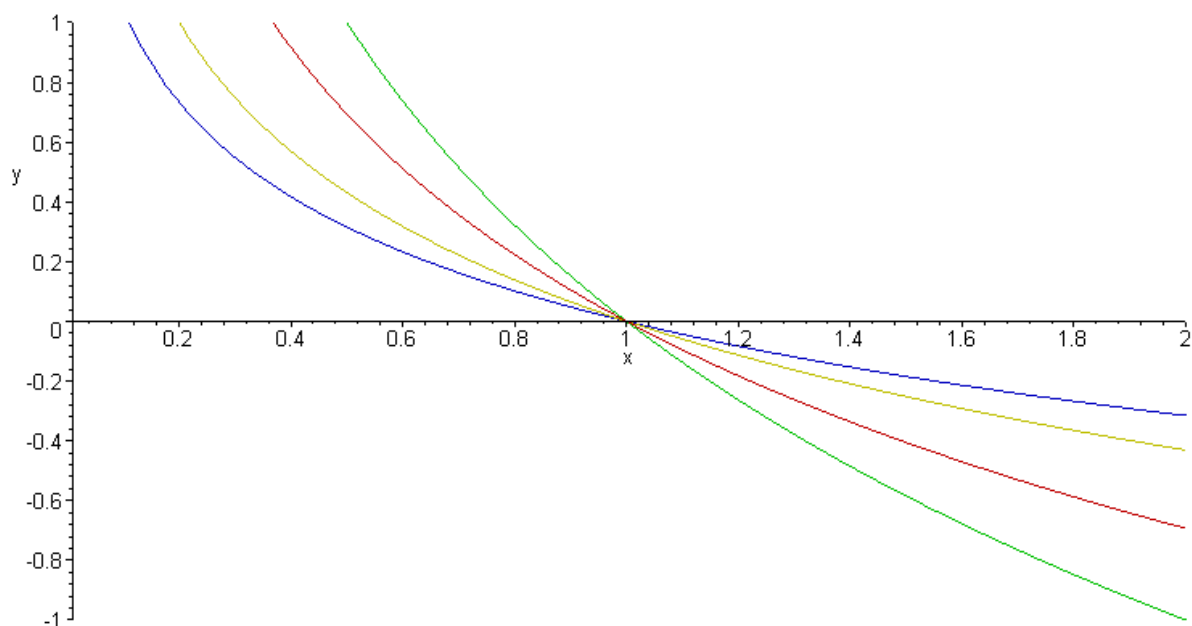


FIGURA 4.10. Logaritmi in base minore di 1

e

$$\log_a(x) = \log_a(b) \log_b(x)$$

□

4. Le funzioni trigonometriche

Ci apprestiamo, per concludere questa parte, a definire le cosiddette funzioni circolari. A questo scopo abbiamo bisogno di definire la lunghezza di un arco di circonferenza.

DEFINIZIONE 4.8. Sia Γ un arco della circonferenza C e siano A e B gli estremi di Γ ; supponiamo che A preceda B (scriveremo in tal caso $A \ll B$) considerando positivo il senso di rotazione antiorario.

Diciamo che è data una poligonale P inscritta in Γ (si veda fig. 4.12) se esistono n punti di Γ , $A_1 = A$, $A_2, \dots, A_n = B$ tali che $A_i \ll A_{i+1}$, $i = 1, 2, \dots, n-1$.

Definiamo lunghezza della poligonale P , $\ell(P)$, come

$$\ell(P) = \sum_{i=1}^{n-1} \overline{A_i A_{i+1}}$$

DEFINIZIONE 4.9. Sia Γ un arco di circonferenza di estremi A e B e sia \mathcal{P} l'insieme delle poligonali inscritte in Γ . Definiamo

$$\ell(\Gamma) = \sup\{\ell(P) : P \in \mathcal{P}\}$$

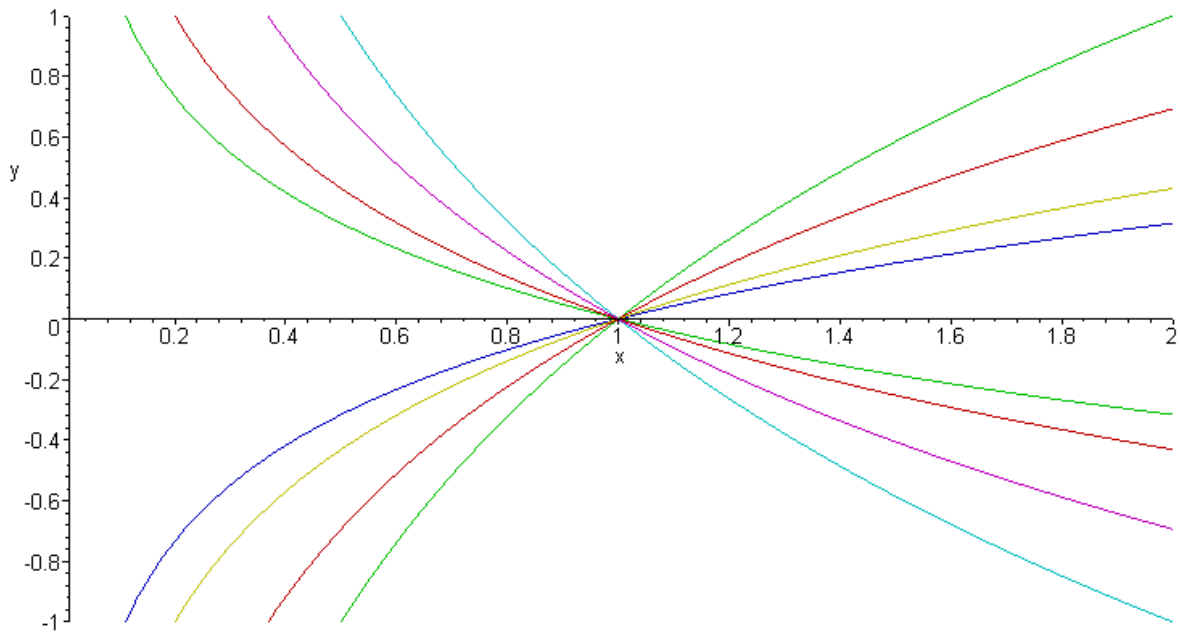


FIGURA 4.11. Logaritmi

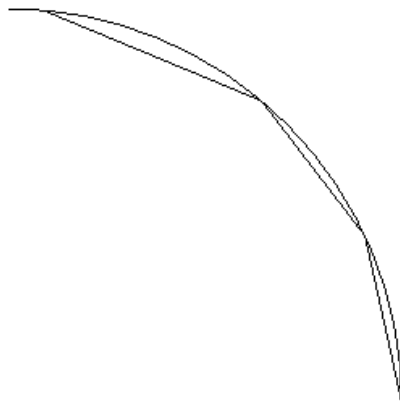


FIGURA 4.12. Una poligonale inscritta

(Si osservi che la definizione è ben posta in quanto $\ell(P) \leq 8R$ per ogni $P \in \mathcal{P}$, dove R è il raggio della circonferenza di cui Γ fa parte).

Usualmente si indica con π la lunghezza di una semicirconferenza di raggio 1; si ha con 51 cifre decimali esatte:

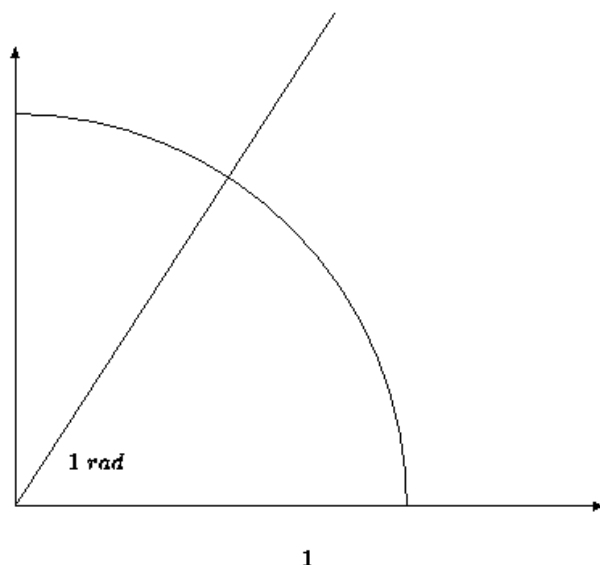


FIGURA 4.13. Definizione di radiante

$$\pi = 3.141592653589793238462643383279502884197169399375105$$

DEFINIZIONE 4.10. Diciamo che un angolo x misura 1 radiante se è l'angolo al centro che sottende un arco di lunghezza R in una circonferenza di raggio R . (vedi fig. 4)

Osservazione. Ovviamente un angolo giro misura 2π radianti, mentre un angolo piatto misura π radianti. Più in generale, se α è la misura di un angolo in gradi sessagesimali e x è la misura dello stesso angolo in radianti, si ha

$$\frac{\alpha}{x} = \frac{180}{\pi}$$

Questa relazione permette di convertire rapidamente la misura di un angolo da gradi in radianti e viceversa. \square

DEFINIZIONE 4.11. Sia $x \in [0, 2\pi]$ e consideriamo su di una circonferenza di raggio 1, centrata nell'origine delle coordinate, un arco Γ di lunghezza x aventi il primo estremo coincidente con il punto $(1, 0)$ (vedi fig. 4).

Definiamo $\cos(x)$ e $\sin(x)$ rispettivamente l'ascissa e l'ordinata del secondo estremo dell'arco.

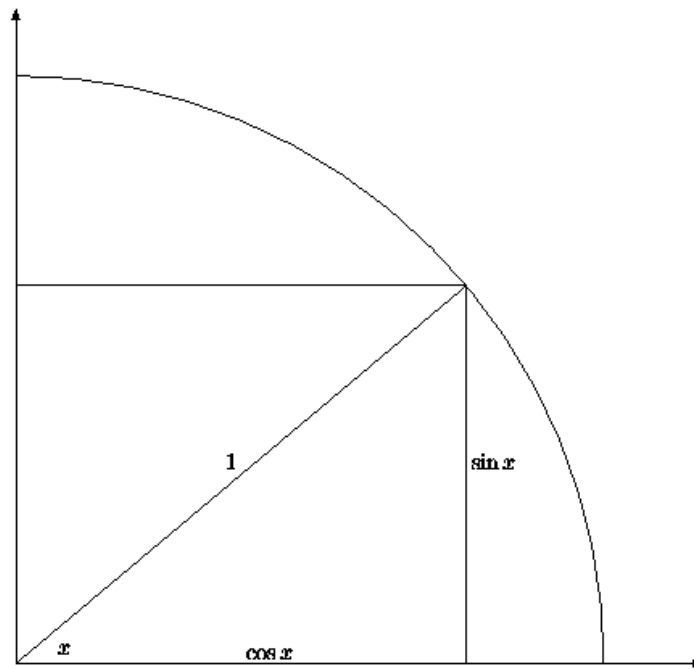


FIGURA 4.14. Definizione di sin e cos

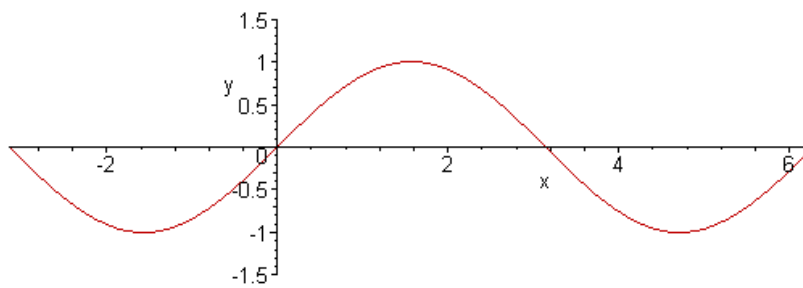
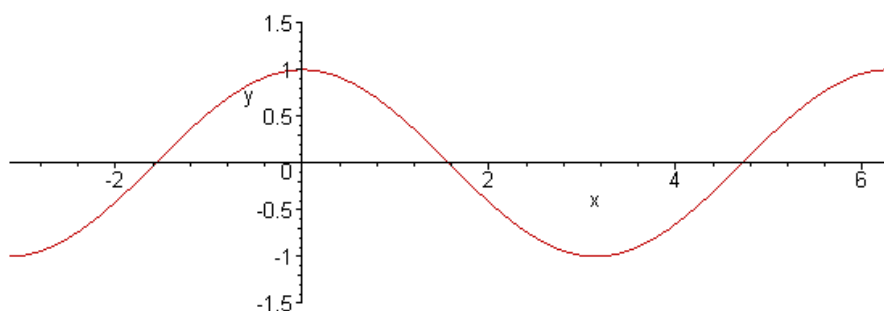


FIGURA 4.15. Grafico della funzione sin

DEFINIZIONE 4.12. Definiamo $\cos : \mathbb{R} \longrightarrow \mathbb{R}$ e $\sin : \mathbb{R} \longrightarrow \mathbb{R}$ nella seguente maniera: sia $x \in \mathbb{R}$ e sia $k = E(x/2\pi)$, allora $2k\pi \leq x < 2(k+1)\pi$.

FIGURA 4.16. grafico della funzione \cos

Poniamo

$$\cos(x) = \cos(x - 2k\pi) \quad e \quad \sin(x) = \sin(x - 2k\pi)$$

TEOREMA 4.14. *Valgono i seguenti fatti*

- (1) $R(\cos) = [-1, 1]$
- (2) $R(\sin) = [-1, 1]$

DIMOSTRAZIONE. E' ovvio dalla definizione che $R(\cos) \subset [-1, 1]$ e $R(\sin) \subset [-1, 1]$. Rimandiamo al seguito la dimostrazione dell'inclusione opposta. \square

DEFINIZIONE 4.13. *Definiamo $\tan : D \longrightarrow \mathbb{R}$, $D = \{x \neq k\pi + \pi/2, k \in \mathbb{Z}\}$, mediante la*

$$\tan(x) = \frac{\sin(x)}{\cos(x)}$$

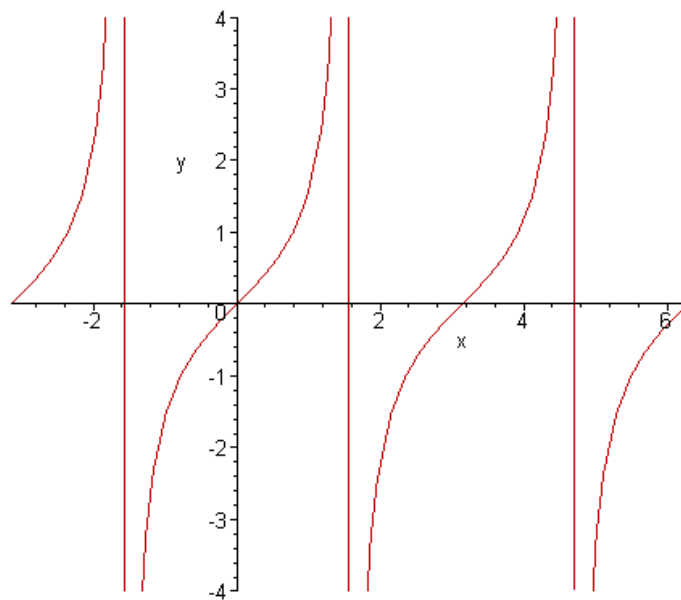
DEFINIZIONE 4.14. *Siano $D \subset \mathbb{R}$ e $T \in \mathbb{R}$ tali che $x \in D \Rightarrow x + T \in D$; sia inoltre $f : D \longrightarrow \mathbb{R}$; f si dice periodica di periodo T se,*

$$\forall x \in D, \quad f(x) = f(x + T)$$

Enunciamo a questo punto, senza dimostrarle, alcune fondamentali proprietà delle funzioni introdotte.

Siano $x, y \in \mathbb{R}$, valgono i seguenti fatti:

- (1) $\cos(-x) = \cos(x)$
- (2) $\sin(-x) = -\sin(x)$
- (3) $\sin^2(x) + \cos^2(x) = 1$
- (4) $\sin(x + y) = \sin(x)\cos(y) + \cos(x)\sin(y)$
- (5) $\cos(x + y) = \cos(x)\cos(y) - \sin(x)\sin(y)$

FIGURA 4.17. grafico della funzione \tan

(6) \sin e \cos sono periodiche di periodo 2π

(7) \tan è periodica di periodo π

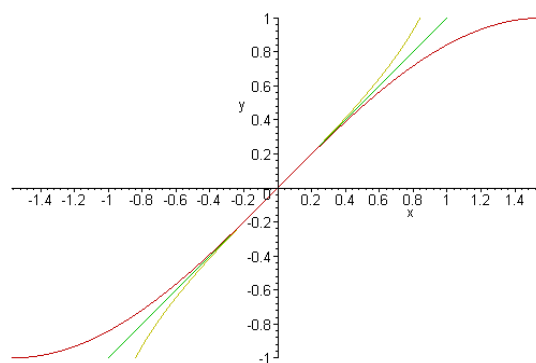
Valgono inoltre i seguenti fatti

(1) $\sin : [-\pi/2, \pi/2] \rightarrow [-1, 1]$ è strettamente crescente e surgettiva.

(2) $\cos : [0, \pi] \rightarrow [-1, 1]$ è strettamente decrescente e surgettiva.

(3) $\tan : (-\pi/2, \pi/2) \rightarrow \mathbb{R}$ è strettamente crescente e surgettiva.

Le verifiche delle proprietà enunciate sono basate su considerazioni geometriche che qui non stiamo ad investigare.

FIGURA 4.18. Grafico della funzione \arcsin

DEFINIZIONE 4.15. *Definiamo*

$$(4.1) \quad \arccos : [-1, 1] \longrightarrow [0, \pi]$$

$$(4.2) \quad \arcsin : [-1, 1] \longrightarrow [-\pi/2, \pi/2]$$

$$(4.3) \quad \arctan : \mathbb{R} \longrightarrow (-\pi/2, \pi/2)$$

mediante le

$$(4.4) \quad \arccos = \cos^{-1}$$

$$(4.5) \quad \arcsin = \sin^{-1}$$

$$(4.6) \quad \arctan = \tan^{-1}$$

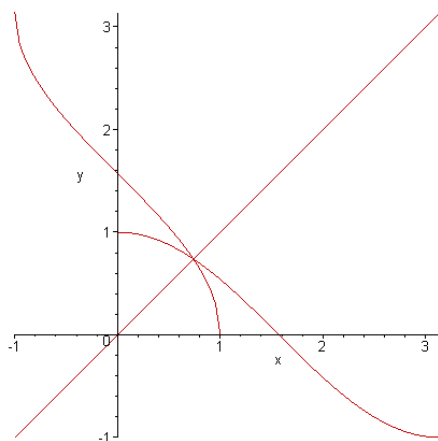


FIGURA 4.19. Grafico della funzione \arccos

La definizione è ben posta e valgono i seguenti fatti:

TEOREMA 4.15. *La funzione \arccos è strettamente decrescente su $[-1, 1]$; la funzione \arcsin è strettamente crescente su $[-1, 1]$; la funzione \arctan è strettamente crescente su \mathbb{R} .*

Osservazione. Occorre ricordare sempre che la denominazione \arcsin , \arccos ed \arctan è riservata alle inverse delle funzioni trigonometriche negli intervalli indicati nella definizione 4.15

Naturalmente, tali intervalli non sono gli unici in cui le funzioni in questione sono invertibili, tuttavia se vogliamo invertire una funzione trigonometrica in intervalli diversi da quelli sopra citati dobbiamo tener presente che le funzioni che otteniamo sono differenti da quelle definite in 4.15.

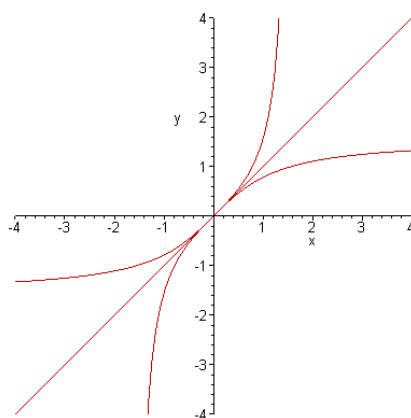


FIGURA 4.20. Grafico della funzione arctan

In particolare è opportuno ricordare che

$$(4.7) \quad \sin(\arcsin(x)) = x \quad \forall x \in [-1, 1]$$

$$(4.8) \quad \cos(\arccos(x)) = x \quad \forall x \in [-1, 1]$$

$$(4.9) \quad \tan(\arctan(x)) = x \quad \forall x \in \mathbb{R}$$

(4.10)

$$\arcsin(\sin(x)) = |x - 2k\pi + \pi/2| - \pi/2 \quad \forall x \in \mathbb{R}, \quad k = E\left(\frac{x + 3\pi/2}{2\pi}\right)$$

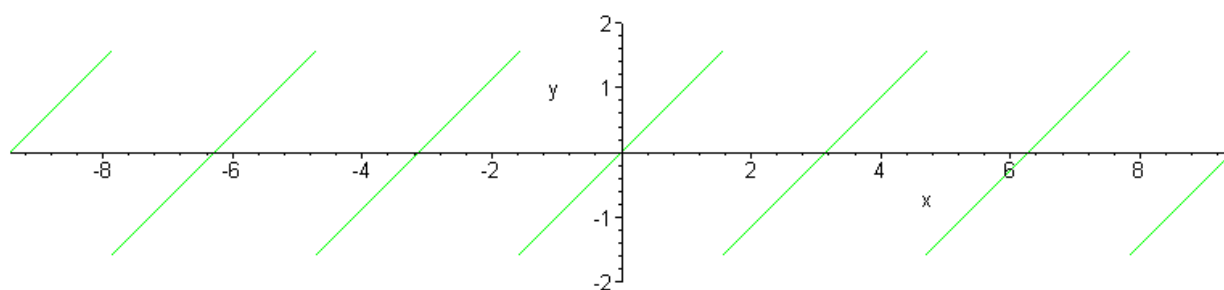
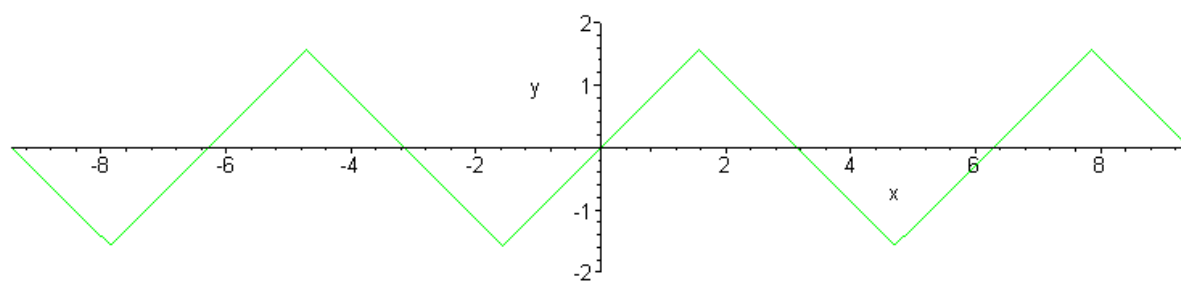
$$(4.11) \quad \arccos(\cos(x)) = |x - 2k\pi| \quad \forall x \in \mathbb{R}, \quad k = E\left(\frac{x + \pi}{2\pi}\right)$$

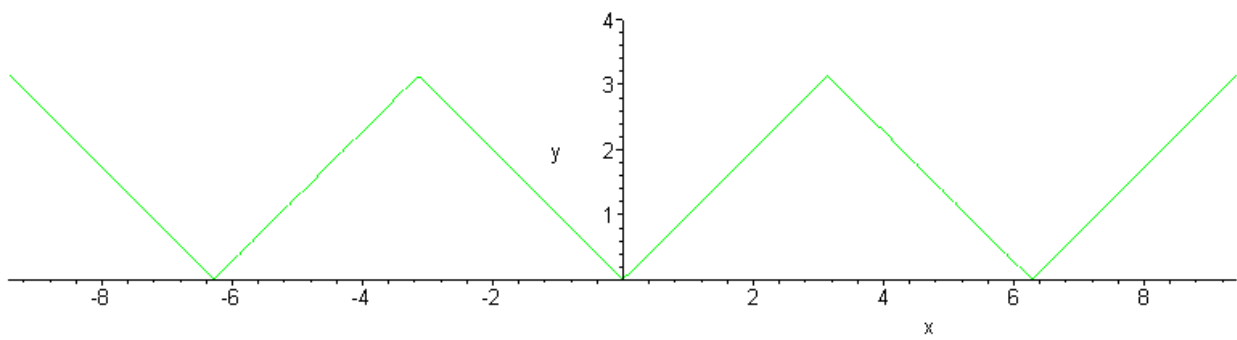
$$(4.12) \quad \arctan(\tan(x)) = x - k\pi \quad \forall x \in \mathbb{R}, \quad k = E\left(\frac{x + \pi/2}{\pi}\right)$$

(4.13)

□

Le notazioni arcsin, arccos, arctan non sono universalmente adottate. Nel seguito useremo anche le notazioni asin, acs, atn, rispettivamente.

FIGURA 4.21. Grafico della funzione $\arctan(\tan)$ FIGURA 4.22. grafico della funzione $\arcsin(\sin)$

FIGURA 4.23. Grafico della funzione $\arccos(\cos)$

CAPITOLO 5

DEFINIZIONE DI LIMITE E SUE CONSEGUENZE

Il concetto di limite è centrale in tutta l'analisi e da esso dipende l'essenza stessa del calcolo infinitesimale.

Si tratta di formalizzare un concetto che consenta di estendere il concetto di uguaglianza algebrica.

A questo scopo è necessario premettere alcuni concetti.

Conveniamo di indicare con \mathbb{R}^* l'insieme $\mathbb{R} \cup \{\pm\infty\}$, che chiameremo \mathbb{R} esteso.

DEFINIZIONE 5.1. Sia $x \in \mathbb{R}^*$ e $\delta > 0$, definiamo intorno di centro x e ampiezza δ l'insieme $I(x, \delta)$ definito da

$$(5.1) \quad I(x, \delta) = (x - \delta, x + \delta)$$

$$(5.2) \quad I(+\infty, \delta) = (\delta, +\infty)$$

$$(5.3) \quad I(-\infty, \delta) = (-\infty, -\delta)$$

Definiamo intorno bucato di centro x e ampiezza δ l'insieme $I^o(x, \delta)$.

$$(5.4) \quad I^o(x, \delta) = I(x, \delta) \setminus \{x\}$$

$$(5.5) \quad I^o(+\infty, \delta) = I(+\infty, \delta)$$

$$(5.6) \quad I^o(-\infty, \delta) = I(-\infty, \delta)$$

DEFINIZIONE 5.2. Sia $A \subset \mathbb{R}$, diciamo che $x \in \mathbb{R}^*$ è un punto di accumulazione per A se

$$(5.7) \quad \forall \delta \in \mathbb{R}_+ \quad A \cap I^o(x, \delta) \neq \emptyset$$

Indichiamo con $\mathcal{D}(A)$ l'insieme dei punti di accumulazione di A ; $\mathcal{D}(A)$ si indica usualmente con il nome di insieme derivato di A .

Osserviamo esplicitamente che $\mathcal{D}(A)$ può non essere un sottoinsieme di \mathbb{R} in quanto $+\infty$ e $-\infty$ possono essere elementi di $\mathcal{D}(A)$.

DEFINIZIONE 5.3. Sia $f : D \longrightarrow \mathbb{R}$, $x_0 \in \mathcal{D}(D)$ ed $\ell \in \mathbb{R}^*$; diciamo che

$$(5.8) \quad \lim_{x \rightarrow x_0} f(x) = \ell$$

se

$$(5.9) \quad \forall \varepsilon > 0 \quad \exists \delta_\varepsilon > 0 \text{ tale che}$$

$$\forall x \in I^o(x_0, \delta_\varepsilon) \cap D \quad \text{si ha} \quad f(x) \in I(\ell, \varepsilon)$$

Osservazione. Nel caso in cui x_0 ed ℓ siano entrambi reali la 5.9 può essere riscritta nella seguente maniera

$$(5.10) \quad \forall x \in D \quad \text{tale che} \quad 0 < |x - x_0| < \delta_\varepsilon \quad \text{si ha} \quad |f(x) - \ell| < \varepsilon$$

se $x_0 = +\infty$ ($-\infty$) ed $\ell \in \mathbb{R}$ la 5.9 diviene

$$(5.11) \quad \forall x \in D \quad \text{tale che} \quad x > \delta_\varepsilon \quad (x < -\delta_\varepsilon) \quad \text{si ha} \quad |f(x) - \ell| < \varepsilon$$

Notiamo che, nel caso $\ell \in \mathbb{R}$, se la 5.9 è verificata per $\varepsilon \in (0, \varepsilon_0)$, essa è automaticamente verificata anche per tutti gli $\varepsilon \geq \varepsilon_0$, pur di definire $\delta_\varepsilon = \delta_{\varepsilon_0}$.

Se $x_0 \in \mathbb{R}$ e $\ell = +\infty$ ($-\infty$) la 5.9 diviene

$$(5.12)$$

$$\forall x \in D \quad \text{tale che} \quad 0 < |x - x_0| < \delta_\varepsilon \quad \text{si ha} \quad f(x) > \varepsilon \quad (f(x) < -\varepsilon)$$

se $x_0 = +\infty$ ($-\infty$) e $\ell = +\infty$ ($-\infty$) la 5.9 diviene

$$(5.13)$$

$$\forall x \in D \quad : \quad x > \delta_\varepsilon \quad (x < -\delta_\varepsilon) \quad \text{si ha} \quad f(x) > \varepsilon \quad (f(x) < -\varepsilon)$$

Notiamo anche qui che, nel caso in cui $\ell = +\infty$ o $\ell = -\infty$, se la 5.9 è verificata per $\varepsilon > \varepsilon_0$, essa è automaticamente verificata pure per tutti gli $\varepsilon \in (0, \varepsilon_0]$, pur di definire $\delta_\varepsilon = \delta_{\varepsilon_0}$.

□

Osservazione. Se esiste $\ell \in \mathbb{R}$ tale che valga la definizione 5.3 si dice che f ammette limite finito per $x \rightarrow x_0$; in caso contrario si dice che f non ammette limite finito.

Se esiste $\in \mathbb{R}^*$ tale che valga la definizione 5.3 si dice che f ammette limite per $x \rightarrow x_0$; in caso contrario si dice che f non ammette limite.

□

DEFINIZIONE 5.4. Sia $f : D \rightarrow \mathbb{R}$, sia $x_0 \in \mathcal{D}(D)$; diciamo che f è localmente (superiormente) [inferiormente] limitata in x_0 se esiste $M \in \mathbb{R}$ ed esiste $\delta > 0$ tale che

$$(5.14)$$

$$|f(x)| \leq M \quad , \quad (f(x) \leq M) \quad , \quad [f(x) \geq M] \text{ Quad} \forall x \in I(x_0, \delta) \cap D$$

Passiamo ora a dimostrare che una funzione che ammette limite finito è localmente limitata.

TEOREMA 5.1. Sia $f : D \rightarrow \mathbb{R}$, sia $x_0 \in \mathcal{D}(D)$ e supponiamo che

$$\lim_{x \rightarrow x_0} f(x) = \ell$$

allora:

- (1) se $\ell \in \mathbb{R}$, allora f è localmente limitata in x_0 ;
- (2) se $\ell > 0$ (eventualmente $\ell = +\infty$) allora esiste $\delta > 0$ ed esiste $M \in \mathbb{R}$ tale che se $x \in I^\circ(x_0, \delta) \cap D$ si ha

$$f(x) \geq M > 0$$

DIMOSTRAZIONE. Proviamo solo la prima affermazione e rimandiamo per la seconda a ?? Sia $\varepsilon > 0$, se $x \in I^o(x_0, \delta_\varepsilon)$ si ha

$$\ell - \varepsilon < f(x) < \ell + \varepsilon$$

e

$$|f(x)| \leq \max\{|\ell + \varepsilon|, |\ell - \varepsilon|, |f(x_0)|\} \quad \forall x \in I(x_0, \delta_\varepsilon) \cap D$$

□

TEOREMA 5.2. Sia $f : D \longrightarrow \mathbb{R}$ e sia $x_0 \in \mathcal{D}(D)$; allora, se f ammette limite per $x \rightarrow x_0$, tale limite è unico.

DIMOSTRAZIONE. Proviamo il teorema nel caso in cui ℓ_1 ed ℓ_2 siano entrambi limiti di f per $x \rightarrow x_0$ e siano entrambi reali;

Si ha $\forall \varepsilon > 0$, se $x \in I^o(x_0, \delta_{\varepsilon/2}) \cap D$,

$$|\ell_1 - \ell_2| \leq |f(x) - \ell_1| + |f(x) - \ell_2| < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

Negli altri casi possiamo procedere come indicato in ??

□

DEFINIZIONE 5.5. Supponiamo $f : D \longrightarrow \mathbb{R}$ e sia $x_0 \in \mathcal{D}(D_+) \cap \mathbb{R}$, $D_+ = \{x \in D : x \geq x_0\}$.

Diciamo che

$$\lim_{x \rightarrow x_0^+} f(x) = \ell \quad , \quad \ell \in \mathbb{R}^*$$

se

$\forall \varepsilon > 0 \exists \delta_\varepsilon > 0$ tale che se $x \in D$ e $x_0 < x < x_0 + \delta_\varepsilon$ si ha

$$f(x) \in I(\ell, \varepsilon)$$

Se $x_0 \in \mathcal{D}(D_-) \cap \mathbb{R}$, $D_- = \{x \in D : x \leq x_0\}$, diciamo che

$$\lim_{x \rightarrow x_0^-} f(x) = \ell \quad , \quad \ell \in \mathbb{R}^*$$

se $\forall \varepsilon > 0 \exists \delta_\varepsilon > 0$ tale che se $x \in D$ e $x_0 - \delta_\varepsilon < x < x_0$ si ha

$$f(x) \in I(\ell, \varepsilon)$$

TEOREMA 5.3. Sia $f : D \longrightarrow \mathbb{R}$ e sia $x_0 \in \mathcal{D}(D_+) \cap \mathcal{D}(D_-) \cap \mathbb{R}$, allora se $\ell \in \mathbb{R}^*$, si ha

$$\lim_{x \rightarrow x_0} f(x) = \ell \quad \Leftrightarrow \quad \lim_{x \rightarrow x_0^+} f(x) = \lim_{x \rightarrow x_0^-} f(x) = \ell$$

DIMOSTRAZIONE. Cominciamo con l'osservare che se il limite esiste $\forall \varepsilon > 0$ esiste $\delta_\varepsilon > 0$ tale che se $x_0 - \delta_\varepsilon < x < x_0 + \delta_\varepsilon$ con $x \neq x_0$ ed $x \in D$ si ha

$$f(x) \in I(\ell, \varepsilon)$$

e ciò implica per la definizione 5.5, la tesi.

Se viceversa esistono i limiti da destra e da sinistra $\forall \varepsilon > 0$ esistono $\delta_\varepsilon^1, \delta_\varepsilon^2 > 0$ tali che se $x_0 < x < x_0 + \delta_\varepsilon^1$, $x \in D$ si ha

$$f(x) \in I(\ell, \varepsilon)$$

e se $x_0 - \delta_\varepsilon^2 < x < x_0$, $x \in D$ si ha

$$f(x) \in I(\ell, \varepsilon)$$

Pertanto se si sceglie

$$\delta_\varepsilon = \min\{\delta_\varepsilon^1, \delta_\varepsilon^2\}$$

la definizione di limite è verificata. \square

A questo punto è conveniente definire in \mathbb{R}^* le operazioni di addizione e di moltiplicazione che fino a questo momento sono definite solamente in \mathbb{R} .

Osserviamo esplicitamente che non sono applicabili a queste operazioni le usuali regole che permettono di svolgere calcoli con i numeri reali. Riterremo pertanto lecite tutte e sole le uguaglianze che coinvolgono gli elementi $+\infty$ e $-\infty$ che elenchiamo qui di seguito.

Definiamo:

$$\begin{aligned} x \pm \infty &= \pm\infty + x = \pm\infty & \forall x \in \mathbb{R} \\ x(\pm\infty) &= (\pm\infty)x = \pm\infty & \forall x \in \mathbb{R}_+ \\ x(\pm\infty) &= (\pm\infty)x = \mp\infty & \forall x \in \mathbb{R}_- \\ \frac{x}{\pm\infty} &= 0 & \forall x \in \mathbb{R} \\ \left|\frac{x}{0}\right| &= +\infty & \forall x \in \mathbb{R} \setminus \{0\} \\ +\infty + \infty &= +\infty, & -\infty - \infty = -\infty \\ (\pm\infty)(+\infty) &= \pm\infty & |\pm\infty| = +\infty \end{aligned}$$

Ricordiamo inoltre che *non* sono definite le seguenti operazioni;

$$+\infty - \infty, \quad 0(\pm\infty), \quad \frac{\pm\infty}{\pm\infty}, \quad \frac{0}{0}$$

in quanto ciò potrebbe dar luogo facilmente ad inconvenienti e ad errate interpretazioni.

TEOREMA 5.4. *Siano $f, g : D \longrightarrow \mathbb{R}$ e sia $x_0 \in \mathcal{D}(D)$; supponiamo che*

$$\lim_{x \rightarrow x_0} f(x) = \ell_1 \quad \text{e} \quad \lim_{x \rightarrow x_0} g(x) = \ell_2, \quad \ell_1, \ell_2 \in \mathbb{R}^*$$

Allora

- (1) $\lim_{x \rightarrow x_0} |f(x)| = |\ell_1|$
- (2) $\lim_{x \rightarrow x_0} (f(x) + g(x)) = \ell_1 + \ell_2$ *tranne che nel caso in cui $\ell_1 = \pm\infty$ e $\ell_2 = \mp\infty$*
- (3) $\lim_{x \rightarrow x_0} f(x)g(x) = \ell_1\ell_2$ *tranne che nel caso in cui $\ell_1 = 0$ e $\ell_2 = \pm\infty$*
- (4) $\lim_{x \rightarrow x_0} \frac{1}{f(x)} = \frac{1}{\ell_1}$ *tranne che nel caso in cui $\ell_1 = 0$*

DIMOSTRAZIONE. Dimostriamo nel caso in cui $x_0 \in \mathbb{R}$, $\ell_1, \ell_2 \in \mathbb{R}$ la seconda e la terza delle asserzioni fatte.

Per ipotesi abbiamo che $\forall \varepsilon > 0 \quad \exists \delta_\varepsilon^1, \delta_\varepsilon^2 > 0$ tali che

$\forall x \in I^o(x_0, \delta_\varepsilon^1) \cap D$ si ha

$$|f(x) - \ell_1| < \frac{\varepsilon}{2}$$

e $\forall x \in I^o(x_0, \delta_\varepsilon^2) \cap D$ si ha

$$|g(x) - \ell_2| < \frac{\varepsilon}{2}$$

Allora

Sia $\delta_\varepsilon = \min\{\delta_{\varepsilon/2}^1, \delta_{\varepsilon/2}^2\}$, se $x \in I^o(x_0, \delta_\varepsilon) \cap D$ si ha

$$|f(x) + g(x) - (\ell_1 + \ell_2)| \leq |f(x) - \ell_1| + |g(x) - \ell_2| < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

Con ciò la seconda affermazione è provata.

Sia δ^3 tale che se $x \in I^o(x_0, \delta^3) \cap D$ si ha

$$|g(x)| \leq M$$

sia $\ell_1 \neq 0$ (il caso $\ell_1 = 0$ risulta banale) e sia

$$\delta_\varepsilon = \min\{\delta_{\varepsilon/(2M)}^1, \delta_{\varepsilon/(2|\ell_1|)}^2, \delta^3\}$$

allora se $x \in I^o(x_0, \delta_\varepsilon) \cap D$ si ha

$$\begin{aligned} (5.15) \quad |f(x)g(x) - \ell_1\ell_2| &= |f(x)g(x) - g(x)\ell_1 + g(x)\ell_1 - \ell_1\ell_2| \leq \\ &\leq |g(x)||f(x) - \ell_1| + |\ell_1||g(x) - \ell_2| < \\ &< M\varepsilon/(2M) + \varepsilon|\ell_1|/(2|\ell_1|) = \varepsilon \end{aligned}$$

Quindi anche la terza è provata. □

Possiamo a questo punto stabilire un utile corollario.

COROLLARIO 5.1. Siano $f, g : D \rightarrow \mathbb{R}$, $x_0 \in \mathcal{D}(D)$ e supponiamo che

$$\lim_{x \rightarrow x_0} f(x) = \ell_1 \quad , \quad \lim_{x \rightarrow x_0} g(x) = \ell_2$$

con $\ell_1, \ell_2 \in \mathbb{R}^*$ ed $\ell_1 < \ell_2$; allora

$$\exists \delta > 0 \quad : \quad \forall x \in I^o(x_0, \delta) \quad , \quad f(x) < g(x)$$

Per completare il quadro di risultati proviamo il seguente

TEOREMA 5.5. Sia $f : D \rightarrow \mathbb{R}$, $D_1 = \{x \in D : f(x) \neq 0\}$, $x_0 \in \mathcal{D}(D_1)$,

$$\lim_{x \rightarrow x_0} f(x) = 0$$

allora

$$(5.16) \quad \lim_{x \rightarrow x_0} \frac{1}{|f(x)|} = +\infty$$

se esiste $\delta > 0$ tale che per $x \in I^o(x_0, \delta) \cap D$ si ha $f(x) > 0$ ($f(x) < 0$),

$$(5.17) \quad \lim_{x \rightarrow x_0} \frac{1}{f(x)} = +\infty \quad (-\infty)$$

DIMOSTRAZIONE. Per ipotesi, Per ogni $\varepsilon > 0$ esiste $\delta_\varepsilon > 0$ tale che se $x \in I^o(x_0, \delta_\varepsilon) \cap D$ si ha $|f(x)| < \varepsilon$.

$$x \in I^o(x_0, \delta_{1/\varepsilon}) \Rightarrow |f(x)| < 1/\varepsilon$$

e

$$\frac{1}{|f(x)|} > \varepsilon$$

(2) Supponiamo per semplicità che f sia localmente positiva in x_0 ; sia

$$\delta'_\varepsilon = \min\{\delta, \delta_{1/\varepsilon}\},$$

allora, se $x \in I^o(x_0, \delta'_\varepsilon) \cap D$

$$0 < f(x) < 1/\varepsilon \quad \text{e} \quad \frac{1}{f(x)} > \varepsilon$$

Il caso in cui f sia localmente negativa si riconduce banalmente al caso sopra descritto. □

Osservazione. Notiamo esplicitamente che è essenziale nella (2) l'ipotesi che f abbia segno localmente costante in x_0 .

Sia infatti $f(x) = x \sin(1/x)$ per $x \neq 0$; allora si può facilmente verificare che

$$\lim_{x \rightarrow 0} f(x) = 0 \quad \text{e} \quad \lim_{x \rightarrow 0} 1/|f(x)| = +\infty;$$

tuttavia è altrettanto immediato verificare che $1/f(x)$ non tende né a $-\infty$ né a $+\infty$, in quanto, se ciò accadesse, per valori di x vicini allo 0 si dovrebbe avere $f(x) < 0$ oppure $f(x) > 0$. □

Sarà pure di fondamentale importanza il seguente teorema:

TEOREMA 5.6. Siano $f, g, h : D \rightarrow \mathbb{R}$, sia $x_0 \in \mathcal{D}(D)$; supponiamo che esista $\delta > 0$ tale che, se $x \in I^o(x_0, \delta) \cap D$

$$f(x) \leq g(x) \leq h(x)$$

siano inoltre $\lim_{x \rightarrow x_0} f(x) = \ell_1$ e $\lim_{x \rightarrow x_0} h(x) = \ell_2$. Allora

$$(5.18) \quad \ell_1, \ell_2 \in \mathbb{R} \Rightarrow \ell_1 \leq \ell_2$$

$$(5.19) \quad \ell_1 = \ell_2 = \ell \Rightarrow \lim_{x \rightarrow x_0} g(x) = \ell$$

$$(5.20) \quad \ell_1 = +\infty \Rightarrow \ell_2 = +\infty \quad \text{e} \quad \lim_{x \rightarrow x_0} g(x) = +\infty$$

$$(5.21) \quad \ell_2 = -\infty \Rightarrow \ell_1 = -\infty \quad \text{e} \quad \lim_{x \rightarrow x_0} g(x) = -\infty$$

DIMOSTRAZIONE. La prima affermazione è una diretta conseguenza del corollario 5.1.

Per quanto riguarda la seconda affermazione, dalle ipotesi si ha che $\forall \varepsilon > 0$ esiste $\delta_\varepsilon > 0$ tale che, se $x \in I^o(x_0, \delta_\varepsilon) \cap D$

$$\ell - \varepsilon < f(x) < \ell + \varepsilon \quad e \quad \ell - \varepsilon < h(x) < \ell + \varepsilon$$

da cui, per gli stessi valori di x si ha

$$\ell - \varepsilon < f(x) < g(x) < h(x) < \ell + \varepsilon$$

Lasciamo al lettore la dimostrazione delle restanti affermazioni. \square

TEOREMA 5.7. Siano $f : D \longrightarrow \mathbb{R}$ e $g : A \longrightarrow D$; siano $x_0 \in \mathcal{D}(D)$ e $t_0 \in \mathcal{D}(A)$; supponiamo che

$$\lim_{x \rightarrow x_0} f(x) = \ell \quad e \quad \lim_{t \rightarrow t_0} g(t) = x_0$$

Supponiamo inoltre che sia verificata una delle seguenti condizioni:

- (1) esiste $\delta > 0$ tale che $g(t) \neq x_0$ per ogni $t \in I^o(t_0, \delta)$;
- (2)

$$f(x_0) = \ell$$

Allora

$$\lim_{t \rightarrow t_0} f(g(t)) = \ell$$

Osserviamo che se tutte e due le condizioni vengono a mancare, il teorema precedente può non essere vero.

Sia ad esempio $D = A = \mathbb{R}$, $g(t) = 0$ ed $f(x) = 0$ se $x \neq 0$, $f(0) = 1$; allora

$$\lim_{t \rightarrow 0} g(t) = 0 \quad , \quad \lim_{x \rightarrow 0} f(x) = 0$$

mentre

$$\lim_{t \rightarrow 0} f(g(t)) = 1$$

Osserviamo inoltre che ognuna delle seguenti condizioni

- (1) $x_0 \notin D$,
- (2) $x_0 = \pm\infty$
- (3) g è iniettiva
- (4) g è strettamente monotona

è sufficiente per la (1) del teorema 5.7

TEOREMA 5.8. Sia $f : (a, b) \longrightarrow \mathbb{R}$, f crescente [decrescente]; allora

$$\lim_{x \rightarrow a^+} f(x) \quad e \quad \lim_{x \rightarrow b^-} f(x)$$

esistono e sono uguali rispettivamente a

$$\inf\{f(x) : x \in (a, b)\} \quad [\sup\{f(x) : x \in (a, b)\}]$$

e

$$\sup\{f(x) : x \in (a, b)\} \quad [\inf\{f(x) : x \in (a, b)\}]$$

Osserviamo esplicitamente che nel teorema precedente è essenziale supporre che l'intervallo in cui si considera la funzione sia aperto.

Sia infatti $f(x) = x$ se $x \in [0, 1)$, $f(1) = 2$; allora

$$\sup\{f(x) : x \in [0, 1]\} = 2 \neq 1 = \lim_{x \rightarrow 1^-} f(x)$$

COROLLARIO 5.2. *Sia $f : D \longrightarrow \mathbb{R}$ una funzione monotona, allora per ogni $x_0 \in \mathcal{D}(D_+) \cap \mathcal{D}(D_-)$ si ha che*

$$\lim_{x \rightarrow x_0^+} f(x) \quad , \quad \lim_{x \rightarrow x_0^-} f(x)$$

esistono.

Per stabilire l'esistenza del limite di una funzione è possibile avvalersi del criterio di convergenza di Cauchy.

TEOREMA 5.9. - Criterio di Cauchy - *Sia $f : D \longrightarrow \mathbb{R}$ e sia $x_0 \in \mathcal{D}(D)$; sono condizioni equivalenti:*

- (1) *esiste $\ell \in \mathbb{R}$ tale che $\lim_{x \rightarrow x_0} f(x) = \ell$*
- (2) *per ogni $\varepsilon > 0$ esiste $\delta_\varepsilon > 0$ tale che se $x, y \in I^0(x_0, \delta_\varepsilon)$ si ha*

$$|f(x) - f(y)| < \varepsilon$$

CAPITOLO 6

LE SUCCESSIONI

Le successioni costituiscono una classe molto particolare di funzioni: si tratta di funzioni definite su un sottoinsieme di \mathbb{R} molto particolare, l'insieme \mathbb{N} dei numeri naturali; questa caratteristica conferisce loro la semplicità che è tipica degli insiemi discreti, mentre impedisce una significativa rappresentazione grafica e rende il concetto di successione apparentemente ostico.

Il concetto di successione, inoltre, interpreta un ruolo di notevole importanza nelle applicazioni pratiche e nelle descrizioni algoritmiche.

DEFINIZIONE 6.1. *Chiamiamo successione di numeri reali una funzione definita sull'insieme \mathbb{N} dei numeri naturali, che assume valori in \mathbb{R}*

$$a : \mathbb{N} \longrightarrow \mathbb{R}$$

Seguendo le consuetudini introdotte per la descrizione di una funzione sarebbe naturale usare il simbolo $a(n)$ per identificare il valore di a calcolato in n tuttavia è normale usare, in luogo di esso il simbolo $a_n = a(n)$.

E' immediato esplicitare per le successioni i concetti di crescita, decrescenza, monotonia, limitatezza, che sono stati introdotti, in generale, per le funzioni.

Nell'estendere il concetto di limite però occorre tenere presente che $\mathcal{D}(\mathbb{N})$ è costituito dal solo elemento $+\infty$, per cui, per una successione, ha senso soltanto considerare il concetto di limite per $n \rightarrow +\infty$.

Più precisamente si dice che

$$(6.1) \quad \lim_{n \rightarrow +\infty} a_n = \ell$$

se $\forall \varepsilon > 0$ esiste $n_\varepsilon \in \mathbb{N}$ tale che, per $n > n_\varepsilon$, si abbia

$$a_n \in I(\ell, \varepsilon)$$

Osserviamo che, dal momento che nessuna ambiguità è possibile, scriveremo spesso

$$\lim_n a_n \quad \text{oppure} \quad \lim a_n$$

in luogo di $\lim_{n \rightarrow +\infty} a_n$.

E' molto importante, quando si trattano le successioni, il concetto di successione estratta da un'altra successione.

Tale concetto è strettamente legato, o meglio è una specializzazione del concetto di composizione di funzioni ed è molto utile per caratterizzare i limiti di una successione.

In altre parole si dice successione estratta dalla successione a_n una nuova successione $a_{n(k)}$.

Naturalmente non ogni funzione $n : \mathbb{N} \longrightarrow \mathbb{N}$ può essere usata, per due ragioni:

- (1) n deve dar luogo, composta con a , ad una nuova successione, per cui deve aversi che il dominio di f è \mathbb{N} ;
- (2) $R(n)$ deve essere contenuto nel dominio di a e perciò deve aversi $R(n) \subset \mathbb{N}$.

Dovrà pertanto essere $n : \mathbb{N} \longrightarrow \mathbb{N}$.

- (3) Inoltre, poichè vogliamo collegare il comportamento al limite della successione a_n con quello delle sue estratte, è necessario che $+\infty$ sia un punto di accumulazione per $R(n)$.

In altre parole n è una particolare successione (particolare in quanto assume valori solo in \mathbb{N}) e pertanto è d'uso far riferimento alla notazione

$$n_k = n(k)$$

DEFINIZIONE 6.2. Sia a_n una successione e sia $n : \mathbb{N} \longrightarrow \mathbb{N}$ strettamente crescente; diciamo che la successione $b_k = a_{n(k)}$ è una successione estratta da a_n .

Sempre a proposito di terminologia, ricordiamo anche che si dice che una successione è convergente se ammette limite reale, mentre si dice che una successione è positivamente (negativamente) divergente se ammette come limite $+\infty$ ($-\infty$).

LEMMA 6.1. Sia $n : \mathbb{N} \longrightarrow \mathbb{N}$, n strettamente crescente; allora

$$\lim_k n_k = +\infty$$

DIMOSTRAZIONE. Dal momento che n è strettamente crescente si ha

$$n_{k+1} > n_k \quad \text{e} \quad n_{k+1} \geq n_k + 1$$

perciò, per induzione, si prova facilmente che $n_k \geq k$ e la tesi. \square

TEOREMA 6.1. Sia a_n una successione e sia

$$\lim_n a_n = \ell$$

allora se b_k è una successione estratta da a_n si ha

$$\lim_k b_k = \ell$$

DIMOSTRAZIONE. Sia $b_k = a_{n_k}$, se $n > n_\varepsilon$ si ha $a_n \in I(\ell, \varepsilon)$, inoltre, dal momento che $n_k \rightarrow +\infty$, se $k > k_\varepsilon$ si ha $n_k > n_\varepsilon$; ne deduciamo che, se $k > k_\varepsilon$,

$$b_k = a_{n_k} \in I(\ell, \varepsilon)$$

□

Si può inoltre dimostrare che

TEOREMA 6.2. *Ogni successione convergente è limitata.*

DIMOSTRAZIONE. Sia a_n una successione e sia

$$\lim_n a_n = \ell$$

allora, se $n > n_\varepsilon$ si ha

$$|a_n| \leq |a_n - \ell| + |\ell| < \varepsilon + |\ell|$$

Perciò se

$$M = \max\{|a_1|, \dots, |a_{n_\varepsilon}|, |\ell| + \varepsilon\}$$

si può affermare che

$$|a_n| \leq M$$

□

TEOREMA 6.3. *Sia a_n una successione crescente e sia*

$$\lambda = \sup\{a_n : n \in \mathbb{N}\}$$

allora

$$\lim_n a_n = \lambda$$

DIMOSTRAZIONE. Distinguiamo due casi.

(1) $\lambda = +\infty$

in tal caso $\{a_n : n \in \mathbb{N}\}$ è un insieme non limitato e $\forall \varepsilon > 0$ esiste $n_\varepsilon \in \mathbb{N}$ tale che

$$a_{n_\varepsilon} > \varepsilon$$

ma allora, dal momento che a_n è crescente, si ha, per $n > n_\varepsilon$

$$a_n > a_{n_\varepsilon} > \varepsilon$$

(2) $\lambda \in \mathbb{R}$

in questo caso, per le proprietà dell'estremo superiore, si ha

$$a_n \leq \lambda \quad \forall n \in \mathbb{N}$$

e

$$\forall \varepsilon > 0 \quad \exists n_\varepsilon \in \mathbb{N} : a_{n_\varepsilon} > \lambda - \varepsilon$$

pertanto, se $n > n_\varepsilon$, si ha, essendo a_n crescente

$$\lambda - \varepsilon < a_{n_\varepsilon} \leq a_n \leq \lambda$$

□

In maniera analoga si può dimostrare il seguente teorema.

TEOREMA 6.4. *Sia a_n una successione decrescente e sia*

$$\lambda = \inf\{a_n : n \in \mathbb{N}\}$$

allora

$$\lim_n a_n = \lambda$$

Il risultato che segue è uno dei più importanti tra quelli che riguardano le successioni di numeri reali.

TEOREMA 6.5. - *Bolzano-Weierstraß* - *Sia a_n una successione limitata, allora esiste $\lambda \in \mathbb{R}$ ed esiste una successione b_k estratta da a_n tale che*

$$\lim_k b_k = \lambda$$

DIMOSTRAZIONE. Per ipotesi esistono $m, M \in \mathbb{R}$ tali che

$$m \leq a_n \leq M \quad \forall n \in \mathbb{N}$$

Consideriamo i due intervalli

$$\left[m, \frac{m+M}{2} \right] \quad \text{e} \quad \left[\frac{m+M}{2}, M \right]$$

almeno uno di essi contiene un numero infinito di termini della successione a_n sia esso $[\alpha_1, \beta_1]$ ovviamente si avrà

$$m \leq \alpha_1 < \beta_1 \leq M \quad \text{e} \quad \beta_1 - \alpha_1 = \frac{M-m}{2}$$

Consideriamo ora gli intervalli

$$\left[\alpha_1, \frac{\alpha_1 + \beta_1}{2} \right] \quad \text{e} \quad \left[\frac{\alpha_1 + \beta_1}{2}, \beta_1 \right]$$

almeno uno di essi contiene un numero infinito di termini di a_n sia esso $[\alpha_2, \beta_2]$ ovviamente si avrà

$$m \leq \alpha_1 \leq \alpha_2 < \beta_2 \leq \beta_1 \leq M \quad \text{e} \quad \beta_2 - \alpha_2 = \frac{M-m}{4}$$

Il procedimento descritto si può iterare e si ottengono così due successioni α_k e β_k soddisfacenti le seguenti proprietà:

$$(6.2) \quad m \leq \alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_k < \beta_k \leq \dots \leq \beta_2 \leq \beta_1 \leq M$$

$$(6.3) \quad \beta_k - \alpha_k = \frac{M-m}{2^k}$$

$$(6.4) \quad \{n : a_n \in [\alpha_k, \beta_k]\} \quad \text{ha infiniti elementi}$$

Possiamo pertanto concludere che le successioni α_k e β_k sono, rispettivamente, crescente e decrescente ed inoltre che sono entrambe limitate. Si ottiene pertanto che

$$\lim_k \alpha_k = \alpha \quad \text{e} \quad \lim_k \beta_k = \beta$$

ove

$$\alpha = \sup\{\alpha_k : k \in \mathbb{N}\} \in \mathbb{R} \quad \text{e} \quad \beta = \inf\{\beta_k : k \in \mathbb{N}\} \in \mathbb{R}$$

Per la 6.3 si ha che

$$\beta - \alpha = \lim_k (\beta_k - \alpha_k) = \lim_k \frac{M - m}{2^k} = 0$$

(si ricordi che è facile provare per induzione che $2^k \geq k$), e perciò si ha

$$\alpha = \beta = \lambda$$

in altre parole chiamiamo λ il valore comune di α e β .

Sia ora

$$\begin{array}{lll} n_1 \in \mathbb{N} & \text{tale che} & \alpha_1 \leq a_{n_1} \leq \beta_1 \\ n_2 \in \mathbb{N} & \text{tale che} & \alpha_2 \leq a_{n_2} \leq \beta_2, \quad n_2 > n_1 \\ & \dots\dots\dots & \\ n_k \in \mathbb{N} & \text{tale che} & \alpha_k \leq a_{n_k} \leq \beta_k, \quad n_k > n_{k-1} \end{array}$$

Allora $b_k = a_{n_k}$ è una successione estratta dalla successione a_n , la cui esistenza è assicurata dalla ?? ed inoltre si ha

$$\alpha_k \leq b_k \leq \beta_k \quad \forall k \in \mathbb{N}$$

e pertanto $\lim_k b_k = \lambda$. □

TEOREMA 6.6. *Sia a_n una successione:*

(1) *se a_n non è limitata superiormente, esiste b_k estratta da a_n tale che*

$$\lim_k b_k = +\infty$$

(2) *se a_n non è limitata inferiormente, esiste b_k estratta da a_n tale che*

$$\lim_k b_k = -\infty$$

DIMOSTRAZIONE. Proviamo ad esempio la prima affermazione. Sia

$$\begin{array}{lll} n_1 \in \mathbb{N} & \text{tale che} & a_{n_1} > 1 \\ n_2 \in \mathbb{N} & \text{tale che} & a_{n_2} > 2, \quad n_2 > n_1 \\ & \dots\dots\dots & \\ n_k \in \mathbb{N} & \text{tale che} & a_{n_k} > k, \quad n_k > n_{k-1} \end{array}$$

Allora $b_k = a_{n_k}$ (l'esistenza di tale estratta è assicurata dall'ipotesi che a_n non è limitata superiormente) tende a $+\infty$ □

Sappiamo bene cosa significa che una successione converge ad un limite $\ell \in \mathbb{R}^*$ cerchiamo ora di stabilire il significato del fatto che a_n non converge ad un limite $\ell \in \mathbb{R}^*$

LEMMA 6.2. *Sia a_n una successione e sia $\ell \in \mathbb{R}^*$; a_n non converge ad ℓ se e solo se esiste $\varepsilon_0 > 0$ ed esiste una successione b_k estratta da a_n tale che $b_k \notin I(\ell, \varepsilon_0)$.*

TEOREMA 6.7. *Sia a_n una successione soddisfacente la seguente proprietà:*

- *esiste $\ell \in \mathbb{R}^*$ tale che per ogni successione b_k estratta da a_n è possibile trovare una successione c_h estratta da b_k con*

$$\lim_h c_h = \ell$$

Allora si ha

$$\lim_n a_n = \ell$$

TEOREMA 6.8. - *Criterio di convergenza di Cauchy* - *Sia a_n una successione; sono fatti equivalenti:*

- (1) *esiste $\ell \in \mathbb{R}$ tale che*

$$\lim_n a_n = \ell$$

- (2) *Per ogni $\varepsilon > 0$ esiste $n_\varepsilon \in \mathbb{N}$ tale che se $n, m > n_\varepsilon$ si ha*

$$|a_n - a_m| < \varepsilon$$

DIMOSTRAZIONE.

- (1) \Rightarrow (2) sia $\varepsilon > 0$, allora esiste $n_\varepsilon \in \mathbb{N}$ tale che per $n > n_\varepsilon$ si ha

$$|a_n - \ell| < \varepsilon$$

Allora se $n, m > n_{\varepsilon/2}$ si ha

$$|a_n - a_m| \leq |a_n - \ell| + |a_m - \ell| < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

- (2) \Rightarrow (1) se $n > n_\varepsilon$ si ha

$$a_{n_\varepsilon} - \varepsilon < a_n < a_{n_\varepsilon} + \varepsilon$$

per cui la successione a_n è limitata.

Sia a_{n_k} una successione estratta da a_n tale che

$$\lim_k a_{n_k} = \ell$$

Se $k > k_\varepsilon$ si ha

$$|a_{n_k} - \ell| < \varepsilon$$

ed inoltre se $k > k_\varepsilon^1$ si ha

$$n_k > n_\varepsilon$$

Ma allora fissato $k > \max\{k_{\varepsilon/2}, k_{\varepsilon/2}^1\}$, se $n > n_{\varepsilon/2}$ si ha

$$|a_n - \ell| \leq |a_n - a_{n_k}| + |a_{n_k} - \ell| < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

□

E' di grande utilità per il seguito provare i due seguenti risultati.

LEMMA 6.3. Sia $A \subset \mathbb{R}$, $A \neq \emptyset$, e siano

$$\lambda = \sup A \quad , \quad \mu = \inf A$$

allora esistono due successioni $a_n, b_n \in A$ tali che

$$\lim_n a_n = \lambda \quad , \quad \lim_n b_n = \mu$$

DIMOSTRAZIONE. Proviamo ad esempio che esiste una successione $b_n \in A$ tale che

$$\lim_n b_n = \mu$$

Occorre distinguere due casi:

- (1) $\mu = -\infty$; in tal caso l'insieme A non è inferiormente limitato e pertanto per ogni $n \in \mathbb{N}$ esiste $b_n \in A$ tale che $b_n < -n$.

Ciò è sufficiente per concludere che

$$\lim_n b_n = -\infty$$

- (2) $\mu \in \mathbb{R}$; in tal caso si ha che:

$$\mu \leq b \quad \forall b \in A$$

$$\forall n \in \mathbb{N} \quad \exists b_n \in A : \mu + 1/n > b_n$$

Pertanto si ha:

$$\mu + 1/n > b_n \geq \mu$$

e la tesi.

□

DEFINIZIONE 6.3. Sia a_n una successione e sia

$$\Phi(a_n) = \{ \ell \in \mathbb{R}^* : \exists a_{n_k} , \lim_k a_{n_k} = \ell \}$$

Definiamo

$$\limsup_n a_n = \sup \Phi(a_n)$$

$$\liminf_n a_n = \inf \Phi(a_n)$$

Riguardo al massimo ed al minimo limite di una successione si possono provare molti risultati, non semplici, che non riportiamo.

LEMMA 6.4. Sia a_n una successione:

- (1) se $a_n > 0$ e $\lim_n \frac{a_{n+1}}{a_n} = \ell < 1$ allora $\lim_n a_n = 0$
- (2) se $a_n \geq 0$ e $\lim_n \sqrt[n]{a_n} = \ell < 1$ allora $\lim_n a_n = 0$;
- (3) se $a_n > 0$ e $\lim_n \frac{a_{n+1}}{a_n} = \ell > 1$ allora $\lim_n a_n = +\infty$;
- (4) se $a_n \leq 0$ e $\lim_n \sqrt[n]{a_n} = \ell > 1$ allora $\lim_n a_n = +\infty$

DIMOSTRAZIONE. Proviamo ad esempio la prima e l'ultima affermazione.

(1) Si ha, fissato $\varepsilon > 0$ in modo che $\ell + \varepsilon < 1$

$$\frac{a_{n+1}}{a_n} < \ell + \varepsilon \quad \forall n > n_\varepsilon$$

pertanto

$$a_{n+1} < a_n(\ell + \varepsilon)$$

e se $n > m > n_\varepsilon$

$$0 < a_n < (\ell + \varepsilon)^{n-m} a_m$$

e si può concludere che (1) è vera.

(4) Fissato ε in modo che $\ell - \varepsilon > 1$ si ha

$$(a_n)^{1/n} > (\ell - \varepsilon) \quad \forall n > n_\varepsilon$$

e

$$a_n > (\ell - \varepsilon)^n \quad \forall n > n_\varepsilon.$$

Ciò è sufficiente per concludere. □

DEFINIZIONE 6.4. Definiamo per $n \in \mathbb{N} \cup \{0\}$

$$0! = 1$$

$$n! = n(n-1)!$$

$n!$ verrà detto n fattoriale.

Osserviamo che, se $n \geq 1$,

$$n! = \prod_{i=1}^n i$$

Definiamo inoltre per $n \in \mathbb{N} \cup \{0\}$ nella seguente maniera:

$$(0)!! = 1$$

$$(1)!! = 1$$

$$n!! = n(n-2)!!$$

$n!!$ verrà indicato con il nome di n semifattoriale.

Osserviamo che

$$(2n)!! = \prod_{i=1}^n 2i \quad , \quad (2n+1)!! = \prod_{i=0}^n (2i+1)$$

Definiamo infine

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

coefficiente binomiale di ordine n e posto k e verrà detta n su k .

I coefficienti binomiali godono di notevoli proprietà: ad esempio possono essere calcolati usando il ben noto triangolo di Tartaglia e consentono di stabilire la formula della potenza di un binomio di Newton.

Si può provare con qualche calcolo che

$$(6.5) \quad \binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}$$

La precedente uguaglianza consente di costruire il così detto triangolo di Tartaglia:

$$\begin{array}{ccccccc}
 & & \binom{1}{0} & \binom{1}{1} & & & \\
 & & \binom{2}{0} & \binom{2}{1} & \binom{2}{2} & & \\
 & & \binom{3}{0} & \binom{3}{1} & \binom{3}{2} & & \binom{3}{3} \\
 & \dots & \dots & \dots & \dots & \dots & \dots \\
 & \binom{n}{0} & \dots & \binom{n}{k-1} & & \binom{n}{k} & \dots & \binom{n}{n} \\
 & \dots & \dots & \dots & \binom{n+1}{k} & \dots & \dots & \dots \\
 & \dots & \dots & \dots & \dots & \dots & \dots & \dots
 \end{array}$$

È importante osservare che ogni elemento del triangolo si può ottenere dalla somma dei due elementi della riga precedente, che occupano la posizione sopra e a sinistra della posizione occupata dall'elemento considerato.

Vale inoltre il seguente risultato che è noto con il nome di

binomio di Newton

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^{n-k} b^k$$

Ricordiamo anche la seguente disuguaglianza

LEMMA 6.5. *Se $a > -1$ allora*

$$(6.6) \quad (1+a)^n \geq 1+na$$

DIMOSTRAZIONE. Si prova per induzione;

(1) La disuguaglianza è banalmente vera per $n = 1$

(2) Inoltre se supponiamo $(1+a)^n \geq 1+na$ avremo che
 (6.7) $(1+a)^{n+1} = (1+a)^n(1+a) > (1+na)(1+a) = 1+(n+1)a+na^2 \geq 1+(n+1)a$

□

Possiamo ora studiare le proprietà di una successione di notevole importanza.

Sia E_n la successione definita da

$$E_n = \left(1 + \frac{1}{n}\right)^n$$

si ha che E_n è una successione strettamente crescente ed inoltre

$$2 \leq E_n < 3$$

Infatti si ha

$$E_n = \left(1 + \frac{1}{n}\right)^n = \sum_{k=0}^n \binom{n}{k} \frac{1}{n^k}$$

per cui

$$E_n \geq 1 + n(1/n) = 2$$

Per dimostrare che E_n è crescente osserviamo che si ha

$$E_n = \left(1 + \frac{1}{n}\right)^n \geq \left(1 + \frac{1}{n-1}\right)^{n-1} = E_{n-1}$$

se e solo se

$$\left(\frac{n+1}{n}\right)^n \geq \left(\frac{n}{n-1}\right)^n \left(\frac{n-1}{n}\right)$$

se e solo se

$$\left(\frac{n^2-1}{n^2}\right)^n \geq \left(\frac{n-1}{n}\right)$$

se e solo se

$$\left(1 - \frac{1}{n^2}\right)^n \geq 1 - \frac{1}{n}$$

e l'ultima disuguaglianza si deduce immediatamente dal lemma 6.5.

Infine, dal momento che si può facilmente provare per induzione che

$$(k+1)! \geq 2^k \quad \text{per } k \geq 0$$

si ha

$$(6.8) \quad E_n < \sum_{k=0}^n \frac{1}{k!} = 1 + \sum_{k=0}^{n-1} \frac{1}{(k+1)!} \leq 1 + \sum_{k=0}^{n-1} \frac{1}{2^k} < 3$$

Pertanto E_n è una successione crescente e limitata per cui possiamo affermare che

$$\lim_n E_n$$

esiste ed è reale e pertanto è lecito definire chiamare e il suo limite.

$$e = \lim_n E_n$$

Se x_n è una successione a termini positivi, $x_n \rightarrow x$ si può provare (si veda il capitolo successivo sulla continuità) che

$$\lim \log_a x_n = \log_a x$$

e pertanto si ha

$$\lim n \log_a \left(1 + \frac{1}{n}\right) = \log_a e$$

Osserviamo anche che

$$\log_a e = 1 \Leftrightarrow a = e$$

per cui si è naturalmente indotti a privilegiare il numero e come base per i logaritmi.

Si ha con 51 cifre decimali esatte

$$e = 2.718281828459045235360287471352662497757247093699959 .$$

DEFINIZIONE 6.5. Definiamo *logaritmo naturale* la funzione \log_e . Più semplicemente scriveremo $\log_e x = \ln x$.

Elenchiamo ora alcune successioni che saranno utili nel seguito e calcoliamone i relativi limiti:

- (1) $\lim n^k = +\infty$, $k > 0$
- (2) $\lim n^k = 0$, $k < 0$
- (3) $\lim a^n = +\infty$, $a > 1$
- (4) $\lim a^n = 0$, $|a| < 1$
- (5) $\lim \sqrt[n]{a} = 1$, $a > 0$
- (6) $\lim \sqrt[n]{n} = 1$
- (7) $\lim \frac{a^n}{n^n} = 0$

Possiamo verificare le affermazioni precedenti mediante le seguenti argomentazioni:

- (1) si prova mediante la definizione di limite.
- (2) si deduce dalla precedente tenendo conto che $n^k = 1/n^{-k}$.

- (3) sia $a = 1 + b$ con $b > 0$, allora $a^n = (1 + b)^n \geq 1 + nb$.
 (4) si deduce dalla precedente tenendo conto che $|a^n| = 1/(1/|a|)^n$.
 (5) se $a > 1$, posto

$$y_n = \sqrt[n]{a} - 1 > 0$$

si ha

$$a = (1 + y_n)^n \geq 1 + ny_n$$

da cui

$$0 \leq y_n \leq (a - 1)/n$$

se $0 < a < 1$ si ha $\sqrt[n]{a} = 1/\sqrt[n]{1/a}$.

(6) posto

$$y_n = \sqrt[n]{n} - 1 > 0$$

si ha

$$n = (1 + y_n)^n \geq 1 + n(n - 1)y_n^2/2 \text{ e } 0 \leq y_n \leq (2/n)$$

(7) Se $n > 2|a|$ si ha

$$0 < |a|/n < 1/2$$

per cui

$$0 < |a|^n/n^n < (1/2)^n$$

Valgono i seguenti fatti:

- (1) $\lim \frac{a^n}{n!} = 0$
 (2) $\lim \frac{n!}{n^n} = 0$
 (3) $\lim \frac{a^n}{n^k} = 0, |a| < 1$
 (4) $\lim \frac{a^n}{n^k} = +\infty, a > 1$
 (5) $\lim \frac{\log_a n}{n^k} = 0, k > 0, a > 0, a \neq 1$

Infatti se indichiamo con b_n ciascuna delle successioni in oggetto si ha

- (1) $\frac{|b_{n+1}|}{|b_n|} = \frac{|a|}{n+1} \longrightarrow 0$
 (2) $\frac{b_{n+1}}{b_n} = \left(\frac{n}{n+1}\right)^n \longrightarrow \frac{1}{e} < 1$
 (3) $\frac{|b_{n+1}|}{|b_n|} = |a| \left(\frac{n}{n+1}\right)^k \longrightarrow |a|$
 (4) si può facilmente dedurre dal fatto che

$$\frac{\log_a n}{n^k} = \frac{1}{k} \log_a (\sqrt[k]{n^k})$$

Anche se a prima vista ciò non appare verosimile, operare con successioni piuttosto che con funzioni è molto più comodo e facile; è pertanto molto utile provare il seguente risultato che permette di ottenere informazioni sul limite di una funzione utilizzando opportune successioni.

TEOREMA 6.9. Sia $f : D \longrightarrow \mathbb{R}$ e siano $x_0 \in \mathcal{D}(D)$, $\ell \in \mathbb{R}^*$; sono fatti equivalenti:

- (1) $\lim_{x \rightarrow x_0} f(x) = \ell$
 (2) per ogni $x_n \in D \setminus \{x_0\}$, $x_n \rightarrow x_0$, si ha

$$\lim f(x_n) = \ell$$

DIMOSTRAZIONE. Se vale la prima condizione avremo che per ogni $\varepsilon > 0$ esiste $\delta_\varepsilon > 0$ tale che, se $x \in I^0(x_0, \delta_\varepsilon)$ si ha

$$f(x) \in I(\ell, \varepsilon)$$

Inoltre esiste $n_\varepsilon \in \mathbb{N}$ tale che per $n > n_\varepsilon$ si abbia $x_n \in I^0(x_0, \delta_\varepsilon)$ e di conseguenza si ha

$$f(x_n) \in I(\ell, \varepsilon)$$

Da cui la seconda asserzione.

Se viceversa la prima asserzione è falsa, allora esiste $\varepsilon_0 > 0$ tale che, per ogni $n \in \mathbb{N}$ esiste $x_n \in D$, $x_n \neq x_0$, con $x_n \in I^0(x_0, 1/n)$ e

$$f(x_n) \notin I(\ell, \varepsilon_0)$$

e quindi la seconda è falsa

□

Si può provare che ogni successione convergente ammette una successione estratta monotona (e convergente allo stesso limite); pertanto nel verificare (2) è sufficiente limitarsi alle sole successioni monotone.

Il criterio di convergenza di Cauchy riveste notevole importanza ed è utile sapere che esso può essere provato anche per le funzioni nella seguente forma.

TEOREMA 6.10. - *Criterio di Cauchy* - Sia $f : D \longrightarrow \mathbb{R}$ e sia $x_0 \in \mathcal{D}(D)$; sono condizioni equivalenti:

- (1) esiste $\ell \in \mathbb{R}$ tale che $\lim_{x \rightarrow x_0} f(x) = \ell$
 (2) per ogni $\varepsilon > 0$ esiste $\delta_\varepsilon > 0$ tale che se $x, y \in I^0(x_0, \delta_\varepsilon)$ si ha

$$|f(x) - f(y)| < \varepsilon$$

DIMOSTRAZIONE. (1) \Rightarrow (2) Per ogni $\varepsilon > 0$ esiste $\delta_\varepsilon > 0$ tale che se $x \in I^0(x_0, \delta_{\varepsilon/2})$ si ha

$$|f(x) - \ell| < \varepsilon/2$$

per cui se $x, y \in I^0(x_0, \delta_{\varepsilon/2})$ si ha

$$|f(x) - f(y)| \leq |f(x) - \ell| + |f(y) - \ell| < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

(2) \Rightarrow (1) Se per assurdo (1) non fosse vera esisterebbero due successioni $x_n, y_n \in D$, convergenti ad x_0 tali che $f(x_n) \longrightarrow \ell_1$ e $f(y_n) \longrightarrow \ell_2$ con $\ell_1, \ell_2 \in \mathbb{R}^*$, $\ell_1 \neq \ell_2$.

Pertanto la condizione (2) non potrebbe essere soddisfatta. □

Mediante le successioni siamo anche in grado di provare i seguenti risultati che caratterizzano gli insiemi aperti, chiusi e compatti in \mathbb{R} e che sono facilmente estendibili a più generali situazioni.

TEOREMA 6.11. Sia $A \subset \mathbb{R}$, allora

- (1) A è aperto se e solo se per ogni $x \in A$ e per ogni successione x_n , tale che $x_n \rightarrow x$, si ha $x_n \in A$ definitivamente;
- (2) A è chiuso se e solo se per ogni $x_n \in A$, $x_n \rightarrow x$ si ha $x \in A$
- (3) A è un insieme compatto se e solo se per ogni $x_n \in A$ esiste $x_{n_k} \rightarrow x$, tale che $x \in A$

1. Infinitesimi ed Infiniti

Se $f(x) \rightarrow \ell$, $\ell \in \mathbb{R}$ allora $f(x) - \ell \rightarrow 0$ per cui per studiare il comportamento di una funzione che ammette limite finito sarà sufficiente considerare funzioni che tendono a 0; tali funzioni si definiscono infinitesime ed è importante cercare di ottenere qualche informazione in più su come una funzione infinitesima tende a 0.

Ad esempio è evidente che x^n diminuisce più o meno velocemente, in dipendenza da n , quando ci si avvicina a 0. È quindi ovvio che sia utile cercare di individuare anche in funzioni più complesse tali comportamenti.

Quanto detto per le funzioni infinitesime si può poi facilmente estendere anche alle funzioni che tendono all'infinito: che chiameremo infinite.

Pertanto introduciamo la definizione di ordine di infinitesimo e di ordine di infinito.

DEFINIZIONE 6.6. Sia $f : (a, b) \rightarrow \mathbb{R}$, diciamo che f è infinitesima in a^+ se

$$\lim_{x \rightarrow a^+} f(x) = 0$$

In maniera analoga si possono dare le definizioni per $x \rightarrow a^-$, $x \rightarrow a$, $x \rightarrow +\infty$ e $x \rightarrow -\infty$.

DEFINIZIONE 6.7. Siano f, g due funzioni infinitesime in a^+ e supponiamo che

$$\lim_{x \rightarrow a^+} \frac{f(x)}{g(x)} = \ell$$

- se $\ell \in \mathbb{R} \setminus \{0\}$ diciamo che f e g hanno lo stesso ;
- se $\ell = 0$ diciamo che f è infinitesima di ordine superiore a g .

DEFINIZIONE 6.8. Chiamiamo infinitesimo campione di ordine $\alpha \in \mathbb{R}_+$ in $a^+, a^-, a, +\infty, -\infty$ rispettivamente la funzione

$$(x - a)^\alpha, (a - x)^\alpha, |x - a|^\alpha, \frac{1}{x^\alpha}, \frac{1}{(-x)^\alpha}$$

Si dice che f è infinitesimo di ordine $\alpha \in \mathbb{R}_+$ se f ha lo stesso ordine dell'infinitesimo campione di ordine α .

Osserviamo esplicitamente che può accadere che f non abbia ordine di infinitesimo reale.

Ad esempio la funzione

$$\frac{1}{\ln x}$$

è infinitesima per $x \rightarrow +\infty$ di ordine inferiore ad ogni $\alpha \in \mathbb{R}_+$.

Infatti per ogni $\alpha \in \mathbb{R}_+$

$$(6.9) \quad \lim_{x \rightarrow +\infty} \frac{\frac{1}{\ln x}}{\frac{1}{x^\alpha}} = 0$$

La definizione di ordine di infinitesimo consente di provare che

TEOREMA 6.12. *Siano f e g due funzioni infinitesime in a^+ di ordine α e β rispettivamente; allora*

- (1) fg ha ordine $\alpha + \beta$
- (2) se $\alpha < \beta$, $f + g$ ha ordine α
- (3) se $\alpha = \beta$, $f + g$ ha ordine maggiore o uguale ad α .

DEFINIZIONE 6.9. *Diciamo che f è infinita in a^+ se $1/f$ è infinitesima in a^+ .*

Diciamo che f è infinita di ordine α se $1/f$ è infinitesima di ordine α .

TEOREMA 6.13. *Siano f e g due funzioni infinite in a^+ di ordine α e β rispettivamente; allora*

- (1) fg ha ordine $\alpha + \beta$;
- (2) se $\alpha < \beta$, $f + g$ ha ordine β ;
- (3) se $\alpha = \beta$, $f + g$ ha ordine minore o uguale ad α

Osserviamo che si potrebbe definire f di ordine $\alpha \in \mathbb{R}$ in a^+ se

$$\lim_{x \rightarrow a^+} \frac{f(x)}{(x - a)^\alpha} \in \mathbb{R} \setminus \{0\}$$

ed osservare che f è infinitesima se $\alpha > 0$ ed è infinita se $\alpha < 0$.

Con queste convenzioni si può provare che se f ha ordine α e g ha ordine β , allora f/g ha ordine $\alpha - \beta$.

CAPITOLO 7

LA CONTINUITÀ

La maggior parte delle situazioni semplici che cerchiamo di rappresentare mediante l'uso di una funzione reale di una variabile reale presentano una caratteristica comune:

**piccoli cambiamenti della variabile (argomento della funzione)
causano piccoli cambiamenti dei valori della funzione stessa.**

Ad esempio se $T(x)$ rappresenta la temperatura di una sbarra di metallo in un punto che dista x da una delle sue estremità, ci aspettiamo che due punti vicini sulla sbarra abbiano temperature non molto dissimili.

Tuttavia non tutti i fenomeni sono facilmente rappresentabili mediante funzioni continue; se ad esempio $L(t)$ rappresenta la luminosità di una stanza nella quale si accende una lampada all'istante t_0 è evidente che in quell'istante il valore della luminosità può subire una brusca variazione, (se la luminosità della lampada è alta in confronto con la luminosità ambiente).

Anche nel linguaggio comune è naturale attribuire l'aggettivo continuo al primo fenomeno ma non al secondo.

In parole povere, una funzione è continua in un punto se il valore che essa assume in tale punto dipende dai valori da essa assunti nei punti vicini, o per meglio dire, se piccole variazioni dell'argomento danno luogo a piccole variazioni dei corrispondenti valori della funzione.

In altri termini una funzione è continua se non ammette repentini cambiamenti, salti, "discontinuità".

Vogliamo allora formalizzare cosa si intende per continuità di una funzione.

Precisamente poniamo la seguente definizione

DEFINIZIONE 7.1. Sia $f : D \longrightarrow \mathbb{R}$, $x_0 \in D$, diciamo che f è continua in x_0 se

$$\forall \varepsilon > 0 \exists \delta_\varepsilon > 0 \text{ tale che se } |x - x_0| < \delta_\varepsilon \quad , \quad x \in D \text{ si ha}$$
$$|f(x) - f(x_0)| < \varepsilon$$

Diciamo che f è continua in D se è continua in ogni punto di D .

E' immediato verificare l'analogia, ma non l'identità, con la definizione di limite, ed è immediato provare che:

TEOREMA 7.1. *Sia $f : D \rightarrow \mathbb{R}$ e sia $x_0 \in D \cap \mathcal{D}(D)$; sono fatti equivalenti:*

- (1) f è continua in x_0 ,
- (2) $\lim_{x \rightarrow x_0} f(x) = f(x_0)$.

I teoremi sui limiti consentono di stabilire alcuni semplici risultati che ci limitiamo ad enunciare.

TEOREMA 7.2. *Siano $f, g : D \rightarrow \mathbb{R}$ continue in $x_0 \in D$, siano inoltre $\alpha, \beta \in \mathbb{R}$; si ha*

- (1) $\alpha f + \beta g$ è continua in x_0 ;
- (2) fg è continua in x_0 ;
- (3) se $f(x_0) \neq 0$, $1/f$ è continua in x_0 .

TEOREMA 7.3. *Sia $f : D \rightarrow \mathbb{R}$, $x_0 \in D$; sono fatti equivalenti:*

- (1) f è continua in x_0 ;
- (2) $\forall x_n \in D, x_n \rightarrow x_0$, si ha $f(x_n) \rightarrow f(x_0)$.

Il precedente teorema consente di caratterizzare la continuità per successioni: nel confronto con il teorema 6.9 mediante il quale sono caratterizzati i limiti si evidenzia il fatto che:

per caratterizzare il limite la successione x_n deve essere scelta in $D \setminus \{x_0\}$, mentre per la continuità x_n assume valori in D .

Osserviamo anche che, come nel teorema 6.9, ci si può limitare a considerare soltanto successioni monotone.

TEOREMA 7.4. *Sia $f : D \rightarrow \mathbb{R}$ continua in $x_0 \in D$, allora se $f(x_0) \neq 0$ esiste $\delta > 0$ tale che se $x \in D \cap I(x_0, \delta)$ si ha $f(x)f(x_0) > 0$.*

TEOREMA 7.5. *Siano $f : D \rightarrow \mathbb{R}$ continua in $x_0 \in D$, $g : A \rightarrow D$ continua in $t_0 \in A$, $x_0 = g(t_0)$; allora $f(g(\cdot)) : A \rightarrow \mathbb{R}$ è continua in t_0 .*

CAPITOLO 8

I TEOREMI SULLA CONTINUITÀ

Dopo questa rapida rassegna di risultati passiamo a studiare le proprietà più importanti ed interessanti delle funzioni continue in un insieme.

La maggior parte delle proprietà che studieremo riguardano le funzioni continue su di un intervallo chiuso e limitato. E' facile vedere, mediante esempi, che se si considerano funzioni continue su insiemi che non soddisfano i requisiti opportuni, tali proprietà possono non essere soddisfatte.

TEOREMA 8.1. - degli zeri - Sia $f : [a, b] \longrightarrow \mathbb{R}$ una funzione continua e supponiamo che $f(a)f(b) < 0$.

Allora esiste $x_0 \in (a, b)$ tale che $f(x_0) = 0$.

DIMOSTRAZIONE.

Definiamo le successioni α_n e β_n nella seguente maniera:

$$(8.1) \quad [\alpha_0, \beta_0] = [a, b]$$

$$(8.2) \quad [\alpha_{n+1}, \beta_{n+1}] = \begin{cases} [\alpha_n, (\alpha_n + \beta_n)/2] & \text{se } f(\alpha_n)f((\alpha_n + \beta_n)/2) < 0 \\ [(\alpha_n + \beta_n)/2, \beta_n] & \text{se } f(\alpha_n)f((\alpha_n + \beta_n)/2) > 0 \\ [(\alpha_n + \beta_n)/2, (\alpha_n + \beta_n)/2] & \text{se } f((\alpha_n + \beta_n)/2) = 0 \end{cases}$$

Se, esiste \bar{n} , $f((\alpha_{\bar{n}} + \beta_{\bar{n}})/2) = 0$ si è trovato lo zero;
in caso contrario, per α_n e β_n si ha:

•

$$(8.3) \quad \alpha_n \text{ è crescente, } \beta_n \text{ è decrescente,}$$

•

$$(8.4) \quad \alpha_n, \beta_n \in [a, b], f(\alpha_n)f(\beta_n) < 0$$

•

$$(8.5) \quad \beta_n - \alpha_n = \frac{b - a}{2^n}$$

Pertanto si può affermare che

$$(8.6) \quad \alpha_n \nearrow \alpha \quad \beta_n \searrow \beta \quad \alpha, \beta \in [a, b]$$

e dalla 8.5 si ricava $\alpha = \beta = c$.

Per la continuità di f e per 8.4 si ha

$$(8.7) \quad 0 \geq \lim f(\alpha_n)f(\beta_n) = (f(c))^2 \quad \text{e} \quad f(c) = 0$$

Si ha anche che $c \in (a, b)$ ed inoltre

$$(8.8) \quad 0 \leq c - \alpha_n \leq \beta_n - \alpha_n \leq \frac{b-a}{2^n}$$

e

$$(8.9) \quad 0 \leq \beta_n - c \leq \beta_n - \alpha_n \leq \frac{b-a}{2^n}$$

□

Il teorema 8.1 ammette come immediato corollario il seguente:

TEOREMA 8.2. - *dei valori intermedi* - Sia $f : [a, b] \rightarrow \mathbb{R}$ una funzione continua e siano $c, d \in R(f)$, $c < d$, allora

$$[c, d] \subset R(f).$$

DIMOSTRAZIONE. Siano $\alpha, \beta \in [a, b]$ tali che $f(\alpha) = c$, $f(\beta) = d$ e consideriamo $y \in (c, d)$; la funzione

$$g(x) = f(x) - y$$

è continua su $[\alpha, \beta]$, $g(\alpha) < 0$, $g(\beta) > 0$ e perciò, per il teorema 8.1, esiste $x_0 \in (\alpha, \beta)$ tale che

$$g(x_0) = f(x_0) - y = 0$$

da cui $f(x_0) = y$ e $y \in R(f)$. □

COROLLARIO 8.1. Sia $f : [a, b] \rightarrow \mathbb{R}$, se f è continua allora $R(f)$ è un intervallo.

Ci proponiamo ora di dimostrare un teorema di esistenza del massimo per una funzione continua su un insieme compatto (cioè chiuso e limitato).

TEOREMA 8.3. - *Weierstraß* - Sia $f : D \rightarrow \mathbb{R}$ una funzione continua, D compatto; allora esistono $\alpha, \beta \in D$ tali che

$$f(\alpha) = \min\{f(x) : x \in D\}$$

$$f(\beta) = \max\{f(x) : x \in D\}.$$

DIMOSTRAZIONE. Proviamo ad esempio l'esistenza del minimo della funzione f . Sia

$$\lambda = \inf\{f(x) : x \in D\} = \inf R(f);$$

per il lemma 6.3 esiste $y_n \in R(f)$ tale che $y_n \rightarrow \lambda$.

Sia $x_n \in D$ tale che $y_n = f(x_n)$; dal momento che D è compatto esiste x_{n_k} estratta da x_n tale che

$$x_{n_k} \rightarrow \alpha \in D.$$

Pertanto

$$y_{n_k} = f(x_{n_k}) \rightarrow f(\alpha)$$

per la continuità di f ed anche $y_{n_k} \rightarrow \lambda$ da cui

$$\lambda = f(\alpha)$$

e la tesi. \square

E' possibile generalizzare il teorema 8.3 senza l'ipotesi di compattezza dell'insieme D , ad esempio possiamo provare:

TEOREMA 8.4. *Sia $f : (a, b) \rightarrow \mathbb{R}$ continua, $a, b \in \mathbb{R}^*$, e supponiamo che esista $\bar{x} \in (a, b)$ tale che*

$$\lim_{x \rightarrow a^+} f(x), \lim_{x \rightarrow b^-} f(x) > f(\bar{x})$$

allora esiste $\alpha \in (a, b)$ tale che

$$f(\alpha) = \min\{f(x) : x \in (a, b)\}.$$

DIMOSTRAZIONE. Supponiamo $a, b \in \mathbb{R}$, essendo la dimostrazione negli altri casi analoga.

Siano

$$\lambda = \inf\{f(x) : x \in (a, b)\}$$

si ha

$$\mu = f(\bar{x}) \geq \lambda, \quad \mu \in \mathbb{R}$$

Se $\lambda = f(\bar{x})$ avremmo che il minimo è assunto.

sia $\delta > 0$ tale che

$$x \in (a, a + \delta) \cup (b - \delta, b) \Rightarrow f(x) > \mu.$$

Sia

$$y_n \in R(f), \quad y_n \rightarrow \lambda$$

poiché $\mu > \lambda$, definitivamente si ha $y_n \leq \mu$ e perciò esiste $x_n \in [a + \delta, b - \delta]$ tale che $f(x_n) = y_n$.

Ne segue che esiste $x_{n_k} \rightarrow \alpha \in [a + \delta, b - \delta]$ e

$$y_{n_k} = f(x_{n_k}) \rightarrow f(\alpha).$$

\square

A questo punto sarebbe ragionevole introdurre il concetto di uniforme continuità, tuttavia poichè si tratta di un concetto fondamentale ma difficile da comprendere rimandiamo chi fosse interessato a quanto è contenuto negli approfondimenti.

In parole povere diciamo che una funzione è uniformemente continua su un intervallo $[a, b]$, se, nella definizione di continuità applicata ad un punto $x \in [a, b]$, il valore di δ si può trovare in funzione di ϵ , ma non dipende anche da x .

Possiamo cioè dire che in un qualunque punto di $x_0 \in [a, b]$ il modo con cui $f(x)$ si avvicina ad $f(x_0)$ quando x si avvicina ad x_0 è in questo senso uniforme.

Abbiamo a suo tempo dimostrato che, se una funzione è strettamente monotona, allora essa è invertibile; vediamo ora che se ci restringiamo alla

classe delle funzioni continue, la stretta monotonia è anche necessaria per l'invertibilità.

TEOREMA 8.5. *Sia $f : [a, b] \longrightarrow \mathbb{R}$ continua, allora f è invertibile se e solo se f è strettamente monotona.*

DIMOSTRAZIONE. Ci limitiamo a provare la parte 'solo se', in quanto la parte 'se' è già stata provata nel teorema 3.15.

Se per assurdo f non fosse monotona, per il teorema 3.14

$$(8.10) \quad \exists x, y, z \in [a, b] : x < y < z, [f(y) - f(x)][f(z) - f(y)] \leq 0$$

se nella 8.10 vale l'uguaglianza, f non è invertibile; se vale la disuguaglianza stretta possiamo, per fissare le idee, supporre che

$$f(z) - f(y) < 0, \quad f(y) - f(x) > 0, \quad f(z) > f(x).$$

Allora, per il teorema dei valori intermedi, poichè $f(x) < f(z) < f(y)$

$$\exists \alpha \in (x, y) : f(\alpha) = f(z)$$

e ciò è contro l'invertibilità di f . □

Per concludere con la continuità proviamo i seguenti risultati.

TEOREMA 8.6. *Sia $f : (a, b) \longrightarrow \mathbb{R}$, strettamente monotona, siano $x_0 \in (a, b)$ e $y_0 = f(x_0)$; allora f^{-1} è continua in y_0 .*

DIMOSTRAZIONE. Supponiamo, per fissare le idee, che f sia strettamente crescente e proviamo che, per ogni $\varepsilon > 0$ esiste $\delta_\varepsilon > 0$ tale che

$$\forall y : |y - y_0| < \delta_\varepsilon, \text{ si ha } |f^{-1}(y) - f^{-1}(y_0)| < \varepsilon.$$

Sia $\varepsilon_0 > 0$ tale che $(x_0 - \varepsilon_0, x_0 + \varepsilon_0) \subset (a, b)$.

Dal momento che f è strettamente crescente si ha, per $0 < \varepsilon \leq \varepsilon_0$

$$f(x_0 - \varepsilon) < y_0 < f(x_0 + \varepsilon).$$

Definiamo

$$\delta_\varepsilon = \min\{y_0 - f(x_0 - \varepsilon), f(x_0 + \varepsilon) - y_0\} > 0;$$

se $|y - y_0| < \delta_\varepsilon$ si ha

$$y_0 + f(x_0 - \varepsilon) - y_0 < y_0 - \delta_\varepsilon < y < y_0 + \delta_\varepsilon < y_0 + f(x_0 + \varepsilon) - y_0$$

e

$$f(x_0 - \varepsilon) < y < f(x_0 + \varepsilon).$$

Poiché f^{-1} è strettamente crescente si ha

$$f^{-1}(f(x_0 - \varepsilon)) < f^{-1}(y) < f^{-1}(f(x_0 + \varepsilon))$$

cioè

$$\begin{aligned} x_0 - \varepsilon &< f^{-1}(y) < x_0 + \varepsilon \\ f^{-1}(y_0) - \varepsilon &< f^{-1}(y) < f^{-1}(y_0) + \varepsilon \end{aligned}$$

e

$$|f^{-1}(y) - f^{-1}(y_0)| < \varepsilon.$$

□

TEOREMA 8.7. *Sia $f : (a, b) \longrightarrow \mathbb{R}$ continua e invertibile; allora f^{-1} è continua.*

DIMOSTRAZIONE. Dal momento che f è continua ed invertibile su (a, b) , essa è strettamente monotona e si può applicare il teorema 8.6. □

Si può verificare che:

- (1) p_a , per $a \in \mathbb{R}$ è continua in \mathbb{R}_+
- (2) p_n , per $n \in \mathbb{N}$ è continua in \mathbb{R}
- (3) \exp_a , per $a \in \mathbb{R}_+$ è continua in \mathbb{R}
- (4) \sin è continua in \mathbb{R}
- (5) \cos è continua in \mathbb{R}
- (6) r_n , per n pari è continua in $\overline{\mathbb{R}}_+$
- (7) r_n , per n dispari è continua in \mathbb{R}
- (8) \log_a , per $a \in \mathbb{R}_+ \setminus \{1\}$ è continua in \mathbb{R}_+ .

La verifica di questi fatti può essere completata usando la definizione di limite.

possiamo altresì verificare, usando il teorema dei valori intermedi che

TEOREMA 8.8. *Si ha*

- (1) $R(p_n) = \overline{\mathbb{R}}_+$, per n pari
- (2) $R(p_n) = \mathbb{R}$, per n dispari
- (3) $R(\exp_a) = \mathbb{R}_+$ per $a \in \mathbb{R}_+$, $a \neq 1$
- (4) $R(p_a) = \mathbb{R}_+$ per $a \in \mathbb{R}$, $a \neq 0$
- (5) $R(\sin) = [-1, 1]$
- (6) $R(\cos) = [-1, 1]$
- (7) $R(\tan) = \mathbb{R}$

CAPITOLO 9

LA DERIVABILITÀ.

Consideriamo una funzione f continua in un punto x_0 , avremo che, quando x si discosta di poco da x_0 , $f(x)$ è poco distante da $f(x_0)$.

È in questo caso importante valutare come varia $f(x) - f(x_0)$ in rapporto a $x - x_0$ cioè il valore del rapporto

$$(9.1) \quad \frac{f(x) - f(x_0)}{x - x_0}$$

Possiamo vedere che 9.1 rappresenta il coefficiente angolare della corda che passa per i punti $(x_0, f(x_0))$ e $(x, f(x))$.

Se x è vicino al punto x_0 il denominatore tende a 0, ma se f è continua anche il numeratore tende a 0 e quindi è significativo considerare il valore limite di 9.1 per $x \rightarrow x_0$.

Si stabilisce quindi la seguente definizione.

DEFINIZIONE 9.1. Sia $f : (a, b) \longrightarrow \mathbb{R}$, $x_0 \in (a, b)$;

$$g(x) = \frac{f(x) - f(x_0)}{x - x_0}$$

è definito per ogni $x \in (a, b) \setminus \{x_0\}$ e si chiama rapporto incrementale relativo alla funzione f nel punto x_0 .

Dal momento che x_0 è in (a, b) ha senso considerare

$$\lim_{x \rightarrow x_0} g(x).$$

Diciamo che f è derivabile in x_0 se

$$\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

esiste finito.

In tal caso chiamiamo il suo valore derivata di f in x_0 e scriviamo

$$f'(x_0) \quad \text{o} \quad \frac{df}{dx}(x_0)$$

Diciamo che f è derivabile in x_0 da destra oppure da sinistra se

$$\lim_{x \rightarrow x_0^+} \frac{f(x) - f(x_0)}{x - x_0}$$

oppure

$$\lim_{x \rightarrow x_0^-} \frac{f(x) - f(x_0)}{x - x_0}$$

esiste ed è finito,

In tal caso chiamiamo tale limite derivata destra o derivata sinistra di f in x_0 e scriviamo $f'_+(x_0)$ o $f'_-(x_0)$, ovvero

$$\frac{d^+ f}{dx}(x_0) \quad \text{o} \quad \frac{d^- f}{dx}(x_0).$$

Diciamo che f è derivabile in (a, b) se è derivabile in ogni punto di (a, b) . In tal caso possiamo definire una funzione

$$f' = \frac{df}{dx} : (a, b) \longrightarrow \mathbb{R}$$

che si chiama derivata di f .

In maniera del tutto analoga si possono definire le funzioni derivata destra e derivata sinistra.

Osserviamo che $f'(x)$ è il limite per x che tende a x_0 del coefficiente angolare della corda secante il grafico di f nei punti $(x, f(x))$, $(x_0, f(x_0))$ e che pertanto è ragionevole supporre che sia il coefficiente angolare della retta tangente al grafico in $(x_0, f(x_0))$.

La derivata di f , fornisce, vicino ad x_0 , una stima del variare di $f(x) - f(x_0)$ rispetto a $x - x_0$.

Poichè

$$(9.2) \quad \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = f'(x)$$

si ha

$$\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} - f'(x) = 0$$

da cui

$$(9.3) \quad \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0) - f'(x_0)(x - x_0)}{x - x_0} = 0$$

Se ora poniamo

$$(9.4) \quad \frac{f(x) - f(x_0) - f'(x_0)(x - x_0)}{x - x_0} = \omega(x - x_0)$$

avremo che

$$(9.5) \quad f(x) - (f(x_0) + f'(x_0)(x - x_0)) = \omega(x - x_0)(x - x_0)$$

con

$$\lim_{x \rightarrow x_0} \omega(x - x_0) = 0$$

In altre parole, alla quantità $f(x)$ è possibile sostituire la quantità

$$(9.6) \quad f(x_0) + f'(x_0)(x - x_0)$$

commettendo un errore

$$\omega(x - x_0)(x - x_0)$$

che tende a 0 più velocemente di $(x - x_0)$

Poichè l'equazione $y = f(x_0) + f'(x_0)(x - x_0)$ rappresenta una retta nel piano con la proprietà di approssimare $f(x)$ con un errore infinitesimo di ordine superiore al primo, per $x \rightarrow x_0$, possiamo definirla retta tangente al grafico di f nel punto x_0 .

Resta così giustificato l'uso di $f'(x_0)$ per identificare il coefficiente angolare della retta tangente al grafico di f in x_0 .

Definiamo pertanto, allo scopo di sviluppare questa idea, la differenziabilità di una funzione.

DEFINIZIONE 9.2. Sia $f : (a, b) \longrightarrow \mathbb{R}$, $x_0 \in (a, b)$; diciamo che f è differenziabile in x_0 se esiste $a \in \mathbb{R}$ tale che

$$\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0) - a(x - x_0)}{x - x_0} = 0.$$

La funzione lineare $L(x) = a(x - x_0)$ si chiama differenziale di f in x_0 e si indica con $df(x_0)(x)$.

TEOREMA 9.1. Sia $f : (a, b) \longrightarrow \mathbb{R}$, $x_0 \in (a, b)$, allora f è derivabile in x_0 se e solo se f è differenziabile in x_0 .

Inoltre

$$df(x_0)(h) = hf'(x_0)$$

per ogni $h \in \mathbb{R}$.

DIMOSTRAZIONE. Se f è derivabile in x_0 è sufficiente definire

$$a = f'(x_0)$$

e si ha

$$\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0) - a(x - x_0)}{x - x_0} = 0.$$

Se viceversa f è differenziabile in x_0 si ha che

$$\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0) - a(x - x_0)}{x - x_0} = 0$$

e

$$\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} - a = 0$$

da cui

$$f'(x_0) = a$$

e

$$df(x_0)(h) = hf'(x_0)$$

□

Dalla 9.5 risulta evidente che se f è derivabile in x_0 si ha:

$$f(x) - f(x_0) = f'(x_0)(x - x_0) + (x - x_0)\omega(x - x_0).$$

Pertanto

$$\lim_{x \rightarrow x_0} [f(x) - f(x_0)] = \lim_{x \rightarrow x_0} [f'(x_0)(x - x_0) + (x - x_0)\omega(x - x_0)] = 0$$

e

$$\lim_{x \rightarrow x_0} f(x) = f(x_0).$$

Si è così provato che

TEOREMA 9.2. *Sia $f : D \rightarrow \mathbb{R}$, $D \subset \mathbb{R}$ aperto, e sia $x_0 \in D$; se f è derivabile in x_0 allora f è continua in x_0 .*

Non è però vero il viceversa; esempi che illustrino questo fatto non sono difficili a trovarsi (basta considerare $f(x) = |x|$), e di più è possibile costruire una funzione continua su un intervallo ed ivi mai derivabile.

In virtù del teorema 9.1 d'ora in poi, per le funzioni di una variabile, useremo indifferentemente i termini derivabilità e differenziabilità; useremo inoltre, per caratterizzare questa proprietà una qualunque delle condizioni enunciate nelle 9.2, 9.3, 9.5.

Proviamo ora alcuni risultati sulla derivabilità.

TEOREMA 9.3. *Siano $f, g : D \rightarrow \mathbb{R}$, $D \subset \mathbb{R}$ aperto, e sia $x_0 \in D$; supponiamo che f e g siano derivabili in x_0 , allora:*

(1) $\alpha f + \beta g$ è derivabile in $x_0 \forall \alpha, \beta \in \mathbb{R}$ e si ha

$$(\alpha f + \beta g)'(x_0) = \alpha f'(x_0) + \beta g'(x_0);$$

(2) fg è derivabile in x_0 e si ha

$$(fg)'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0).$$

(3) Se $f(x_0) \neq 0$ allora $(1/f)$ è derivabile in x_0 e si ha:

$$\left(\frac{1}{f}\right)'(x_0) = -\frac{f'(x_0)}{f^2(x_0)}.$$

DIMOSTRAZIONE.

(1) è banale conseguenza della definizione di derivata e dei risultati provati sui limiti.

(2) si può provare osservando che

$$\frac{f(x)g(x) - f(x_0)g(x_0)}{x - x_0} = \frac{f(x)[g(x) - g(x_0)]}{x - x_0} + \frac{[f(x) - f(x_0)]g(x_0)}{x - x_0}$$

e passando al limite.

(3) Dal momento che f è derivabile in x_0 , f è ivi continua e per il teorema della permanenza del segno $\exists \delta > 0$ tale che $f(x) \neq 0$ se $|x - x_0| < \delta$.

Possiamo pertanto considerare la funzione $1/f$ in $|x - x_0| < \delta$ e costruire il suo rapporto incrementale

$$\frac{1/f(x) - 1/f(x_0)}{x - x_0} = -\frac{\frac{f(x) - f(x_0)}{x - x_0}}{\frac{1}{f(x)f(x_0)}}.$$

Passando al limite per $x \rightarrow x_0$ si ottiene che

$$\left(\frac{1}{f}\right)'(x_0) = -\frac{f'(x_0)}{f^2(x_0)}$$

□

TEOREMA 9.4. Siano $f : (a, b) \rightarrow \mathbb{R}$, $g : (c, d) \rightarrow (a, b)$; sia $t_0 \in (c, d)$, g derivabile in t_0 ; sia $x_0 = g(t_0)$ e sia f derivabile in x_0 .

Allora se $\varphi = f(g(\cdot))$, φ è derivabile in t_0 e si ha

$$\varphi'(t_0) = f'(x_0)g'(t_0).$$

DIMOSTRAZIONE. Per la 9.5 si ha che

$$(9.7) \quad f(x) = f(x_0) + f'(x_0)(x - x_0) + (x - x_0)\omega_1(x - x_0)$$

$$(9.8) \quad g(t) = g(t_0) + g'(t_0)(t - t_0) + (t - t_0)\omega_2(t - t_0).$$

Si ha

$$\begin{aligned} (9.9) \quad \varphi(t) &= f(g(t)) = f(g(t_0)) + f'(g(t_0))[g(t) - g(t_0)] + \\ &\quad + [g(t) - g(t_0)]\omega_1(g(t) - g(t_0)) = \\ &= f(x_0) + f'(x_0)[g'(t_0)(t - t_0) + (t - t_0)\omega_2(t - t_0)] + \\ &\quad + [g(t) - g(t_0)]\omega_1(g(t) - g(t_0)) = \\ &= f(x_0) + f'(x_0)g'(t_0)(t - t_0) + (t - t_0)\omega_3(t - t_0). \end{aligned}$$

E dal momento che

$$\lim_{t \rightarrow t_0} \omega_3(t - t_0) = 0$$

si ha la tesi

□

TEOREMA 9.5. Sia $f : (a, b) \rightarrow \mathbb{R}$, $x_0 \in (a, b)$, $y_0 = f(x_0)$; supponiamo f strettamente monotona in (a, b) , derivabile in x_0 ed $f'(x_0) \neq 0$; allora f^{-1} è derivabile in y_0 e si ha

$$(f^{-1})'(y_0) = \frac{1}{f'(x_0)}.$$

DIMOSTRAZIONE. Consideriamo il rapporto incrementale

$$G(y) = \frac{f^{-1}(y) - f^{-1}(y_0)}{y - y_0},$$

avremo che

$$G(y) = F(f^{-1}(y))$$

ove

$$F(x) = \frac{x - x_0}{f(x) - f(x_0)} = \frac{x - f^{-1}(y_0)}{f(x) - y_0}.$$

Ora

$$\lim_{x \rightarrow x_0} \frac{x - x_0}{f(x) - f(x_0)} = \frac{1}{f'(x_0)}$$

e

$$\lim_{y \rightarrow y_0} f^{-1}(y) = f^{-1}(y_0) = x_0$$

(si ricordi che f^{-1} è continua in y_0 in quanto inversa di una funzione monotona continua).

Inoltre x_0 non appartiene al campo di definizione di F ; pertanto possiamo applicare il teorema che consente di calcolare il limite di una funzione composta per concludere che

$$\lim_{y \rightarrow y_0} G(y) = \frac{1}{f'(x_0)}$$

e la tesi. □

Calcoliamo ora le derivate di alcune funzioni elementari;

- $\frac{d}{dx} x^a = ax^{a-1}$
- $\frac{d}{dx} a^x = a^x \ln a$
- $\frac{d}{dx} \log_a x = \frac{1}{x \ln a}$
- $\frac{d}{dx} \sin x = \cos x$
- $\frac{d}{dx} \cos x = -\sin x$
- $\frac{d}{dx} \tan x = 1 + \tan^2 x$
- $\frac{d}{dx} \arcsin x = \frac{1}{\sqrt{1-x^2}}$
- $\frac{d}{dx} \arccos x = -\frac{1}{\sqrt{1-x^2}}$
- $\frac{d}{dx} \arctan x = \frac{1}{1+x^2}$.

Ciascuna delle formule vale per quegli x per cui ha senso e può essere provata usando la definizione di derivata.

Abbiamo con ciò introdotto quello che si chiama derivata prima di una funzione f ; ovviamente applicando successivamente più volte lo stesso procedimento, otterremo quelle che si chiamano derivata seconda, terza, ..., n -esima di f .

Indichiamo con

$$\begin{array}{ccc} f''(x_0) & \text{o} & \frac{d^2}{dx^2} f(x_0) \\ f^{(3)}(x_0) & \text{o} & \frac{d^3}{dx^3} f(x_0) \\ \dots\dots\dots & & \dots\dots\dots \\ f^{(n)}(x_0) & \text{o} & \frac{d^n}{dx^n} f(x_0) \end{array}$$

la derivata seconda, terza, ..., n -esima di f in x_0 .

Discorsi e notazioni analoghe vanno bene per le derivate successive destre e sinistre.

Indichiamo infine con $\mathcal{C}^n(I)$, $I \subset \mathbb{R}$, l'insieme delle funzioni $f : I \longrightarrow \mathbb{R}$ derivabili almeno n volte, con derivata n -esima continua.

In particolare $\mathcal{C}^0(I)$ è l'insieme delle funzioni continue, mentre $\mathcal{C}^\infty(I)$ è l'insieme delle funzioni che ammettono derivate di ogni ordine in ogni punto di I .

Si può facilmente verificare che ognuno di questi insiemi è uno spazio vettoriale su \mathbb{R} .

CAPITOLO 10

I TEOREMI DI ROLLE, LAGRANGE E CAUCHY.

Le derivate forniscono un'importante strumento per lo studio delle proprietà e del grafico di una funzione.

L'applicazione di tale strumento si concretizza attraverso alcuni risultati dimostrati nel corso del '700, dei quali ci occupiamo di seguito.

Cominciamo con il provare il seguente lemma.

LEMMA 10.1. Sia $f : [a, b] \longrightarrow \mathbb{R}$ derivabile, sia $x_1 \in [a, b]$ tale che

$$f(x_1) = \min\{f(x) : x \in [a, b]\}$$

si ha che

(1) se $x_1 \in (a, b)$ allora $f'(x_1) = 0$

(2) Se $x_1 = a$ allora $f'_+(x_1) \geq 0$

(3) Se $x_1 = b$ allora $f'_-(x_1) \leq 0$

DIMOSTRAZIONE. L'esistenza del punto di minimo è assicurata dalla continuità di f in $[a, b]$.

Proviamo la prima affermazione; sia

$$g(x) = \frac{f(x) - f(x_1)}{x - x_1}$$

si ha

$$\begin{aligned} g(x) &\leq 0 & \text{se} & \quad x < x_1 \\ g(x) &\geq 0 & \text{se} & \quad x > x_1. \end{aligned}$$

Pertanto

$$0 \geq \lim_{x \rightarrow x_1^-} g(x) = f'(x_1) = \lim_{x \rightarrow x_1^+} g(x) \geq 0.$$

Per quanto riguarda la seconda affermazione: se $x_1 = a$ si ha che

$$f'_+(x_1) = \lim_{x \rightarrow x_1^+} g(x) \geq 0.$$

La terza affermazione si dimostra in modo simile. \square

È chiaro che un risultato simile si può provare per i punti di massimo.

A questo punto siamo in grado di provare un risultato che è, pur nella sua semplicità, fondamentale per lo sviluppo del calcolo differenziale.

TEOREMA 10.1. - *Rolle* - Sia $f : [a, b] \longrightarrow \mathbb{R}$ continua in $[a, b]$ e derivabile in (a, b) ; allora

$$f(a) = f(b) \quad \Rightarrow \quad \exists c \in (a, b) : f'(c) = 0.$$

DIMOSTRAZIONE. Per il teorema di Weierstraß f ammette massimo e minimo assoluti in $[a, b]$; se entrambi sono assunti negli estremi si ha

$$\max\{f(x) : x \in [a, b]\} = \min\{f(x) : x \in [a, b]\}$$

ed f è costante, da cui $f'(x) = 0 \quad \forall x \in (a, b)$.

Se invece il massimo o il minimo è assunto in un punto interno c , dal lemma 10.1 si ha $f'(c) = 0$. \square

Ne segue che

TEOREMA 10.2. - *Lagrange* - Sia $f : [a, b] \longrightarrow \mathbb{R}$ continua in $[a, b]$ e derivabile in (a, b) ; allora esiste $c \in (a, b)$ tale che

$$f(b) - f(a) = f'(c)(b - a).$$

DIMOSTRAZIONE. Consideriamo la funzione

$$g : [a, b] \longrightarrow \mathbb{R}$$

definita da

$$g(x) = f(x) - \left(f(a) + \frac{f(b) - f(a)}{b - a}(x - a) \right).$$

g è continua in $[a, b]$ e derivabile in (a, b) ed inoltre $g(a) = g(b) = 0$.

Per il teorema di Rolle esiste $c \in (a, b)$ tale che

$$0 = g'(c) = f'(c) - \frac{f(b) - f(a)}{b - a}$$

e

$$f(b) - f(a) = f'(c)(b - a).$$

\square

TEOREMA 10.3. - *Peano* - Siano $f, g : [a, b] \longrightarrow \mathbb{R}$, f, g continue in $[a, b]$ e derivabili in (a, b) ; allora esiste $c \in (a, b)$ tale che

$$(10.1) \quad \det \begin{pmatrix} f(b) - f(a) & f'(c) \\ g(b) - g(a) & g'(c) \end{pmatrix} = 0$$

DIMOSTRAZIONE. Consideriamo la funzione $h : [a, b] \longrightarrow \mathbb{R}$ definita da

$$(10.2) \quad h(x) = \det \begin{pmatrix} f(b) - f(a) & f(x) \\ g(b) - g(a) & g(x) \end{pmatrix} = [f(b) - f(a)]g(x) - [g(b) - g(a)]f(x)$$

h soddisfa le ipotesi del teorema di Rolle e pertanto si può affermare che esiste $c \in (a, b)$ tale che

$$(10.3) \quad h'(c) = \det \begin{pmatrix} f(b) - f(a) & f'(c) \\ g(b) - g(a) & g'(c) \end{pmatrix} = 0$$

\square

TEOREMA 10.4. - Cauchy - Siano $f, g : [a, b] \longrightarrow \mathbb{R}$ continue in $[a, b]$ e derivabili in (a, b) ; sia inoltre $g'(x) \neq 0$ per ogni $x \in (a, b)$. Allora esiste $c \in (a, b)$ tale che

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(c)}{g'(c)}$$

DIMOSTRAZIONE. Per il teorema di Peano si ha che esiste $c \in (a, b)$ tale che

$$[f(b) - f(a)]g'(c) = [g(b) - g(a)]f'(c).$$

Ma $g'(x) \neq 0$ per ogni $x \in (a, b)$ e pertanto anche $g(b) - g(a) \neq 0$ (se così non fosse ci sarebbe, per il teorema di Rolle, un punto $\xi \in (a, b)$ tale che $g'(\xi) = 0$). Possiamo allora dividere per $g'(c)$ e per $g(b) - g(a)$ ed ottenere la tesi. \square

I teoremi appena dimostrati forniscono tutta una serie di risultati molto utili per lo studio del grafico di una funzione.

TEOREMA 10.5. Sia $f : [a, b] \longrightarrow \mathbb{R}$ continua in $[a, b]$ e derivabile in (a, b) ; allora f è costante in $[a, b]$ se e solo se $f'(x) = 0$ per ogni $x \in (a, b)$.

DIMOSTRAZIONE. Se f è costante in $[a, b]$ è immediato provare che f' è identicamente nulla.

Proviamo viceversa che se $f'(x) = 0$ per ogni $x \in (a, b)$ si ha che f è costante: sia $x \in (a, b)$ ed applichiamo il teorema di Lagrange all'intervallo $[a, x]$. Per un opportuno valore di $c \in (a, x)$ si ha

$$f(x) - f(a) = f'(c)(x - a) = 0$$

e si deduce che $f(x) = f(a)$ \square

COROLLARIO 10.1. Siano $f, g : [a, b] \longrightarrow \mathbb{R}$ continue in $[a, b]$, derivabili in (a, b) e tali che

$$f'(x) = g'(x) \quad \forall x \in (a, b) ;$$

allora

$$\exists c \in \mathbb{R} : f(x) = g(x) + c \quad \forall x \in [a, b].$$

TEOREMA 10.6. Sia $f : (a, b) \longrightarrow \mathbb{R}$, derivabile; allora f è crescente (decrecente) in (a, b) se e solo se $f'(x) \geq 0$ ($f'(x) \leq 0$) per ogni $x \in (a, b)$.

DIMOSTRAZIONE. E' intanto ovvio che se f è crescente allora si ha $f'(x) \geq 0$ per ogni $x \in (a, b)$; supponiamo viceversa che $f'(x) \geq 0$ per ogni $x \in (a, b)$, allora se $x_1, x_2 \in (a, b)$, $x_1 < x_2$, si ha

$$f(x_2) - f(x_1) = f'(c)(x_2 - x_1) \quad , \quad x_1 < c < x_2$$

e pertanto

$$f(x_2) - f(x_1) \geq 0$$

\square

Sottolineiamo che i risultati provati funzionano soltanto per funzioni definite su un intervallo.

È infatti facile trovare esempi che contraddicano gli enunciati precedenti se si rinuncia alla condizione di intervallo:

Ad esempio consideriamo le funzioni

$$(10.4) \quad f(x) = \begin{cases} 1 & , x > 0 \\ -1 & , x < 0 \end{cases}$$

oppure

$$(10.5) \quad g(x) = 1/x \quad \text{su} \quad \mathbb{R} \setminus \{0\}$$

Si può anche provare che:

TEOREMA 10.7. *Sia $f : (a, b) \longrightarrow \mathbb{R}$ derivabile e supponiamo che $f'(x) > 0$ per ogni $x \in (a, b)$; allora f è strettamente crescente in (a, b) .*

La funzione $f(x) = x^3$ ci convince inoltre che possono esistere funzioni strettamente crescenti la cui derivata non è sempre strettamente maggiore di zero.

LA REGOLA DI DE L'HÔPITAL

Dal teorema di Cauchy è possibile ricavare un risultato molto importante usualmente identificato come regola di De L'Hôpital, dal nome del marchese che pubblicò un trattato che la contiene, ma è più probabilmente dovuta a Johann Bernoulli.

Il suo scopo è fornire uno strumento atto a risolvere, in certi casi, il problema di trovare il limite di una forma indeterminata.

E' importante ricordare che l'applicazione di tale regola è subordinata, come sempre, alla verifica di alcune ipotesi, in assenza delle quali si possono ottenere dei risultati sbagliati.

La regola di De l'Hôpital è un raffinamento del seguente fatto del tutto elementare.

TEOREMA 11.1. *Siano $f, g : D \rightarrow \mathbb{R}$ derivabili in $x_0 \in D$, D aperto; allora, se $f(x_0) = g(x_0) = 0$ e $g'(x_0) \neq 0$, si ha*

$$\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = \frac{f'(x_0)}{g'(x_0)}.$$

DIMOSTRAZIONE. E' sufficiente osservare che

$$\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} \frac{x - x_0}{g(x) - g(x_0)} = \frac{f'(x_0)}{g'(x_0)}$$

□

Il risultato appena enunciato si può generalizzare al caso in cui non sia possibile considerare

$$\frac{f'(x_0)}{g'(x_0)}$$

ma soltanto

$$\lim_{x \rightarrow x_0} \frac{f'(x)}{g'(x)}.$$

Naturalmente tutto ciò è fatto allo scopo di determinare il

$$\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)}$$

nel caso in cui

$$\lim_{x \rightarrow x_0} f(x) = \lim_{x \rightarrow x_0} g(x) = 0.$$

Non sarà ovviamente restrittivo trattare solo il caso in cui $x \rightarrow x_0^+$.

TEOREMA 11.2. Siano $f, g : (a, b) \longrightarrow \mathbb{R}$ derivabili; supponiamo che

$$g'(x) \neq 0 \quad \forall x \in (a, b)$$

$$\lim_{x \rightarrow a^+} f(x) = \lim_{x \rightarrow a^+} g(x) = 0.$$

Allora, se

$$\lim_{x \rightarrow a^+} \frac{f'(x)}{g'(x)}$$

esiste, si ha

$$\lim_{x \rightarrow a^+} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a^+} \frac{f'(x)}{g'(x)}.$$

DIMOSTRAZIONE. Sia

$$\lim_{x \rightarrow a^+} \frac{f'(x)}{g'(x)} = \ell \in \mathbb{R}^*.$$

Se $a < x < a + \delta_\varepsilon$ si ha

$$\frac{f'(x)}{g'(x)} \in I(\ell, \varepsilon).$$

Ora, se prolunghiamo f e g per continuità in a ponendo

$$f(a) = g(a) = 0,$$

si può applicare il teorema di Cauchy nell'intervallo $[a, x]$ con $x \in (a, b)$ ed ottenere che

$$\frac{f(x)}{g(x)} = \frac{f(x) - f(a)}{g(x) - g(a)} = \frac{f'(c)}{g'(c)} \quad \text{con } a < c < x$$

Perciò, se $a < x < a + \delta_\varepsilon$ si ha $a < c < x < a + \delta_\varepsilon$ e

$$\frac{f(x)}{g(x)} = \frac{f'(c)}{g'(c)} \in I(\ell, \varepsilon).$$

□

Il teorema 11.2 può ovviamente essere ritenuto anche considerando limiti per $x \rightarrow a^-$ e per $x \rightarrow a$.

Restano fuori da questa trattazione i limiti per $x \rightarrow +\infty$ e per $x \rightarrow -\infty$.

Osserviamo che in tali casi può essere utilizzato il fatto che

$$\lim_{x \rightarrow +\infty} \frac{f(x)}{g(x)} = \lim_{t \rightarrow 0^+} \frac{f(1/t)}{g(1/t)}.$$

A quest'ultimo limite può essere applicato il teorema 11.2 non appena si siano verificate le ipotesi in esso richieste.

Enunciamo, per comodità, il risultato che si ottiene seguendo questa via.

COROLLARIO 11.1. Siano $f, g : (a, +\infty) \longrightarrow \mathbb{R}$ derivabili; supponiamo

$$g'(x) \neq 0 \quad \forall x \in (a, +\infty)$$

$$\lim_{x \rightarrow +\infty} f(x) = \lim_{x \rightarrow +\infty} g(x) = 0.$$

Allora, se

$$\lim_{x \rightarrow +\infty} \frac{f'(x)}{g'(x)}$$

esiste, si ha

$$\lim_{x \rightarrow +\infty} \frac{f(x)}{g(x)} = \lim_{x \rightarrow +\infty} \frac{f'(x)}{g'(x)}.$$

Il caso in cui $g \rightarrow +\infty$ oppure $g \rightarrow -\infty$ è molto più tecnico e complicato; pertanto ci limitiamo ad enunciare il risultato.

TEOREMA 11.3. *Siano $f, g : (a, b) \longrightarrow \mathbb{R}$ derivabili; supponiamo*

$$g'(x) \neq 0 \quad \forall x \in (a, b)$$

$$\lim_{x \rightarrow a^+} g(x) = +\infty.$$

Allora, se

$$\lim_{x \rightarrow a^+} \frac{f'(x)}{g'(x)}$$

esiste, si ha

$$\lim_{x \rightarrow a^+} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a^+} \frac{f'(x)}{g'(x)}.$$

La regola di De L'Hôpital permette di ricavare un risultato molto utile per calcolare la derivata di una funzione in punti che presentino qualche criticità.

COROLLARIO 11.2. *Sia $f : D \longrightarrow \mathbb{R}$, derivabile in $D \setminus \{x_0\}$ e continua in $x_0 \in D$, D aperto, con*

$$\lim_{x \rightarrow x_0^+} f'(x) = \lambda \quad , \quad \lim_{x \rightarrow x_0^-} f'(x) = \mu.$$

Allora

- (1) *se $\lambda \in \mathbb{R}$ allora $f'_+(x_0) = \lambda$*
- (2) *Se $\mu \in \mathbb{R}$ allora $f'_-(x_0) = \mu$*
- (3) *Se $\lambda = \pm\infty$ allora f non è derivabile da destra in x_0*
- (4) *Se $\mu = \pm\infty$ allora f non è derivabile da sinistra in x_0*

CAPITOLO 12

LA FORMULA DI TAYLOR

La formula di Taylor nasce dall'esigenza di trovare buone approssimazioni, facilmente calcolabili, per le funzioni elementari.

Si tratta essenzialmente dello sviluppo del concetto di approssimazione lineare che è stato introdotto con la definizione di derivata. Infatti se supponiamo che f sia una funzione derivabile in x_0 ; abbiamo visto che

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + (x - x_0)\omega(x - x_0)$$

dove

$$\lim_{x \rightarrow x_0} \omega(x - x_0) = 0 = \omega(0).$$

Possiamo pertanto affermare che in tale occasione abbiamo trovato un polinomio di primo grado che approssima la funzione f con un errore che può essere espresso nella forma $(x - x_0)\omega(x - x_0)$, con $\omega(x - x_0) \rightarrow 0$ se $x \rightarrow x_0$, tale errore quindi risulta essere infinitesimo di ordine superiore ad 1 cioè di ordine superiore al grado del polinomio approssimante.

Poniamoci ora il problema di approssimare la funzione f con un polinomio di grado n , commettendo un errore che sia infinitesimo di ordine superiore ad n , cioè che possa essere espresso nella forma

$$(x - x_0)^n \omega(x - x_0) \quad \text{ove} \quad \lim_{x \rightarrow x_0} \omega(x - x_0) = 0.$$

Sia pertanto

$$P_n(x) = \sum_{i=0}^n a_i (x - x_0)^i$$

un tale polinomio; dovrà aversi

$$(12.1) \quad f(x) = \sum_{i=0}^n a_i (x - x_0)^i + (x - x_0)^n \omega(x - x_0)$$

con $\omega(x - x_0) \rightarrow 0$ se $x \rightarrow x_0$.

Se supponiamo f derivabile n volte, affinché la 12.1 sia vera dovrà essere

$$f(x_0) = a_0$$

per cui si avrà

$$f(x) = f(x_0) + \sum_{i=1}^n a_i (x - x_0)^i + (x - x_0)^n \omega(x - x_0)$$

e

$$\frac{f(x) - f(x_0)}{x - x_0} = a_1 + \sum_{i=2}^n a_i (x - x_0)^{i-1} + (x - x_0)^{n-1} \omega(x - x_0).$$

Passando al limite per $x \rightarrow x_0$ si ottiene

$$f'(x_0) = a_1$$

e si avrà

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \sum_{i=2}^n a_i (x - x_0)^i + (x - x_0)^n \omega(x - x_0)$$

da cui

$$\begin{aligned} (12.2) \quad \frac{f(x) - f(x_0) - f'(x_0)(x - x_0)}{(x - x_0)^2} &= \\ &= a_2 + \sum_{i=3}^n a_i (x - x_0)^{i-2} + (x - x_0)^{n-2} \omega(x - x_0) \end{aligned}$$

per cui, applicando la regola di De L'Hôpital, si ottiene che

$$(12.3) \quad \lim_{x \rightarrow x_0} \frac{f'(x) - f'(x_0)}{2(x - x_0)} = a_2$$

e

$$\frac{f''(x_0)}{2!} = a_2.$$

Così procedendo si ottiene che

$$\frac{f^{(n)}(x_0)}{n!} = a_n$$

e pertanto, affinché il nostro scopo sia raggiunto, sarà necessario che

$$P(x) = \sum_{i=0}^n \frac{f^{(i)}(x_0)}{i!} (x - x_0)^i.$$

Riassumendo possiamo dire che

Affinchè si abbia

$$(12.4) \quad f(x) = \sum_{i=0}^n a_i (x - x_0)^i + (x - x_0)^n \omega(x - x_0)$$

con $\omega(x - x_0) \rightarrow 0$ se $x \rightarrow x_0$. deve essere

$$(12.5) \quad a_n = \frac{f^{(n)}(x_0)}{n!}$$

Ci resta ora da provare che tale polinomio soddisfa effettivamente le condizioni richieste.

Ciò sarà fatto provando il seguente risultato:

TEOREMA 12.1. - *Formula di Taylor con il resto di Peano* - Sia $f : (a, b) \longrightarrow \mathbb{R}$ derivabile $n-1$ volte in (a, b) ed n volte in $x_0 \in (a, b)$; allora

$$(12.6) \quad f(x) = \sum_{i=0}^n \frac{f^{(i)}(x_0)}{i!} (x - x_0)^i + (x - x_0)^n \omega(x - x_0)$$

con

$$\lim_{x \rightarrow x_0} \omega(x - x_0) = 0 = \omega(0) .$$

DIMOSTRAZIONE. Definiamo

$$P(x) = \sum_{i=0}^n \frac{f^{(i)}(x_0)}{i!} (x - x_0)^i$$

e chiamiamo

$$\omega(x - x_0) = \frac{f(x) - P(x)}{(x - x_0)^n};$$

proviamo che

$$\lim_{x \rightarrow x_0} \omega(x - x_0) = 0$$

Allo scopo di applicare la regola di De L'Hôpital calcoliamo

$$(12.7) \quad \lim_{x \rightarrow x_0} \frac{f'(x) - P'(x)}{n(x - x_0)^{n-1}} =$$

$$= \lim_{x \rightarrow x_0} \frac{1}{n(x - x_0)^{n-1}} \left(f'(x) - \sum_{i=1}^n \frac{f^{(i)}(x_0)}{(i-1)!} (x - x_0)^{i-1} \right)$$

e proseguendo calcoliamo

$$(12.8) \quad \lim_{x \rightarrow x_0} \frac{f''(x) - P''(x)}{n(n-1)(x - x_0)^{n-2}} =$$

$$= \lim_{x \rightarrow x_0} \frac{1}{n(n-1)(x - x_0)^{n-2}} \left(f''(x) - \sum_{i=2}^n \frac{f^{(i)}(x_0)}{(i-2)!} (x - x_0)^{i-2} \right)$$

fino ad arrivare a

$$(12.9) \quad \lim_{x \rightarrow x_0} \frac{f^{(n-1)}(x) - P^{(n-1)}(x)}{n!(x - x_0)} =$$

$$= \lim_{x \rightarrow x_0} \frac{f^{(n-1)}(x) - f^{(n-1)}(x_0) - f^{(n)}(x_0)(x - x_0)}{n!(x - x_0)} =$$

$$= \lim_{x \rightarrow x_0} \frac{1}{n!} \left(\frac{f^{(n-1)}(x) - f^{(n-1)}(x_0)}{x - x_0} - f^{(n)}(x_0) \right) = 0$$

Si può pertanto dedurre che

$$(12.10) \quad \lim_{x \rightarrow x_0} \omega(x - x_0) = 0$$

□

La formula di Taylor con il resto nella forma di Peano permette di estendere la possibilità di approssimare una funzione f con un polinomio di primo grado, fino ad ottenere la possibilità di approssimarla con un polinomio di grado n arbitrario.

Ovviamente il fatto più importante è la valutazione dell'errore commesso e, se consideriamo il resto nella forma di Peano, tale valutazione è di tipo qualitativo.

Se vogliamo una valutazione dell'errore di tipo quantitativo ci occorre seguire un procedimento diverso dalla definizione di differenziabilità. Un rapido sguardo ai risultati di calcolo differenziale fino ad ora provati ci convincerà ben presto che il risultato da estendere è il teorema di Lagrange.

Cercheremo in altre parole di valutare la differenza

$$f(x) - \sum_{i=0}^n \frac{f^{(i)}(x_0)}{i!} (x - x_0)^i$$

in funzione di maggioranti di $|f^{(n+1)}(x)|$.

TEOREMA 12.2. *Formula di Taylor con il resto di Lagrange - Sia $f : (a, b) \rightarrow \mathbb{R}$ derivabile $n + 1$ volte in (a, b) ; siano $x, x_0 \in (a, b)$, allora esiste c tra x_0 ed x , tale che*

$$f(x) = \sum_{i=0}^n \frac{f^{(i)}(x_0)}{i!} (x - x_0)^i + \frac{f^{(n+1)}(c)(x - x_0)^{n+1}}{(n + 1)!}.$$

DIMOSTRAZIONE. Proviamo il teorema nel caso in cui $n = 2$; dovremo in questo caso provare che esiste c tra x_0 ed x , tale che

$$(12.11) \quad f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2 + \frac{f'''(c)}{3!}(x - x_0)^3$$

Sia

(12.12)

$$F(x) = f(x) - f(x_0) - f'(x_0)(x - x_0) - \frac{f''(x_0)}{2}(x - x_0)^2 - R(x - x_0)^3$$

Ovviamente R dipende dal fatto che abbiamo fissato $n = 3$ oltre che da x e da x_0 , che comunque sono essi pure fissati,

Se consideriamo F sull'intervallo di estremi x_0 ed x , possiamo affermare che è derivabile almeno tre volte e si ha

$$(12.13) \quad F'(x) = f'(x) - f'(x_0) - f''(x_0)(x - x_0) - 3R(x - x_0)^2$$

$$(12.14) \quad F''(x) = f''(x) - f''(x_0) - 6R(x - x_0)$$

$$(12.15) \quad F'''(x) = f'''(x) - 6R$$

Poichè $F(x) = F(x_0) = 0$ per il teorema di Rolle esiste un punto α tra x_0 ed x tale che

$$F'(\alpha) = 0$$

Poichè inoltre $F'(x_0) = 0$, sempre per il teorema di Rolle si ha che esiste un punto β tra x_0 ed α tale che

$$F''(\beta) = 0$$

Ed ancora per il teorema di Rolle, poichè ancora $F''(x_0) = 0$ esiste un punto c tra x_0 ed β tale che

$$F'''(c) = 0$$

Ne ricaviamo infine che

$$F'''(c) = f'''(c) - 6R = 0$$

e ne deduciamo che

$$R = \frac{f'''(c)}{6}$$

□

CAPITOLO 13

QUALCHE SVILUPPO DI TAYLOR NOTEVOLE

Alcuni sviluppi di funzioni elementari ricorrono spesso e quindi è molto comodo fare una breve raccolta di risultati in merito

Nel seguito indichiamo con ω una funzione infinitesima per $x \rightarrow x_0$

1. Lo sviluppo di McLaurin di e^x

Sia

$$f(x) = e^x$$

Avremo che $f \in \mathcal{C}^{+\infty}(\mathbb{R})$ e si ha

$$(13.1) \quad f(x) = e^x \quad f(0) = 1$$

$$(13.2) \quad f'(x) = e^x \quad f'(0) = 1$$

$$(13.3) \quad f''(x) = e^x \quad f''(0) = 1$$

$$(13.4) \quad f'''(x) = e^x \quad f'''(0) = 1$$

$$(13.5) \quad \dots\dots \quad \dots\dots$$

$$(13.6) \quad f^{(n)}(x) = e^x \quad f^{(n)}(0) = 1$$

da cui si ricava che il polinomio di McLaurin P_n di e^x di grado n è

$$P_n(x) = \sum_{k=0}^n \frac{x^k}{k!}$$

ed il resto di Lagrange R_n assume la forma

$$R_n(x) = \frac{e^c}{(n+1)!} x^{n+1} \quad |c| \leq |x|$$

Possiamo pertanto concludere che

$$(13.7) \quad e^x = \sum_{k=0}^n \frac{x^k}{k!} + x^n \omega(x)$$

$$(13.8) \quad e^x = \sum_{k=0}^n \frac{x^k}{k!} + \frac{e^c}{(n+1)!} x^{n+1} \quad |c| \leq |x|$$

2. Lo sviluppo di McLaurin di $\sin x$

Sia

$$f(x) = \sin x$$

Avremo che $f \in \mathcal{C}^{+\infty}(\mathbb{R})$ e si ha

$$(13.9) \quad f(x) = \sin x \quad f^{(iv)}(x) = \sin x$$

$$(13.10) \quad f'(x) = \cos x \quad f^{(v)}(x) = \cos x$$

$$(13.11) \quad f''(x) = -\sin x \quad f^{(vi)}(x) = -\sin x$$

$$(13.12) \quad f'''(x) = -\cos x \quad f^{(vii)}(x) = -\cos x$$

Pertanto le derivate di f si ripetono di 4 in 4 e si ha

$$(13.13) \quad f(0) = 0 \quad f^{(iv)}(0) = 0$$

$$(13.14) \quad f'(0) = 1 \quad f^{(v)}(0) = 1$$

$$(13.15) \quad f''(0) = 0 \quad f^{(vi)}(0) = 0$$

$$(13.16) \quad f'''(0) = -1 \quad f^{(vii)}(0) = -1$$

da cui si ricava che il polinomio di McLaurin P_n di $\sin x$ di grado $2n+1$ è

$$P_{2n+1}(x) = \sum_{k=0}^n \frac{x^{2k+1}}{(2k+1)!}$$

ed il resto di Lagrange R_{2n+1} assume la forma

$$R_{2n+1}(x) = \frac{f^{(2n+3)}(c)}{(2n+3)!} \quad |c| \leq |x|$$

Ricordiamo che il termine di grado $2n+2$ è nullo.

Possiamo pertanto concludere che

$$(13.17) \quad \sin x = \sum_{k=0}^n \frac{x^{2k+1}}{(2k+1)!} + x^{2n+3} \omega(x)$$

$$(13.18) \quad \sin x = \sum_{k=0}^n \frac{x^{2k+1}}{(2k+1)!} + \frac{f^{(2n+3)}(c)}{(2n+3)!} x^{2n+3} \quad |c| \leq |x|$$

3. Lo sviluppo di McLaurin di $\cos x$

Sia

$$f(x) = \cos x$$

Avremo che $f \in \mathcal{C}^{+\infty}(\mathbb{R})$ e si ha

$$(13.19) \quad f(x) = \cos x \quad f^{(iv)}(x) = \cos x$$

$$(13.20) \quad f'(x) = -\sin x \quad f^{(v)}(x) = -\sin x$$

$$(13.21) \quad f''(x) = -\cos x \quad f^{(vi)}(x) = -\cos x$$

$$(13.22) \quad f'''(x) = \sin x \quad f^{(vii)}(x) = \sin x$$

Pertanto le derivate di f si ripetono di 4 in 4 e si ha

$$(13.23) \quad f(0) = 1 \quad f^{(iv)}(0) = 1$$

$$(13.24) \quad f'(0) = 0 \quad f^{(v)}(0) = 0$$

$$(13.25) \quad f''(0) = -1 \quad f^{(vi)}(0) = -1$$

$$(13.26) \quad f'''(0) = 0 \quad f^{(vii)}(0) = 0$$

da cui si ricava che il polinomio di McLaurin P_n di $\cos x$ di grado $2n$ è

$$P_{2n}(x) = \sum_{k=0}^n \frac{x^{2k}}{(2k)!}$$

ed il resto di Lagrange R_{2n} assume la forma

$$R_{2n}(x) = \frac{f^{(2n+2)}(c)}{(2n+2)!} \quad |c| \leq |x|$$

Ricordiamo che il termine di grado $2n+1$ è nullo.

Possiamo pertanto concludere che

$$(13.27) \quad \cos x = \sum_{k=0}^n \frac{x^{2k}}{(2k)!} + x^{2n+1}\omega(x)$$

$$(13.28) \quad \cos x = \sum_{k=0}^n \frac{x^{2k}}{(2k)!} + \frac{f^{(2n+2)}(c)}{(2n+2)!} x^{2n+3} \quad |c| \leq |x|$$

4. Lo sviluppo di McLaurin di $\ln(1+x)$

Sia

$$f(x) = \ln(1+x)$$

Avremo che $f \in \mathcal{C}^{+\infty}((-1, +\infty))$ e si ha

$$(13.29) \quad f(x) = \ln(1+x)$$

$$(13.30) \quad f'(x) = \frac{1}{1+x}$$

$$(13.31) \quad f''(x) = -\frac{1}{(1+x)^2}$$

$$(13.32) \quad f'''(x) = \frac{2}{(1+x)^3}$$

$$(13.33) \quad f^{(iv)}(x) = -\frac{3 \cdot 2}{(1+x)^4}$$

Possiamo quindi congetturare che

$$(13.34) \quad f^{(n)}(x) = (-1)^{n+1} \frac{(n-1)!}{(1+x)^n}$$

La 13.36 si dimostra per induzione, infatti:

(1) per $n = 1$

$$f'(x) = \frac{1}{1+x}$$

e la 13.36 è vera.

(2) se la 13.36 è vera per n allora è vera anche per $n+1$ infatti:

$$(13.35) \quad f^{(n+1)}(x) = \frac{d}{dx} f^{(n)}(x) = \frac{d}{dx} (-1)^{n+1} \frac{(n-1)!}{(1+x)^n} =$$

$$(-1)(-1)^{n+1} \frac{(n-1)!n(1+x)^{n-1}}{(1+x)^{2n}} (-1)^{n+2} \frac{n!}{(1+x)^{n+1}}$$

Pertanto

$$(13.36) \quad f^{(n)}(0) = (-1)^{n+1} (n-1)!$$

e quindi

$$P_n(x) = \sum_{k=0}^n (-1)^{k+1} (k-1)! \frac{x^k}{k!} = \sum_{k=0}^n (-1)^{k+1} \frac{x^k}{k}$$

ed il resto di Lagrange R_{2n} assume la forma

$$R_n(x) = (-1)^{n+2} \frac{(n)!}{(1+c)^{n+1}} \frac{x^{n+1}}{(n+1)!} = (-1)^{n+2} \frac{x^{n+1}}{(n+1)(1+c)^{n+1}} \quad |c| \leq |x|$$

Possiamo pertanto concludere che

(13.37)

$$\ln(1+x) = \sum_{k=0}^n (-1)^{k+1} \frac{x^k}{k} + x^n \omega(x)$$

(13.38)

$$\ln(1+x) = \sum_{k=0}^n (-1)^{k+1} \frac{x^k}{k} + (-1)^{n+2} \frac{x^{n+1}}{(n+1)(1+c)^{n+1}} \quad |c| \leq |x|$$

5. Lo sviluppo di McLaurin di $\sqrt{1+x}$

Sia

$$f(x) = \sqrt{1+x}$$

. Avremo che $f \in \mathcal{C}^{+\infty}((-1, +\infty))$ e si ha

$$(13.39) \quad f(x) = \sqrt{1+x} = (1+x)^{1/2}$$

$$(13.40) \quad f'(x) = \frac{1}{2}(1+x)^{-1/2}$$

$$(13.41) \quad f''(x) = \frac{1}{2} \left(-\frac{1}{2}\right) (1+x)^{-3/2}$$

$$(13.42) \quad f'''(x) = \frac{1}{2} \left(-\frac{1}{2}\right) \left(-\frac{3}{2}\right) (1+x)^{-5/2}$$

$$(13.43) \quad f^{(iv)}(x) = \frac{1}{2} \left(-\frac{1}{2}\right) \left(-\frac{3}{2}\right) \left(-\frac{5}{2}\right) (1+x)^{-7/2}$$

Possiamo quindi congetturare che

$$(13.44) \quad f^{(n)}(x) = (-1)^{n+1} \frac{(2n-3)!!}{2^n} (1+x)^{-\frac{2n-1}{2}}$$

La 13.44 si dimostra per induzione, infatti:

(1) per $n = 1$

$$f'(x) = \frac{1}{2}(1+x)^{-1/2}$$

e la 13.44 è vera.

(2) se la 13.44 è vera per n allora è vera anche per $n+1$ infatti:

$$\begin{aligned}
 (13.45) \quad f^{(n+1)}(x) &= \frac{d}{dx} f^{(n)}(x) = \frac{d}{dx} (-1)^{n+1} \frac{(2n-3)!!}{2^n} (1+x)^{-\frac{2n-1}{2}} = \\
 &= (-1)^{n+1} \frac{(2n-3)!!}{2^n} \left(-\frac{2n-1}{2} \right) (1+x)^{-\frac{2n-1}{2}-1} = \\
 &= (-1)^{n+2} \frac{(2n-1)!!}{2^{n+1}} (1+x)^{-\frac{2n+1}{2}}
 \end{aligned}$$

Pertanto

$$(13.46) \quad f^{(n)}(0) = (-1)^{n+1} \frac{(2n-3)!!}{2^n}$$

e quindi

$$P_n(x) = \sum_{k=0}^n (-1)^{k+1} \frac{(2k-3)!!}{2^k} \frac{x^k}{k!} = \sum_{k=0}^n (-1)^{k+1} \frac{(2k-3)!!}{k! 2^k} x^k$$

ed il resto di Lagrange R_{2n} assume la forma

$$R_n(x) = (-1)^{n+2} \frac{(2n-1)!!}{(n+1)! 2^{n+1}} (1+c)^{-\frac{2n+1}{2}} \quad |c| \leq |x|$$

Possiamo pertanto concludere che

(13.47)

$$\sqrt{1+x} = \sum_{k=0}^n (-1)^{k+1} \frac{(2k-3)!!}{k! 2^k} x^k + x^n \omega(x)$$

(13.48)

$$\sqrt{1+x} = \sum_{k=0}^n (-1)^{k+1} \frac{(2k-3)!!}{2^k} \frac{x^k}{k!} + (-1)^{n+2} \frac{(2n-1)!!}{(n+1)! 2^{n+1}} (1+c)^{-\frac{2n+1}{2}} \quad |c| \leq |x|$$

6. Lo sviluppo di McLaurin di $\frac{1}{1-x}$

Sia

$$f(x) = \frac{1}{1-x}$$

Avremo che $f \in \mathcal{C}^{+\infty}((-1, 1))$ e si ha

Definiamo

$$S_n(x) = \sum_{k=0}^n x^k$$

ed osserviamo che

$$(13.49) \quad S_n(x) = \sum_{k=0}^n x^k = 1 + x + x^2 + x^3 + \cdots + x^n$$

$$(13.50) \quad xS_n(x) = x \sum_{k=0}^n x^k = x + x^2 + x^3 + x^4 + \cdots + x^{n+1}$$

Sommando le due uguaglianze otteniamo

$$(13.51) \quad (1-x)S_n = 1 - x^{n+1}$$

$$(13.52) \quad S_n = \frac{1 - x^{n+1}}{1 - x}$$

e

$$(13.53) \quad S_n = \frac{1}{1-x} - \frac{x^{n+1}}{1-x}$$

Ne deduciamo che

$$(13.54) \quad \frac{1}{1-x} = S_n + \frac{x^{n+1}}{1-x} = \sum_{k=0}^n x^k + \frac{x^{n+1}}{1-x}$$

ed osservando che

$$(13.55) \quad \lim_{x \rightarrow 0} \frac{x^{n+1}}{1-x} = 0$$

di ordine $n+1 \in \mathbb{N}$ possiamo concludere ricordando la 12.4 che

$$(13.56) \quad P_n = \sum_{k=0}^n x^k$$

è il polinomio di McLaurin di $f(x) = \frac{1}{1-x}$.

Pertanto

$$(13.57) \quad \frac{1}{1-x} = \sum_{k=0}^n x^k + x^n \omega(x)$$

e

$$(13.58) \quad \frac{1}{1-x} = \sum_{k=0}^n x^k + \frac{x^{n+1}}{1-x}$$

Allo stesso risultato si può pervenire dimostrando per induzione che

$$(13.59) \quad f^{(n)}(x) = \frac{1}{(1-x)^{n+1}} \quad . \quad f^{(n)}(0) = 1$$

In questo modo si trova che che

$$(13.60) \quad R_n = \frac{1}{(1-c)^{n+1}} \quad |c| \leq |x|$$

7. Come ricavare altri sviluppi

Le precedenti formule possono essere utilizzate per ricavare nuovi sviluppi di Taylor mediante semplice sostituzione.

Ad esempio dalla 13.7 possiamo ricavare, sostituendo x con $-x^2$ che

$$(13.61) \quad e^{-x^2} = \sum_{k=0}^n \frac{(-1)^k x^{2k}}{k!} + x^{2n} \omega(x)$$

$$(13.62) \quad e^{-x^2} = \sum_{k=0}^n \frac{(-1)^k x^{2k}}{k!} + (-1)^{n+1} \frac{e^c}{(n+1)!} x^{2n+2} \quad |c| \leq |x^2|$$

Da quest'ultima, osservando che

$$x^{2n} \omega(x)$$

è un infinitesimo di ordine superiore ad $2n$ e ricordando la 12.4 possiamo affermare che

$$\sum_{k=0}^n \frac{(-1)^k x^{2k}}{k!}$$

è il polinomio di McLaurin di e^{-x^2} di grado n .

L'affermazione è giustificata dal fatto che $\sum_{k=0}^n \frac{(-1)^k x^{2k}}{k!}$ differisce da e^{-x^2} per infinitesimi di ordine superiore a $2n$.

Si capisce quindi che può essere utile disporre di criteri che consentano di affermare che la differenza tra un polinomio ed una funzione è infinitesima di ordine superiore al grado del polinomio.

Possiamo a questo proposito dire che

Se f è derivabile e se

$$(13.63) \quad f(x) = P_n(x) + R_n(x)$$

allora

$$(13.64) \quad f'(x) = (P_n(x))' + (R_n(x))'$$

(R_n è derivabile perchè $R_n = f - P_n$ e quindi è la differenza di due funzioni derivabili.)

Ora se $(R_n(x))'$ è un infinitesimo di ordine superiore ad $n - 1$ si ha

$$(13.65) \quad \lim_{x \rightarrow 0} \frac{(R_n(x))'}{x^{n-1}} = 0$$

e, per la regola di De l'Hopital

$$(13.66) \quad \lim_{x \rightarrow 0} \frac{(R_n(x))}{x^n} = \lim_{x \rightarrow 0} \frac{(R_n(x))'}{nx^{n-1}} = 0$$

CAPITOLO 14

LA CONVESSITÀ

Con le definizioni e gli strumenti che abbiamo introdotto fino a questo punto siamo in grado di distinguere una funzione il cui grafico sia del tipo illustrato in figura 14.1.1 da una il cui grafico sia quello illustrato nella figura 14.1.2

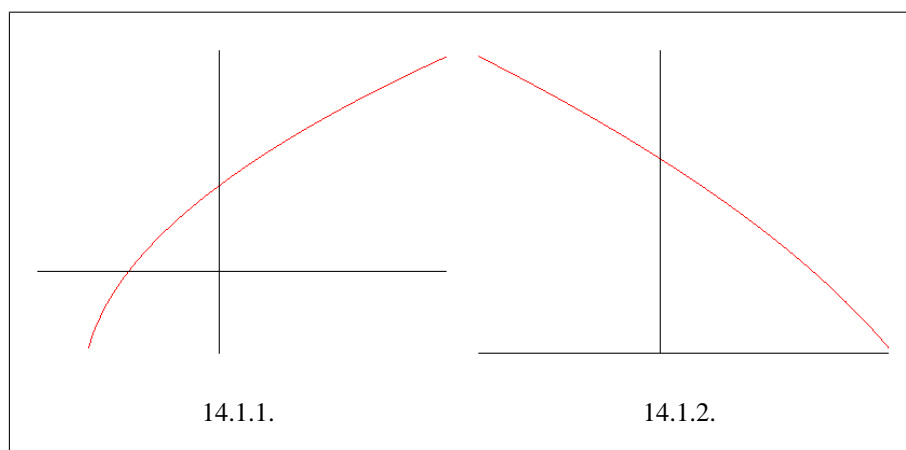


FIGURA 14.1.

Possiamo infatti osservare che il primo è il grafico di una funzione crescente mentre il secondo rappresenta una funzione decrescente.

Abbiamo inoltre già sviluppato strumenti (studio del segno della derivata prima) che ci consentono di stabilire se una funzione è crescente o decrescente.

Non siamo tuttavia ancora in grado di distinguere tra i grafici delle tre seguenti funzioni in quanto, ad un primo esame, possiamo osservare che tutte e tre sono funzioni crescenti; è tuttavia chiaro che si tratta di funzioni il cui grafico presenta caratteristiche molto diverse, così come è evidente quale è la differenza tra una scodella ed un ombrello.

Onde cercare di definire una proprietà che ci consenta di distinguere tra i tre grafici cominciamo ad esaminare il più semplice dei tre cioè il secondo. Chiaramente si tratta di una retta e quindi il suo grafico è individuato da due punti.

Indichiamo con ℓ la funzione e con $(x, \ell(x))$, $(y, \ell(y))$ due punti del suo grafico. Possiamo individuare il valore di ℓ in z semplicemente usando la proporzionalità tra i triangoli indicati in figura.

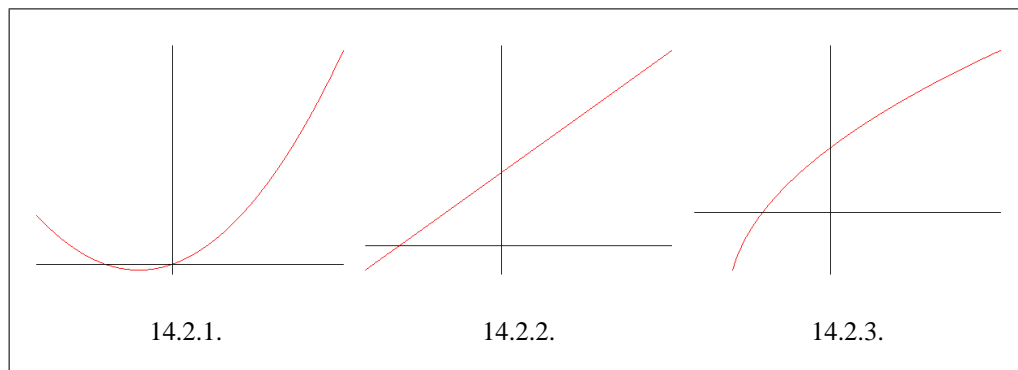


FIGURA 14.2.

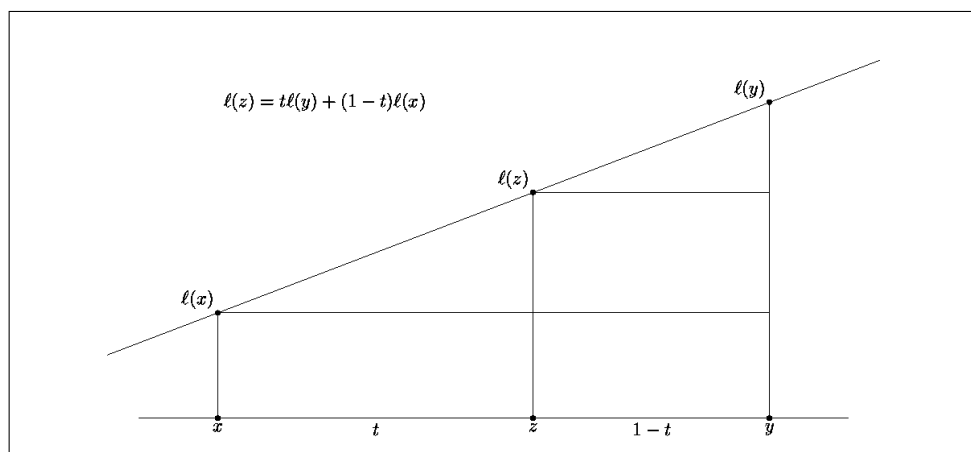


FIGURA 14.3.

Avremo infatti che

$$(14.1) \quad \frac{\ell(z) - \ell(x)}{z - x} = \frac{\ell(y) - \ell(x)}{y - x}$$

Poichè

$$\frac{\ell(z) - \ell(x)}{z - x} = \frac{\ell(x) - \ell(z)}{x - z}, \quad \frac{\ell(x) - \ell(y)}{x - y} = \frac{\ell(y) - \ell(x)}{y - x}$$

la 14.1 non cambia anche nel caso in cui z non sia, come in figura, interno all'intervallo di estremi x ed y . Inoltre non è restrittivo considerare $x < y$.

Avremo pertanto che il valore di ℓ in z è dato da

$$(14.2) \quad \ell(z) = \ell(x) + (z - x) \frac{\ell(y) - \ell(x)}{y - x}$$

La 22.1 è semplicemente l'equazione di una retta che passa per il punto $(x, \ell(x))$ ed ha coefficiente angolare $\frac{\ell(y) - \ell(x)}{y - x}$.

È utile osservare che, se poniamo

$$t = \frac{z - x}{y - x}$$

esprimiamo, nel contempo, la proporzionalità

$$\frac{t}{1} = \frac{z - x}{y - x}$$

tra le lunghezze dei segmenti di $[x, z]$ e $[x, y]$ ed i valori t ed 1.

Pertanto il rapporto tra i segmenti $[z, y]$ e $[x, y]$, sarà uguale a $1 - t$.

Un semplice calcolo mostra infatti che

$$1 - t = 1 - \frac{z - x}{y - x} = \frac{y - x - z + x}{y - x} = \frac{y - z}{y - x}$$

Inoltre se poniamo

$$(14.3) \quad t = \frac{z - x}{y - x}$$

avremo

$$(14.4) \quad z - x = t(y - x)$$

e quindi

$$(14.5) \quad z = x + t(y - x)x = ty + (1 - t)x$$

Per $t \in (0, 1)$ la 14.5 individua un punto z che si trova all'interno dell'intervallo di estremi x ed y , mentre per $t > 1$ si hanno punti a destra di y e per $t < 0$ si hanno punti a sinistra di x .

Similmente possiamo scrivere la 22.1 come

$$(14.6) \quad \begin{aligned} \ell(z) &= \ell(x) + (z - x) \frac{\ell(y) - \ell(x)}{y - x} = \ell(x) + (\ell(y) - \ell(x)) \frac{z - x}{y - x} \\ &\quad \ell(x) + t(\ell(y) - \ell(x)) = t\ell(y) + (1 - t)\ell(x) \end{aligned}$$

ed infine possiamo scrivere

$$(14.7) \quad \ell(ty + (1 - t)x) = t\ell(y) + (1 - t)\ell(x)$$

ed osservare che al variare di t la 14.7 consente di esprimere il fatto che tutti i valori $\ell(z) = \ell(ty + (1 - t)x)$ si trovano sulla retta di cui abbiamo studiato il grafico.

Se ora sovrapponiamo i primi due grafici della figura 14.2 risulta evidente che, se chiamiamo f la funzione del primo grafico ed x e y i punti di intersezione tra il grafico e la retta, avremo che, all'interno dell'intervallo $[x, y]$, il grafico di f sta sotto il grafico della retta.

Chiamiamo una tale funzione **convessa** ed esprimiamo il fatto che abbiamo appena individuato semplicemente chiedendo che

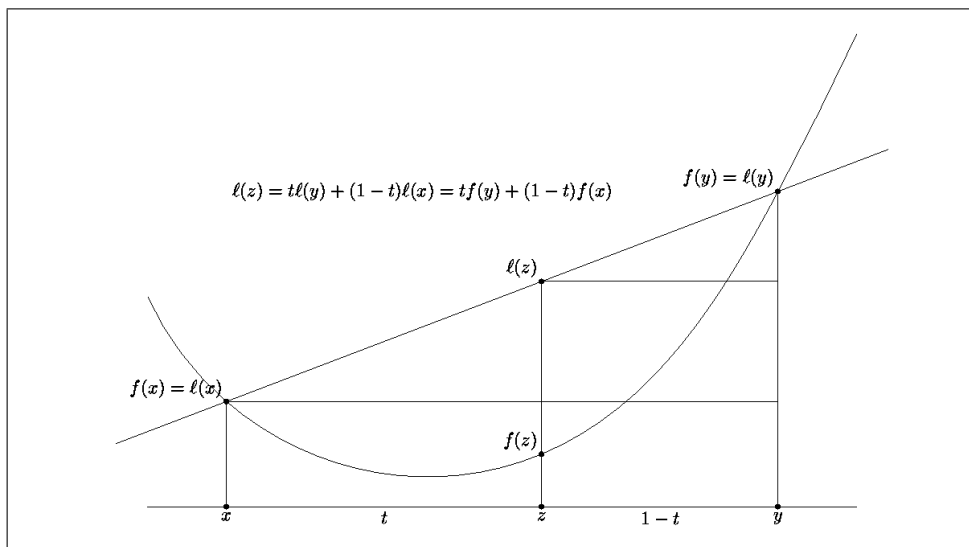


FIGURA 14.4.

$$f(ty + (1-t)x) \leq tf(y) + (1-t)f(x) \quad \forall t \in (0, 1)$$

Poniamo in altre parole la seguente definizione

Sia $f : (a, b) \longrightarrow \mathbb{R}$; f si dice convessa in (a, b) se

$$(14.8) \quad f(ty + (1-t)x) \leq tf(y) + (1-t)f(x)$$

per ogni $x, y \in (a, b)$ e per ogni $t \in (0, 1)$

Inoltre

Diciamo che f è strettamente convessa

$$(14.9) \quad f(ty + (1-t)x) < tf(y) + (1-t)f(x)$$

per ogni $x, y \in (a, b)$ e per ogni $t \in (0, 1)$

È utile osservare che la 14.8 può essere scritta in diversi modi tutti utili per comprendere le proprietà delle funzioni convesse.

$$(14.10) \quad f(ty + (1-t)x) \leq tf(y) + (1-t)f(x)$$

$$(14.11) \quad f(z) \leq tf(y) + (1-t)f(x)$$

$$(14.12) \quad f(z) \leq f(x) + (z-x) \frac{f(y) - f(x)}{y-x}$$

Dalla definizione di convessità si ricava sottraendo ad ambo i membri $f(y)$

$$(14.13) \quad f(z) - f(y) \leq (t-1)(f(y) - f(x))$$

$$(14.14) \quad f(z) - f(y) \leq \frac{z-y}{y-x}(f(y) - f(x))$$

$$(14.15) \quad \frac{f(z) - f(y)}{z-y} \geq \frac{f(y) - f(x)}{y-x}$$

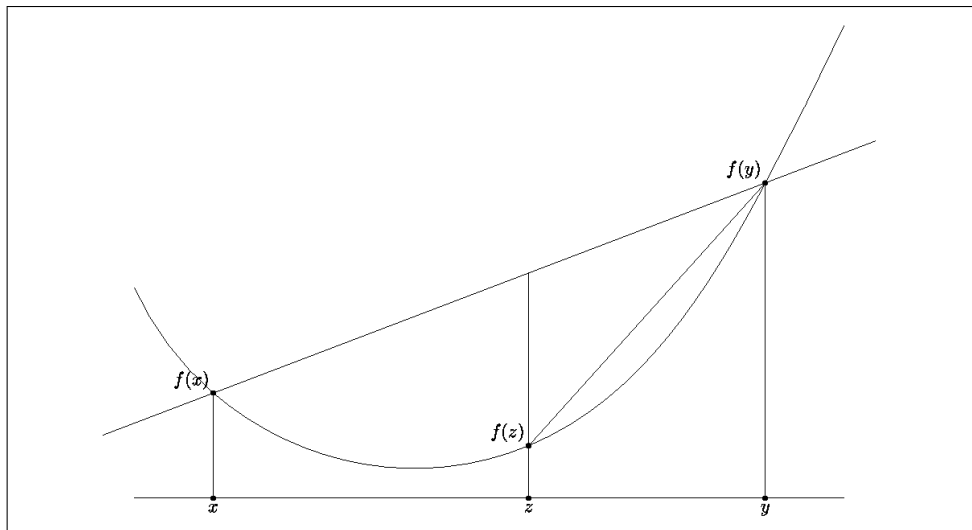


FIGURA 14.5.

Possiamo pertanto concludere, osservando che abbiamo sempre operato trasformando una disuguaglianza in una equivalente, che

Sono fatti equivalenti (si veda la figura 14.5):

- f è convessa in (a, b)
- In ogni punto $y \in (a, b)$ il rapporto incrementale

$$t \mapsto \frac{f(t) - f(y)}{t - y}$$

è una funzione crescente

D'altro canto, se f è convessa si ha:

$$(14.16) \quad f(z) \leq tf(y) + (1-t)f(x)$$

$$(14.17) \quad f(z)(t + (1-t)) \leq tf(y) + (1-t)f(x)$$

$$(14.18) \quad t(f(z) - f(y)) \leq (1-t)(f(x) - f(z))$$

$$(14.19) \quad (z-x)(f(z) - f(y)) \leq (y-z)(f(x) - f(z))$$

$$(14.20) \quad \frac{f(y) - f(z)}{y - z} \geq \frac{f(z) - f(x)}{z - x}$$

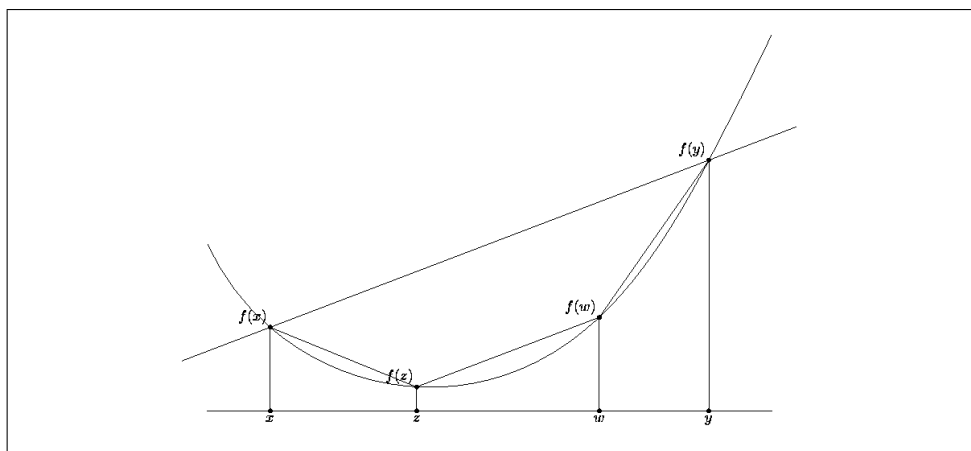


FIGURA 14.6.

Ora, se $x < z < w < y$ si ha

$$(14.21) \quad \frac{f(z) - f(x)}{z - x} \leq \frac{f(w) - f(z)}{w - z} \leq \frac{f(y) - f(w)}{y - w}$$

Passando al limite per $x \rightarrow z^-$ e per $y \rightarrow w^+$ se f è convessa e derivabile allora

$$(14.22) \quad f'(z) \leq f'(w)$$

e quindi f' è crescente.

Viceversa se f è derivabile ed f' è crescente allora, usando il teorema di Lagrange si può affermare che

$$(14.23) \quad \frac{f(z) - f(x)}{z - x} = f'(\xi) \geq f'(\eta) = \frac{f(y) - f(w)}{y - w}$$

e quindi f è convessa.

Ne concludiamo che se f è derivabile, allora

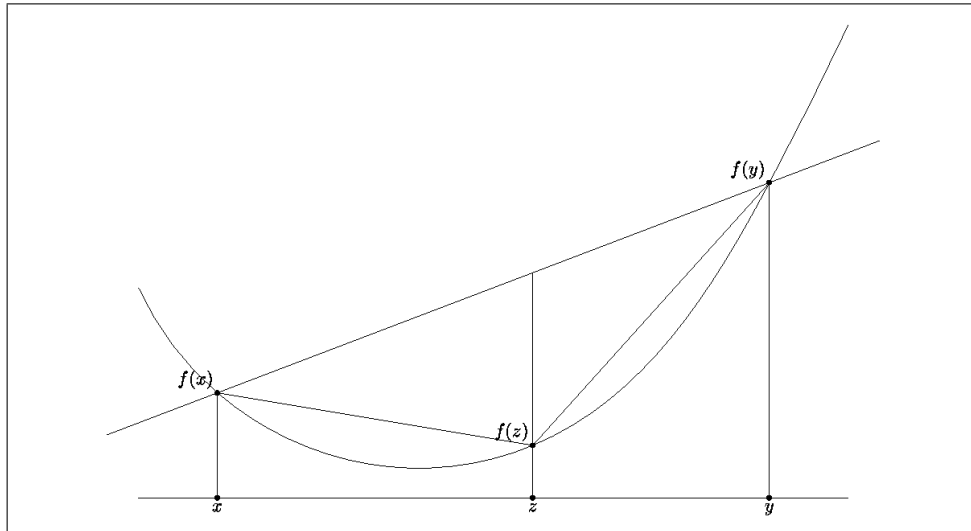


FIGURA 14.7.

Sono fatti equivalenti (si veda 14.7):

- f è convessa in (a, b)
- f' è una funzione crescente in (a, b)

Osserviamo infine che, se f è convessa, allora

$$(14.24) \quad f(y) - f(z) \geq (y - z) \frac{f(z) - f(x)}{z - x}$$

$$(14.25) \quad f(y) \geq f(z) + (y - z) \frac{f(z) - f(x)}{z - x}$$

e passando al limite per $x \rightarrow z$

$$(14.26) \quad f(y) \geq f(z) + f'(z)(y - z)$$

$$(14.27)$$

e pertanto il grafico di f sta' sopra al grafico di ogni sua retta tangente,

Se viceversa il grafico di f sta' sopra al grafico di ogni sua retta tangente, allora

$$(14.28) \quad f(y) \geq f(z) + f'(z)(y - z)$$

e

$$(14.29) \quad f(x) \geq f(z) + f'(z)(x - z)$$

da cui, tenendo conto che $y - z > 0$, e $x - z < 0$

$$(14.30) \quad \frac{f(y) - f(z)}{y - z} \geq f'(z) \geq \frac{f(z) - f(x)}{z - x}$$

e

$$(14.31) \quad f(y) - f(z) \geq (y - z) \frac{f(z) - f(x)}{z - x}$$

e quindi f è convessa.

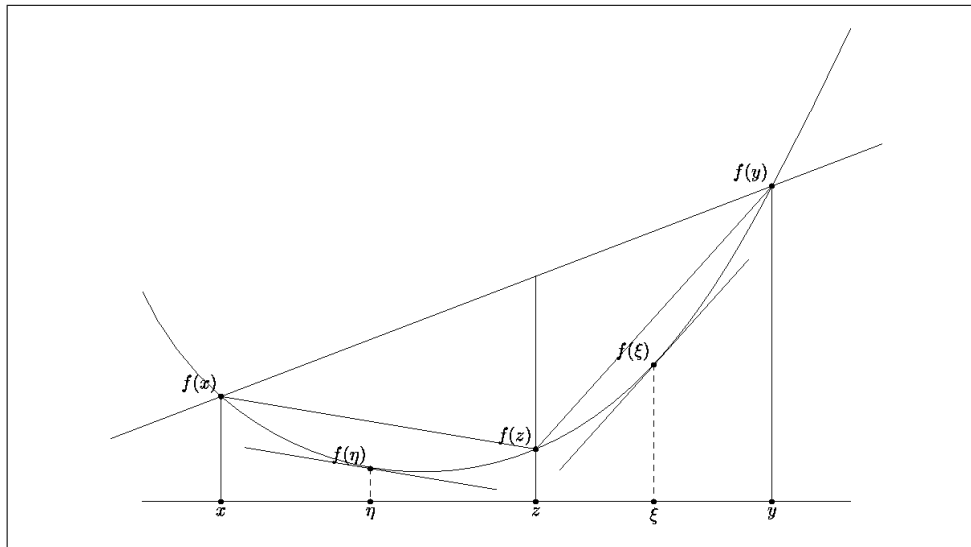


FIGURA 14.8.

Ne concludiamo che se f è derivabile, allora

Sono fatti equivalenti (si veda la figura 14.8):

- f è convessa in (a, b)
- il grafico di f sta' sopra al grafico di ogni sua retta tangente

I risultati che legano segno della derivata e crescita della funzione permettono poi di concludere che

Sia f una funzione derivabile due volte in (a, b) ; sono condizioni equivalenti:

- f è convessa in (a, b) ;
- f' è crescente in (a, b) ;
- f'' è non negativa in (a, b) .

DEFINIZIONE 14.1. Sia $f : (a, b) \longrightarrow \mathbb{R}$, diciamo che f è concava in (a, b) se $-f$ è convessa in (a, b) .

DEFINIZIONE 14.2. Diciamo che $f : (a, b) \longrightarrow \mathbb{R}$ ha un punto di flesso in $x_0 \in (a, b)$ se esiste $\delta > 0$ tale che f è convessa (concava) in $(x_0 - \delta, x_0)$ e concava (convessa) in $(x_0, x_0 + \delta)$.

Semplici esempi mostrano come sia possibile per una funzione avere un punto di flesso in 0 e

- non essere derivabile in 0 ($f(x) = \sqrt[3]{x}$)
- avere derivata non nulla in 0 ($f(x) = \sin x$)
- avere derivata nulla in 0 ($f(x) = x^3$).

TEOREMA 14.1. Sia $f : (a, b) \longrightarrow \mathbb{R}$ e sia $x_0 \in (a, b)$, supponiamo f derivabile in (a, b) ; allora x_0 è un punto di flesso se e solo se f' è crescente (decrecente) in un intorno destro di x_0 e decrescente (crescente) in un intorno sinistro.

E' pertanto evidente che non è possibile caratterizzare un punto di flesso facendo uso soltanto della derivata prima nel punto.

Possiamo tuttavia provare nel successivo paragrafo condizioni in grado di caratterizzare i punti di flesso.

CAPITOLO 15

ESTREMI RELATIVI E ASINTOTI.

Abbiamo già visto cosa si intende per minimo e massimo assoluto di una funzione e abbiamo già trovato condizioni necessarie e sufficienti per l'esistenza di un minimo o un massimo assoluto. (Si veda il lemma 9.1 ed il teorema 7.10).

In questo paragrafo ci occuperemo di stabilire la definizione di massimo e minimo relativo per una funzione e daremo condizioni necessarie e sufficienti per l'esistenza di un punto di minimo o di massimo relativo.

DEFINIZIONE 15.1. Sia $f : D \longrightarrow \mathbb{R}$ diciamo che $x_0 \in D$ è un punto di minimo (massimo) relativo per la funzione f se $\exists \delta > 0$ tale che se $x \in D \cap (x_0 - \delta, x_0 + \delta)$ si ha

$$f(x) \geq f(x_0) \quad (f(x) \leq f(x_0))$$

Usando la formula di Taylor possiamo ottenere uno strumento utile ad identificare i punti di massimo e di minimo relativo per una funzione. Tutto si fonda sul fatto che il polinomio di Taylor approssima una funzione a meno di infinitesimi di ordine superiore al grado del polinomio stesso.

Infatti, sia P_n il polinomio di Taylor di f centrato in x_0 di grado n , (ricordiamo che per scrivere il polinomio di Taylor di f , f deve essere derivabile almeno n volte); per il teorema 12.1 possiamo allora affermare che

$$(15.1) \quad f(x) = P_n(x) + (x - x_0)^n \omega(x - x_0)$$

dove, come al solito, qui e nel seguito supponiamo

$$\lim_{x \rightarrow x_0} \omega(x - x_0) = 0$$

e se definiamo

$$P_n^1(x) = \sum_{k=1}^n \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k$$

si avrà

$$(15.2) \quad f(x) - f(x_0) = P_n^1(x) + (x - x_0)^n \omega(x - x_0)$$

mentre se

$$P_n^2(x) = \sum_{k=2}^n \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k$$

si avrà

$$(15.3) \quad f(x) - f(x_0) - f'(x_0)(x - x_0) = P_n^2(x) + (x - x_0)^n \omega(x - x_0)$$

Osserviamo che P^1 e P^2 sono, rispettivamente, i polinomi di Taylor di $f(x) - f(x_0)$ e $f(x) - f(x_0) - f'(x_0)(x - x_0)$.

Dividendo le 15.1, 15.2, 15.3 per P , P^1 e P^2 , rispettivamente, otteniamo

$$(15.4) \quad \frac{f(x)}{P_n(x)} = 1 + \frac{(x - x_0)^n}{P_n(x)} \omega(x - x_0)$$

$$(15.5) \quad \frac{f(x) - f(x_0)}{P_n^1(x)} = 1 + \frac{(x - x_0)^n}{P_n^1(x)} \omega(x - x_0)$$

$$(15.6) \quad \frac{f(x) - f(x_0) - f'(x_0)(x - x_0)}{P_n^2(x)} = 1 + \frac{(x - x_0)^n}{P_n^2(x)} \omega(x - x_0)$$

(15.7)

Poichè P_n, P_n^1, P_n^2 , sono polinomi di grado n e quindi sono infinitesimi, per $x \rightarrow x_0$ di ordine al più n , tenendo conto che ω è a sua volta infinitesima, possiamo dedurre che

$$(15.8) \quad \frac{(x - x_0)^n}{P_n(x)} \omega(x - x_0) \quad \frac{(x - x_0)^n}{P_n^1(x)} \omega(x - x_0) \quad \frac{(x - x_0)^n}{P_n^2(x)} \omega(x - x_0)$$

sono infinitesimi per $x \rightarrow x_0$.

Il teorema della permanenza del segno permette quindi di affermare che
In un intorno di x_0

- (1) f ha lo stesso segno di P
- (2) $f(x) - f(x_0)$ ha lo stesso segno di P^1
- (3) $f(x) - f(x_0) - f'(x_0)(x - x_0)$ ha lo stesso segno di P^2

Poichè il segno di P_n in un intorno di x_0 è quello di $f(x_0)$, la prima affermazione si riduce semplicemente alla riaffermazione del teorema della permanenza del segno, tuttavia le altre due forniscono utili informazioni su crescenza e convessità.

Infatti poichè $f(x) - f(x_0)$ ha lo stesso segno di P_n^1 in un intorno di x_0 possiamo dire che x_0 è un punto di minimo relativo se siamo in grado di stabilire che P_n^1 è positivo in un intorno di x_0 , viceversa possiamo dire che x_0 non è di minimo relativo se il polinomio P_n^1 cambia segno in un intorno di x_0 .

Ora se supponiamo che f sia derivabile almeno n volte in $(a, b) \ni x_0$ e che $f^{(n)}(x_0)$ sia la prima derivata non nulla di f in x_0 possiamo considerare il polinomio P_n^1 che risulta essere definito da

$$P_n^1(x) = \frac{f^{(n)}(x_0)}{n!} (x - x_0)^n$$

e quindi risulta evidente che P_n^1 mantiene segno costante o cambia segno in un intorno di x_0 a seconda che n sia pari o dispari; nel caso che n sia pari il segno di P_n^1 è determinato dal segno di $f^{(n)}(x_0)$

Possiamo allora enunciare il seguente risultato

TEOREMA 15.1. *Sia $f : (a, b) \rightarrow \mathbb{R}$ una funzione derivabile almeno n volte e sia $x_0 \in (a, b)$; sia $f^{(n)}(x_0) \neq 0$ la prima derivata che non si annulla, $n \geq 1$; allora x_0 è punto di minimo relativo per f se e solo se n è pari e $f^{(n)}(x_0) > 0$.*

In maniera simile $f(x) - f(x_0) - f'(x_0)(x - x_0)$ ha lo stesso segno di P_n^2 in un intorno di x_0 e quindi si ha che x_0 è un punto di flesso se P_n^2 cambia segno in un intorno di x_0 , viceversa possiamo dire che x_0 non è un punto di flesso se il polinomio P_n^2 è positivo in un intorno di x_0 ,

Ora se, come prima, supponiamo che f sia derivabile almeno n volte in $(a, b) \ni x_0$ e che $f^{(n)}(x_0)$ sia la prima derivata non nulla di f in x_0 possiamo considerare il polinomio P_n^2 che risulta essere definito da

$$P_n^2(x) = \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n$$

e quindi risulta evidente che P_n^2 mantiene segno costante o cambia segno in un intorno di x_0 a seconda che n sia pari o dispari; nel caso che n sia pari il segno di P_n^2 è determinato dal segno di $f^{(n)}(x_0)$

Possiamo allora enunciare il seguente risultato

TEOREMA 15.2. *Sia $f : (a, b) \rightarrow \mathbb{R}$ una funzione derivabile almeno n volte e sia $x_0 \in (a, b)$; sia $f^{(n)}(x_0) \neq 0$ la prima derivata che non si annulla, $n \geq 2$; allora x_0 è punto di flesso per f se e solo se n è dispari. Il segno di $f^{(n)}(x_0)$ fornisce poi informazioni sul fatto che il grafico di f sia sopra (funzione localmente convessa) o sotto (funzione localmente concava) la retta tangente al suo grafico*

DEFINIZIONE 15.2. *Siano $f, g : (a, +\infty) \rightarrow \mathbb{R}$; diciamo che f e g sono asintotiche se*

$$\lim_{x \rightarrow +\infty} f(x) - g(x) = 0.$$

Nel caso in cui sia

$$g(x) = \alpha x + \beta$$

diciamo che g è un asintoto per f .

TEOREMA 15.3. *Sia $f : (a, +\infty) \rightarrow \mathbb{R}$; la retta di equazione*

$$y = \alpha x + \beta$$

è un asintoto per f se e solo se

$$(15.9) \quad \alpha = \lim_{x \rightarrow +\infty} \frac{f(x)}{x}, \quad \beta = \lim_{x \rightarrow +\infty} f(x) - \alpha x$$

DIMOSTRAZIONE. E' immediato verificare che le 15.9 sono sufficienti affinché la retta sia asintoto.

Viceversa, se la retta è un asintoto, si ha

□

$$(15.10) \quad \lim_{x \rightarrow +\infty} \frac{f(x) - \alpha x - \beta}{x} = \lim_{x \rightarrow +\infty} \frac{f(x)}{x} - \alpha = 0$$

DEFINIZIONE 15.3. Sia $f : (a, b) \longrightarrow \mathbb{R}$, diciamo che la retta di equazione $x = c$ è un asintoto verticale per f se

$$\lim_{x \rightarrow c} |f(x)| = +\infty$$

CAPITOLO 16

RICERCA NUMERICA DI ZERI E MINIMI.

Una delle applicazioni più tipiche della convessità consiste nella ricerca approssimata degli zeri di una funzione.

Il più semplice dei metodi di ricerca degli zeri è indubbiamente il metodo di bisezione di cui abbiamo già dato una dimostrazione in 8.1

Il metodo di bisezione offre indubbi vantaggi di semplicità di applicazione e necessita di ipotesi ridotte alla sola continuità della funzione f ; tuttavia, in presenza di migliori condizioni, si possono trovare metodi che convergono alla soluzione molto più velocemente.

Tali metodi, usualmente utilizzano la convessità della funzione, e sono tanto più importanti quanto è più grande la difficoltà di svolgere calcoli.

Chiaramente, con tempi di calcolo sempre più ridotti, tali metodi perdono parte della loro attrattiva anche se rimangono interessanti per la loro eleganza ed efficienza.

E' questo il caso del metodo di Newton (o delle tangenti) e del metodo della 'regula falsi'; essi convergono se le funzioni di cui si ricercano gli zeri sono convesse e possono essere generalizzati al caso non convesso purché le derivate prime e seconde della funzione f siano opportunamente maggiorabili o minorabili.

TEOREMA 16.1. - *Metodo di Newton (o delle tangenti)- Supponiamo $f : (\alpha, \beta) \longrightarrow \mathbb{R}$, convessa e derivabile due volte in (α, β) ; supponiamo inoltre che $\alpha < a < b < \beta$ e sia*

$$f(a) < 0 \quad , \quad f(b) > 0 \quad .$$

Allora esiste uno ed un solo punto $c \in (a, b)$ tale che

$$f(c) = 0 \quad , \quad f'(x) \geq f'(c) > 0 \quad \forall x \in [c, b].$$

Definiamo la successione x_n nella seguente maniera:

$$x_0 = b$$
$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)};$$

allora:

- x_n è decrescente e inferiormente limitata,
- $\lim x_n = c$
- se $0 \leq f''(x) \leq M$ per ogni $x \in [a, b]$ e se $f'(a) = P > 0$ si ha

$$0 \leq x_n - c \leq \frac{2P}{M} \left(\frac{M}{2P} (b - a) \right)^{2^n}$$

Il precedente metodo può essere generalizzato al caso in cui la funzione non sia convessa, ma siano verificate opportune condizioni.

TEOREMA 16.2. -*Metodo della regula falsi* - Sia $f : (\alpha, \beta) \longrightarrow \mathbb{R}$ una funzione convessa e derivabile due volte in (α, β) e siano $a, b \in (\alpha, \beta)$, $a < b$, tali che $f(a) < 0$, $f(b) > 0$.

Allora esiste uno ed un solo $c \in (a, b)$, tale che $f(c) = 0$ e $f'(x) \geq f'(c) > 0 \quad \forall x \in [c, b]$.

Inoltre, se definiamo una successione x_n nella seguente maniera:

$$(16.1) \quad x_0, x_1 \in [c, b], \quad x_1 < x_0$$

$$(16.2) \quad x_{n+1} = x_n - f(x_n) \frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})} \\ = x_{n-1} - f(x_{n-1}) \frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})}$$

si ha

- x_n è decrescente ed inferiormente limitata,
- $\lim x_n = c$
- se $M, P \in \mathbb{R}$ sono tali che

$$0 \leq f''(x) \leq M, \quad f'(x) \geq P > 0 \quad \forall x \in [a, b]$$

allora

$$0 \leq x_n - c \leq \frac{2P}{M} \left(\frac{M}{2P} (b - a) \right)^{\delta_n}$$

ove δ_n è la successione di Fibonacci.

CAPITOLO 17

INTEGRAZIONE.

Consideriamo un punto materiale P che si muove lungo l'asse x di un sistema di riferimento cartesiano, ed è sottoposto ad una forza di richiamo costante a tratti verso un punto O della retta, che assumiamo come origine degli assi coordinati.

Più precisamente se x è lo spostamento da O del punto P la forza di richiamo R sarà espressa da:

$$R(x) = k_i \quad \text{se} \quad i \leq x < i + 1 \quad \text{con} \quad i = 0, 1, 2, \dots$$

Il lavoro svolto per muovere un punto su cui agisce una forza costante, si calcola moltiplicando l'intensità della forza per lo spostamento che il punto ha subito, pertanto il lavoro che occorre per spostare il punto P dall'origine è dato da:

$$\Lambda(x) = \sum_{j=0}^{i-1} k_j + k_i(x - i) \quad \text{se} \quad i \leq x < i + 1 \quad \text{con} \quad i = 0, 1, 2, \dots$$

Se supponiamo che la forza di richiamo R anzichè costante a tratti sia proporzionale alla distanza x di P da O , come ad esempio accade nel caso in cui su P agisca una forza elastica, cioè se ipotizziamo che

$$R(x) = kx \quad \text{con} \quad k \in \mathbb{R}_+$$

avremo qualche problema in più per il calcolo del lavoro che non è più svolto da una forza costante, o costante a tratti. Possiamo allora tentare di calcolare il lavoro approssimando la forza di richiamo con una forza costante su tratti abbastanza piccoli.

Siano

$$0 = x_0 < x_1 < \dots < x_n = x$$

n punti che conveniamo di indicare come

$$P = \{x_0, x_1, \dots, x_n\}$$

e possiamo chiamare partizione dell'intervallo $[0, x]$. Possiamo approssimare $\Lambda(x)$ con le quantità

$$\Lambda^+(P, x) \quad \text{e} \quad \Lambda^-(P, x)$$

definite mediante le

$$(17.1) \quad \Lambda^-(P, x) = \sum_{i=1}^n kx_{i-1}(x_i - x_{i-1})$$

$$(17.2) \quad \Lambda^+(P, x) = \sum_{i=1}^n kx_i(x_i - x_{i-1})$$

Per come sono state definite si ha

$$\Lambda^-(P, x) \leq \Lambda(x) \leq \Lambda^+(P, x).$$

ed inoltre se consideriamo le partizioni

$$P_n = \{ix/n, i = 0, 1, 2, \dots, n\}$$

si ha che

$$(17.3) \quad \frac{kx^2}{n^2} \frac{(n-1)n}{2} = \frac{kx^2}{n^2} \sum_{i=1}^n (i-1) = \Lambda^-(P_n, x) \leq \\ \leq \sup\{\Lambda^-(P, x) : P\} \leq \inf\{\Lambda^+(P, x) : P\} \leq \\ \leq \Lambda^+(P_n, x) = \frac{kx^2}{n^2} \sum_{i=1}^n i = \frac{kx^2}{n^2} \frac{n(n+1)}{2}$$

Per cui passando al limite per $n \rightarrow +\infty$ si ottiene che

$$(17.4) \quad \frac{kx^2}{2} \leq \sup\{\Lambda^+(P, x) : P\} \leq \inf\{\Lambda^-(P, x) : P\} \leq \frac{kx^2}{2}$$

ed è lecito definire

$$(17.5) \quad \Lambda(x) = \inf\{\Lambda^+(P, x) : P\} = \sup\{\Lambda^-(P, x) : P\} = \frac{kx^2}{2}$$

Lo stesso problema si pone non appena cerchiamo di definire l'area dell'insieme

$$D = \{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq 1, 0 \leq y \leq x^2\}$$

Siano $0 = x_0 < x_1 < \dots < x_n = 1$ e definiamo

$$P = \{x_0, x_1, \dots, x_n\};$$

possiamo approssimare, rispettivamente per eccesso e per difetto, l'area di D mediante le

$$A(P) = \sum_{i=1}^n x_i^2 (x_i - x_{i-1}) \quad , \quad a(P) = \sum_{i=1}^n x_{i-1}^2 (x_i - x_{i-1})$$

e possiamo definire l'area di D come l'eventuale valore comune di $\inf\{A(P) : P\}$ e $\sup\{a(P) : P\}$ dichiarando che D non è misurabile se tali valori non risultano coincidenti.

Considerata la partizione $P_n = \{i/n : i = 0, 1, 2, \dots, n\}$ si calcola che

$$\begin{aligned}
 (17.6) \quad \frac{1}{n^3} \frac{(n-1)n(2n-1)}{6} &= \frac{1}{n^3} \sum_{i=1}^n (i-1)^2 \leq \\
 &\leq \sup\{a(P) : P\} \leq \inf\{A(P) : P\} \leq \\
 &\leq \frac{1}{n^3} \sum_{i=1}^n i^2 \leq \frac{1}{n^3} \frac{n(n+1)(2n+1)}{6}
 \end{aligned}$$

e per $n \rightarrow +\infty$ si ottiene

$$(17.7) \quad \frac{1}{3} \leq \sup\{a(P) : P\} \leq \inf\{A(P) : P\} \leq \frac{1}{3}$$

onde è lecito definire

$$\text{area}(D) = \inf\{A(P) : P\} = \sup\{a(P) : P\} = \frac{1}{3}$$

La definizione di integrale nasce dall'esigenza di formalizzare procedimenti del tipo che abbiamo esposto; in sostanza si tratta di definire l'estensione del concetto di somma discreta al caso in cui la somma sia fatta su insieme continuo di indici.

DEFINIZIONE 17.1. Sia $[a, b] \subset \mathbb{R}$, chiamiamo *partizione di $[a, b]$ un insieme*

$$P = \{x_0, x_1, \dots, x_n\}$$

di punti di $[a, b]$ tali che

$$a = x_0 < x_1 < \dots < x_n = b.$$

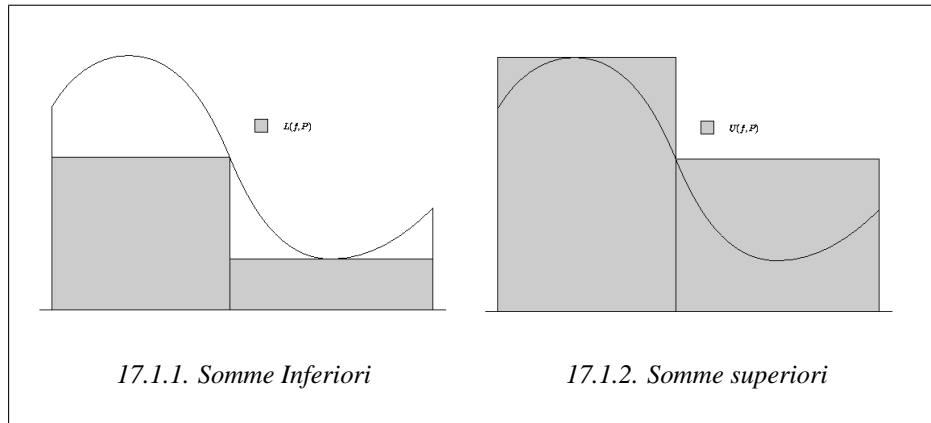


FIGURA 17.1.

Indichiamo con $\mathcal{P}(a, b)$ l'insieme delle partizioni di $[a, b]$.

Definiamo

$$(17.8) \quad I_k = [x_k, x_{k+1}] \quad , \quad \Delta I_k = x_{k+1} - x_k$$

$$(17.9) \quad I = [a, b] \quad , \quad \Delta I = b - a$$

ovviamente si avrà

$$(17.10) \quad I = \bigcup_{k=0}^{n-1} I_k, \quad [a, b] = \bigcup_{k=0}^{n-1} [x_k, x_{k+1}]$$

Definiamo inoltre, per ogni $P \in \mathcal{P}(a, b)$,

$$\Delta(P) = \max\{\Delta I_k, k = 0..n-1\}$$

DEFINIZIONE 17.2. Sia $P \in \mathcal{P}(a, b)$ e sia $f : [a, b] \longrightarrow \mathbb{R}$ una funzione limitata che supporremo sempre;

$$m \leq f(x) \leq M \quad \forall x \in [a, b]$$

poniamo

$$P = \{x_0, x_1, \dots, x_n\}$$

e definiamo

$$m_k = \inf\{f(x) : x \in [x_k, x_{k+1}]\}$$

$$M_k = \sup\{f(x) : x \in [x_k, x_{k+1}]\}.$$

Definiamo inoltre

$$(17.11) \quad L(f, P) = \sum_{k=0}^{n-1} m_k \Delta I_k$$

$$(17.12) \quad U(f, P) = \sum_{k=0}^{n-1} M_k \Delta I_k$$

$$(17.13) \quad R(f, P, \mathcal{S}) = \sum_{k=0}^{n-1} f(c_k) \Delta I_k$$

ove si indichi con $\mathcal{S} = \{c_1, \dots, c_n\}$ una scelta di punti tale che $x_k \leq c_k \leq x_{k+1}$.

$L(f, P)$ ed $U(f, P)$ si dicono, rispettivamente, somme inferiori e somme superiori di f rispetto alla partizione P , mentre $R(f, P, \mathcal{S})$ si dice somma di Cauchy-Riemann.

Vale la pena di osservare che le somme di Cauchy-Riemann dipendono dalla scelta dei punti \mathcal{S} oltre che dai punti c_k .

DEFINIZIONE 17.3. Siano $P, Q \in \mathcal{P}(a, b)$; diciamo che P è una partizione più fine di Q , e scriviamo $P \ll Q$, se $P \supset Q$.

Diciamo inoltre che $P_n \in \mathcal{P}(a, b)$ è una successione ordinata di partizioni se

$$(17.14) \quad \begin{cases} P_{n+1} \ll P_n \\ \lim \Delta(P_n) = 0 \end{cases}$$

È evidente dalle figure che valgono i seguenti fatti la cui dimostrazione può essere scritta formalizzando ciò che è suggerito da esse.

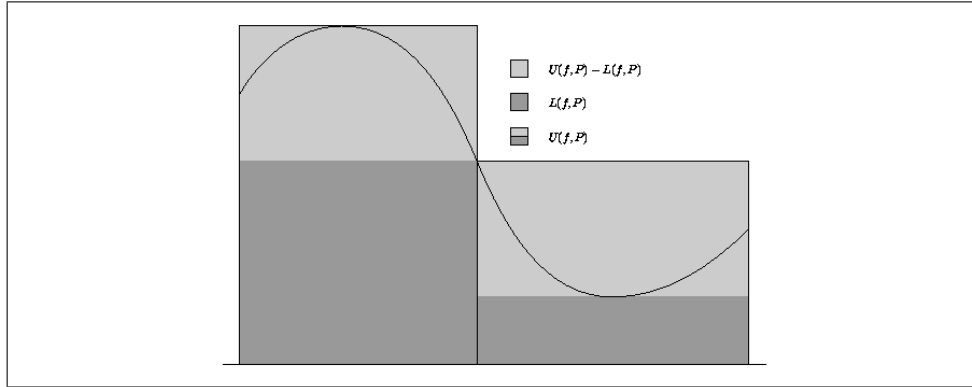


FIGURA 17.2. Confronto tra somme superiori e somme inferiori

LEMMA 17.1. *Siano $P, Q \in \mathcal{P}(a, b)$, $Q \ll P$, Allora*

$$(17.15) \quad m(b-a) \leq L(f, P) \leq L(f, Q) \leq R(f, Q, \mathcal{S}) \leq U(f, Q) \leq U(f, P) \leq M(b-a)$$

Inoltre, comunque si scelgano $R, S \in \mathcal{P}(a, b)$, si ha

$$L(f, R) \leq U(f, S).$$

DEFINIZIONE 17.4. *Definiamo*

$$(17.16) \quad \int_a^b f(x)dx = \inf\{U(f, P) : P \in \mathcal{P}(a, b)\}$$

$$(17.17) \quad \int_a^b f(x)dx = \sup\{L(f, P) : P \in \mathcal{P}(a, b)\}$$

Le precedenti quantità si dicono, rispettivamente, integrale superiore e integrale inferiore di f in $[a, b]$

È immediato verificare che

$$m(b-a) \leq \int_a^b f(x)dx \leq \int_a^b f(x)dx \leq M(b-a).$$

DEFINIZIONE 17.5. *Diciamo che f è integrabile in $[a, b]$ se*

$$\int_a^b f(x)dx = \int_a^b f(x)dx .$$

In tal caso chiamiamo il valore comune ottenuto integrale di f tra a e b e lo denotiamo con il simbolo

$$\int_a^b f(x)dx .$$

Definiamo

$$\int_b^a f(x)dx = - \int_a^b f(x)dx$$

ed osserviamo che

$$\int_a^a f(x)dx = 0.$$

DEFINIZIONE 17.6. Diciamo che f soddisfa la condizione di integrabilità in $[a, b]$ se $\forall \varepsilon > 0$ esiste una partizione $P_\varepsilon \in \mathcal{P}(a, b)$ tale che

$$(17.18) \quad 0 \leq U(f, P_\varepsilon) - L(f, P_\varepsilon) < \varepsilon$$

Dal momento che la quantità $U(f, P) - L(f, P)$ decresce al raffinarsi della partizione, restando non negativa, la precedente condizione è equivalente alla seguente

$$\begin{aligned} \forall \varepsilon > 0 \text{ esiste } P_\varepsilon \in \mathcal{P}(a, b) \text{ tale che} \\ 0 \leq U(f, P) - L(f, P) < \varepsilon. \\ \forall P \in \mathcal{P}(a, b), P \ll P_\varepsilon \end{aligned}$$

DEFINIZIONE 17.7. Diciamo che f è integrabile secondo Cauchy-Riemann in $[a, b]$ se esiste $I \in \mathbb{R}$ per cui esiste una partizione $P_\varepsilon \in \mathcal{P}(a, b)$ tale che $\forall \varepsilon > 0$ si ha che

$$(17.19) \quad |R(f, P, \mathcal{S}) - I| < \varepsilon$$

per ogni $P \in \mathcal{P}(a, b)$, $P \ll P_\varepsilon$ e per ogni scelta di punti \mathcal{S}

TEOREMA 17.1. Sono fatti equivalenti:

- (1) f è integrabile su $[a, b]$
- (2) f soddisfa la condizione di integrabilità in $[a, b]$

DIMOSTRAZIONE. Se f è integrabile allora

$$\int_a^b f(x)dx = \int_a^b f(x)dx = \int_a^b f(x)dx$$

Ma per definizione

$$(17.20) \quad \int_a^b f(x)dx = \sup_{\mathcal{P}} L(f, P) \quad , \quad \int_a^b f(x)dx = \inf_{\mathcal{P}} U(f, P)$$

e quindi possiamo trovare due partizioni P_ϵ e Q_ϵ tali che

$$(17.21) \quad \int_a^b f(x)dx - \frac{\epsilon}{2} \leq L(f, Q_\epsilon) \leq \int_a^b f(x)dx$$

$$(17.22) \quad \int_a^b f(x)dx \leq U(f, P_\epsilon) \leq \int_a^b f(x)dx + \frac{\epsilon}{2}$$

Se ne deduce che

$$(17.23) \quad U(f, P_\epsilon) - L(f, Q_\epsilon) \leq \epsilon$$

Se $R_\epsilon = Q_\epsilon \cup P_\epsilon$, si ottiene che

$$(17.24) \quad U(f, R_\epsilon) - L(f, R_\epsilon) \leq U(f, P_\epsilon) - L(f, Q_\epsilon) \leq \epsilon$$

e quindi vale la condizione di integrabilità.

Se viceversa si ha

$$(17.25) \quad U(f, P_\epsilon) - L(f, P_\epsilon) \leq \epsilon$$

allora

$$(17.26) \quad U(f, P_\epsilon) \leq L(f, P_\epsilon) + \epsilon$$

e pertanto

$$(17.27) \quad \int_a^{\bar{b}} f(x) dx \leq \int_a^b f(x) dx + \epsilon$$

e, passando al limite per $\epsilon \rightarrow 0$ si ottiene

$$(17.28) \quad \int_a^{\bar{b}} f(x) dx \leq \int_a^b f(x) dx$$

poichè è ovvio che

$$(17.29) \quad \int_a^b f(x) dx \leq \int_a^{\bar{b}} f(x) dx$$

si può concludere che

$$(17.30) \quad \int_a^{\bar{b}} f(x) dx = \int_a^b f(x) dx$$

e l'integrabilità di f è dimostrata. □

TEOREMA 17.2. *Se f è integrabile su $[a, b]$ allora f è integrabile secondo Cauchy-Riemann ed il valore dell'integrale è lo stesso.*

DIMOSTRAZIONE. Poichè

$$(17.31) \quad L(f, P) \leq R(f, P, \mathcal{S}) \leq U(f, P)$$

ed anche

$$(17.32) \quad L(f, P) \leq \int_a^b f(x) dx \leq U(f, P)$$

si ha

$$(17.33) \quad |R(f, P, \mathcal{S}) - \int_a^b f(x) dx| \leq U(f, P) - L(f, P)$$

Quando f è integrabile $U(f, P) - L(f, P)$ può essere reso piccolo quanto si vuole, pur di raffinare la partizione e quindi per la precedente disuguaglianza è possibile verificare la definizione di integrale secondo Cauchy-Riemann. \square

Nel teorema 17.2 può essere dimostrata anche l'implicazione opposta per cui

TEOREMA 17.3. *f è integrabile su $[a, b]$ se e solo se f è integrabile secondo Cauchy-Riemann ed il valore dell'integrale è lo stesso.*

DIMOSTRAZIONE. Dal momento che vale la definizione 17.7 avremo che se P è abbastanza fine allora

$$(17.34) \quad I - \varepsilon \leq R(f, P, \mathcal{S}) \leq I + \varepsilon$$

per ogni scelta di punti \mathcal{S} .

Poichè si ha

$$(17.35) \quad \begin{aligned} U(f, p) &= \sum_{k=0}^{n-1} M_k \Delta I_k \leq \sum_{k=0}^{n-1} (f(c_k) + \varepsilon) \Delta I_k = \\ &= R(f, P, \mathcal{S}_1) + \varepsilon(b-a) \leq I + \varepsilon(b-a) + \varepsilon \end{aligned}$$

$$(17.36) \quad \begin{aligned} L(f, p) &= \sum_{k=0}^{n-1} m_k \Delta I_k \geq \sum_{k=0}^{n-1} (f(d_k) - \varepsilon) \Delta I_k = \\ &= R(f, P, \mathcal{S}_2) - \varepsilon(b-a) \leq I - \varepsilon(b-a) - \varepsilon \end{aligned}$$

Ne viene allora che

$$(17.37) \quad I - \varepsilon(b-a) - \varepsilon \leq L(f, P) \leq U(f, P) \leq I + \varepsilon(b-a) + \varepsilon$$

e quindi vale la condizione di integrabilità e

$$\int_a^b f(x) dx = I$$

\square

TEOREMA 17.4. *Se f è una funzione integrabile e sia P_n una successione ordinata di partizioni di, allora*

$$(17.38) \quad \int_a^b f(x)dx = \lim_n U(f, P_n) = \lim_n L(f, P_n) = \lim_n R(f, P_n, S)$$

DIMOSTRAZIONE. Dal momento che f è integrabile, possiamo trovare una partizione P_ϵ tale che

$$U(f, P_\epsilon) - L(f, P_\epsilon) < \epsilon$$

Se P_ϵ è costituita da N punti, dalla figura 17.3

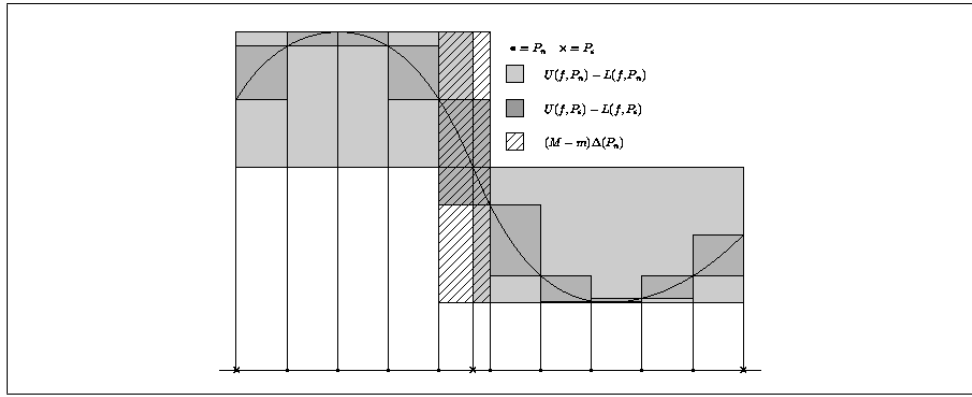


FIGURA 17.3. Confronto tra $U(f, P_n) - L(f, P_n)$ e $U(f, P_\epsilon) - L(f, P_\epsilon)$

si vede che

$$(17.39) \quad U(f, P_n) - L(f, P_n) \leq U(f, P_\epsilon) - L(f, P_\epsilon) + N(M - m)\Delta P_n$$

infatti è evidente che **non si può affermare semplicemente** che

$$U(f, P_n) - L(f, P_n) \leq U(f, P_\epsilon) - L(f, P_\epsilon)$$

a causa del fatto messo in evidenza dalla zona tratteggiata in figura 17.3.

Tuttavia l'area di tale zona può essere maggiorata con

$$(M - m)\Delta P_n$$

e tale evenienza ha luogo al più in tanti casi quanti sono i punti (N) di P_ϵ .

Pertanto, fissando opportunamente P_ϵ e scegliendo n abbastanza grande, si ottiene che $U(f, P_n) - L(f, P_n)$ diventa piccola quanto si vuole e quindi

$$U(f, P_n) - L(f, P_n) \rightarrow 0$$

Ora, se ricordiamo che

$$(17.40) \quad n \mapsto L(f, P_n) \quad \text{e} \quad n \mapsto U(f, P_n)$$

sono successioni crescenti per il fatto che la successione di partizioni P è ordinata, che f è integrabile e che

$$(17.41) \quad L(f, P_n) \leq R(f, P_n, S) \leq U(f, P_n)$$

possiamo concludere che

$$(17.42) \quad \lim_n U(f, P_n) \quad , \quad \lim_n L(f, P_n) \quad , \quad \lim_n R(f, P_n, S)$$

esistono e quindi poichè si ha

$$L(f, P_n) \leq \int_a^b f(x) dx \leq U(f, P_n)$$

$$(17.43) \quad \lim_n U(f, P_n) = \lim_n L(f, P_n) = \lim_n R(f, P_n, S) = \int_a^b f(x) dx$$

□

TEOREMA 17.5. *Se $f, g : [a, b] \longrightarrow \mathbb{R}$ sono integrabili su $[a, b]$, e se $\alpha, \beta \in \mathbb{R}$; allora $\alpha f + \beta g$ e fg sono integrabili su $[a, b]$ e*

$$\int_a^b [\alpha f(x) + \beta g(x)] dx = \alpha \int_a^b f(x) dx + \beta \int_a^b g(x) dx .$$

TEOREMA 17.6. *Se f è integrabile in $[a, b]$; e se $c \in (a, b)$, allora f è integrabile in $[a, c]$ ed in $[c, b]$ e*

$$\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx .$$

TEOREMA 17.7. *Se $f, g : [a, b] \longrightarrow \mathbb{R}$ sono integrabili in $[a, b]$ e se $f(x) \geq g(x) \forall x \in [a, b]$; allora*

$$\int_a^b f(x) dx \geq \int_a^b g(x) dx .$$

DIMOSTRAZIONE. Sarà sufficiente provare che, se $f(x) \geq 0$,

$$\int_a^b f(x) dx \geq 0,$$

ma questo è ovvia conseguenza del fatto che $m = \inf\{f(x) : x \in [a, b]\} \geq 0$. □

TEOREMA 17.8. *Se $f : [a, b] \longrightarrow \mathbb{R}$ e se definiamo*

$$f_+(x) = \max\{f(x), 0\} \quad , \quad f_-(x) = \min\{f(x), 0\}.$$

Allora f_+ ed f_- sono integrabili su $[a, b]$ se e solo se f è integrabile su $[a, b]$ e si ha

$$\int_a^b f(x) dx = \int_a^b f_+(x) dx + \int_a^b f_-(x) dx.$$

DIMOSTRAZIONE. Sia $P \in \mathcal{P}(a, b)$, allora

$$U(f_+, P) - L(f_+, P) \leq U(f, P) - L(f, P)$$

in quanto

$$\sup\{f_+(x) - f_+(y) : x, y \in [x_{i-1}, x_i]\} \leq \sup\{f(x) - f(y) : x, y \in [x_{i-1}, x_i]\}.$$

Inoltre si ha $f = f_+ + f_-$. □

COROLLARIO 17.1. *Se $f : [a, b] \longrightarrow \mathbb{R}$ è integrabile in $[a, b]$, allora anche $|f|$ è integrabile in $[a, b]$.*

DIMOSTRAZIONE. Basta osservare che $|f| = f_+ + f_-$ □

Osserviamo che $|f|$ può essere integrabile senza che f sia tale; ad esempio

$$f(x) = \begin{cases} -1 & , x \in \mathbb{Q} \\ 1 & , x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

TEOREMA 17.9. *Se f è integrabile, allora*

$$\left| \int_a^b f(x) dx \right| \leq \int_a^b |f(x)| dx.$$

DIMOSTRAZIONE. Si ha $-|f(x)| \leq f(x) \leq |f(x)|$ e la tesi segue dai risultati precedenti. □

TEOREMA 17.10. *Se f è integrabile in $[a, b]$, non negativa, e se $[\alpha, \beta] \subset [a, b]$; allora*

$$\int_\alpha^\beta f(x) dx \leq \int_a^b f(x) dx.$$

TEOREMA 17.11. *Siano $f, g : [a, b] \longrightarrow \mathbb{R}$, e sia f limitata ed integrabile in $[a, b]$; se*

$$f(x) = g(x) \quad \forall x \in [a, b] \setminus N, \quad N = \{y_1, \dots, y_k\}$$

allora anche g è integrabile in $[a, b]$ e si ha

$$\int_a^b f(x) dx = \int_a^b g(x) dx.$$

DIMOSTRAZIONE. È sufficiente provare che la funzione

$$h_c(x) = \begin{cases} 1 & , x = c \\ 0 & , x \neq c \end{cases}$$

con $c \in [a, b]$, è integrabile in $[a, b]$ e

$$\int_a^b h_c(x) dx = 0.$$

Sia $P \in \mathcal{P}(a, b)$; si ha

$$L(h_c, P) = 0, \quad U(h_c, P) \leq 2\Delta(P).$$

Pertanto

$$\inf\{U(h_c, P) : P \in \mathcal{P}(a, b)\} = 0.$$

La tesi segue tenendo conto del fatto che

$$f(x) = g(x) + \sum_{i=1}^k h_{y_i}(x) [f(y_i) - g(y_i)].$$

□

Ci proponiamo ora di dare alcune condizioni sufficienti per l'integrabilità.

TEOREMA 17.12. *Se f è monotona su $[a, b]$, allora f è integrabile in $[a, b]$.*

DIMOSTRAZIONE. Il teorema segue da quanto illustrato nella figura

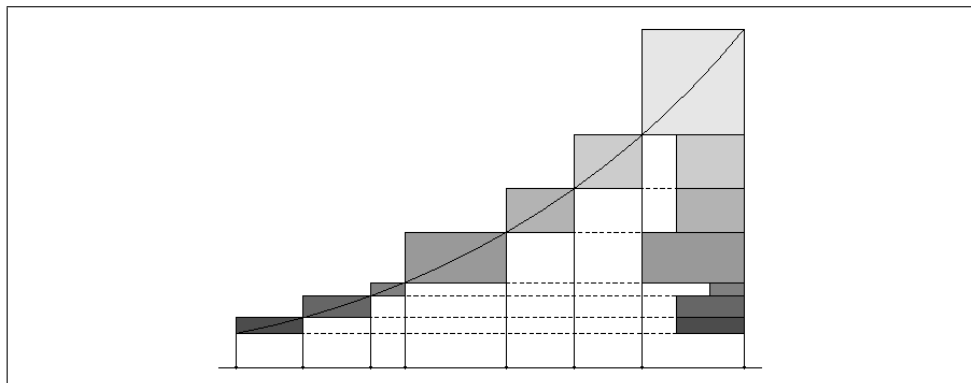


FIGURA 17.4. Integrabilità delle funzioni monotone

Supponiamo ad esempio che f sia crescente in $[a, b]$; si ha

$$m_i = f(x_{i-1}) \quad , \quad M_i = f(x_i)$$

per cui

$$U(f, P) - L(f, P) = \sum_{i=1}^n [f(x_i) - f(x_{i-1})](x_i - x_{i-1})$$

e, scelta $P_\varepsilon \in \mathcal{P}(a, b)$ in modo che $\Delta(P_\varepsilon) < \varepsilon$, si ha

$$U(f, P_\varepsilon) - L(f, P_\varepsilon) < \varepsilon \sum_{i=1}^n [f(x_i) - f(x_{i-1})] = \varepsilon [f(b) - f(a)].$$

□

TEOREMA 17.13. *Se $f : [a, b] \longrightarrow \mathbb{R}$ è continua, allora f è integrabile in $[a, b]$.*

DIMOSTRAZIONE. La dimostrazione di questo teorema si fonda sul concetto di uniforme continuità.

Poichè non si tratta di un concetto semplice, tanto che agli albori del calcolo esso era ignorato, è conveniente illustrare la dimostrazione per le funzioni lipschitziane, cioè per le funzioni per cui si può affermare che

$$|f(x) - f(y)| \leq L|x - y|$$

In tal caso si ha

(17.44)

$$U(f, P_\varepsilon) - L(f, P_\varepsilon) = \sum_{k=0}^{n-1} (M_k - m_k) \Delta I_k = \sum_{k=0}^{n-1} (f(x_k) - f(y_k)) \Delta I_k \leq \sum_{k=0}^{n-1} L |x_k - y_k| \Delta I_k$$

con $x_k, y_k \in I_k$ e se scegliamo $\Delta P < \frac{\varepsilon}{L}$ otteniamo che

$$U(f, P_\varepsilon) - L(f, P_\varepsilon) \leq \varepsilon \sum_{k=0}^{n-1} \Delta I_k = \varepsilon(b-a)$$

□

Possiamo anche dimostrare che

TEOREMA 17.14. *Se f è limitata in $[a, b]$ e continua in $[a, b] \setminus N$, $N = \{y_1, \dots, y_k\}$; allora f è integrabile in $[a, b]$.*

TEOREMA 17.15. *Se f è continua e non negativa e se*

$$\int_a^b f(x) dx = 0$$

allora $f(x) = 0$ per ogni $x \in [a, b]$

DIMOSTRAZIONE. Supponiamo per assurdo che esista $\alpha \in [a, b]$ tale che $f(\alpha) > 0$; allora esiste $[c, d] \subset [a, b]$ in modo che $f(x) \geq m > 0$ in $[c, d]$.

Pertanto

$$\int_a^b f(x) dx \geq \int_c^d f(x) dx \geq m(d-c) > 0.$$

□

Gli sviluppi del calcolo integrale e le sue applicazioni dipendono dal legame strettissimo tra il concetto di integrale e quello di derivata che è espresso dal teorema fondamentale del calcolo integrale e dalle proprietà di cui gode la funzione integrale $x \mapsto \int_{x_0}^x f(t) dt$.

Definiamo pertanto

$$(17.45) \quad F(x) = \int_{x_0}^x f(t) dt$$

e dedichiamo un po' di attenzione allo studio della continuità e della derivabilità di F

TEOREMA 17.16. *Sia f integrabile in $[a, b]$ e consideriamo, per ogni $x \in [a, b]$, la funzione definita da*

$$F(x) = \int_a^x f(t) dt.$$

Allora F è lipschitziana in $[a, b]$.

DIMOSTRAZIONE. Siano $x, y \in [a, b]$ e sia $|f(x)| \leq M \forall x \in [a, b]$; si ha

$$(17.46) \quad |F(x) - F(y)| = \left| \int_x^y f(t) dt \right| \leq \left| \int_x^y |f(t)| dt \right| \leq M|y - x|.$$

□

TEOREMA 17.17. *Sia f integrabile in $[a, b]$; consideriamo la funzione definita da*

$$F(x) = \int_a^x f(t) dt.$$

Se f è continua in $x_0 \in (a, b)$, allora F è derivabile in x_0 e

$$F'(x_0) = f(x_0).$$

DIMOSTRAZIONE. Per ogni $\varepsilon > 0$ esiste $\delta_\varepsilon > 0$ tale che se $|t - x_0| < \delta_\varepsilon$ si ha $|f(t) - f(x_0)| < \varepsilon$.

Perciò se $|h| < \delta_\varepsilon$ si ha

$$\left| \frac{F(x_0 + h) - F(x_0)}{h} - f(x_0) \right| \leq \frac{1}{|h|} \left| \int_{x_0}^{x_0+h} |f(t) - f(x_0)| dt \right| \leq \varepsilon.$$

□

TEOREMA 17.18. *della media - Sia f continua, allora esiste $c \in [a, b]$ tale che*

$$\int_a^b f(x) dx = f(c)(b - a).$$

DIMOSTRAZIONE. Possiamo applicare il teorema di Lagrange alla Funzione $F(x)$ su $[a, b]$ □

Il precedente teorema può essere generalizzato nella seguente forma

TEOREMA 17.19. *- della media - Siano f, g continue, g non negativa; allora esiste $c \in [a, b]$ tale che*

$$\int_a^b f(x)g(x)dx = f(c) \int_a^b g(x)dx.$$

DIMOSTRAZIONE. Il teorema è banale se g è identicamente nulla. In caso contrario si ha

$$\int_a^b g(x)dx > 0$$

e, posto

$$m = \min\{f(x) : x \in [a, b]\} \quad , \quad M = \max\{f(x) : x \in [a, b]\}$$

si ha

$$m \int_a^b g(x)dx \leq \int_a^b f(x)g(x)dx \leq M \int_a^b g(x)dx$$

e

$$m \leq \frac{\int_a^b f(x)g(x)dx}{\int_a^b g(x)dx} \leq M.$$

Pertanto la tesi segue dal fatto che f è continua in $[a, b]$ ed assume tutti i valori compresi tra il suo minimo m ed il suo massimo M . \square

I precedenti risultati indicano la necessità di introdurre un nuovo concetto: quello di funzione la cui derivata è assegnata.

DEFINIZIONE 17.8. Diciamo che F è una primitiva di f in (a, b) se F è ivi derivabile e risulta

$$F'(x) = f(x) \quad \forall x \in (a, b).$$

Definiamo integrale indefinito di f e lo indichiamo con il simbolo

$$\int f(x)dx$$

l'insieme delle primitive di f .

TEOREMA 17.20. Supponiamo che F e G siano due primitive di f in (a, b) , allora esiste $k \in \mathbb{R}$ tale che

$$F(x) = G(x) + k \quad \forall x \in (a, b).$$

DIMOSTRAZIONE. Dal momento che $(F - G)'(x) = 0$ in (a, b) si può applicare il corollario del teorema di Lagrange che assicura che se una funzione ha derivata nulla allora è costante. \square

COROLLARIO 17.2. Sia F una primitiva di f in (a, b) e sia f continua in (a, b) ; allora esiste $k \in \mathbb{R}$ tale che

$$F(x) = \int_a^x f(t)dt + k.$$

Il precedente corollario permette di determinare l'integrale indefinito di una funzione. Ricordiamo tuttavia che la sua validità è limitata a funzioni definite su un intervallo.

Anche il calcolo dell'integrale di f in $[a, b]$ beneficia di questo risultato vale infatti il seguente teorema.

TEOREMA 17.21. Sia f integrabile in $[a, b]$, sia F una primitiva di f in (a, b) e sia $[\alpha, \beta] \subset (a, b)$; allora

$$\int_{\alpha}^{\beta} f(x)dx = F(\beta) - F(\alpha).$$

DIMOSTRAZIONE. Sia $P \in \mathcal{P}(a, b)$, $P = \{x_0, \dots, x_n\}$, si ha

$$\begin{aligned}
 (17.47) \quad \left| \int_{\alpha}^{\beta} f(x) dx - [F(\beta) - F(\alpha)] \right| &= \\
 &= \left| \int_{\alpha}^{\beta} f(x) dx - \sum_{i=1}^n [F(x_i) - F(x_{i-1})] \right| = \\
 &= \left| \int_{\alpha}^{\beta} f(x) dx - \sum_{i=1}^n F'(\xi_i)(x_i - x_{i-1}) \right| = \\
 &= \left| \int_{\alpha}^{\beta} f(x) dx - R(f, P, \Xi) \right|
 \end{aligned}$$

e si può concludere scegliendo una partizione sufficientemente fine. \square

Il teorema precedente consente di usare le primitive di una funzione per calcolare il valore di un integrale definito. È pertanto importante conoscere le primitive di alcune funzioni elementari. Rimandando all'appendice per una informazione più completa, ci limitiamo qui ad osservare che la tabella di derivate data nel paragrafo 8, letta da destra verso sinistra, fornisce le primitive delle principali funzioni elementari.

Le funzioni che sono primitive di qualche altra funzione godono di una certa regolarità che è precisata nel seguente enunciato.

TEOREMA 17.22. *Se f ha una primitiva in (a, b) , per ogni $c \in (a, b)$ non è possibile che*

$$\lim_{x \rightarrow c^+} f(x) \neq \lim_{x \rightarrow c^-} f(x).$$

DIMOSTRAZIONE. Sia F una primitiva di f in (a, b) ; se i limiti in oggetto esistessero, si avrebbe, per le proprietà della derivabilità,

$$F'(c) = \lim_{x \rightarrow c^+} f(x) = \lim_{x \rightarrow c^-} f(x).$$

\square

Per il calcolo degli integrali definiti è molto utile servirsi delle seguenti regole di integrazione.

TEOREMA 17.23. - *integrazione per parti* - Siano f, g di classe \mathcal{C}^1 e sia $[\alpha, \beta] \subset (a, b)$; allora

$$\int_{\alpha}^{\beta} f'(x)g(x) dx = f(\beta)g(\beta) - f(\alpha)g(\alpha) - \int_{\alpha}^{\beta} f(x)g'(x) dx.$$

DIMOSTRAZIONE. Si ha

$$(fg)'(x) = f'(x)g(x) + f(x)g'(x);$$

pertanto

$$\int_{\alpha}^{\beta} f'(x)g(x) dx = \int_{\alpha}^{\beta} (fg)'(x) dx - \int_{\alpha}^{\beta} f(x)g'(x) dx$$

ed osservando che fg è una primitiva di $(fg)'$ in (a, b) si deduce la tesi. \square

TEOREMA 17.24. - *integrazione per sostituzione* - Se $f \in \mathcal{C}^0$, $g \in \mathcal{C}^1$; sia $[\alpha, \beta] \subset (c, d)$, allora

$$\int_{\alpha}^{\beta} f(g(x))g'(x)dx = \int_{g(\alpha)}^{g(\beta)} f(x)dx.$$

DIMOSTRAZIONE. Sia F una primitiva di f in (a, b) , allora $F(g(\cdot))$ è una primitiva di $f(g(\cdot))g'(\cdot)$ in (c, d) e si ha

$$\int_{\alpha}^{\beta} f(g(x))g'(x)dx = F(g(\beta)) - F(g(\alpha)) = \int_{g(\alpha)}^{g(\beta)} f(x)dx$$

□

Se f è una funzione, indichiamo

$$f(x) \Big|_a^b = f(b) - f(a)$$

In tal modo risultano semplificati molti enunciati che coinvolgono gli integrali definiti.

Ad esempio si può dire che, se F è una primitiva della funzione continua f , allora

$$\int_a^b f(x)dx = F(x) \Big|_a^b$$

Ci siamo occupati fino ad ora del problema di integrare una funzione limitata su di un intervallo limitato che, in genere, è anche supposto chiuso.

Nella pratica è spesso necessario integrare funzioni non limitate o su intervalli non limitati, a questo scopo è necessario definire una estensione del concetto di integrale, che permetta di considerare anche questi casi.

La definizione non è strettamente collegata col procedimento di integrazione definita dato precedentemente, anche se da esso dipende in maniera essenziale, e, per questa ragione, viene denominato procedimento di integrazione impropria.

DEFINIZIONE 17.9. Sia f integrabile in $[x, b]$, per ogni $x \in (a, b]$. Diciamo che f ammette integrale improprio (finito) in $(a, b]$ se esiste (finito)

$$\lim_{x \rightarrow a^+} \int_x^b f(t)dt$$

In tal caso definiamo il suo valore

$$\int_a^b f(x)dx$$

Definizioni analoghe permettono di considerare facilmente l'integrale improprio su di un intervallo $[a, b]$ di una funzione f limitata e integrabile in ogni intervallo $[x, y] \subset [a, b] \setminus N$, dove N è un insieme finito di punti

DEFINIZIONE 17.10. Sia f integrabile in $[a, x]$, per ogni $x \geq a$. Diciamo che f ammette integrale improprio (finito) in $[a, +\infty)$ se esiste (finito)

$$\lim_{x \rightarrow +\infty} \int_a^x f(t)dt$$

In tal caso definiamo il suo valore

$$\int_a^{+\infty} f(x)dx$$

Definizioni analoghe permettono di considerare l'integrale improprio di una funzione limitata e integrabile su ogni intervallo $[x, y]$, in $(-\infty, a]$ o $(-\infty, +\infty)$.

In entrambe le due precedenti definizioni diciamo che f ammette integrale improprio convergente o divergente, a seconda che il suo valore sia finito o infinito rispettivamente.

Per semplicità, nel seguito faremo riferimento solo ai casi contemplati nelle definizioni 17.9, 17.10, tuttavia i risultati che proveremo possono essere facilmente rinunciati e ridimostrati negli altri casi.

Possiamo considerare qualche esempio per illustrare i concetti introdotti
Sia

$$f_\alpha(x) = \frac{1}{x^\alpha}$$

allora:

$$\begin{aligned} \int_0^1 f_\alpha(x)dx &= \frac{1}{1-\alpha} & \text{se } 0 < \alpha < 1 \\ \int_0^1 f_\alpha(x)dx &= +\infty & \text{se } \alpha \geq 1 \\ \int_1^{+\infty} f_\alpha(x)dx &= \frac{1}{\alpha-1} & \text{se } \alpha > 1 \\ \int_1^{+\infty} f_\alpha(x)dx &= +\infty & \text{se } 0 < \alpha \leq 1 \end{aligned}$$

I risultati esposti si possono ricavare applicando semplicemente le definizioni e le regole elementari di integrazione. Partendo da questi semplici punti fermi possiamo ricavare dei criteri che consentono di stabilire se una funzione è integrabile in senso improprio.

A questo scopo dobbiamo considerare una conseguenza del criterio di convergenza di Cauchy.

TEOREMA 17.25. *Sia f integrabile in $[x, b]$, per ogni $x \in (a, b]$; allora f ammette integrale improprio convergente in $(a, b]$ se e solo se per ogni $\varepsilon > 0$ esiste $\delta_\varepsilon > 0$ tale che se si considerano $x', x'' \in (a, a + \delta_\varepsilon)$ allora*

$$\left| \int_{x'}^{x''} f(t)dt \right| < \varepsilon$$

TEOREMA 17.26. *Sia f integrabile in $[a, x]$, per ogni $x \geq a$; allora f ammette integrale improprio convergente in $[a, +\infty)$ se e solo se per ogni $\varepsilon > 0$ esiste $\delta_\varepsilon > 0$ tale che se $x', x'' > \delta_\varepsilon$ allora*

$$\left| \int_{x'}^{x''} f(t) dt \right| < \varepsilon$$

I due precedenti teoremi seguono immediatamente dal criterio di convergenza di Cauchy.

TEOREMA 17.27. *Siano f, g integrabili in ogni $[x, y] \subset I$; valgono i seguenti fatti:*

- (1) *se $|f|$ ammette integrale improprio convergente in I , allora anche f ammette integrale improprio convergente in I ;*
- (2) *se $|f| \leq g$ e g ammette integrale improprio convergente in I , allora anche $|f|$ (e quindi f) ammette integrale improprio convergente in I .*

DIMOSTRAZIONE. Si ha

$$\left| \int_{x'}^{x''} f(x) dx \right| \leq \left| \int_{x'}^{x''} |f(x)| dx \right| \leq \left| \int_{x'}^{x''} g(x) dx \right|$$

□

Non è tuttavia vero che se f ammette integrale improprio convergente anche $|f|$ ammette integrale improprio convergente. Per esempio si consideri

$$f(x) = \frac{\sin x}{x}$$

sull'intervallo $[0, +\infty)$, (si veda teorema 17.31).

Diamo ora un risultato che sarà di grande utilità per stabilire l'integrabilità in senso improprio di una funzione.

TEOREMA 17.28. *Sia f integrabile in $[x, b]$, per ogni $x \in (a, b]$, e supponiamo che f sia infinita per $x \rightarrow a^+$ di ordine β .*

- (1) *Se esiste $\alpha \in \mathbb{R}$, con $\beta \leq \alpha < 1$ allora f ammette integrale improprio convergente in $(a, b]$.*
- (2) *se $\beta \geq 1$ allora f ammette integrale improprio divergente in $(a, b]$.*

DIMOSTRAZIONE. Supponiamo ad esempio $f(x) \rightarrow +\infty$ per $x \rightarrow a^+$ e proviamo (1). Si ha

$$\lim_{x \rightarrow a^+} \frac{f(x)}{1/(x-a)^\alpha} = \ell \in \mathbb{R}_+$$

e pertanto esistono $k, \delta > 0$ tali che

$$0 \leq f(x) \leq \frac{k}{(x-a)^\alpha} \quad \forall x \in (a, a+\delta)$$

e la tesi segue dal teorema 17.27 e dalle 17.48, non appena si sia tenuto conto del fatto che

$$\lim_{x \rightarrow a^+} \int_x^b f(t) dt = \int_{a+\delta}^b f(t) dt + \lim_{x \rightarrow a^+} \int_x^{a+\delta} f(t) dt.$$

Proviamo ora (2); se $x \in (a, a+\delta)$, con $k, \delta > 0$ opportunamente scelti, si ha

$$f(x) \geq \frac{k}{x-a}$$

e come prima segue la tesi. \square

TEOREMA 17.29. *Sia f integrabile in $[a, x]$, per ogni $x \geq a$; supponiamo che f ammetta integrale improprio convergente in $[a, +\infty)$ e che*

$$\lim_{x \rightarrow +\infty} f(x) = \ell.$$

Allora $\ell = 0$.

DIMOSTRAZIONE. Supponiamo ad esempio che sia $\ell > 0$; allora, se $x > \delta$, $f(x) > \ell/2$. Pertanto

$$\begin{aligned} (17.48) \quad \lim_{x \rightarrow +\infty} \int_a^x f(t) dt &= \int_a^\delta f(t) dt + \lim_{x \rightarrow +\infty} \int_\delta^x f(t) dt \geq \\ &\geq \int_a^\delta f(t) dt + \lim_{x \rightarrow +\infty} (x - \delta) \ell/2 = +\infty \end{aligned}$$

\square

Osserviamo che f può ammettere integrale improprio convergente in $[a, +\infty)$ senza che esista il limite di f per $x \rightarrow +\infty$, se ad esempio

$$(17.49) \quad f(x) = \begin{cases} 1 & i < x \leq i + 1/2^i \\ 0 & i + 1/2^i \leq x \leq i + 1 \end{cases}, \quad i \in \mathbb{N}$$

si ha

$$\int_1^{+\infty} f(t) dt = \lim_n \sum_{i=1}^n \frac{1}{2^i} = \lim_n \frac{1}{2} \frac{1 - 1/2^n}{1 - 1/2} = 1$$

Si può analogamente provare che

TEOREMA 17.30. *Sia f integrabile in $[a, x]$, per ogni $x \geq a$, e supponiamo che f sia infinitesima di ordine β per $x \rightarrow +\infty$.*

- (1) *Se esiste $\alpha \in \mathbb{R}$ tale che $\beta \geq \alpha > 1$, allora f ammette integrale improprio convergente in $[a, +\infty)$.*
- (2) *Se $\beta \leq 1$ allora $|f|$ ammette integrale improprio divergente in $[a, +\infty)$.*

Osserviamo, a proposito dei teoremi 17.28,?? che qualora $\beta \in \mathbb{R}$, in (1) è sufficiente prendere $\alpha = \beta$.

TEOREMA 17.31. *Siano $f, g, f \in \mathcal{C}^0$, $g \in \mathcal{C}^1$, e sia F una primitiva di f ; allora le seguenti condizioni sono sufficienti per la convergenza dell'integrale improprio di fg in $[a, +\infty)$:*

- (1) *F limitata in $[a, +\infty)$ e g monotona a 0 per $x \rightarrow +\infty$;*
- (2) *F convergente per $x \rightarrow +\infty$ e g monotona e limitata.*

Segue da

$$\int_a^x f(t)g(t)dt = F(x)g(x) - F(a)g(a) - \int_a^x F(t)g'(t)dt.$$

QUALCHE STUDIO DI FUNZIONE INTEGRALE

Lo studio di una funzione che sia data mediante un integrale ricorre in molti casi: ad esempio quando si studiano le soluzioni di equazioni differenziali in cui compaiono funzioni che non ammettono primitive elementari o che ammettono primitive elementari non facilmente calcolabili.

Per funzione integrale si intende una funzione definita da

$$(18.1) \quad F(x) = \int_{x_0}^x f(t)dt$$

I risultati che occorre tener ben presenti quando si studia una funzione integrale sono i seguenti:

- (1) I risultati che sono sufficienti a garantire l'integrabilità di una funzione: sono quelli contenuti nei teoremi ?? e si possono brevemente riassumere dicendo che:

- f è integrabile su ogni intervallo su cui è continua
- f è integrabile su ogni intervallo su cui è monotona
- f è integrabile su ogni intervallo su cui è limitata e continua a meno di un numero finito di punti
- f è integrabile su ogni intervallo su cui differisce da una funzione integrabile a meno di un insieme finito di punti

- (2) Il risultato che assicura che se f è limitata allora F è continua.
 (3) il teorema fondamentale del calcolo che assicura che se f è continua in x allora F è derivabile in x e

$$F'(x) = f(x)$$

- (4) la definizione di integrale improprio per cui:

- Se f è integrabile in senso improprio in c^+

$$\lim_{x \rightarrow c^+} F(x) = \int_{x_0}^c f(t)dt$$

- Se f è integrabile in senso improprio a $+\infty$

$$\lim_{x \rightarrow +\infty} F(x) = \int_{x_0}^{+\infty} f(t)dt$$

Vale inoltre la pena di ricordare che se

$$(18.2) \quad G(x) = \int_{x_0}^{\beta(x)} f(t) dt$$

allora

$$(18.3) \quad G(x) = F(\beta(x))$$

e quindi, se le funzioni in gioco sono derivabili

$$(18.4) \quad G'(x) = F'(\beta(x))\beta'(x) = f(\beta(x))\beta'(x)$$

Inoltre se

$$(18.5) \quad G(x) = \int_{\alpha(x)}^{\beta(x)} f(t) dt$$

allora se c è scelto nel campo di integrabilità di f , si ha

$$(18.6) \quad G(x) = \int_{\alpha(x)}^{\beta(x)} f(t) dt = \int_{\alpha(x)}^c f(t) dt + \int_c^{\beta(x)} f(t) dt = \int_c^{\beta(x)} f(t) dt - \int_c^{\alpha(x)} f(t) dt$$

e quindi

$$(18.7) \quad G'(x) = F'(\beta(x))\beta'(x) - F'(\alpha(x))\alpha'(x) = f(\beta(x))\beta'(x) - f(\alpha(x))\alpha'(x)$$

1. Esempio

Si consideri la funzione

$$f(x) = \int_0^x \frac{e^{t^2}}{\sqrt[3]{1-e^t}(t-1)\sqrt{t+2}} dt$$

Cominciamo a determinare il dominio di f .

La funzione integranda risulta definita e continua (e quindi integrabile) per $t \in (-2, 0) \cup (0, 1) \cup (1, +\infty)$.

Inoltre

$$\lim_{t \rightarrow -2} \frac{e^{t^2}}{\sqrt[3]{1-e^t}(t-1)\sqrt{t+2}} = -\infty$$

di ordine $\frac{1}{2}$ in quanto l'integranda è infinita a causa del fattore $\sqrt{t+2}$ presente nel denominatore;

$$\lim_{t \rightarrow 0^-} \frac{e^{t^2}}{\sqrt[3]{1-e^t}(t-1)\sqrt{t+2}} = -\infty$$

di ordine $\frac{1}{3}$ a causa del fattore a denominatore $\sqrt[3]{1-e^t}$ (si ricordi che $1-e^t$ è infinitesimo in zero di ordine 1); analogamente

$$\lim_{t \rightarrow 0^+} \frac{e^{t^2}}{\sqrt[3]{1-e^t}(t-1)\sqrt{t+2}} = +\infty$$

di ordine $\frac{1}{3}$
 Infine

$$\lim_{t \rightarrow 1^-} \frac{e^{t^2}}{\sqrt[3]{1-e^t}(t-1)\sqrt{t+2}} = +\infty$$

di ordine 1 a causa del fattore $t-1$ a denominatore.

Ne segue che la funzione integranda è integrabile (eventualmente in senso improprio) in $[-2, 1) \cup (1, +\infty)$.

Poichè gli estremi di integrazione sono 0 ed x dovrà essere $x \in [-2, 1)$.

Dal momento che

- la funzione integranda è continua per $t \in (-2, 0) \cup (0, 1) \cup (1, +\infty)$
- f è definita in $[-2, 1)$

il teorema fondamentale del calcolo assicura che

$$f'(x) = \frac{e^{x^2}}{\sqrt[3]{1-e^x}(x-1)\sqrt{x+2}}$$

Per ogni $x \in (-2, 0) \cup (0, 1)$

Per quanto riguarda i punti $x = -2$ e $x = 0$ si è già visto che

$$\lim_{x \rightarrow -2} f'(x) = -\infty$$

$$\lim_{x \rightarrow 0^-} f'(x) = -\infty$$

$$\lim_{x \rightarrow 0^+} f'(x) = +\infty$$

cui f non è derivabile per $x = -1$ ed $x = 0$.

Per tracciare il grafico di f dobbiamo tenere conto che

- $f'(x) > 0$ per $x \in (0, 1)$
- $\lim_{x \rightarrow 1^-} f(x) = +\infty$,
- f non è derivabile in -2 ed in 0
- in -2 ed in 0 il grafico ha tangente verticale

Pertanto il grafico di f risulta:

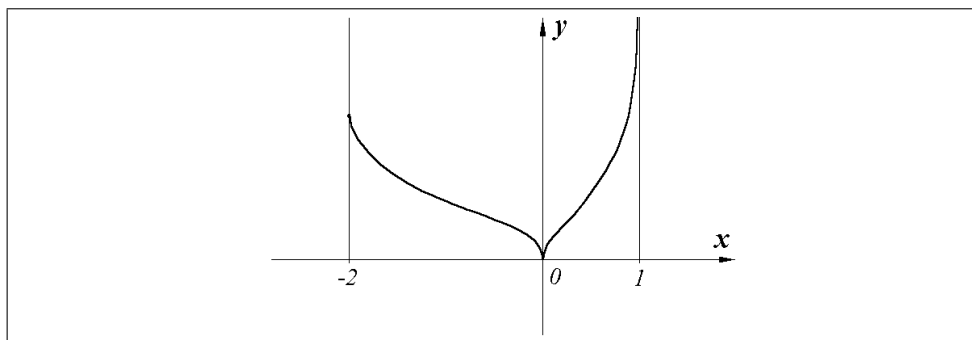


FIGURA 18.1. Grafico di $f(x)$

Consideriamo ora

$$g(x) = \int_0^{|x|} \frac{e^{t^2}}{\sqrt[3]{1-e^t}(t-1)\sqrt{t+2}} dt$$

Dal momento che

$$g(x) = f(|x|)$$

il grafico di g sarà uguale a quello di f per gli $x \in [0, 1)$ ed il simmetrico rispetto all'asse y per gli $x \in (-1, 0]$.

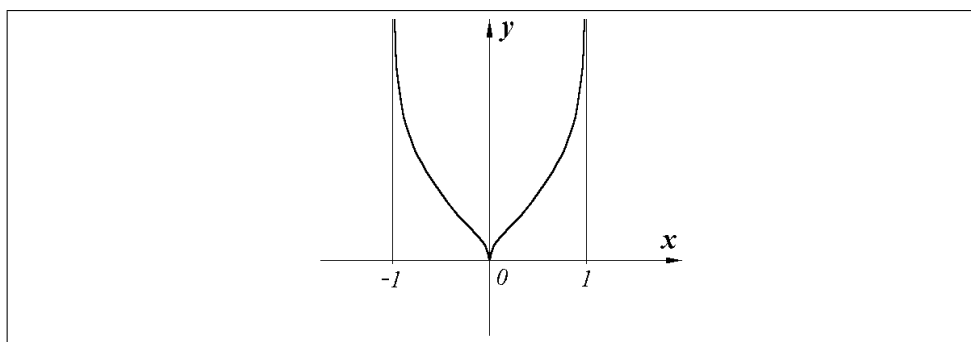


FIGURA 18.2. Grafico di $g(x)$

Se infine consideriamo

$$h(x) = \int_{x^3}^{x^2+2} \frac{e^{t^2}}{\sqrt[3]{1-e^t}(t-1)\sqrt{t+2}} dt$$

per quanto visto ai punti precedenti, (f è integrabile in $[-2, 1) \cup (1, +\infty)$) L'intervallo di integrazione dovrà essere contenuto nell'insieme in cui è possibile calcolare l'integrale; dovrà cioè risultare che

$$[x^3, x^2+2] \subset [-2, 1) \cup (1, +\infty)$$

e quindi la funzione h risulta definita per $x^3 > 1$ ovvero per $x > 1$.

Poiché l'integranda è continua per $t > 1$ e gli estremi di integrazione sono derivabili, si ha, per $x \in (1, +\infty)$

$$h'(x) = \frac{2xe^{(x^2+2)^2}}{\sqrt[3]{1-e^{x^2+2}}(x^2-1)\sqrt{x^2+4}} - \frac{3x^2e^{x^6}}{\sqrt[3]{1-e^{x^3}}(x^3-1)\sqrt{x^3+2}}$$

2. Esempio

Consideriamo la funzione

$$f(x) = \int_x^{4-x} \frac{3 + \cos(t)}{(t-4)\sqrt[3]{t^5-1}} dt$$

La funzione integranda è definita e continua (e quindi integrabile) in $(-\infty, 1) \cup (1, 4) \cup (4, +\infty)$.

Inoltre

$$\lim_{t \rightarrow 1} \left| \frac{3 + \cos(t)}{(t-4)\sqrt[3]{t^5-1}} \right| = +\infty \quad \text{di ordine } \frac{1}{3}$$

mentre

$$\lim_{t \rightarrow 4} \left| \frac{3 + \cos(t)}{(t-4)\sqrt[3]{t^5-1}} \right| = +\infty \quad \text{di ordine 1}$$

Pertanto l'integranda risulta integrabile (anche in senso improprio) in $(-\infty, 4) \cup (4, +\infty)$.

Dovrà allora essere

$$\begin{cases} x < 4 \\ 4 - x < 4 \end{cases} \quad \text{oppure} \quad \begin{cases} x > 4 \\ 4 - x > 4 \end{cases}$$

ovvero $0 < x < 4$.

Essendo gli estremi di integrazione due funzioni continue e derivabili, f risulta continua in tutto il suo dominio (perché l'integranda è integrabile) e derivabile per $x \neq 1$ e $4 - x \neq 1$ essendo l'integranda continua per $t \neq 1$ (per $t = 1$ l'integranda è infinita).

Pertanto l'insieme di continuità è $(0, 4)$ e l'insieme di derivabilità è $(0, 1) \cup (1, 3) \cup (3, 4)$.

Dal teorema fondamentale del calcolo integrale e dalla formula di derivazione delle funzioni composte si ha, se $x \in (0, 1) \cup (1, 3) \cup (3, 4)$

$$f'(x) = -\frac{3 + \cos(4-x)}{(-x)\sqrt[3]{(4-x)^5-1}} - \frac{3 + \cos(x)}{(x-4)\sqrt[3]{x^5-1}}$$

3. Esempio

Si considerino le funzioni

$$h(x) = x^4 + 8x + k \quad \text{e} \quad g(x) = \frac{1}{h(x)}$$

Si ha, per ogni $k \in \mathbb{R}$,

$$h : \mathbb{R} \rightarrow \mathbb{R}, \quad \text{continua} \quad , \quad \lim_{x \rightarrow \pm\infty} h(x) = +\infty$$

per cui h non è limitata superiormente (e quindi non ammette massimo globale), mentre, per il teorema di Weierstrass generalizzato, ammette minimo assoluto, per ogni $k \in \mathbb{R}$.

Se g risulta continua allora ha primitive in \mathbb{R} ed inoltre, (essendo infinita dove non è continua, g ammette primitive se e solo se $h(x) = x^4 + 8x + k \neq 0$ per ogni $x \in \mathbb{R}$).

Poiché come visto nel punto precedente h ha minimo assoluto, e tale valore è assunto nel punto $x = -\sqrt[3]{2}$ (ove si annulla $h'(x) = 4x^3 + 8$) e si ha $h(-\sqrt[3]{2}) = k - 6\sqrt[3]{2}$, si conclude che g ha primitive in \mathbb{R} se e solo se

$$k - 6\sqrt[3]{2} > 0 \quad \text{ovvero} \quad k > 6\sqrt[3]{2}$$

Similmente g ha primitive in $[-1, +\infty)$ se e solo se risulta continua in tale intervallo ovvero se e solo se $h(x) = x^4 + 8x + k \neq 0$ per ogni $x \geq -1$.

Essendo h crescente per $x \geq -\sqrt[3]{2}$, e quindi per $x \geq -1$, g ha primitive in $[-1, +\infty)$ se e solo se $h(-1) = k - 7 > 0$ ovvero

$$k > 7$$

Posto $k = 0$ si ha $g(x) = \frac{1}{x^4+8x}$, che è definita e continua in $(-\infty, -2) \cup (-2, 0) \cup (0, +\infty)$.

Utilizzando la decomposizione in fratti semplici si ha

$$\frac{1}{x^4+8x} = \frac{a}{x} + \frac{b}{x+2} + \frac{cx+d}{x^2-2x+4} = \frac{(a+b+c)x^3 + (2c-2b+d)x^2 + (4b+2d)x + 8a}{x^4+8x}$$

da cui

$$\begin{cases} a+b+c=0 \\ 2c-2b+d=0 \\ 4b+2d=0 \\ 8a=1 \end{cases}$$

che risolto fornisce $a = \frac{1}{8}$, $b = -\frac{1}{24}$, $c = -\frac{1}{12}$, $d = \frac{1}{12}$; pertanto

$$\frac{1}{x^4+8x} = \frac{1}{24} \left(\frac{3}{x} - \frac{1}{x+2} - \frac{2x-2}{x^2-2x+4} \right)$$

Una primitiva di g è quindi

$$\frac{1}{24} \ln \left| \frac{x^3}{(x+2)(x^2-2x+4)} \right| = \frac{1}{24} \ln \left| \frac{x^3}{x^3+8} \right|$$

Tutte le primitive di g in $(-\infty, -2) \cup (-2, 0) \cup (0, +\infty)$ sono pertanto

$$\begin{cases} \frac{1}{24} \ln \frac{x^3}{x^3+8} + c_1 & , \text{ se } x < -2 \\ \frac{1}{24} \ln \frac{-x^3}{x^3+8} + c_2 & , \text{ se } -2 < x < 0 \\ \frac{1}{24} \ln \frac{x^3}{x^3+8} + c_3 & , \text{ se } x > 0 \end{cases}$$

Si consideri ora il problema

$$\begin{cases} y''(x) = g(x) \\ y(0) = 0 \\ y'(0) = 0 \end{cases}$$

Il problema ha soluzioni in \mathbb{R} se e solo se g ha primitive in \mathbb{R} , poiché y' è la primitiva di g che soddisfa $y'(0) = 0$ e di conseguenza y è la primitiva di $\int_0^x g(t)dt$ che soddisfa $y(0) = 0$ cioè

$$y(x) = \int_0^x \left(\int_0^s g(t)dt \right) ds$$

Pertanto il problema dato ha una ed una sola soluzione per $k > 6\sqrt[3]{2}$.

INTRODUZIONE AI MODELLI DIFFERENZIALI

Uno degli argomenti più interessanti del calcolo differenziale è costituito dalle equazioni differenziali: si tratta di equazioni in cui l'incognita è una funzione $y(x)$ di cui sono noti i valori iniziali ed il fatto che deve essere verificata, per ogni x , una relazione tra la funzione stessa e la sua derivata prima $y'(x)$.

L'esempio più semplice e naturale di un problema di questo genere è dato dal modello che descrive la caduta di un grave.

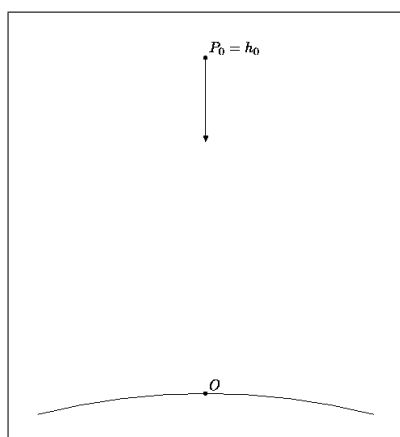


FIGURA 19.1. Un punto materiale soggetto alla gravità

Se consideriamo un punto di massa m posto ad un'altezza h dalla superficie terrestre e trascuriamo gli effetti della resistenza dell'aria, avremo che sul punto agisce solo la forza di gravità $F = mg$.

L'esperienza mostra che il punto materiale P si muove verso il basso; per descrivere il suo moto possiamo considerare un sistema di riferimento che coincide con la retta che il punto percorre cadendo.

Assumiamo l'origine in corrispondenza del suolo e consideriamo positive le altezze misurate dal suolo.

La velocità con cui il punto P si muove verso il basso lungo la retta scelta come asse di riferimento è

$$v(t) = \dot{x}(t)$$

e la sua accelerazione è

$$a(t) = \ddot{x}(t)$$

Come già detto, sul punto agisce la sola forza gravitazionale $F = mg$.

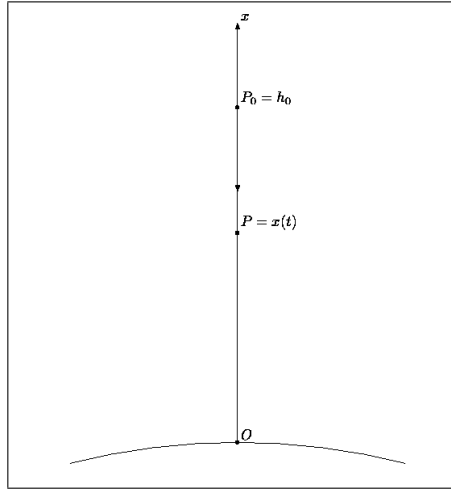


FIGURA 19.2. Il sistema di riferimento

Per le leggi di Newton si avrà allora

$$ma(t) = -mg$$

e quindi

$$(19.1) \quad \ddot{x}(t) = g$$

La 19.1 è un semplicissimo esempio di equazione differenziale: essa impone una relazione che coinvolge una funzione e le sue derivate.

Il moto del punto si può ricavare integrando due volte tra t e $t_0 = 0$, e assumiamo che il moto inizi all'istante $t_0 = 0$.

Si ottiene

$$(19.2) \quad \dot{x}(t) = -gt + c_1$$

e

$$(19.3) \quad x(t) = -\frac{1}{2}gt^2 + c_1t + c_0$$

e si vede che per determinare in maniera unica il moto dovremo procurarci dei valori per c_0 e c_1 . Questo si può fare utilizzando informazioni sulla velocità e sulla posizione iniziale del punto. È subito visto infatti dalla 19.2 e dalla 19.3 rispettivamente che

$$(19.4) \quad v_0 = \dot{x}(0) = c_1 \quad h_0 = x(0) = c_0$$

Possiamo osservare che per determinare il moto abbiamo cioè bisogno di conoscere posizione e velocità iniziale del punto P e ciò corrisponde anche all'intuizione.

Se teniamo conto di tali dati, possiamo affermare che il punto P si muove sull'asse x seguendo la legge

$$(19.5) \quad x(t) = -\frac{1}{2}gt^2 + v_0t + h_0$$

Possiamo descrivere lo stesso fenomeno anche usando il principio di conservazione dell'energia.

L'energia potenziale del punto P , soggetto al solo campo gravitazionale è, in ogni istante t ,

$$U(t) = mgx(t)$$

mentre la sua energia cinetica è

$$\frac{1}{2}m\dot{x}^2(t)$$

e la sua energia totale

$$E(t) = \frac{1}{2}m\dot{x}^2(t) + mgx(t)$$

si mantiene costante durante il moto

$$(19.6) \quad \frac{1}{2}m\dot{x}^2(t) + mgx(t) = mk$$

Se conosciamo le condizioni iniziali v_0 ed h_0 siamo anche in grado di calcolare

$$k = \frac{1}{2}mv_0^2 + mgh_0$$

La 19.6 è una equazione differenziale, che è in grado di descrivere la posizione $x(t)$ del punto P in ogni istante t , tuttavia ricavare x da tale relazione è più difficile.

Possiamo riscrivere la 19.6 come

$$(19.7) \quad \frac{1}{2}\dot{x}^2(t) = k - gx(t)$$

e da questa uguaglianza possiamo ricavare una prima informazione:

la quantità $k - gx(t)$ deve mantenersi positiva e quindi $x(t) \leq \frac{k}{g}$.

Abbiamo così ricavato una limitazione per la soluzione dell'equazione senza risolverla, abbiamo ottenuto cioè una limitazione a priori per la soluzione dell'equazione.

Osserviamo anche che

$x(t) = \frac{k}{g}$ è una soluzione costante dell'equazione 19.7

Per cercare soluzioni non costanti possiamo applicare la radice ad entrambi i membri

$$(19.8) \quad \dot{x}(t) = \pm \sqrt{2k - 2gx(t)}$$

e dividere per il secondo membro

$$(19.9) \quad \frac{\dot{x}(t)}{\sqrt{2k - 2gx(t)}} = \pm 1$$

Ora, se moltiplichiamo per g

$$(19.10) \quad \frac{g\dot{x}(t)}{\sqrt{2k - 2gx(t)}} = \pm g$$

ed integriamo tra $t_0 = 0$ e t , otteniamo

$$(19.11) \quad \int_0^t \frac{g\dot{x}(s)}{\sqrt{2k - 2gx(s)}} ds = \pm gt$$

dove tuttavia il primo integrale non può essere calcolato in quanto la funzione integranda dipende dalla funzione incognita $x(t)$.

Possiamo integrare per sostituzione ponendo

$$u = x(s) \quad , \quad du = \dot{x}(s)ds$$

osservando che per $s = 0$ e $s = t$ avremo $x(s) = x(0) = h_0$ e $x(s) = x(t)$, da cui si ricava che $v_0 = \pm\sqrt{2k - 2gx_0}$, avremo

$$(19.12) \quad \int_{x_0}^{x(t)} \frac{gdu}{\sqrt{2k - 2gu}} = \pm gt$$

A questo punto possiamo calcolare l'integrale a sinistra ed ottenere che

$$(19.13) \quad \sqrt{2k - 2gx(t)} - \sqrt{2k - 2gx_0} = \pm gt$$

$$(19.14) \quad \sqrt{2k - 2gx(t)} = \pm gt + v_0$$

$$(19.15) \quad 2k - 2gx(t) = (\pm gt + v_0)^2$$

$$(19.16) \quad x(t) = \frac{k}{g} - \frac{1}{2g} (\pm gt + v_0)^2$$

La 19.16 descrive il moto del punto negli stessi termini ottenuti in precedenza; il segno \pm di $\pm gt$ si può determinare dalla 19.8: poichè il moto avviene con continuità il segno dovrà essere lo stesso di v_0 .

La scelta del segno e la validità dell'equazione si mantengono fino a quando la derivata di $x(t)$, cioè la velocità non si annulla; questa eventualità non si verifica mai se $v_0 < 0$ mentre ha luogo per $t_0 = \frac{v_0}{g}$ nel caso in cui $v_0 > 0$.

In tal caso dobbiamo riconsiderare le condizioni iniziali che diventano

$$x(t_0) = \dot{x}(t_0) = 0$$

e quindi non forniscono indicazioni sul segno da attribuire alla radice che rappresenta la velocità nella 19.8.

Dobbiamo quindi esaminare tutti i casi disponibili:

(1) se supponiamo che il moto abbia velocità positive

$$(19.17) \quad \dot{x}(t) = +\sqrt{2k - 2gx(t)}$$

(2) se supponiamo che il moto abbia velocità negative

$$(19.18) \quad \dot{x}(t) = -\sqrt{2k - 2gx(t)}$$

(3) se supponiamo che il moto abbia velocità nulla entrambe le precedenti sono accettabili.

Osserviamo che a questo punto occorre distinguere tra risultato del modello e soluzione dell'equazione differenziale: infatti

È evidente che per $t > t_0$ la 19.17 non può più rappresentare il moto del punto materiale P in quanto il moto avviene con velocità negativa, il che non è consentito dalla 19.17.

L'unica soluzione prevista dalla 19.17 è quella costante che tuttavia è in contrasto con l'evidenza del fenomeno.

Dovremo pertanto considerare le soluzioni dell'equazione 19.18 per trovare la descrizione del seguito del movimento.

CAPITOLO 20

EQUAZIONI DIFFERENZIALI A VARIABILI SEPARABILI.

Risolvere una equazione differenziale a variabili separabili, significa trovare una funzione y , che sia derivabile e per cui si abbia

$$y'(x) = f(x)g(y(x))$$

con f, g assegnate.

Più precisamente possiamo dire che

Se $I, J \subset \mathbb{R}$ sono intervalli aperti e non vuoti ed $f : I \longrightarrow \mathbb{R}$, $g : J \longrightarrow \mathbb{R}$ sono due funzioni, diciamo che risolviamo l'equazione differenziale a variabili separabili

$$(20.1) \quad y'(x) = f(x)g(y(x))$$

se troviamo un intervallo $I' \subset I$ ed una funzione $y : I' \longrightarrow J$ tale che la 20.1 sia soddisfatta per ogni $x \in I'$

Quando si cercano soluzioni di un'equazione differenziale che soddisfino anche un dato iniziale, si parla di problema di Cauchy.

Precisamente se

$I, J \subset \mathbb{R}$ sono intervalli aperti $x_0 \in I, y_0 \in J$, $f : I \longrightarrow \mathbb{R}$ e $g : J \longrightarrow \mathbb{R}$ sono funzioni; chiamiamo problema di Cauchy a variabili separabili il problema di trovare $I' \subset I$ ed $y : I' \longrightarrow \mathbb{R}$, derivabile, tali che

$$(20.2) \quad \begin{cases} y'(x) = f(x)g(y(x)) & , \quad \forall x \in I' \\ y(x_0) = y_0 \end{cases}$$

Vale il seguente teorema di esistenza ed unicità della soluzione del problema di Cauchy a variabili separabili, per dimostrare il quale procediamo in maniera costruttiva utilizzando un metodo che, di fatto, consente di risolvere l'equazione.

La dimostrazione è, in questo caso, molto più utile dell'enunciato, ma anche le condizioni di esistenza ed unicità della soluzione sono di fondamentale importanza.

Nei teoremi che segue giocano un ruolo fondamentale il fatto che I e J siano intervalli aperti e che $g(y) \neq 0 \forall y \in J$.

Quest'ultima condizione è certamente soddisfatta se g è continua e se $g(y_0) \neq 0$ a meno di considerare un intervallo J più piccolo.

TEOREMA 20.1. *Siano $I, J \subset \mathbb{R}$, intervalli aperti, siano $x_0 \in I, y_0 \in J$ e siano $f : I \rightarrow \mathbb{R}, g : J \rightarrow \mathbb{R}$ due funzioni continue, supponiamo inoltre che $g(y) \neq 0$, per ogni $y \in J$.*

Allora esiste un intervallo $I' \subset I$ e una ed una sola soluzione $y : I' \rightarrow J$ del problema di Cauchy 20.2.

DIMOSTRAZIONE.

y è soluzione del problema assegnato se e solo se

$$(20.3) \quad \begin{cases} \frac{y'(x)}{g(y(x))} = f(x) \\ y(x_0) = y_0 \end{cases}$$

e ciò si verifica se e solo se

$$(20.4) \quad \int_{x_0}^x \frac{y'(t)}{g(y(t))} dt = \int_{x_0}^x f(t) dt$$

se e solo se

$$(20.5) \quad \int_{y_0}^{y(x)} \frac{ds}{g(s)} = \int_{x_0}^x f(t) dt$$

se e solo se, dette F e G due primitive di f ed $1/g$ su I e J rispettivamente,

$$(20.6) \quad G(y(x)) - G(y_0) = F(x) - F(x_0)$$

$R(G - G(y_0))$ e $R(F - F(x_0))$ sono intervalli per la continuità delle medesime, entrambi contengono 0 e $R(G)$ contiene 0 al suo interno in virtù del fatto che G è strettamente monotona in quanto $g = G'$ ha segno costante inoltre G è invertibile.

Ciò assicura che esiste un intervallo I' , aperto e contenente x_0 in cui l'uguaglianza vale ed in tale intervallo si può scrivere che

$$(20.7) \quad y(x) = G^{-1}(F(x) + G(y_0) - F(x_0)).$$

□

È importante anche ricordare due risultati di esistenza e di unicità la cui dimostrazione non è opportuna a questo punto, che possiamo tuttavia utilizzare per ottenere informazioni sull'esistenza e l'unicità della soluzione di un problema di Cauchy.

Siano $I, J \subset \mathbb{R}$, intervalli aperti, siano $x_0 \in I, y_0 \in J$ e siano $f : I \rightarrow \mathbb{R}, g : J \rightarrow \mathbb{R}$ due funzioni continue.

Allora esiste un intervallo $I' \subset I$ e una soluzione $y : I' \rightarrow J$ del problema di Cauchy 20.2.

Se inoltre $g \in \mathcal{C}^1$, cioè se ammette derivata prima continua, allora la soluzione è anche unica.

L'unicità è anche assicurata dalla lipschitzianità di g cioè dalla condizione

$$(20.8) \quad |g(x) - g(y)| \leq L|x - y|$$

Vale la pena di ricordare che, usando il teorema di Lagrange, si può dimostrare che una funzione che abbia derivata prima limitata è lipschitziana: infatti se $|g'(c)| \leq L$ si ha

$$(20.9) \quad |g(x) - g(y)| = |g'(c)||x - y| \leq L|x - y|$$

Ricordiamo anche che se $g \in \mathcal{C}^1$, il teorema di Weierstraß assicura che $|g'|$ (che è continua) ammette massimo su ogni intorno chiuso e limitato di x_0

Possiamo procedere alla soluzione dell'equazione differenziale a variabili separabili anche senza precisi riferimenti ai dati iniziali seguendo essenzialmente gli stessi passi percorsi in precedenza

Siano f e g continue sugli intervalli aperti I e J e supponiamo che $g(y) \neq 0$ su J ;

Consideriamo l'equazione a variabili separabili

$$(20.10) \quad y'(x) = f(x)g(y(x))$$

Dal momento che $g(y) \neq 0$ in J , avremo che la **20.10** è soddisfatta in I' se e solo se

$$\frac{y'(x)}{g(y(x))} = f(x)$$

e, dette F e G due primitive in I e J di f ed $1/g$ rispettivamente, l'ultima uguaglianza è equivalente a

$$(20.11) \quad G(y(x)) = F(x) + c$$

con $c \in \mathbb{R}$ (ricordiamo che stiamo lavorando su intervalli e quindi due primitive differiscono per costante).

Ora, se fissiamo x_0 interno ad I e chiamiamo $y(x_0) = y_0 \in J$, posto

$$c = G(y_0) - F(x_0)$$

avremo che la **20.11** diventa

$$(20.12) \quad G(y(x)) - G(y_0) = F(x) - F(x_0)$$

ed è verificata almeno in un intervallo $I' \subset I$.

Infatti $R(G - G(y_0))$ e $R(F - F(x_0))$ sono intervalli per la continuità delle medesime, entrambi contengono 0 e $R(G)$ contiene 0 al suo interno in virtù del fatto che G è strettamente monotona in quanto $g = G'$ ha segno costante inoltre G è invertibile.

Pertanto possiamo ricavare

$$y(x) = G^{-1}(F(x) + c)$$

per $x \in I'$.

Il procedimento sopra esposto fornisce, al variare di c , l'insieme di tutte le soluzioni dell'equazione differenziale a variabili separabili considerata. Allorquando necessiti trovare le soluzioni dell'equazione considerata, che soddisfino di più la condizione $y(x_0) = y_0$, $x_0 \in I$, $y_0 \in J$, è sufficiente considerare $c = G(y_0) - F(x_0)$ ed osservare che tale scelta di c consente di determinare $I' \subset I$ tale che $F(I') + c \subset G(J)$. In tal caso si risolve un problema di Cauchy.

Per una corretta risoluzione di un'equazione a variabili separabili non va trascurato di considerare quanto accade se g si annulla in qualche punto.

Ricordiamo che per separare le variabili occorre dividere per g e quindi in questo caso non si può procedere già dall'inizio.

È ragionevole limitarci al caso in cui y_0 è uno zero isolato di g , cioè se esiste un intorno di y_0 in cui g non si annulla altre volte.

In tal caso possiamo osservare che la funzione

$$y(x) = y_0$$

è una soluzione dell'equazione, che in presenza di condizioni che assicurino l'unicità è anche la sola soluzione possibile.

Qualora non sussistano tali condizioni occorre indagare l'esistenza di altre soluzioni; a questo scopo si procede studiando l'equazione per $y \neq y_0$ e, giunti al punto di considerare

$$(20.13) \quad \int_{y_0}^{y(x)} \frac{ds}{g(s)} = \int_{x_0}^x f(t) dt$$

prima di procedere, occorre studiare l'esistenza in senso improprio dell'integrale a sinistra.

Le informazioni che abbiamo sull'integrazione impropria ci consentono allora di capire che:

- se g è infinitesima in y_0 di ordine $\alpha \geq 1$. la primitiva G di $1/g$ non può essere prolungata per continuità in y_0 e pertanto la soluzione costante è l'unica possibile.
- Se invece g è infinitesima in y_0 di ordine $\alpha \leq \beta < 1$, $\beta \in \mathbb{R}$. Allora G può essere prolungata per continuità in y_0 e, si può procedere oltre.

Proviamo infine un risultato riguardante una disequazione differenziale che è spesso utile per trovare limitazioni a priori per soluzioni di equazioni differenziali che non si è in grado di risolvere.

LEMMA 20.1. - di Gronwall - Siano $y, f : I \longrightarrow \overline{\mathbb{R}}_+$ funzioni continue, I intervallo, e siano $c > 0$, $x_0 \in I$; allora se

$$y(x) \leq \left| \int_{x_0}^x f(t)y(t) dt \right| + c$$

per ogni $x \in I$ si ha

$$0 \leq y(x) \leq ce^{\left| \int_{x_0}^x f(t) dt \right|}$$

per ogni $x \in I$.

DIMOSTRAZIONE. Supponiamo $x \geq x_0$; dividendo ambo i membri per il secondo e moltiplicando poi per $f(x)$ si ottiene (si ricordi che $f \geq 0$, $c > 0$)

$$\frac{y(x)f(x)}{c + \int_{x_0}^x f(t)y(t) dt} \leq f(x)$$

da cui

$$\frac{d}{dx} \left[\ln \left(c + \int_{x_0}^x f(t)y(t) dt \right) \right] \leq f(x).$$

Integrando ora tra x_0 ed x si ha

$$\ln \left(c + \int_{x_0}^x f(t)y(t) dt \right) - \ln c \leq \int_{x_0}^x f(t) dt$$

onde

$$c + \int_{x_0}^x f(t)y(t)dt \leq ce^{\int_{x_0}^x f(t)dt}.$$

e

$$y(x) \leq c + \int_{x_0}^x f(t)y(t)dt \leq ce^{\int_{x_0}^x f(t)dt}$$

Se $x \leq x_0$ si procede in modo analogo solo tenendo conto di un cambiamento di segno. \square

COROLLARIO 20.1. *Siano $y, f : I \longrightarrow \overline{\mathbb{R}}_+$ continue, I intervallo, e sia $x_0 \in I$; allora se*

$$y(x) \leq \left| \int_{x_0}^x f(t)y(t)dt \right| \quad \forall x \in I$$

si ha

$$y(x) = 0 \quad \forall x \in I$$

DIMOSTRAZIONE. Si ha

$$y(x) \leq \left| \int_{x_0}^x f(t)y(t)dt \right| + c \quad \forall c > 0$$

e pertanto

$$0 \leq y(x) \leq ce^{\left| \int_{x_0}^x f(t)dt \right|} \quad \forall c > 0$$

per cui, al limite per $c \rightarrow 0^+$, si ha $y(x) \equiv 0$. \square

Se nel lemma di Gronwall si suppone

$$y(x) \leq \left| \int_{x_0}^x f(t)y(t)dt \right| + c(x)$$

con c , si prova che

$$0 \leq y(x) \leq c(x)e^{\left| \int_{x_0}^x f(t)dt \right|}$$

CAPITOLO 21

ESEMPI NOTEVOLI DI PROBLEMI DI CAUCHY

1. Esempio

Consideriamo l'equazione

$$(21.1) \quad y'(x) = y^2(x)$$

Osserviamo innanzi tutto che $y(x) \equiv 0$ è soluzione dell'equazione.

Se $y(x) \neq 0$ possiamo separare le variabili

$$(21.2) \quad \frac{y'(x)}{y^2(x)} = 1$$

ed integrando tra x_0 ed x

$$(21.3) \quad \int_{x_0}^x \frac{y'(t)}{y^2(t)} dt = x - x_0$$

posto $s = y(t)$, avremo $ds = y'(t)dt$ e

$$(21.4) \quad \int_{y(x_0)=y_0}^{y(x)} \frac{ds}{s^2} = x - x_0$$

Poichè $\frac{1}{s^2}$ è infinita in $s = 0$ di ordine 2, non è integrabile in $s = 0$ (intendiamo con ciò che non è integrabile in intervalli che contengano 0). Pertanto y ed y_0 dovranno avere sempre lo stesso segno: soluzioni che partono con valori y_0 positivi (negativi), rimangono positive (negative).

Sotto tale condizione avremo che

$$(21.5) \quad -\frac{1}{y} + \frac{1}{y_0} = x - x_0$$

$$(21.6) \quad \frac{1}{y} = \frac{1}{y_0} + x_0 - x = c - x$$

dove si sia definito

$$c = \frac{1}{y_0} + x_0$$

Osserviamo inoltre che al variare di x_0 ed y_0 c può assumere tutti i valori reali.

Le soluzioni dell'equazione saranno pertanto date da

$$(21.7) \quad y(x) = \frac{1}{c - x}$$

ed il loro grafico è indicato in figura 21.1.

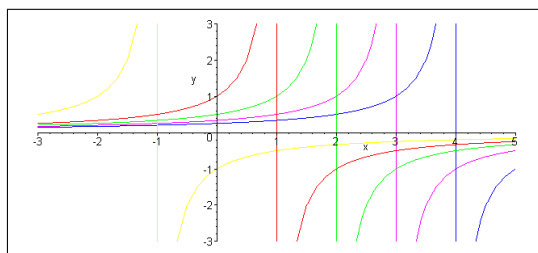


FIGURA 21.1.

2. Esempio

Consideriamo l'equazione

$$(21.8) \quad y'(x) = \sqrt{y(x)}$$

Osserviamo innanzi tutto che deve essere $y(x) \geq 0$ e che $y(x) \equiv 0$ è soluzione dell'equazione.

Se $y(x) \neq 0$ possiamo separare le variabili

$$(21.9) \quad \frac{y'(x)}{\sqrt{y(x)}} = 1$$

ed integrando tra x_0 ed x

$$(21.10) \quad \int_{x_0}^x \frac{y'(t)}{\sqrt{y(t)}} dt = x - x_0$$

posto $s = y(t)$, avremo $ds = y'(t)dt$ e

$$(21.11) \quad \int_{y(x_0)=y_0}^{y(x)} \frac{ds}{\sqrt{s}} = x - x_0$$

Poichè $\frac{1}{\sqrt{s}}$ è infinita in $s = 0$ di ordine $1/2$, è integrabile in $s = 0$ (intendiamo con ciò che è integrabile in intervalli che contengano 0). Pertanto y ed y_0 potranno assumere anche il valore 0. Avremo

$$(21.12) \quad 2\sqrt{y} - 2\sqrt{y_0} = x - x_0$$

$$(21.13) \quad \sqrt{y} = \frac{1}{2}(x - x_0 + 2\sqrt{y_0}) = \frac{1}{2}(x + c)$$

dove si sia definito

$$c = 2\sqrt{y_0} - x_0$$

Osserviamo inoltre che la 21.13 impone che deve essere

$$\frac{1}{2}(x+c) \geq 0 \quad \text{cioè} \quad x \geq -c$$

Osserviamo che al variare di x_0 ed y_0 c può assumere tutti i valori reali. Le soluzioni dell'equazione saranno pertanto date da

$$(21.14) \quad y(x) = \frac{1}{4}(x+c)^2 \quad \text{per} \quad x \geq -c$$

ed il loro grafico è indicato in figura 21.2.

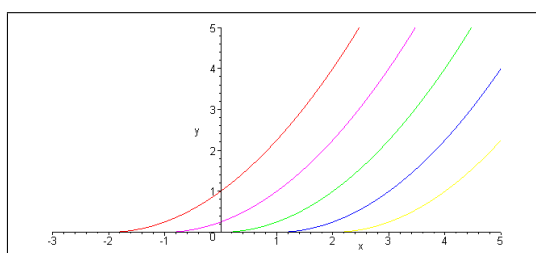


FIGURA 21.2.

3. Esempio

Consideriamo l'equazione

$$(21.15) \quad y'(x) = x\sqrt{y(x)}$$

Osserviamo innanzi tutto che deve essere $y(x) \geq 0$ e che $y(x) \equiv 0$ è soluzione dell'equazione.

Se $y(x) \neq 0$ possiamo separare le variabili

$$(21.16) \quad \frac{y'(x)}{\sqrt{y(x)}} = x$$

ed integrando tra x_0 ed x

$$(21.17) \quad \int_{x_0}^x \frac{y'(t)}{\sqrt{y(t)}} dt = \int_{x_0}^x t dt$$

posto $s = y(t)$, avremo $ds = y'(t)dt$ e

$$(21.18) \quad \int_{y(x_0)=y_0}^{y(x)} \frac{ds}{\sqrt{s}} = \frac{x^2}{2} - \frac{x_0^2}{2}$$

Poichè $\frac{1}{\sqrt{s}}$ è infinita in $s = 0$ di ordine $1/2$, è integrabile in $s = 0$ (intendiamo con ciò che è integrabile in intervalli che contengano 0). Pertanto y ed y_0 potranno assumere anche il valore 0. Avremo

$$(21.19) \quad 2\sqrt{y} - 2\sqrt{y_0} = \frac{x^2}{2} - \frac{x_0^2}{2}$$

$$(21.20) \quad \sqrt{y} = \frac{x^2}{4} + \left(\sqrt{y_0} - \frac{x_0^2}{2}\right) = \frac{x^2}{4} + c$$

dove si sia definito

$$c = \sqrt{y_0} - \frac{x_0^2}{2}$$

Osserviamo che la 21.19 impone che deve essere

$$\frac{x^2}{4} + c \geq 0 \quad \text{cioè} \quad \begin{cases} \text{sempre} & \text{se } c > 0 \\ |x| \geq -2c & \text{se } c < 0 \end{cases}$$

Osserviamo che al variare di x_0 ed y_0 c può assumere tutti i valori reali. Le soluzioni dell'equazione saranno pertanto date da

$$(21.21) \quad y(x) = \left(\frac{x^2}{4} + c\right)^2$$

sotto le condizioni indicate per x ed il loro grafico è indicato in figura 21.3.

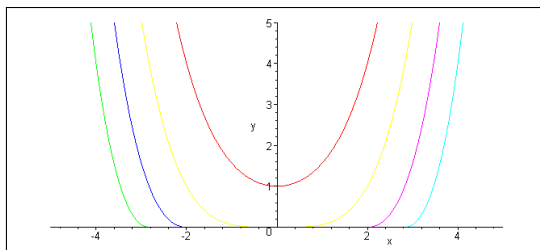


FIGURA 21.3.

4. Esempio

Consideriamo l'equazione

$$(21.22) \quad y'(x) = -x\sqrt{y(x)}$$

Osserviamo innanzi tutto che deve essere $y(x) \geq 0$ e che $y(x) \equiv 0$ è soluzione dell'equazione.

Se $y(x) \neq 0$ possiamo separare le variabili

$$(21.23) \quad \frac{y'(x)}{\sqrt{y(x)}} = -x$$

ed integrando tra x_0 ed x

$$(21.24) \quad \int_{x_0}^x \frac{y'(t)}{\sqrt{y(t)}} dt = - \int_{x_0}^x t dt$$

posto $s = y(t)$, avremo $ds = y'(t)dt$ e

$$(21.25) \quad \int_{y(x_0)=y_0}^{y(x)} \frac{ds}{\sqrt{s}} = -\frac{x^2}{2} + \frac{x_0^2}{2}$$

Poichè $\frac{1}{\sqrt{s}}$ è infinita in $s = 0$ di ordine $1/2$, è integrabile in $s = 0$ (intendiamo con ciò che è integrabile in intervalli che contengano 0). Pertanto y ed y_0 potranno assumere anche il valore 0. Avremo

$$(21.26) \quad 2\sqrt{y} - 2\sqrt{y_0} = -\frac{x^2}{2} + \frac{x_0^2}{2}$$

$$(21.27) \quad \sqrt{y} = -\frac{x^2}{4} + (\sqrt{y_0} + \frac{x_0^2}{2}) = -\frac{x^2}{4} + c$$

dove si sia definito

$$c = \sqrt{y_0} + \frac{x_0^2}{2}$$

Osserviamo che la 21.27 impone che deve essere

$$-\frac{x^2}{4} + c \geq 0 \quad \text{cioè} \quad \begin{cases} \text{mai} & \text{se } c < 0 \\ |x| \leq -2c & \text{se } c < 0 \end{cases}$$

Osserviamo che al variare di x_0 ed y_0 c può assumere solo valori positivi. Le soluzioni dell'equazione saranno pertanto date da

$$(21.28) \quad y(x) = \left(-\frac{x^2}{4} + c \right)^2$$

sotto le condizioni indicate per x ed il loro grafico è indicato in figura 21.4.

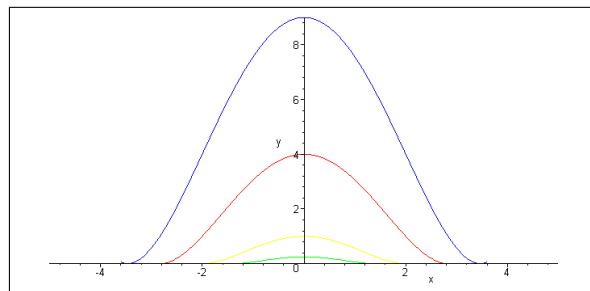


FIGURA 21.4.

5. Esempio

Consideriamo l'equazione

$$(21.29) \quad y'(x) = \sqrt{1 - y^2(x)}$$

Osserviamo innanzi tutto che deve essere $|y(x)| \leq 1$ e che $y(x) \equiv \pm 1$ è soluzione dell'equazione.

Se $y(x) \neq \pm 1$ possiamo separare le variabili

$$(21.30) \quad \frac{y'(x)}{\sqrt{1 - y^2(x)}} = 1$$

ed integrando tra x_0 ed x

$$(21.31) \quad \int_{x_0}^x \frac{y'(t)}{\sqrt{1 - y^2(t)}} dt = x - x_0$$

posto $s = y(t)$, avremo $ds = y'(t)dt$ e

$$(21.32) \quad \int_{y(x_0)=y_0}^{y(x)} \frac{ds}{\sqrt{1 - s^2}} = x - x_0$$

Poichè $\frac{1}{\sqrt{1-s^2}}$ è infinita in $s = \pm 1$ di ordine $1/2$, è integrabile in $s = \pm 1$ (intendiamo con ciò che è integrabile in intervalli che contengano ± 1). Pertanto y ed y_0 potranno assumere anche il valore ± 1 . Avremo

$$(21.33) \quad \arcsin y(x) - \arcsin y_0 = x - x_0$$

$$(21.34) \quad \arcsin y(x) = x - x_0 + \arcsin y_0 = x + c$$

dove si sia definito

$$c = \arcsin y_0 - x_0$$

Osserviamo che la 21.34 impone che deve essere

$$|x + c| \leq \frac{\pi}{2}$$

Osserviamo che al variare di x_0 ed y_0 c può assumere tutti i valori reali. Le soluzioni dell'equazione saranno pertanto date da

$$(21.35) \quad y(x) = \sin(x + c)$$

sotto le condizioni indicate per x ed il loro grafico è indicato in figura 21.5.

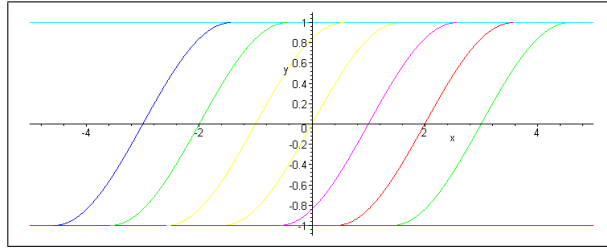


FIGURA 21.5.

6. Esempio

Consideriamo il problema di Cauchy

$$(21.36) \quad \begin{cases} y'(x) = e^{-(y(x))^4} - 1 \\ y(x_0) = y_0 \end{cases}$$

Possiamo scrivere

$$y'(x) = f(x)g(y(x))$$

se definiamo $f(x) = 1$ e $g(y) = e^{-y^4} - 1$;

Si ha $f \in C^0(\mathbb{R})$ e $g \in C^1(\mathbb{R})$, e quindi si avrà una ed una sola soluzione per ogni $x_0 \in \mathbb{R}$ ed $y_0 \in \mathbb{R}$.

L'equazione ammette soluzioni costanti che possono essere trovate ponendo $y(x) = c$ e sostituendo; avremo

$$0 = e^{-c^4} - 1$$

per cui la sola soluzione costante è $y(x) = c = 0$.

Nel caso in cui $y_0 = 0$ la soluzione costante è anche l'unica soluzione del problema di Cauchy.

Se fissiamo $x_0 = 0$ ed $y_0 = 1$, possiamo supporre $y(x) \neq 0$ in un intorno di 0 e separando le variabili ed integrando tra 0 ed x si ottiene

$$(21.37) \quad \frac{y'(x)}{e^{-(y(x))^4} - 1} = 1$$

$$(21.38) \quad \int_0^x \frac{y'(t)}{e^{-(y(t))^4} - 1} dt = \int_0^x dt$$

ovvero

$$\int_1^{y(x)} \frac{ds}{e^{-s^4} - 1} = x$$

Studiamo ora la funzione integrale a primo membro $h(y) = \int_1^y \frac{ds}{e^{-s^4} - 1}$.

Poiché l'integranda è definita e continua per $s \neq 0$ e

$$\lim_{s \rightarrow 0} \frac{1}{e^{-s^4} - 1} = -\infty$$

di ordine 4, l'integrale è divergente per $y < 0$; ne segue che, essendo il primo estremo di integrazione positivo, la funzione è definita per $y > 0$.

Inoltre

$$\lim_{s \rightarrow +\infty} \frac{1}{e^{-s^4} - 1} = -1$$

da cui l'integrale è divergente anche per $y \rightarrow +\infty$.

Si ha infine $h(1) = 0$ e $h'(y) = \frac{1}{e^{-y^4} - 1}$ essendo l'integranda continua per $y > 0$, e tale derivata risulta sempre negativa.

Possiamo anche osservare che

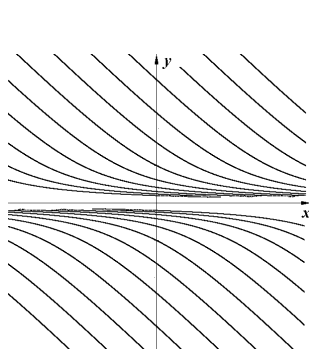
$$h''(y) = \frac{4y^3 e^{-y^4}}{(e^{-y^4} - 1)^2} > 0$$

per ogni $y > 0$, per cui la funzione risulterà convessa; inoltre, poiché

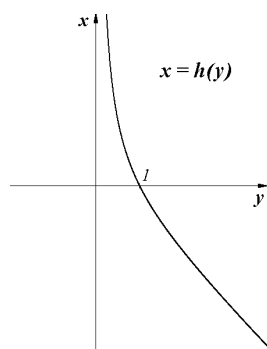
$$\lim_{y \rightarrow +\infty} h'(y) = -1$$

il grafico della funzione tenderà a diventare parallelo alla bisettrice del secondo e quarto quadrante)

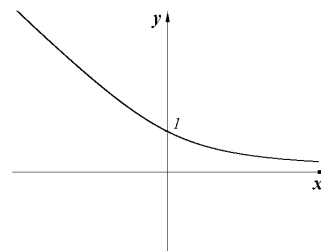
Il grafico della funzione h è indicato nella figura;



21.6.1. Grafico 1



21.6.2. Grafico2



21.6.3. Grafico3

FIGURA 21.6.

Poichè deve aversi

$$h(y(x)) = x$$

il grafico della soluzione del problema di Cauchy sarà quello dell'inversa di h , come riportato nella figura 21.6.6.

Per disegnare il grafico delle soluzioni del problema di Cauchy dato al variare dei dati iniziali $x_0, y_0 \in \mathbb{R}$, possiamo osservare che l'equazione data è un'equazione differenziale autonoma, e quindi se $y(x)$ è soluzione, anche $y(x + a)$ è soluzione per ogni $a \in \mathbb{R}$.

Pertanto tutte le traslate (in orizzontale) della soluzione trovata sono ancora soluzioni, per $y > 0$.

Per quanto riguarda le soluzioni per $y < 0$, ripetendo i calcoli fatti, ad esempio con $x_0 = 0$ e $y_0 = -1$, si ha

$$\int_{-1}^{y(x)} \frac{ds}{e^{-s^4} - 1} = x$$

e con considerazioni analoghe si ottengono le curve indicate in figura 6.6

(Si noti che, se $y(x)$ è soluzione dell'equazione differenziale, tale è pure $-y(-x)$, ovvero i grafici delle soluzioni sono simmetrici rispetto all'origine).

7. Esempio

Si consideri il problema di Cauchy

$$\begin{cases} y'(x) = 6x^2 \sqrt{y(x)} \\ y(x_0) = 1 \end{cases}$$

Si tratta di un problema a variabili separabili con $f(x) = 6x^2$ definita e continua su tutto \mathbb{R} , e $g(y) = \sqrt{y}$ definita e di classe C^1 per $y > 0$; pertanto essendo $y_0 = 1$, per il teorema di esistenza ed unicità, esiste una ed una sola soluzione del problema dato, per ogni $x_0 \in \mathbb{R}$.

Separando le variabili, per $y(x) > 0$, si ottiene

$$\frac{y'(x)}{\sqrt{y(x)}} = 6x^2$$

ed integrando tra 0 ed x

$$\int_0^x \frac{y'(t)}{\sqrt{y(t)}} dt = \int_0^x 6t^2 dt$$

ovvero

$$2\sqrt{y(x)} - 2\sqrt{y(0)} = 2x^3 \quad \text{da cui} \quad \sqrt{y(x)} = 1 + x^3$$

Elevando al quadrato i due membri, dopo aver osservato che $1 + x^3 > 0$ e cioè $x > -1$, si ottiene

$$y(x) = (1 + x^3)^2, \quad x > -1$$

(si noti che la soluzione è prolungabile, in modo unico, con $y(x) = 0$ per $x \leq -1$).

Il grafico delle soluzioni è riportato in figura 21.7

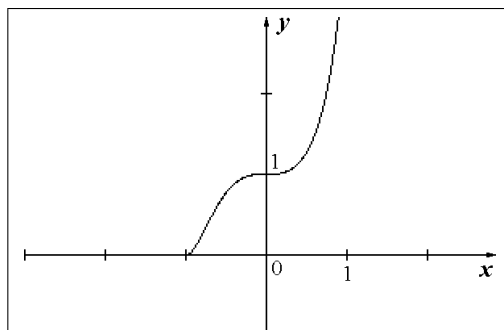


FIGURA 21.7.

SISTEMI ED EQUAZIONI DIFFERENZIALI LINEARI

Un altro tipo importante di equazioni differenziali è costituito dalle equazioni lineari. La più semplice equazione lineare può essere scritta nella forma

$$(22.1) \quad y'(x) = a(x)y(x) + b(x)$$

Se $a, b \in C^0(I)$, l'equazione 22.1 ammette una ed una sola soluzione definita su tutto I ; questa è forse una delle più importanti caratteristiche di questo tipo di equazioni e si può facilmente verificare, in questo caso, direttamente.

Sia $x_0 \in I$, ed $y_0 \in \mathbb{R}$, e sia A una primitiva di a in I . L'esistenza di A è assicurata dalla continuità di a ; ad esempio possiamo porre $A(x) = \int_{x_0}^x a(t)dt$.

La 22.1 è vera se e solo se

$$e^{-A(x)}y'(x) - e^{-A(x)}a(x)y(x) = b(x)e^{-A(x)}$$

e ciò è equivalente a

$$\frac{d}{dx} (e^{-A(x)}y(x)) = b(x)e^{-A(x)}.$$

Integrando tra x_0 ed x , si ottiene

$$e^{-A(x)}y(x) = y_0 + \int_{x_0}^x b(t)e^{-A(t)}dt$$

ed infine

$$(22.2) \quad y(x) = e^{A(x)} \left(y_0 + \int_{x_0}^x b(t)e^{-A(t)} dt \right)$$

Quanto abbiamo esposto consente di affermare che tutte le soluzioni dell'equazione 22.1 si ottengono, al variare di $y_0 \in \mathbb{R}$, dalla 22.2.

Osserviamo anche che la 22.2 stessa può essere riscritta nella seguente maniera:

$$y(x) = y_0 e^{\int_{x_0}^x a(t)dt} + e^{\int_{x_0}^x a(t)dt} \int_{x_0}^x b(t) e^{-\int_{x_0}^t a(s)ds} dt$$

in accordo con i risultati che proveremo nel seguito per il caso più generale.

La 22.2 costituisce, al variare di y_0 , l'integrale generale dell'equazione 22.1.

I passi successivi consistono nel considerare equazioni lineari di ordine superiore oppure sistemi di equazioni del primo ordine.

Un'equazione lineare di ordine n si può scrivere nella forma

$$(22.3) \quad y^{(n)}(x) = \sum_{i=1}^n a_i y^{(i-1)}(x) + b(x)$$

dove $a_i, b \in \mathcal{C}^0$ mentre un sistema lineare di ordine n si scrive nella forma

$$(22.4) \quad Y'(x) = A(x)Y(x) + B(x)$$

dove $A(x) = \{a_{ij}(x)\}$ e $B(x) = \{b_i(x)\}$ sono una matrice ed un vettore i cui elementi sono funzioni continue su un intervallo I ; (scriviamo $A \in \mathcal{C}^k(I)$, $B \in \mathcal{C}^k(I)$ quando intendiamo pertanto affermare che $a_{ij} \in \mathcal{C}^k(I)$, $b_i \in \mathcal{C}^k(I)$ per $i, j = 1, \dots, n$).

Il sistema può essere riscritto usando le componenti di Y , A , B , nella seguente maniera

$$(22.5) \quad \begin{pmatrix} y_1'(x) \\ y_2'(x) \\ \vdots \\ y_n'(x) \end{pmatrix} = \begin{pmatrix} a_{11}(x) & a_{12}(x) & \dots & a_{1n}(x) \\ a_{21}(x) & a_{22}(x) & \dots & a_{2n}(x) \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}(x) & a_{n2}(x) & \dots & a_{nn}(x) \end{pmatrix} \begin{pmatrix} y_1(x) \\ y_2(x) \\ \vdots \\ y_n(x) \end{pmatrix} + \begin{pmatrix} b_1(x) \\ b_2(x) \\ \vdots \\ b_n(x) \end{pmatrix}$$

ed anche, in forma più compatta

$$(22.6) \quad y_i'(x) = \sum_{j=1}^n a_{ij}(x)y_j(x) + b_i(x) \quad , \quad i = 1, \dots, n$$

Qualora $B \equiv 0$ il sistema si dice omogeneo e assume la forma

$$(22.7) \quad Y'(x) = A(x)Y(x)$$

Quando $n = 1$ il sistema si riduce ad una sola equazione differenziale lineare del primo ordine che, posto $A = (a_{11}) = a$ e $B = b_1 = b$, si scrive nella forma

$$y'(x) = a(x)y(x) + b(x)$$

L'insieme \mathcal{T} di tutte le soluzioni di 22.4 si chiama integrale generale del sistema .

Quando si associa al sistema o all'equazione differenziale un opportuno insieme di condizioni iniziali parliamo di problema di Cauchy

$$(22.8) \quad \begin{cases} Y'(x) = A(x)Y(x) + B(x) & , \quad \forall x \in I \\ Y(x_0) = Y_0 \end{cases}$$

$$(22.9) \quad \begin{cases} y^{(n)}(x) = a_n(x)y^{(n-1)}(x) + \dots + a_1(x)y(x) + b(x) & , \quad \forall x \in I \\ y(x_0) = y_0, y'(x_0) = y_1, \dots, y^{(n-1)}(x_0) = y_{n-1} \end{cases}$$

sono problemi di Cauchy.

Lo studio di un sistema consente di trovare risultati anche per l'equazione di ordine n ; sia infatti

$$(22.10) \quad y^{(n)}(x) = a_n(x)y^{(n-1)}(x) + \dots + a_1(x)y(x) + b(x)$$

una equazione differenziale lineare di ordine n e poniamo

$$(22.11) \quad y_i(x) = y^{(i-1)}(x) \quad , \quad i = 1, \dots, n.$$

(Per chiarire le idee osserviamo che si avrà $y_1(x) = y(x)$, ..., $y_n(x) = y^{(n-1)}(x)$).

Possiamo riscrivere l'equazione nella seguente forma

$$(22.12) \quad \begin{cases} y_1'(x) = y_2(x) \\ y_2'(x) = y_3(x) \\ \dots \\ \dots \\ y_n'(x) = a_n(x)y_n(x) + \dots + a_1(x)y_1(x) + b(x) \end{cases}$$

ed anche come

$$Y'(x) = A(x)Y(x) + B(x)$$

non appena si sia definito

$$A(x) = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_1(x) & a_2(x) & a_3(x) & \dots & a_n(x) \end{pmatrix} \quad B(x) = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ b(x) \end{pmatrix}$$

Vale il seguente teorema di cui è importante in questo contesto solo l'enunciato.

TEOREMA 22.1. *Siano $A: I \longrightarrow \mathcal{M}^n$, $B: I \longrightarrow \mathbb{R}^n$ continue e siano $x_0 \in I$, $Y_0 \in \mathbb{R}^n$.*

Allora esiste una ed una sola soluzione del problema di Cauchy

$$(22.13) \quad \begin{cases} Y'(x) = A(x)Y(x) + B(x) & , \quad \forall x \in I \\ Y(x_0) = Y_0 \end{cases}$$

Il teorema precedente consente di provare un risultato di esistenza anche per le equazioni differenziali lineari di ordine n .

TEOREMA 22.2. *Siano $a_i, b \in C^0(I)$, $i = 1, \dots, n$ e siano $x_0 \in I$, $y_i \in \mathbb{R}$, $i = 0, \dots, n-1$. Allora esiste una ed una sola soluzione $y: I \longrightarrow \mathbb{R}$ del problema di Cauchy*

$$(22.14) \quad \begin{cases} y^{(n)}(x) = \sum_{i=1}^n a_i(x)y^{(i-1)}(x) + b(x) \\ y^{(i)}(x_0) = y_i & , \quad i = 0, \dots, n-1 \end{cases}$$

Proviamo ora che l'insieme delle soluzioni di un sistema differenziale lineare, cioè l'integrale generale di un sistema differenziale omogeneo del primo ordine è uno spazio vettoriale avente dimensione uguale al numero di equazioni del sistema stesso.

TEOREMA 22.3. *Sia $A \in C^0(I)$ e consideriamo il sistema differenziale lineare del primo ordine*

$$Y'(x) = A(x)Y(x);$$

sia \mathcal{S} il suo integrale generale. Allora \mathcal{S} è uno spazio vettoriale di dimensione n .

DIMOSTRAZIONE. E' immediato verificare che \mathcal{S} è uno spazio vettoriale in quanto si vede subito che se y e z sono soluzioni del sistema assegnato tali risultano anche $\alpha y + \beta z$ ove α, β sono scalari.

Per provare che $\dim \mathcal{S} = n$ è sufficiente osservare che, per il teorema di esistenza ed unicità della soluzione l'applicazione lineare

$$\Gamma : \mathcal{S} \longrightarrow \mathbb{R}^n$$

definita da

$$\Gamma(Y) = Y(x_0) \quad , \quad x_0 \in I$$

è un isomorfismo. \square

In base al teorema precedente è possibile affermare che ogni soluzione di un sistema differenziale lineare omogeneo di n equazioni in n incognite può essere espressa mediante un combinazione lineare di n soluzioni linearmente indipendenti del sistema stesso.

Siano esse Y_1, \dots, Y_n e sia $(y_i)_j$ la componente j -esima della i -esima soluzione.

Possiamo allora costruire la matrice

$$(22.15) \quad G = \begin{pmatrix} (y_1)_1 & (y_2)_1 & \cdots & (y_n)_1 \\ (y_1)_2 & (y_2)_2 & \cdots & (y_n)_2 \\ \vdots & \vdots & \ddots & \vdots \\ (y_1)_n & (y_2)_n & \cdots & (y_n)_n \end{pmatrix}$$

che indicheremo spesso come

$$G = (Y_1, Y_2, \dots, Y_n)$$

considerando gli Y_i come vettori colonna, e che si chiama matrice fondamentale del sistema assegnato.

È possibile verificare che se G è una matrice fondamentale del sistema omogeneo 22.7 allora si ha

$$(22.16) \quad G'(x) = A(x)G(x)$$

Il sistema 22.16 è un sistema differenziale lineare di n^2 equazioni in n^2 incognite.

Ogni soluzione del nostro sistema potrà allora essere scritta nella forma

$$Y(x) = G(x)C \quad , \quad C \in \mathbb{R}^n$$

ovvero, considerando le componenti,

$$y_i(x) = \sum_{j=1}^n (y_j)_i c_j.$$

Anche lo spazio delle soluzioni di un sistema differenziale lineare ordinario del primo ordine non omogeneo è strutturato in maniera molto precisa.

TEOREMA 22.4. *Siano $A \in \mathcal{C}^0(I)$ $B \in \mathcal{C}^0(I)$ e consideriamo il sistema differenziale lineare non omogeneo del primo ordine*

$$Y'(x) = A(x)Y(x) + B(x)$$

Sia \mathcal{T} l'integrale generale del sistema assegnato e sia \mathcal{S} l'integrale generale del sistema omogeneo ad esso associato

$$Y'(x) = A(x)Y(x)$$

sia ancora $z \in \mathcal{C}^0(I)$ tale che

$$Z'(x) = A(x)Z(x) + B(x)$$

Allora

$$\mathcal{T} = \mathcal{Z} + \mathcal{S}$$

e \mathcal{T} è uno spazio lineare affine di dimensione n .

DIMOSTRAZIONE. E' evidente che $\mathcal{T} \supset \mathcal{Z} + \mathcal{S}$; sia viceversa $Y \in \mathcal{T}$, è facile verificare che $Y - Z$ soddisfa il sistema omogeneo associato e pertanto $Y - Z \in \mathcal{S}$ da cui $Y \in \mathcal{Z} + \mathcal{S}$. \square

DEFINIZIONE 22.1. *Siano Y_1, Y_2, \dots, Y_n n soluzioni del sistema differenziale lineare omogeneo*

$$Y'(x) = A(x)Y(x)$$

Chiamiamo determinante wronskiano, o più semplicemente wronskiano, associato alle n soluzioni assegnate il determinante della matrice

$$(Y_1, Y_2, \dots, Y_n)$$

In altri termini

$$(22.17) \quad W(x) = \det \begin{pmatrix} (y_1(x))_1 & (y_2(x))_1 & \dots & (y_n(x))_1 \\ (y_1(x))_2 & (y_2(x))_2 & \dots & (y_n(x))_2 \\ \vdots & \vdots & \ddots & \vdots \\ (y_1(x))_n & (y_2(x))_n & \dots & (y_n(x))_n \end{pmatrix}$$

Proviamo ora una interessante proprietà del wronskiano.

TEOREMA 22.5. *Siano verificate le ipotesi del teorema di esistenza ed unicità per il sistema differenziale lineare omogeneo*

$$Y'(x) = A(x)Y(x)$$

e siano Y_1, Y_2, \dots, Y_n n soluzioni del sistema stesso.

Sono fatti equivalenti:

- (1) Y_1, \dots, Y_n sono linearmente indipendenti;
- (2) $W(x) \neq 0$ per ogni $x \in I$
- (3) esiste $x_0 \in I$ tale che $W(x_0) \neq 0$.

DIMOSTRAZIONE. Consideriamo, per ogni x fissato in I l'applicazione lineare

$$\Gamma_x : \mathcal{S} \longrightarrow \mathbb{R}^n$$

definita da $\Gamma_x(Y) = Y(x)$. Per il teorema di esistenza ed unicità Γ_x è un isomorfismo.

- (1) \Rightarrow (2)

Se Y_1, \dots, Y_n sono linearmente indipendenti in \mathcal{S} , allora

$$\Gamma_x(Y_1), \dots, \Gamma_x(Y_n)$$

sono linearmente indipendenti in \mathbb{R}^n e perciò

$$0 \neq \det(\Gamma_x(Y_1), \dots, \Gamma_x(Y_n)) = \det(Y_1(x), \dots, Y_n(x)) = W(x)$$

per ogni $x \in I$

- (2) \Rightarrow (3)

È ovvio.

- (3) \Rightarrow (1)

$W(x_0) \neq 0$ implica che $Y_1(x_0), \dots, Y_n(x_0)$ sono linearmente indipendenti in \mathbb{R}^n e perciò

$$Y_1 = \Gamma_{x_0}^{-1}(Y_1(x_0)), \dots, Y_n = \Gamma_{x_0}^{-1}(Y_n(x_0))$$

sono linearmente indipendenti in \mathcal{S}

□

Per il teorema precedente è essenziale che Y_1, \dots, Y_n siano soluzioni del sistema; se ciò non fosse, sarebbe vero solo che (2) \Rightarrow (3) \Rightarrow (1)

Che le altre implicazioni siano false è facilmente visto se si considera il wronskiano associato alle funzioni $Y_{1,2} : \mathbb{R} \longrightarrow \mathbb{R}^2$ definite da

$$Y_1(x) = (x^2, 2x) \quad , \quad Y_2(x) = (2x, 2)$$

oppure

$$Y_1(x) = \begin{cases} (x^2, 2x) & x \geq 0 \\ 0 & x < 0 \end{cases} \quad , \quad Y_2(x) = \begin{cases} (x^2, 2x) & x \leq 0 \\ 0 & x > 0 \end{cases}$$

Altrettanti risultati possono essere ottenuti per le equazioni di ordine n .

TEOREMA 22.6. Siano $a_i, b \in \mathcal{C}^0(I)$, $i = 1, \dots, n$, e consideriamo l'equazione differenziale lineare di ordine n

$$y^{(n)}(x) = \sum_{i=1}^n a_i(x) y^{(i-1)}(x)$$

Sia \mathcal{S} il suo integrale generale, allora \mathcal{S} è uno spazio vettoriale di dimensione n .

Sia

$$y^{(n)}(x) = \sum_{i=1}^n a_i y^{(i-1)}(x) + b(x)$$

la corrispondente equazione differenziale lineare di ordine n non omogenea, e sia \mathcal{T} il suo integrale generale.

\mathcal{T} è uno spazio lineare affine di dimensione n ed inoltre

$$\mathcal{T} = z + \mathcal{S}$$

dove z è una soluzione della equazione non omogenea.

Il teorema precedente consente di affermare che ogni soluzione dell'equazione differenziale lineare omogenea di ordine n si può esprimere come combinazione lineare di n soluzioni y_1, \dots, y_n dell'equazione stessa che siano linearmente indipendenti.

L'insieme y_1, \dots, y_n si chiama sistema fondamentale di soluzioni per l'equazione data; in altre parole ogni soluzione y può essere espressa mediante la

$$y(x) = \sum_{i=1}^n c_i y_i(x)$$

dove $c_i \in \mathbb{R}$

DEFINIZIONE 22.2. Siano y_1, \dots, y_n n soluzioni dell'equazione differenziale lineare di ordine n , omogenea

$$y^{(n)}(x) = \sum_{i=1}^n a_i(x) y^{(i-1)}(x)$$

Chiamiamo wronskiano associato alle soluzioni y_1, \dots, y_n il determinante

$$(22.18) \quad W(x) = \det \begin{pmatrix} y_1(x) & y_2(x) & \dots & y_n(x) \\ y_1'(x) & y_2'(x) & \dots & y_n'(x) \\ \vdots & \vdots & \ddots & \vdots \\ y_1^{(n-1)}(x) & y_2^{(n-1)}(x) & \dots & y_n^{(n-1)}(x) \end{pmatrix}$$

TEOREMA 22.7. Siano verificate le ipotesi del teorema di esistenza ed unicità e siano y_1, \dots, y_n n soluzioni dell'equazione differenziale omogenea di ordine n

$$y^{(n)}(x) = \sum_{i=1}^n a_i(x) y^{(i-1)}(x)$$

Sono fatti equivalenti:

- (1) y_1, \dots, y_n sono linearmente indipendenti;
- (2) $W(x) \neq 0$ per ogni $x \in I$;
- (3) esiste $x_0 \in I$ tale che $W(x_0) \neq 0$.

Come in precedenza, usando lo stesso esempio, si vede che, qualora y_1, \dots, y_n non siano soluzioni dell'equazione, le uniche implicazioni ancora vere sono $(2) \Rightarrow (3) \Rightarrow (1)$

I risultati precedenti assicurano la possibilità di trovare l'integrale generale di un sistema non omogeneo non appena siano noti l'integrale generale del sistema omogeneo ad esso associato ed una soluzione del sistema non omogeneo; è pertanto molto importante avere a disposizione uno strumento che consenta, noto l'integrale generale del sistema omogeneo, di trovare una soluzione del sistema non omogeneo.

Sia G una matrice fondamentale del sistema lineare omogeneo

$$Y'(x) = A(x)Y(x)$$

e $x_0 \in I$. Una soluzione del sistema non omogeneo

$$Y'(x) = A(x)Y(x) + B(x)$$

è data da

$$Z(x) = G(x) \int_{x_0}^x G^{-1}(t)B(t)dt.$$

Infatti se cerchiamo soluzioni del sistema non omogeneo della forma

$$Z(x) = G(x)\lambda(x)$$

dove $\lambda : I \longrightarrow \mathbb{R}^n$ è derivabile, dovrà aversi

$$Z'(x) = A(x)Z(x) + B(x)$$

e pertanto, poiché si può verificare che la regola di derivazione del prodotto può essere estesa anche al prodotto righe per colonne, si ha

$$Z'(x) = G'(x)\lambda(x) + G(x)\lambda'(x)$$

deve essere

$$G'(x)\lambda(x) + G(x)\lambda'(x) = A(x)G(x)\lambda(x) + B(x)$$

Ma G è una matrice fondamentale e quindi,

$$G(x)\lambda'(x) = B(x) \quad e \quad \lambda'(x) = G^{-1}(x)B(x).$$

Se ne deduce che se

$$\lambda(x) = \int_{x_0}^x G^{-1}(t)B(t)dt$$

Z è soluzione del sistema completo.

Osserviamo inoltre che, essendo $G(x)\lambda'(x) = B(x)$, per il teorema di Cramer si ha

$$\lambda'_i(x) = \frac{W_i(x)}{W(x)}$$

essendo

(22.19)

$$W_i = \det \begin{pmatrix} (y_1)_1 & (y_2)_1 & \dots & (y_{i-1})_1 & b_1 & (y_{i+1})_1 & \dots & (y_n)_1 \\ (y_1)_2 & (y_2)_2 & \dots & (y_{i-1})_2 & b_2 & (y_{i+1})_2 & \dots & (y_n)_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ (y_1)_n & (y_2)_n & \dots & (y_{i-1})_n & b_n & (y_{i+1})_n & \dots & (y_n)_n \end{pmatrix}$$

e una soluzione del sistema non omogeneo è data da

$$Y(x) = \sum_{i=1}^n \lambda_i(x) Y_i(x).$$

Come conseguenza se G è una matrice fondamentale del sistema lineare omogeneo

$$Y'(x) = A(x)Y(x)$$

l'integrale generale del sistema lineare non omogeneo

$$Y'(x) = A(x)Y(x) + B(x)$$

è dato da

$$Y(x) = G(x) \left(C + \int_{x_0}^x G^{-1}(t)B(t)dt \right), \quad C \in \mathbb{R}^n$$

Dove $x_0 \in I$ mentre la soluzione del problema di Cauchy relativo ai dati $Y(x_0) = Y_0$ è

$$Y(x) = G(x) \left(G^{-1}(x_0)Y_0 + \int_{x_0}^x G^{-1}(t)B(t)dt \right)$$

Il metodo esposto si chiama della metodo di Lagrange di variazione delle costanti arbitrarie e può ovviamente essere applicato anche alle equazioni differenziali di ordine n non appena le si sia trasformate in un sistema. Tuttavia per le equazioni è più conveniente procedere direttamente; illustriamo qui di seguito, il caso di una equazione del secondo ordine.

Siano $a, b, c \in C^0(I)$ e consideriamo l'equazione lineare del secondo ordine

$$y''(x) = a(x)y'(x) + b(x)y(x) + c(x).$$

Supponiamo note due soluzioni linearmente indipendenti dell'equazione differenziale omogenea associata; avremo allora a disposizione l'integrale generale dell'equazione omogenea nella forma

$$y(x) = c_1 y_1(x) + c_2 y_2(x)$$

Cerchiamo soluzioni per l'equazione non omogenea nella forma

$$z(x) = \lambda_1(x)y_1(x) + \lambda_2(x)y_2(x)$$

Avremo

$$z' = \lambda_1' y_1 + \lambda_2' y_2 + \lambda_1 y_1' + \lambda_2 y_2'$$

e posto

$$\lambda_1' y_1 + \lambda_2' y_2 = 0$$

si ha

$$z'' = \lambda_1' y_1' + \lambda_2' y_2' + \lambda_1 y_1'' + \lambda_2 y_2''.$$

Sostituendo si ottiene

$$\lambda_1' y_1' + \lambda_2' y_2' + \lambda_1 y_1'' + \lambda_2 y_2'' = \lambda_1 a y_1' + \lambda_2 a y_2' + \lambda_1 b y_1 + \lambda_2 b y_2 + c$$

e, tenuto conto che y_1 e y_2 sono soluzioni dell'omogenea,

$$\lambda_1' y_1' + \lambda_2' y_2' = c.$$

Ne viene che λ_1' e λ_2' devono soddisfare il seguente sistema

$$(22.20) \quad \begin{cases} \lambda_1' y_1 + \lambda_2' y_2 = 0 \\ \lambda_1' y_1' + \lambda_2' y_2' = c \end{cases}$$

da cui si possono ricavare λ_1' e λ_2' e per integrazione λ_1 e λ_2 .

Ricordiamo infine, per sommi capi, un metodo che consente di ridurre l'ordine di una equazione differenziale lineare, qualora sia nota una soluzione dell'equazione stessa.

Ci occuperemo qui di mostrare come esso funziona nel caso di una equazione del secondo ordine, essendo l'estensione del metodo del tutto ovvia per equazioni lineari di ordine superiore.

Consideriamo pertanto $a, b \in C^0(I)$ e l'equazione differenziale di ordine 2

$$y''(x) = a(x)y'(x) + b(x)y(x).$$

Supponiamo nota una soluzione z dell'equazione, tale che $z(x) \neq 0 \forall x \in I$.

Cerchiamo soluzioni dell'equazione nella forma $y(x) = u(x)z(x)$

Derivando e sostituendo nell'equazione otteniamo che

$$u''z + 2u'z' + uz'' = au'z + auz' + buz$$

e, tenuto conto che z è soluzione,

$$u''z + 2u'z' - au'z = 0$$

Posto $v = u'$ si ha

$$v'z + v(2z' - az) = 0$$

e quindi, poiché $z \neq 0$,

$$v' + v\left(2\frac{z'}{z} - a\right) = 0.$$

Se ne deduce che deve essere

$$v(x) = e^{-\int_{x_0}^x 2\frac{z'(t)}{z(t)} dt + \int_{x_0}^x a(t) dt}$$

e quindi

$$v(x) = \left(\frac{z(x_0)}{z(x)}\right)^2 e^{\int_{x_0}^x a(t) dt}.$$

Pertanto una soluzione sarà

$$y(x) = \frac{1}{(z(x))^2} e^{\int_{x_0}^x a(t) dt}$$

e

$$u(x) = \int_{x_0}^x \frac{1}{(z(t))^2} e^{\int_{x_0}^t a(s) ds} dt$$

da cui si può ricavare la soluzione cercata.

La soluzione trovata risulta linearmente indipendente da z . Se infatti

$$c_1 z(x) + c_2 z(x) \int_{x_0}^x \frac{1}{(z(t))^2} e^{\int_{x_0}^t a(s) ds} dt = 0$$

per ogni $\forall x$ si ha, per $x = x_0$

$$c_1 z(x_0) = 0 \quad \text{e} \quad c_1 = 0$$

Ne viene anche che

$$c_2 z(x) \int_{x_0}^x \frac{1}{(z(t))^2} e^{\int_{x_0}^t a(s) ds} dt = 0$$

e $c_2 = 0$ in quanto il secondo fattore non può mai annullarsi, se $x \neq x_0$.

Possiamo pertanto scrivere l'integrale generale dell'equazione data come

$$y(x) = z(x) \left(c_1 + c_2 \int_{x_0}^x \frac{1}{(z(t))^2} e^{\int_{x_0}^t a(s) ds} dt \right).$$

Ci occupiamo ora della soluzione di equazioni e sistemi differenziali lineari a coefficienti costanti della forma

$$Y'(x) = AY(x) + B(x)$$

$$Y'(x) = AY(x)$$

$$y^{(n)}(x) = \sum_{k=1}^n a_k y^{(k-1)}(x) + b(x)$$

$$y^{(n)}(x) = \sum_{k=1}^n a_k y^{(k-1)}(x)$$

In pratica l'integrale generale di un'equazione differenziale lineare di ordine n

$$y^{(n)}(x) = \sum_{k=1}^n a_k y^{(k-1)}(x)$$

si può determinare come segue

(1) si considera il polinomio caratteristico associato all'equazione data

$$P(\lambda) = \lambda^n - \sum_{k=1}^n a_k \lambda^{k-1}$$

che si ottiene sostituendo formalmente la quantità algebrica λ^k ad $y^{(k)}(x)$

- (2) si trovano le n soluzioni, reali o complesse e coniugate, dell'equazione (a coefficienti reali)

$$P(\lambda) = 0$$

Consideriamo ogni soluzione $\lambda_1, \dots, \lambda_r$ con la sua molteplicità μ_1, \dots, μ_r

- (3) in corrispondenza ad ogni valore λ , avente molteplicità μ ,

- se λ è reale si considerano le funzioni

$$(22.21) \quad y_1(x) = e^{\lambda x} \quad y_2(x) = x e^{\lambda x} \quad \dots \quad y_\mu(x) = x^{\mu-1} e^{\lambda x}$$

- se $\lambda = \alpha + i\beta$ è complesso, allora anche il suo complesso coniugato $\lambda = \alpha - i\beta$ è autovalore in quanto i coefficienti dell'equazione sono reali, e si considerano le funzioni

$$(22.22)$$

$$u_1(x) = e^{\alpha x} \sin \beta x \quad u_2(x) = x e^{\alpha x} \sin \beta x \quad \dots \quad u_\mu(x) = x^{\mu-1} e^{\alpha x} \sin \beta x$$

$$(22.23)$$

$$v_1(x) = e^{\alpha x} \cos \beta x \quad v_2(x) = x e^{\alpha x} \cos \beta x \quad \dots \quad v_\mu(x) = x^{\mu-1} e^{\alpha x} \cos \beta x$$

Si verifica che le soluzioni trovate sono tra loro linearmente indipendenti.

- (4) Si trovano così

- in corrispondenza di ogni soluzione reale λ , μ soluzioni del sistema linearmente indipendenti
- in corrispondenza di ogni soluzione complessa e della sua coniugata, 2μ soluzioni del sistema linearmente indipendenti

- (5) siano

$$y_1, y_2, y_3, \dots, y_n$$

le soluzioni trovate nei punti precedenti.

Avremo che le soluzioni sono proprio n in quanto la somma del numero delle soluzioni, contate con la loro molteplicità, è proprio n per il teorema fondamentale dell'algebra.

La soluzione dell'equazione sarà pertanto

$$y(x) = \sum_{i=1}^n c_i y_i(x)$$

In pratica l'integrale generale del sistema $Y' = AY$ si può determinare come segue

- (1) si trovano gli autovalori della matrice A , $\lambda_1, \dots, \lambda_r$ e la loro molteplicità μ_1, \dots, μ_r ;
- (2) in corrispondenza ad ogni valore λ di A , avente molteplicità μ ,
 - se λ è reale si considerano le funzioni

$$(22.24) \quad y_1(x) = e^{\lambda x} \quad y_2(x) = x e^{\lambda x} \quad \dots \quad y_\mu(x) = x^{\mu-1} e^{\lambda x}$$

- se λ è complesso, allora anche il suo complesso coniugato è autovalore in quanto i coefficienti del sistema sono reali, e si considerano le funzioni

(22.25)

$$u_1(x) = e^{\alpha x} \sin \beta x \quad u_2(x) = x e^{\alpha x} \sin \beta x \quad \cdots \quad u_\mu(x) = x^{\mu-1} e^{\alpha x} \sin \beta x$$

(22.26)

$$v_1(x) = e^{\alpha x} \cos \beta x \quad v_2(x) = x e^{\alpha x} \cos \beta x \quad \cdots \quad v_\mu(x) = x^{\mu-1} e^{\alpha x} \cos \beta x$$

Si verifica che le soluzioni trovate sono tra loro linearmente indipendenti.

(3) Si trovano così

- in corrispondenza di ogni autovalore reale λ , μ soluzioni del sistema linearmente indipendenti
- in corrispondenza di ogni autovalore complesso e del suo coniugato, 2μ soluzioni del sistema linearmente indipendenti

(4) siano

$$y_1, y_2, y_3, \dots, y_n$$

le soluzioni trovate nei punti precedenti.

Avremo che le soluzioni sono proprio n in quanto la somma del numero delle soluzioni, contate con la loro molteplicità, è proprio n per il teorema fondamentale dell'algebra e possiamo cercare soluzioni

$$Y = (Y_j)$$

del sistema omogeneo che abbiano come componenti delle combinazioni lineari delle funzioni y_i cioè

$$Y_j(x) = \sum_{i=1}^n c_{i,j} y_i(x)$$

- (5) Le costanti introdotte $c_{i,j}$ sono in numero di n^2 e quindi superiore al numero di costanti n necessario e sufficiente per descrivere l'integrale generale del sistema differenziale lineare omogeneo di ordine n ; onde determinare solo n costanti si procede quindi sostituendo nel sistema ed usando le uguaglianze trovate per ridurre il numero di costanti libere ad n

Abbiamo con ciò gli strumenti per risolvere ogni equazione differenziale ed ogni sistema differenziale lineare omogeneo, a coefficienti costanti; per risolvere i corrispondenti problemi non omogenei sarà sufficiente trovare una soluzione particolare dei problemi non omogenei stessi. Ciò può essere fatto, in generale, usando il metodo di variazione delle costanti di Lagrange, ma, nel caso dei coefficienti costanti, possiamo, se inoltre il termine noto è di forma particolarmente semplice, trovare una soluzione particolare di forma similmente semplice.

Più precisamente possiamo affermare che:

- (1) Se consideriamo l'equazione differenziale non omogenea 22.3 e se

$$b(x) = q(x)e^{\lambda x}$$

dove $\lambda \in \mathbb{C}$ e q è un polinomio di grado m a coefficienti complessi, si può trovare un polinomio r di grado al più m tale che, se μ è la molteplicità di λ come radice del polinomio caratteristico P ,

$$y(x) = x^\mu r(x)e^{\lambda x}$$

sia soluzione dell'equazione 22.3.

- (2) Se consideriamo il sistema differenziale non omogeneo 22.4 e se

$$B(x) = Q(x)e^{\lambda x}$$

dove Q è un vettore colonna i cui elementi sono polinomi a coefficienti complessi, di grado minore o uguale ad m , si può trovare un vettore colonna R i cui elementi sono polinomi a coefficienti complessi di grado al più $m + \mu$, dove μ è la molteplicità di λ come radice del polinomio caratteristico P della matrice A , tale che

$$Y(x) = R(x)e^{\lambda x}$$

risolve il sistema 22.4.

Si può inoltre provare che, nel caso in cui i coefficienti siano reali,

- (1) Se

$$b(x) = e^{\alpha x} [q_1(x) \cos(\beta x) + q_2(x) \sin(\beta x)]$$

dove q_1 e q_2 sono polinomi a coefficienti reali di grado massimo m e $\alpha \pm i\beta$ è radice del polinomio caratteristico P di molteplicità μ , si possono trovare due polinomi r_1, r_2 di grado al più m tali che

$$y(x) = x^\mu e^{\alpha x} [r_1(x) \cos(\beta x) + r_2(x) \sin(\beta x)]$$

sia soluzione della 22.3.

- (2) Se

$$B(x) = e^{\alpha x} [Q_1(x) \cos(\beta x) + Q_2(x) \sin(\beta x)]$$

dove Q_1 e Q_2 sono vettori colonna i cui elementi sono polinomi a coefficienti reali di grado al più m e $\alpha \pm i\beta$ è radice del polinomio caratteristico della matrice A con molteplicità μ , si possono trovare R_1 ed R_2 , vettori colonna i cui elementi sono polinomi a coefficienti reali di grado al più $m + \mu$, tali che

$$Y(x) = e^{\alpha x} [R_1(x) \cos(\beta x) + R_2(x) \sin(\beta x)]$$

sia soluzione del sistema 22.4.

1. L'oscillatore armonico

Un esempio molto importante di modello matematico che utilizza la teoria delle equazioni differenziali lineari è costituito dall'oscillatore armonico.

Si consideri l'equazione del secondo ordine

$$(22.27) \quad x''(t) + 2hx'(t) + \omega^2 x(t) = K \sin(\alpha t)$$

dove $h, K, \alpha > 0$.

Essa può descrivere il comportamento di diversi sistemi reali quali,

- (1) un punto materiale soggetto ad una forza di richiamo proporzionale alla distanza ed ad una forza di attrito proporzionale alla velocità, sollecitato da una forza esterna sinusoidale di ampiezza K e di frequenza α .
- (2) l'intensità di corrente che circola in un circuito RLC alimentato da una forza elettromotrice sinusoidale.

Le soluzioni dell'equazione sono date da:

- (1) Se $h > \omega$

$$x(t) = c_1 e^{(-h+\theta)t} + c_2 e^{(-h-\theta)t} + \hat{x}(t)$$

I grafici di possibili soluzioni sono riportati nelle figure

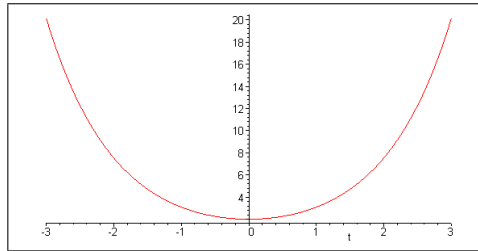


FIGURA 22.1. Soluzioni del polinomio caratteristico reali distinte una positiva ed una negativa

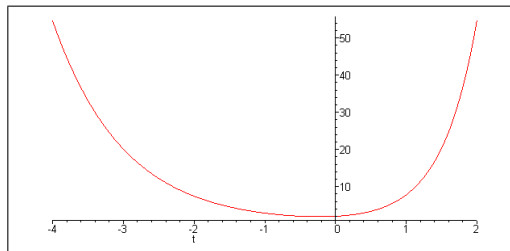


FIGURA 22.2. Soluzioni del polinomio caratteristico reali distinte una positiva ed una negativa

- (2) Se $h = \omega$

$$x(t) = c_1 e^{-ht} + c_2 t e^{-ht} + \hat{x}(t)$$

I grafici di possibili soluzioni sono riportati nelle figure

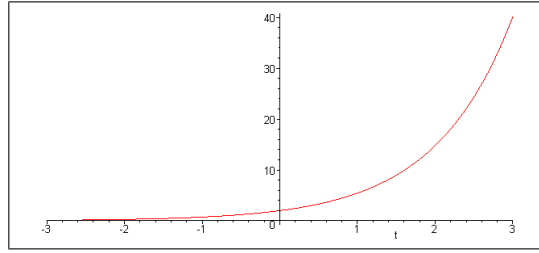


FIGURA 22.3. Soluzioni del polinomio caratteristico reali distinte entrambe positive

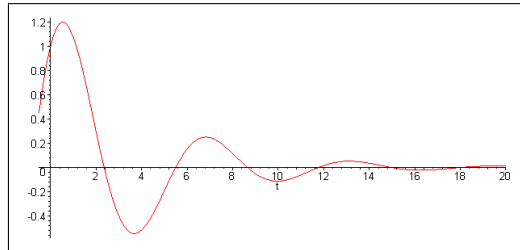


FIGURA 22.4. Soluzioni del polinomio caratteristico complesse e coniugate con parte reale negativa

(3) Se $h < \omega$

$$x(t) = e^{-ht}(c_1 \sin(\theta t) + c_2 \cos(\theta t)) + \hat{x}(t)$$

I grafici di possibili soluzioni sono riportati nelle figure

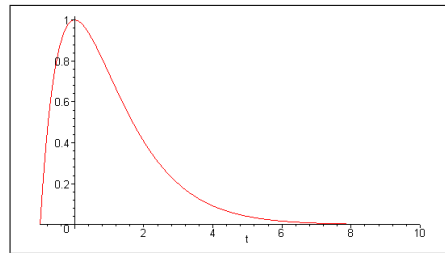


FIGURA 22.5. Soluzioni del polinomio caratteristico reali coincidenti negative

dove

$$\hat{x}(t) = \alpha \sin(\alpha t) + b \cos(\alpha t) = A \sin(\alpha t - \phi)$$

ed inoltre si è posto

$$\begin{aligned} \theta &= |h^2 - \omega^2|^{1/2} \\ a &= K \frac{\omega^2 - \alpha^2}{4h^2\alpha^2 + (\omega^2 - \alpha^2)^2} \\ b &= -K \frac{2h\alpha}{4h^2\alpha^2 + (\omega^2 - \alpha^2)^2} \end{aligned}$$

$$A = \frac{K}{\sqrt{4h^2\alpha^2 + (\omega^2 - \alpha^2)^2}}$$

$$\phi = \arccos\left(\frac{\alpha}{\omega}\right)$$

Nel caso in cui $h = 0$ l'equazione diventa

$$x''(t) + \omega^2 x(t) = K \sin(\alpha t)$$

con $k, \alpha > 0$ e rappresenta un oscillatore armonico non smorzato sollecitato da una forza esterna sinusoidale.

Le soluzioni in questo caso sono

(1) Se $\alpha \neq \omega$

$$x(t) = c_1 \sin(\omega t) + c_2 \cos(\omega t) + \frac{K}{\omega^2 - \alpha^2} \sin(\alpha t)$$

(2) Se $\alpha = \omega$

$$x(t) = c_1 \sin(\omega t) + c_2 \cos(\omega t) - \frac{K}{2\omega} t \cos(\omega t)$$

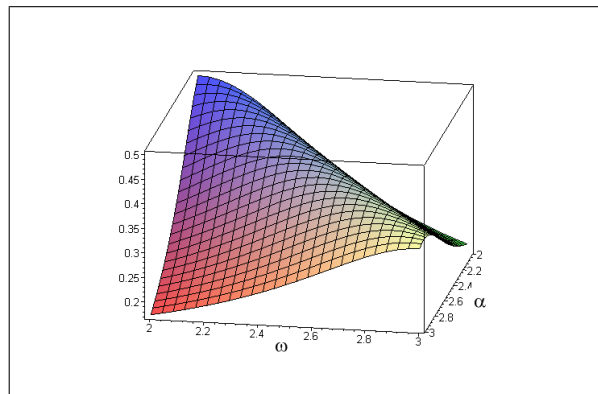


FIGURA 22.6. Grafico di A in funzione di α ed ω

$$(22.28) \quad A = \frac{K}{(4h^2\alpha^2 + (\omega^2 - \alpha^2)^2)^{1/2}} =$$

$$= \frac{K/\omega^2}{(4(h/\omega)^2(\alpha/\omega)^2 + (1 - (\alpha/\omega)^2)^2)^{1/2}}$$

con

$$K/\omega^2 = .5$$

FUNZIONI DI DUE VARIABILI

I modelli matematici spesso devono tenere conto di molti parametri e per questa ragione non è sufficiente considerare funzioni di una sola variabile reale; spesso anzi il numero di parametri in gioco è molto alto e quindi bisogna ricorrere all'uso di funzioni di molte variabili reali.

Dal punto di vista concettuale non c'è grande differenza tra lo studio di una funzione di 2, 3 o 100 variabili reali, ma la differenza tra lo studio di una funzione di 1 variabile reale ed una funzione di 2 variabili reali è grande e va considerata attentamente.

Sviluppiamo pertanto lo studio di una funzione di 2 variabili reali per introdurre gli strumenti necessari al trattamento delle funzioni di più variabili reali a valori reali.

DEFINIZIONE 23.1. Diciamo che è data una funzione di due variabili reali se sono assegnati un sottoinsieme $D \subset \mathbb{R}^2$ ed una corrispondenza f che ad ogni elemento $P = (x, y) \in D$ associa uno ed un solo elemento $z \in \mathbb{R}$.

Diciamo che D è il dominio della funzione e denotiamo con

$$z = f(x, y) = f(P)$$

il corrispondente di $P = (x, y)$ secondo la legge assegnata f ; scriviamo anche

$$P = (x, y) \mapsto z = f(x, y) = f(P)$$

Chiamiamo rango di f l'insieme

$$R(f) = \{z \in \mathbb{R} : \exists (x, y) \in D, z = f(x, y)\}$$

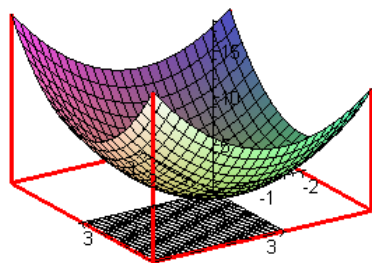
Chiamiamo grafico di f l'insieme

$$G(f) = \{(x, y, z) \in \mathbb{R}^3 : (x, y) \in D, z = f(x, y)\}$$

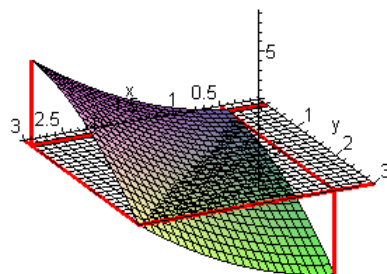
Osservazione. Il grafico di una funzione di 2 variabili è pertanto un sottoinsieme di \mathbb{R}^3 che descrive qualcosa che è immediato identificare come una superficie nello spazio. \square

Restrizione e composizione di funzioni sono definite come nel caso reale e parimenti simile è la definizione di iniettività, surgettività, bigettività.

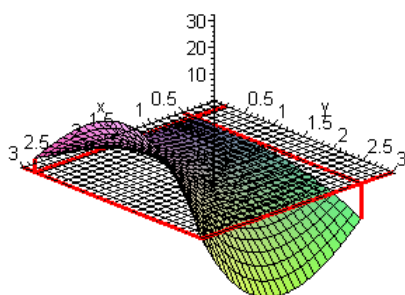
Per avere un'idea del comportamento della funzione sarebbe comodo poter disporre del suo grafico, che nel caso di funzioni di 2 variabili si rappresenta in uno spazio a 3 dimensioni \mathbb{R}^3 ; dobbiamo però tenere presente che:



23.1.1. Grafico 1



23.1.2. Grafico2



23.1.3. Grafico3

FIGURA 23.1. Grafici di funzioni di due variabili

- (1) Non è possibile rappresentare il grafico di funzioni che dipendano da 3 o più variabili
- (2) La rappresentazione in \mathbb{R}^3 di una funzione di due variabili passa attraverso tecniche di prospettiva.
- (3) La proprietà che risulta di maggiore interesse per tracciare il grafico qualitativo di una funzione di 1 variabile è la crescita o la decrescenza, che per le funzioni di 2 o più variabili non può più essere considerata dal momento che il dominio \mathbb{R}^2 (o \mathbb{R}^n) non ammette un ordine completo.

Non sarà pertanto semplice disegnare il grafico qualitativo di una funzione di 2 variabili e per farci un'idea del suo andamento dovremo ricorrere a rappresentazioni nel piano.

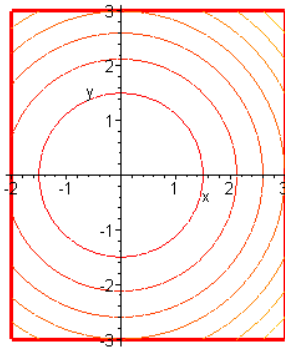
Un modo efficace di rappresentare una superficie è disegnare nel piano (x, y) le curve di livello della funzione.

DEFINIZIONE 23.2. Se $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ chiamiamo *curve od insiemi di livello di f di altezza c gli insiemi*

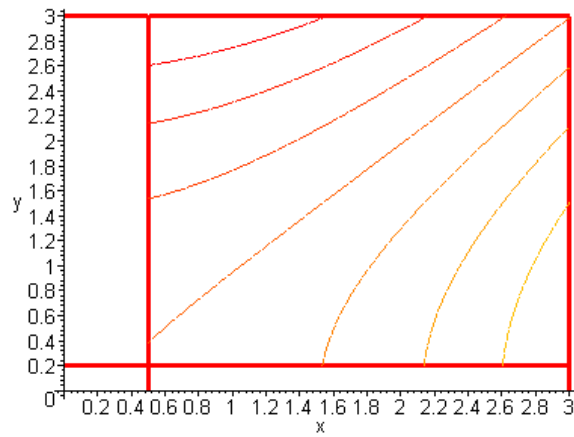
$$L_c = \{(x, y) \in \mathbb{R}^2 : f(x, y) = c\}$$

Le curve di livello di f consentono, in pratica, di rappresentare una mappa della superficie in esame. Esse definiscono i punti in cui la superficie assume quota costante uguale a c e, se le quote c sono scelte ad intervalli regolari, permettono di individuare le zone in cui la superficie è più ripida (le curve di livello sono più ravvicinate).

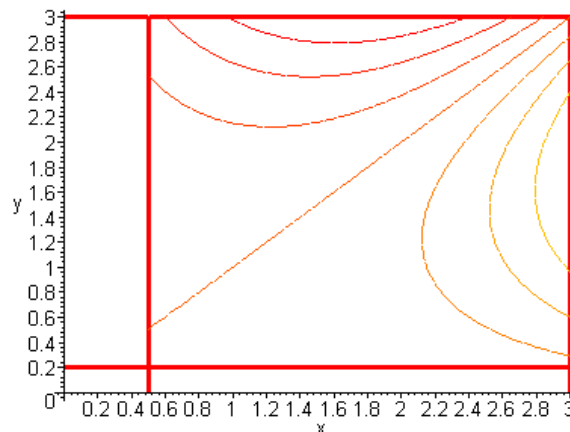
Le superfici prese in considerazione nella figura 23.1 hanno le curve di livello mostrate nella figura 23.2



23.2.1. Grafico 1



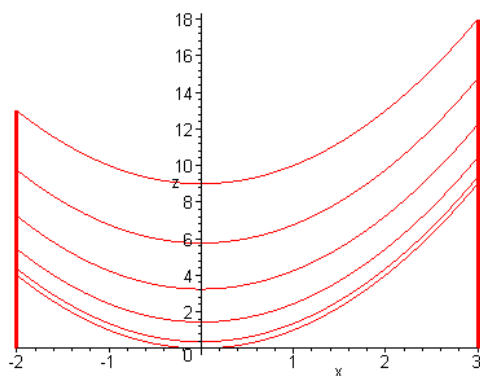
23.2.2. Grafico2



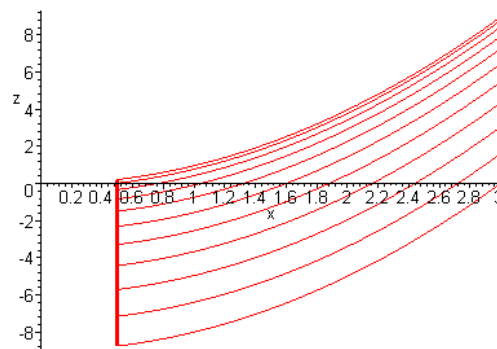
23.2.3. Grafico3

FIGURA 23.2. Curve di livello delle superfici in figura 23.1

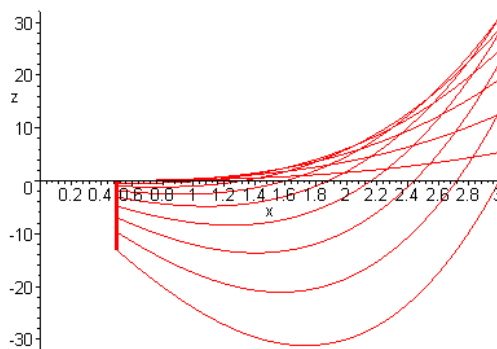
Per farci un'idea del grafico possiamo anche considerare l'andamento delle funzioni di x che si ottengono considerando fissati i valori di y ; chiamiamo questi grafici sezioni lungo l'asse x , si veda figura 23.3, e delle funzioni di y che si ottengono considerando fissati i valori di x ; chiamiamo questi grafici sezioni lungo l'asse y , si veda figura 23.4.



23.3.1. Grafico 1



23.3.2. Grafico2



23.3.3. Grafico3

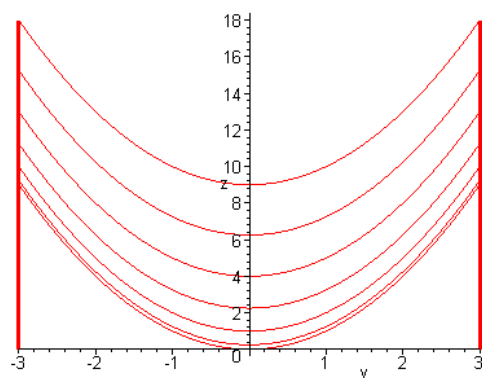
FIGURA 23.3. Sezioni, per y fissato, dei grafici di figura 23.1

Come per le funzioni di una variabile è importante studiare la continuità e la derivabilità di una funzione di 2 o più variabili. Ovviamente per poter considerare la continuità è necessario conoscere la definizione di limite e ancora prima la definizione di intorno e la struttura dello spazio \mathbb{R}^2 in cui stiamo lavorando.

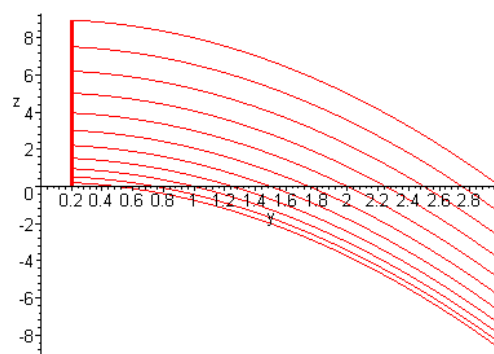
1. La struttura di \mathbb{R}^2 .

Indichiamo con \mathbb{R}^2 lo spazio vettoriale costituito dalla coppie ordinate di numeri reali; in altre parole

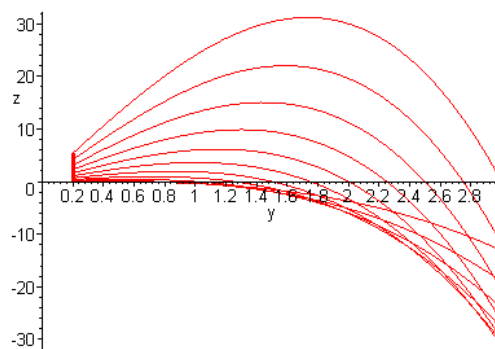
$$P \in \mathbb{R}^2 \Leftrightarrow P = (x, y) \quad x, y \in \mathbb{R}$$



23.4.1. Grafico 1



23.4.2. Grafico2



23.4.3. Grafico3

FIGURA 23.4. Sezioni, per x fissato, dei grafici di figura 23.1

In \mathbb{R}^2 si definiscono le operazioni di somma e di prodotto per uno scalare mediante le

$$P_1 + P_2 = (x_1 + x_2, y_1 + y_2)$$

e, se $\alpha \in \mathbb{R}$,

$$\alpha P = (\alpha x, \alpha y)$$

L'insieme dei vettori

$$e_1 = (1, 0) \quad , \quad e_2 = (0, 1)$$

costituisce una base di \mathbb{R}^2 ; si avrà pertanto che, se $P \in \mathbb{R}^2$,

$$P = xe_1 + ye_2 = x(1, 0) + y(0, 1) = (x, y)$$

DEFINIZIONE 23.3. Si definisce norma in \mathbb{R}^2 una funzione che si indica con

$$\| \cdot \| : \mathbb{R}^2 \rightarrow \mathbb{R}$$

che verifica le seguenti proprietà:

$$\bullet \quad \|P\| \geq 0 \quad \forall P \in \mathbb{R}^2$$

- $\|P\| = 0 \Leftrightarrow P = 0$
- $\|\alpha P\| = |\alpha| \|P\| \quad \forall \alpha \in \mathbb{R}, \forall P \in \mathbb{R}^2$
- $\|P + Q\| \leq \|P\| + \|Q\| \quad \forall P, Q \in \mathbb{R}^2$

Si definisce prodotto scalare in \mathbb{R}^2 una funzione

$$\langle \cdot, \cdot \rangle : \mathbb{R}^2 \times \mathbb{R}^2 \longrightarrow \mathbb{R}$$

tale che

- $\langle P, P \rangle \geq 0 \quad \forall P \in \mathbb{R}^2$
- $\langle P, Q \rangle = \langle Q, P \rangle \quad \forall P, Q \in \mathbb{R}^2$
- $\langle P, P \rangle = 0 \Leftrightarrow P = 0$
- $\langle \alpha P + \beta Q, R \rangle = \alpha \langle P, R \rangle + \beta \langle Q, R \rangle \quad \forall P, Q, R \in \mathbb{R}^2, \forall \alpha, \beta \in \mathbb{R}.$

Un esempio notevole di norma in \mathbb{R}^2 è

$$\|P\| = \sqrt{x^2 + y^2}$$

La norma di P indica la distanza di P dall'origine $O = (0, 0)$; se $P = (x, y), P_0 = (x_0, y_0) \in \mathbb{R}^2$

$$\|P - P_0\|$$

indica la distanza tra i punti P e P_0 .

Un esempio notevole di prodotto scalare in \mathbb{R}^2 è definito da

$$\langle P_1, P_2 \rangle = x_1 x_2 + y_1 y_2$$

Se $\rho > 0$ chiamiamo intorno del punto $P_0 = (x_0, y_0)$, l'insieme

$$S(P_0, \rho) = \{P \in \mathbb{R}^2 : \|P - P_0\| < \rho\}$$

$S(P_0, \rho)$ è la sfera di centro P_0 e raggio ρ .

Definiamo inoltre intorno di ∞ il complementare di ogni sfera centrata nell'origine.

$$S(\infty, \rho) = \{P \in \mathbb{R}^2 : \|P\| > \rho\}$$

Diciamo che due vettori $P, Q \in \mathbb{R}^2$ sono ortogonali se $\langle P, Q \rangle = 0$.

Diciamo che sono paralleli se esiste $\lambda \in \mathbb{R}$ tale che $P = \lambda Q$.

Altri esempi di norme in \mathbb{R}^2 sono i seguenti

$$\|P\|_k = (|x|^k + |y|^k)^{1/k} \quad k \geq 1$$

$$\|P\|_\infty = \max\{|x|, |y|\}$$

Norme euclidea e prodotto scalare sono legati dalla seguente

Disuguaglianza di Schwarz

Per $P, Q \in \mathbb{R}^2$ si ha

$$|\langle P, Q \rangle| \leq \|P\| \|Q\|$$

La disuguaglianza di Schwarz può essere dedotta osservando che, per ogni $t \in \mathbb{R}$

$$0 \leq \|xP + tQ\|^2 = \langle P + tQ, P + tQ \rangle = t^2 \|Q\|^2 + 2t \langle P, Q \rangle + \|P\|^2$$

Ciò implica infatti che

$$\langle P, Q \rangle^2 - \|P\|^2 \|Q\|^2 \leq 0$$

Dalla disuguaglianza di Schwarz possiamo anche ricavare la disuguaglianza triangolare; infatti

$$\|P + Q\|^2 = \|P\|^2 + \|Q\|^2 + 2\langle P, Q \rangle \leq \|P\|^2 + \|Q\|^2 + 2\|P\|\|Q\|.$$

Osserviamo infine che

$$|\langle P, Q \rangle| = \|P\|\|Q\|$$

se e solo se esiste $t \in \mathbb{R}$ tale che $P + tQ = 0$, ovvero P e Q sono paralleli.

Da quanto detto si può dedurre che

$$\|P\| = \sup\{\langle P, Q \rangle : \|Q\| \leq 1\} = \max\{|\langle P, Q \rangle| : \|Q\| \leq 1\}$$

2. Limiti e continuità per le funzioni di 2 variabili.

DEFINIZIONE 23.4. Sia $f : A \rightarrow \mathbb{R}$, $A \subset \mathbb{R}^2$ e sia P_0 un punto tale che ogni intorno di P_0 abbia intersezione non vuota con A (chiamiamo P_0 punto di accumulazione per A); diciamo che

$$\lim_{P \rightarrow P_0} f(P) = \ell$$

se $\forall \varepsilon > 0$ esiste $\delta(\varepsilon) > 0$ tale che per $P \in S(P_0, \delta(\varepsilon)) \cap A$, $P \neq P_0$ si ha

$$f(x) \in I(\ell, \varepsilon)$$

È possibile verificare che

- (1) ogni funzione che ammette limite finito è localmente limitata;
- (2) il limite di una funzione, se esiste, è unico;
- (3) vale il teorema della permanenza del segno;
- (4) il limite di una somma è uguale alla somma dei limiti, se questi esistono finiti;
- (5) il limite del prodotto di due funzioni è uguale al prodotto dei limiti, se questi esistono finiti;
- (6) il limite del reciproco di una funzione è uguale al reciproco del limite della funzione stessa, se non è nullo
- (7) valgono i risultati sul confronto dei limiti, in analogia a quanto già visto per le funzioni di una variabile
- (8) il limite di una funzione può essere caratterizzato per successioni
- (9) il limite di una funzione composta si calcola seguendo quanto fatto per le funzioni di una variabile

DEFINIZIONE 23.5. Diciamo che f è una funzione continua in P_0 se $\forall \varepsilon > 0$ esiste $\delta(\varepsilon) > 0$ tale che se $x \in A$, $\|P - P_0\| < \delta(\varepsilon)$ si ha

$$\|f(P) - f(P_0)\| < \varepsilon$$

Nel caso in cui $P_0 \in A$, sia un punto di accumulazione per A la condizione sopra espressa è equivalente alla

$$\lim_{P \rightarrow P_0} f(P) = f(P_0)$$

Ovviamente f si dice continua in A se è continua in ogni punto di A

Come nel caso delle funzioni reali di una variabile reale si prova che:

- (1) la somma di funzioni continue è continua;
- (2) il prodotto di una funzione a valori vettoriali per una funzione a valori scalari, entrambe continue, è continuo;
- (3) il reciproco di una funzione continua è continuo dove ha senso definirlo;
- (4) il prodotto scalare di due funzioni a valori vettoriali continue, è continuo;
- (5) vale la caratterizzazione della continuità per successioni
- (6) la composta di funzioni continue è una funzione continua.

La conoscenza della continuità delle funzioni elementari e le regole precedentemente enunciate permettono di stabilire in modo semplice la continuità in un gran numero di casi: ad esempio, poichè $(x, y) \mapsto x^2$ e $(x, y) \mapsto y^2$ sono continue possiamo anche affermare che

$$(x, y) \mapsto x^2 + y^2$$

è continua, se poi ricordiamo che l'esponenziale è continua avremo anche che

$$(x, y) \mapsto e^{x^2+y^2}$$

è continua.

Come per le funzioni continue di una variabile si possono provare importanti teoremi, tra i quali ricordiamo i seguenti risultati.

TEOREMA 23.1. - di Weierstraß - *Se f è una funzione continua su un insieme A che sia chiuso (contiene i limiti di ogni successione convergente di suoi punti) e limitato (è contenuto in una sfera) allora f ammette massimo e minimo assoluto su A*

TEOREMA 23.2. - degli zeri - Se f è una funzione continua su un insieme A connesso (cioè, in parole semplici, fatto di un solo pezzo) e se esistono due punti $P_+, P_- \in A$ tali che

$$f(P_+) > 0, \quad f(P_-) < 0$$

allora esiste un punto $P_0 \in A$ tale che

$$f(P_0) = 0$$

Un semplice ragionamento assicura, utilizzando il teorema degli zeri, che se una curva di livello di f

$$L_c = \{(x, y) \in U : f(x, y) = c\} = \{(x, y) \in \mathbb{R}^2 : x \in I, y = \varphi(x)\}$$

divide il piano in due parti connesse allora $f(x, y) > 0$ in una delle due parti e $f(x, y) < 0$ nell'altra.

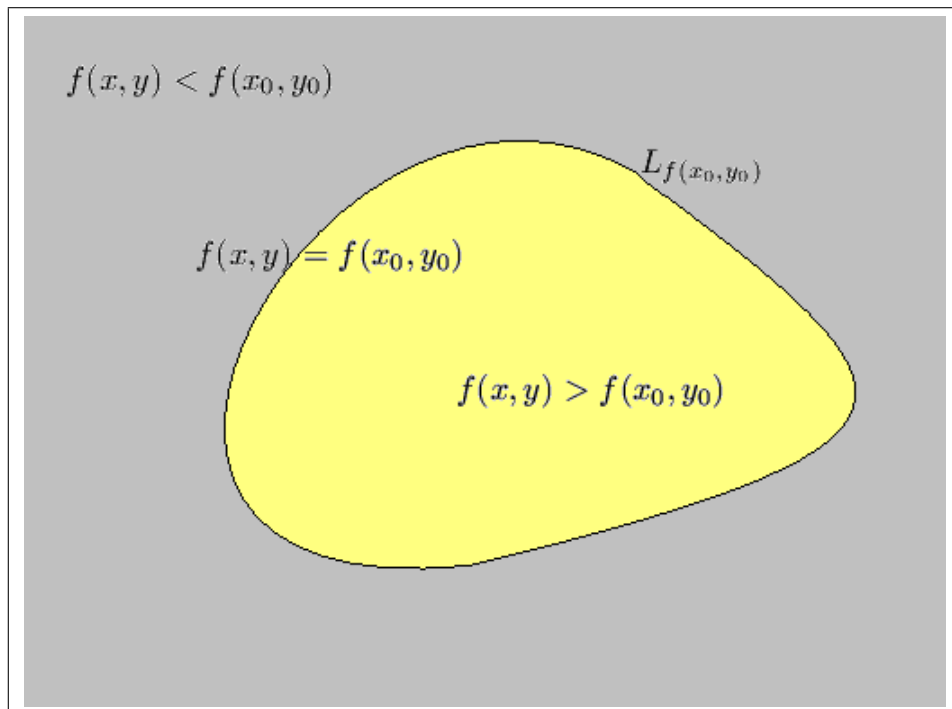


FIGURA 23.5. Curve di livello e segno di f

Se infatti in una parte connessa ci fossero due punti P_+, P_- tali che

$$f(P_+) > 0, \quad f(P_-) < 0$$

esisterebbe in quella parte P_0 tale che

$$f(P_0) = 0$$

ma in quella parte si può solo avere $f(P) > 0$ oppure $f(P) < 0$.

3. Derivabilità e differenziabilità per funzioni di 2 variabili.

Come per le funzioni di 1 variabile è necessario considerare il problema della approssimazione mediante funzioni lineari, cioè il problema della differenziazione.

È molto naturale porre la seguente definizione

DEFINIZIONE 23.6. Diciamo che f è derivabile parzialmente se le funzioni

$$\phi(x) = f(x, y) \quad \psi(y) = f(x, y)$$

sono derivabili.

Chiamiamo $\phi'(x) = f_x(x, y)$ derivata parziale rispetto ad x e $\psi'(y) = f_y(x, y)$ derivata parziale rispetto ad y ; definiamo inoltre gradiente di f e scriviamo $\nabla f(x, y)$ il vettore (punto di \mathbb{R}^2) definito da

$$\nabla f(x, y) = (f_x(x, y), f_y(x, y))$$

Di fatto in tal modo si opera derivando rispetto ad x (o ad y) con y (o x) fissati.

Va osservato che, pur essendo molto naturale, l'uso delle derivate parziali non consente, da solo, di ricavare informazioni utili sulla funzione in esame.

Si pensi ad esempio che la funzione

$$f(x, y) = \begin{cases} 1 & \text{se } xy = 0 \\ 0 & \text{se } xy \neq 0 \end{cases}$$

il cui grafico, si veda la figura ?? è costituito dal piano $z = 0$ privato degli assi x ed y e dalle due rette parallele agli assi x ed y poste a quota $z = 1$, non è continua in $(0, 0)$ pur avendo derivate parziali nulle in $(0, 0)$.

Occorre quindi definire cosa si intende per differenziabile e per questo serve parlare di applicazioni lineari.

DEFINIZIONE 23.7. Si chiama applicazione lineare in \mathbb{R}^2 una funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ tale che

$$f(\alpha P + \beta Q) = \alpha f(P) + \beta f(Q) \quad \forall P, Q \in \mathbb{R}^2, \quad \forall \alpha, \beta \in \mathbb{R}$$

L'insieme delle applicazioni lineari su \mathbb{R}^2 si chiama anche spazio duale di \mathbb{R}^2 .

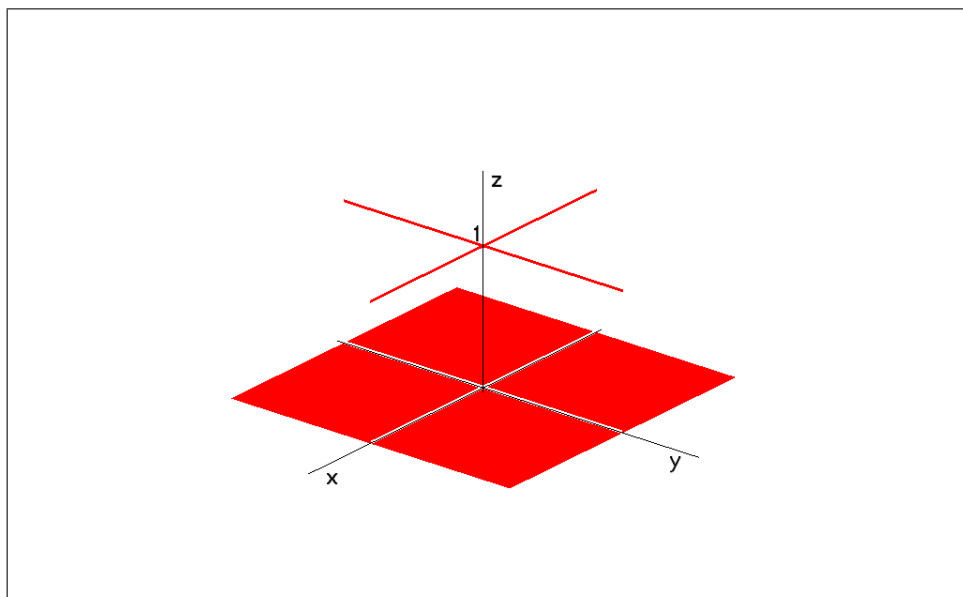


FIGURA 23.6. Il grafico di una funzione derivabile parzialmente, non continua.

Ogni applicazione lineare in \mathbb{R}^2 si può identificare con un punto P^* di \mathbb{R}^2 mediante la seguente uguaglianza

$$f(P) = \langle P, P^* \rangle$$

In altre parole le applicazioni lineari su \mathbb{R}^2 sono tutte e sole le funzioni che si possono scrivere nella forma

$$f(P) = \langle P, P^* \rangle \quad \text{con} \quad P^* \in \mathbb{R}^2$$

È anche utile ricordare che per funzioni lineari possiamo provare che

Se f è una applicazione lineare su \mathbb{R}^2 allora

$$|f(P)| = |\langle P, P^* \rangle| \leq \|P\| \|P^*\|$$

Diciamo che $f \in \mathcal{C}^1(A)$ se f ammette derivate parziali continue in A .

DEFINIZIONE 23.8. Diciamo infine che f è differenziabile in P_0 se esiste $(\alpha, \beta) \in \mathbb{R}^2$ tale che

$$\lim_{P \rightarrow P_0} \frac{f(P) - (f(P_0) + \alpha(x - x_0) + \beta(y - y_0))}{\|P - P_0\|} = 0$$

Pertanto una funzione è differenziabile se

$$f(P) = f(P_0) + \alpha(x - x_0) + \beta(y - y_0) + \|P - P_0\| \omega(P - P_0)$$

dove ω è una funzione infinitesima per $P \rightarrow P_0$

$$\omega(P - P_0) = \frac{f(P) - (f(P_0) + \alpha(x - x_0) + \beta(y - y_0))}{\|P - P_0\|}$$

Questa proprietà si esprime dicendo che $f(P)$ si può approssimare con una funzione lineare affine

$$t(P) = f(P_0) + \alpha(x - x_0) + \beta(y - y_0)$$

a meno di un infinitesimo

$$\|P - P_0\| \omega(P - P_0)$$

di ordine superiore al primo rispetto alla distanza $\|P - P_0\|$.

La funzione $t(p)$ si definisce piano tangente al grafico di f nel punto P_0

Se f è differenziabile in P_0 allora f è anche derivabile parzialmente e si può verificare che risulta

$$\alpha = f_x(P_0) \quad \beta = f_y(P_0)$$

pertanto

Il piano tangente al grafico di una funzione f in P_0 è dato da

$$t(P) = f(P_0) + f_x(P_0)(x - x_0) + f_y(P_0)(y - y_0)$$

DEFINIZIONE 23.9. Se $Q \in \mathbb{R}^2$, diciamo che f è derivabile in P_0 rispetto al vettore Q o che f ammette derivata in P_0 lungo la direzione Q se

$$\lim_{t \rightarrow 0^+} \frac{f(P_0 + tQ) - f(P_0)}{t}$$

esiste finito. In tal caso denotiamo il valore di tale limite con $f'(P_0, Q)$ e lo chiamiamo derivata direzionale di f in P_0 lungo la direzione Q .

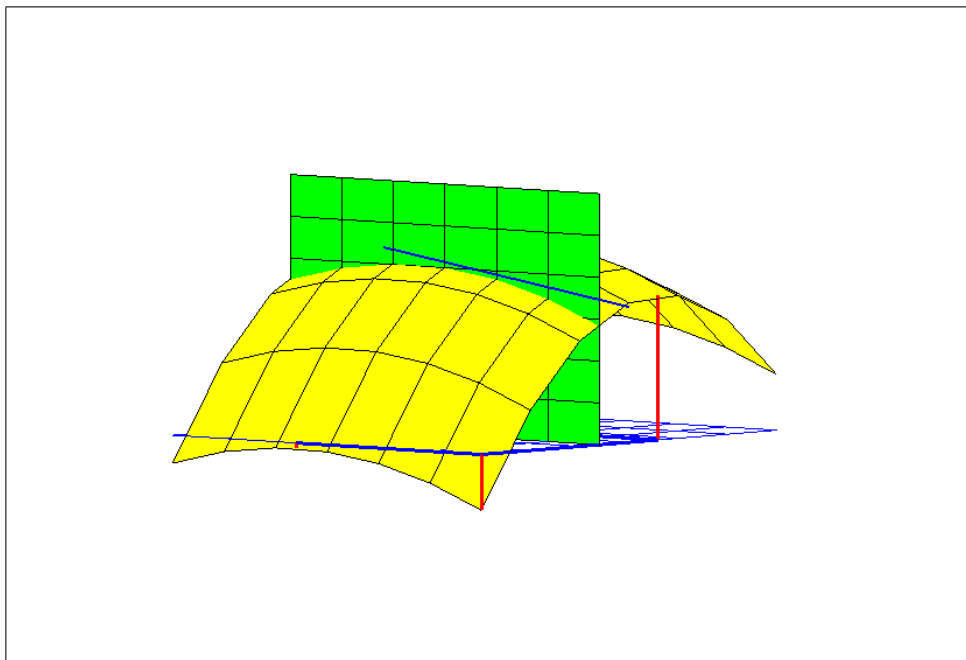


FIGURA 23.7. Definizione di derivata direzionale

Si può vedere che f è derivabile rispetto alla prima variabile se e solo se $f'(P_0, e_1)$ ed $f'(P_0, -e_1)$ esistono finiti e

$$f'(P_0, e_1) = -f'(P_0, -e_1)$$

Analogamente f è derivabile rispetto alla seconda variabile se e solo se $f'(P_0, e_2)$ ed $f'(P_0, -e_2)$ esistono finiti e

$$f'(P_0, e_2) = -f'(P_0, -e_2)$$

Si dimostra che

TEOREMA 23.3. *Se f è differenziabile in P_0 ; allora f è derivabile in P_0 lungo ogni direzione Q e si ha*

$$f'(P_0, Q) = \langle \nabla f(P_0), Q \rangle$$

È utile estendere alle funzioni di più variabili la regola di derivazione delle funzioni composte; ci limitiamo qui a considerare solo due casi particolari.

Siano

$$f : \mathbb{R}^2 \rightarrow \mathbb{R} \quad , \quad g : \mathbb{R} \rightarrow \mathbb{R}^2$$

$$\mathbb{R} \ni t \mapsto g(t) = (x(t), y(t)) \mapsto f(g(t)) = f(x(t), y(t)) \in \mathbb{R}^2$$

Se f e g sono differenziabili (non solo derivabili!) allora

$$\frac{d}{dt}f(g(t)) = f_x(x(t), y(t))\dot{x}(t) + f_y(x(t), y(t))\dot{y}(t)$$

Se viceversa consideriamo

$$f : \mathbb{R} \rightarrow \mathbb{R} \quad , \quad g : \mathbb{R}^2 \rightarrow \mathbb{R}$$

$$\mathbb{R}^2 \ni (x, y) \mapsto g(x, y) \mapsto f(g(x, y)) \in \mathbb{R}$$

e se f e g sono anche qui differenziabili avremo che

$$\begin{aligned} \frac{\partial f}{\partial x} &= f'(g(x, y))g_x(x, y) \\ \frac{\partial f}{\partial y} &= f'(g(x, y))g_y(x, y) \end{aligned}$$

Abbiamo già visto che

se f è differenziabile in $P_0 \in \mathbb{R}^2$ e se Q è una direzione in \mathbb{R}^2 , allora

$$f'(P_0, Q) = \langle \nabla f(P_0), Q \rangle = \|\nabla f(P_0)\| \|Q\| \cos \alpha$$

dove α è l'angolo formato dai vettori $\nabla f(P_0)$ e Q nel piano da essi individuato.

Ne possiamo dedurre che

la derivata direzionale è

- **massima quando $\cos \alpha = 1$ e cioè quando $\alpha = 0$ e $Q = \nabla f(P_0)$,**
- **nulla quando $\cos \alpha = 0$ e cioè quando $\alpha = \frac{\pi}{2}$ e $Q \perp \nabla f(P_0)$,**
- **minima quando $\cos \alpha = -1$ e cioè quando $\alpha = \pi$ e $Q = -\nabla f(P_0)$.**

Consideriamo ora una curva di livello di f

$$L_c = \{(x, y) \in \mathbb{R}^2 : f(x, y) = c\}$$

e supponiamo che sia rappresentabile, almeno localmente, mediante il grafico di una funzione $y = \varphi(x)$. In termini un po' più precisi supponiamo che

$$L_c = \{(x, y) \in U : f(x, y) = c\} = \{(x, y) \in \mathbb{R}^2 : x \in I, y = \varphi(x)\}$$

o più semplicemente

$$f(x, y) = c \iff f(x, \varphi(x)) = c \iff y = \varphi(x)$$

Da $f(x, \varphi(x)) = c$, derivando e tenendo presenti le regole di derivazione delle funzioni composte, otteniamo che:

$$f_x(x, \varphi(x)) + f_y(x, \varphi(x))\varphi'(x) = 0$$

da cui

$$\langle \nabla f(x, \varphi(x)), (1, \varphi'(x)) \rangle = 0$$

e possiamo ricavare che

$$\nabla f(x, \varphi(x)) \perp (1, \varphi'(x))$$

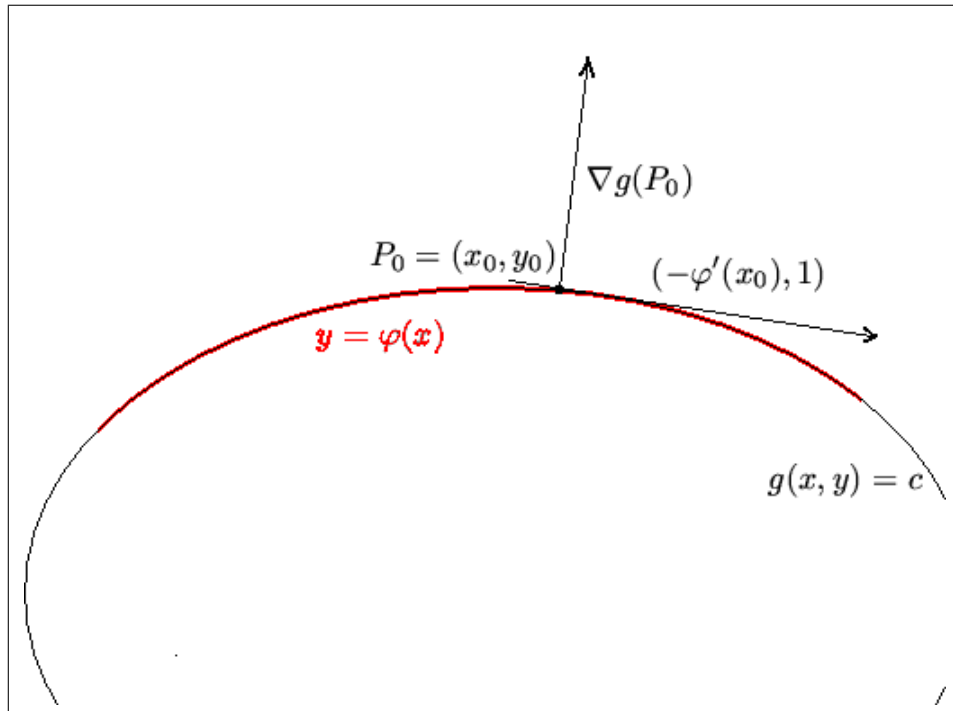


FIGURA 23.8. Curve di livello e gradiente

D'altro canto la retta tangente τ al grafico di φ nel punto $P_0 = (x_0, y_0)$ è data da

$$y - y_0 = \varphi'(x_0)(x - x_0)$$

e si può scrivere nella forma

$$\langle (x - x_0, y - y_0), (\varphi'(x_0), -1) \rangle = 0$$

dalla quale risulta evidente che

$$(\varphi'(x_0), -1) \perp P \quad \forall P \in \tau$$

Se ora teniamo conto che, evidentemente,

$$\langle (\varphi'(x_0), -1), (1, \varphi'(x_0)) \rangle = 0$$

e quindi

$$(\varphi'(x_0), -1) \perp (1, \varphi'(x_0))$$

possiamo ricavare che

$$(23.1) \quad \nabla f(x, \varphi(x)) \perp P \quad \forall P \in \tau$$

Poichè τ è la retta tangente in P_0 al grafico della funzione φ che rappresenta vicino al punto P_0 (localmente in P_0) la curva di livello L_c , esprimeremo la 23.1 dicendo che

il gradiente di f , cioè il vettore $\nabla f(x, y)$, è ortogonale alle curve di livello di f ($L_c = \{(x, y) \in \mathbb{R}^2 : f(x, y) = c\}$)

4. Derivate del secondo ordine: forma quadratica Hessiana.

Possiamo anche considerare le derivate seconde rispetto ad x due volte, ad y due volte, ad x e ad y , ad y e ad x ; chiamiamo queste derivate

$$f_{xx}(P_0) \quad f_{yy}(P_0) \quad f_{xy}(P_0) \quad f_{yx}(P_0)$$

Si può dimostrare che, nel caso in cui $f_{x,y}(P_0)$, o $f_{y,x}(P_0)$ sia continua allora (**teorema di Scharwz**)

$$f_{x,y}(P_0) = f_{y,x}(P_0)$$

Ciò si esprime dicendo che le derivate seconde miste sono uguali.

Chiamiamo matrice Hessiana la matrice i cui elementi sono le derivate seconde di f . Cioè

$$Hf(P_0) = \begin{pmatrix} f_{xx}(P_0) & f_{xy}(P_0) \\ f_{yx}(P_0) & f_{yy}(P_0) \end{pmatrix}$$

Nel caso in cui le derivate miste siano uguali, la matrice Hessiana è simmetrica.

Ad ogni matrice simmetrica, e quindi anche alla matrice Hessiana, possiamo associare un polinomio di secondo grado in 2 variabili (e.g. h, k) omogeneo che chiamiamo forma quadratica associata.

La forma quadratica Hessiana è, posto $R = \begin{pmatrix} h \\ k \end{pmatrix}$

$$Q(R) = Q(h, k) = \begin{pmatrix} h & k \end{pmatrix} \begin{pmatrix} f_{xx}(P_0) & f_{xy}(P_0) \\ f_{yx}(P_0) & f_{yy}(P_0) \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix} =$$

$$R^T H f(P_0) R = f_{xx}(P_0)h^2 + 2f_{xy}(P_0)hk + f_{yy}(P_0)k^2$$

Diciamo che la forma quadratica Q è semidefinita positiva se

$$Q(h, k) = f_{xx}(P_0)h^2 + 2f_{xy}(P_0)hk + f_{yy}(P_0)k^2 \geq 0$$

per ogni $(h, k) \in \mathbb{R}^2$.

Diciamo che Q è definita positiva se

$$Q(h, k) = f_{xx}(P_0)h^2 + 2f_{xy}(P_0)hk + f_{yy}(P_0)k^2 > 0$$

per ogni $(h, k) \in \mathbb{R}^2 \setminus \{(0, 0)\}$.

Ovviamente per identificare una forma quadratica semidefinita o definita negativa è sufficiente cambiare il segno delle disuguaglianze.

Semplici considerazioni sul segno di un trinomio di secondo grado permettono di ottenere condizioni per studiare il carattere di una forma quadratica.

La forma quadratica Q è definita positiva se

$$\det \begin{pmatrix} f_{xx}(P_0) & f_{xy}(P_0) \\ f_{yx}(P_0) & f_{yy}(P_0) \end{pmatrix} > 0$$

e $f_{xx}(P_0) > 0$, oppure $f_{yy}(P_0) > 0$

Osservazione. Se

$$\det \begin{pmatrix} f_{xx}(P_0) & f_{xy}(P_0) \\ f_{yx}(P_0) & f_{yy}(P_0) \end{pmatrix} = f_{xx}(P_0)f_{yy}(P_0) - (f_{xy}(P_0))^2 > 0$$

allora

$$f_{xx}(P_0)f_{yy}(P_0) \geq (f_{xy}(P_0))^2 > 0$$

e quindi $f_{xx}(P_0)$ ed $f_{yy}(P_0)$ hanno lo stesso segno

□

La forma quadratica Q è semidefinita positiva se

$$\det \begin{pmatrix} f_{xx}(P_0) & f_{xy}(P_0) \\ f_{yx}(P_0) & f_{yy}(P_0) \end{pmatrix} \geq 0$$

e $f_{xx}(P_0) \geq 0$, o equivalentemente $f_{yy}(P_0) \geq 0$

Si può inoltre dimostrare che

Se λ_1, λ_2 sono gli autovalori della matrice

$$\det \begin{pmatrix} f_{xx}(P_0) & f_{xy}(P_0) \\ f_{yx}(P_0) & f_{yy}(P_0) \end{pmatrix} \geq 0$$

allora, per la simmetria della matrice, essi sono reali ed inoltre

- **se λ_1, λ_2 sono entrambi positivi (negativi) la forma quadratica Q è definita positiva (negativa)**
- **se λ_1, λ_2 sono entrambi positivi (negativi) o nulli la forma quadratica Q è semidefinita positiva (negativa)**
- **se λ_1, λ_2 hanno segni discordi la forma quadratica Q è non definita**

Osservazione. Se

$$\det \begin{pmatrix} f_{xx}(P_0) & f_{xy}(P_0) \\ f_{yx}(P_0) & f_{yy}(P_0) \end{pmatrix} < 0$$

la forma quadratica può assumere sia valori positivi che negativi e quindi non è definita. \square

5. Massimi e minimi per le funzioni di 2 variabili.

DEFINIZIONE 23.10. Diciamo che P_0 è un punto di minimo (massimo) relativo per f se esiste una sfera $S(P_0, \rho)$, $\rho > 0$, tale che

$$f(P) \geq f(P_0) \quad (f(P) \leq f(P_0))$$

per ogni $P \in S(P_0, \rho)$

Utilizzando tecniche che sfruttano i risultati noti per le funzioni di una variabile possiamo provare le seguenti condizioni necessarie per l'esistenza di un punto di minimo o massimo relativo.

TEOREMA 23.4. Se P_0 è un punto di minimo (massimo) relativo per f interno al suo dominio ed f è differenziabile in P_0 . Allora

- $\nabla f(x) = 0$;

se inoltre f ammette derivate seconde continue in P_0 ,

- $Hf(x)$ è semidefinita positiva (negativa).

Osservazione. Se $\nabla f(x) = 0$ e se $Hf(x)$ non è definito, allora P_0 non è né punto di massimo relativo, né punto di minimo relativo per f ; un punto siffatto viene solitamente indicato con il nome di 'punto sella'. \square

TEOREMA 23.5. Se $f \in \mathcal{C}^2(A)$; e se P_0 è interno al suo dominio e se

- $\nabla f(P_0) = 0$
- $Hf(P_0)$ è definita positiva (negativa)

allora P_0 è punto di minimo (massimo) relativo per f .

Anche per le funzioni di due variabili si può definire e studiare la convessità:

Sia $f : A \rightarrow \mathbb{R}$, e sia $A \subset \mathbb{R}^2$ convesso, cioè supponiamo che A contenga ogni segmento di retta i cui estremi siano contenuti in A ; diciamo che f è convessa se

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) \quad \forall x, y \in A, \quad \forall \lambda \in (0, 1)$$

Inoltre f si dice strettamente convessa se vale la disuguaglianza stretta.

Osservazione. si può dimostrare che se f è convessa allora i suoi insiemi di livello L_c sono a loro volta convessi un insieme convesso \square
Inoltre possiamo anche dimostrare che

TEOREMA 23.6. *Sia $f : A \rightarrow \mathbb{R}$ convessa, A aperto; allora*

- f è continua in A
- $f'(P, Q)$ esiste $\forall P \in A, \forall Q \in \mathbb{R}^2$.

Come per le funzioni di una variabile la convessità si può caratterizzare utilizzando le derivate come si vede dall'enunciato del teorema seguente.

TEOREMA 23.7. *Sia $f : A \rightarrow \mathbb{R}$, $A \subset \mathbb{R}^2$ convesso, aperto; supponiamo inoltre $f \in C^2(A)$, allora sono condizioni equivalenti:*

- f è convessa

•

$$f(y) \geq f(P_0) + \langle \nabla f(P_0), P - P_0 \rangle \quad \forall P, P_0 \in A$$

- $Hf(P)$ è semidefinita positiva.

Inoltre

Ciascuna delle seguenti condizioni è sufficiente per la successiva:

- $Hf(P)$ è definita positiva $\forall P \in A$;
- $f(P) > f(P_0) + \langle \nabla f(P_0), P - P_0 \rangle \quad \forall P, P_0 \in A, P \neq P_0$;
- f è strettamente convessa.

Si può inoltre vedere che se f è strettamente convessa e se $f(P) \rightarrow +\infty$

per $P \rightarrow \infty$; allora esiste uno ed un solo punto $P_0 \in \mathbb{R}^n$ tale che

$$f(P_0) = \min\{f(P) : P \in \mathbb{R}^2\}$$

6. Massimi e minimi vincolati.

Le condizioni fin qui trovate per caratterizzare i punti di massimo e di minimo relativo sono utilizzabili soltanto nel caso in cui si cerchino massimi e minimi di f all'interno di un determinato insieme; nel caso in cui si vogliano cercare massimi e minimi su insiemi che contengano anche punti non interni, questi ultimi andranno considerati a parte esattamente come a parte debbono essere considerati gli estremi di un intervallo se si considerano funzioni di una variabile.

Questo scopo si può raggiungere considerando le restrizioni di f ai punti non interni; tali restrizioni sono funzioni che dipendono da una sola variabile e si può cercare di trattarle con i risultati noti per tal caso.

Ovviamente lo scopo è individuare eventuali massimi o minimi per mezzo di condizioni necessarie e, se si è certi della loro esistenza, tra essi scegliere massimi e minimi assoluti.

A questo scopo è utile considerare il problema di trovare massimi e minimi di una funzione $f(x, y)$ sull'insieme dei punti del piano che soddisfano l'equazione $g(x, y) = 0$

In questo modo, infatti, è possibile identificare in molti casi l'insieme dei punti di frontiera (e quindi non interni) di un insieme.

Più precisamente ci riferiremo a questo problema come al problema di

Cercare massimi e minimi relativi di f vincolati a $g = 0$

6.1. funzioni definite implicitamente. Per studiare il problema è necessario conoscere qualche cosa in più sull'insieme

$$G = \{(x, y) \in \mathbb{R}^2 : g(x, y) = 0\}$$

Più precisamente è necessario rendersi conto che G può essere rappresentato localmente mediante il grafico di una funzione φ .

Per chiarire il concetto consideriamo un semplice esempio.

Sia

$$g(x, y) = x^2 + y^2 - 1$$

ovviamente $g \in \mathcal{C}^1$ ed inoltre

$$\nabla(x, y) = (2x, 2y) \neq (0, 0)$$

per ognuno dei punti tali che

$$g(x, y) = 0$$

È ben noto che l'equazione

$$g(x, y) = x^2 + y^2 - 1 = 0$$

identifica una circonferenza di raggio unitario centrata in $(0, 0)$.

Per illustrare la possibilità di rappresentare la circonferenza localmente in un punto P_0 mediante una funzione φ possiamo considerare i seguenti casi

- se $P_0 = (0, 1)$ possiamo rappresentare la circonferenza mediante la funzione

$$y = \sqrt{1 - x^2}$$

- se $P_0 = (0, -1)$ possiamo rappresentare la circonferenza mediante la funzione

$$y = -\sqrt{1 - x^2}$$

- se $P_0 = (1, 0)$ possiamo rappresentare la circonferenza mediante la funzione

$$x = \sqrt{1 - y^2}$$

- se $P_0 = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ possiamo rappresentare la circonferenza sia mediante la funzione

$$y = \sqrt{1 - x^2}$$

sia mediante la funzione

$$x = \sqrt{1 - y^2}$$

- se $P_0 = (-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ possiamo rappresentare la circonferenza sia mediante la funzione

$$y = \sqrt{1 - x^2}$$

sia mediante la funzione

$$x = -\sqrt{1 - y^2}$$

In generale non è, tuttavia, possibile trovare esplicitamente la funzione φ , come abbiamo fatto nell'esempio appena visto, tuttavia è per taluni scopi sufficiente sapere che questa funzione esiste.

A questo proposito si può dimostrare che

TEOREMA 23.8. - *delle funzioni implicite di U. Dini* - Se g è sufficientemente regolare ($g \in C^1$, $\nabla g(x, y) \neq (0, 0)$) l'insieme

$$\{(x, y) \in \mathbb{R}^2, : g(x, y) = 0\}$$

può essere rappresentato localmente, (cioè in un intorno di ogni suo punto), come grafico di una funzione

$$y = \varphi(x)$$

6.2. il principio dei moltiplicatori di Lagrange. Si può trovare una condizione necessaria affinché un punto P_0 sia di minimo o di massimo per f vincolato a $g = 0$; possiamo enunciare tale condizione come segue

TEOREMA 23.9. - dei moltiplicatori di Lagrange- Se $f, g \in C^1$, $\nabla f(P_0) \neq (0, 0)$ e $P_0 = (x_0, y_0)$ è un punto di minimo o di massimo per f vincolato a $g(x, y) = 0$, allora

$$\nabla f(P_0) \parallel \nabla g(P_0)$$

o equivalentemente esiste $\lambda \in \mathbb{R}$ tale che

$$\nabla f(P_0) = \lambda \nabla g(P_0)$$

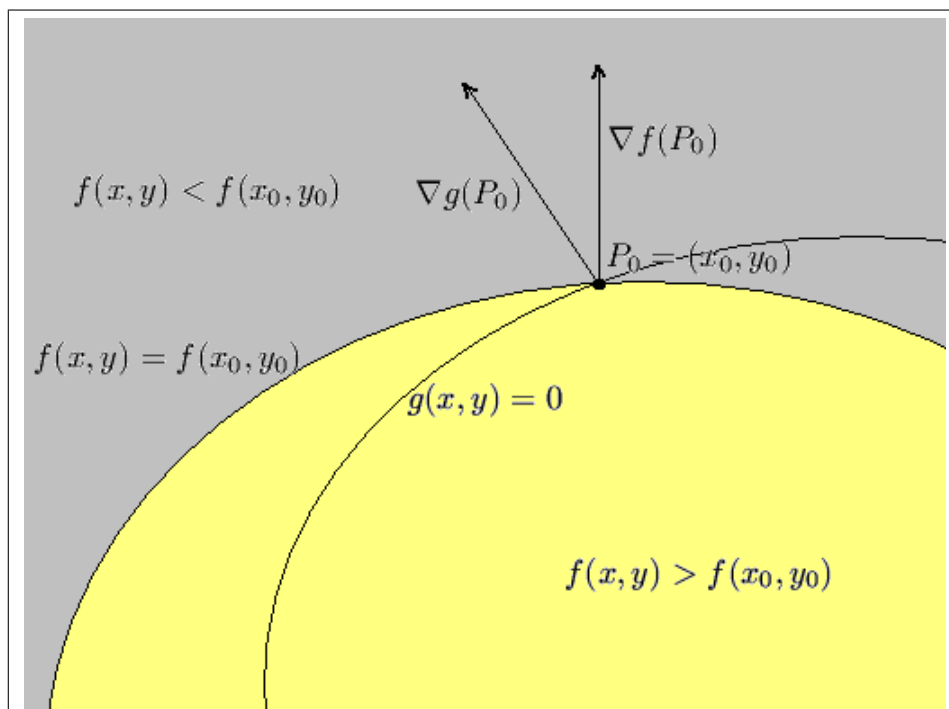


FIGURA 23.9. Principio dei moltiplicatori di Lagrange

Se infatti $\nabla f(P_0)$ e $\nabla g(P_0)$ non fossero paralleli, tenendo conto del fatto che $\nabla f(P_0)$ è ortogonale alla curva definita da $f(x, y) = f(x_0, y_0)$ mentre $\nabla g(P_0)$ è ortogonale alla curva definita da $g(x, y) = 0$ avremmo una situazione simile a quella illustrata nella figura 23.9.

Dalla figura si vede che ci sarebbero punti soddisfacenti l'equazione $g(x, y) = 0$ tali che $f(x, y) > f(x_0, y_0)$ ed anche punti tali che $f(x, y) < f(x_0, y_0)$.

Ciò escluderebbe che P_0 sia un punto di minimo o di massimo di f vincolato a $g = 0$

Possiamo dimostrare con maggior precisione il risultato come segue.

Siano $f, g : A \rightarrow \mathbb{R}$, $P_0 = (x_0, y_0) \in A \subset \mathbb{R}^2$, A aperto, $f, g \in C^1(A)$, e supponiamo che $g(x_0, y_0) = 0$. Supponiamo inoltre che $\nabla g(x_0, y_0) \neq 0$, il che significa, a meno di cambiare il nome delle variabili, che si può supporre $g_y(x_0, y_0) \neq 0$; allora si può dimostrare che esiste una funzione φ definita in un intorno di x_0 che assume valori in un intorno di y_0 e per la quale si ha

$$g(x, \varphi(x)) = 0$$

Pertanto la funzione $f(x, \varphi(x))$ ammette in x_0 un punto di minimo relativo se e solo se $P_0 = (x_0, y_0)$ è un punto di minimo per f vincolato a $g = 0$.

Di conseguenza, se $P_0 = (x_0, y_0)$ è un minimo relativo per f vincolato a $g = 0$ si ha

$$\frac{d}{dx} f(x, \varphi(x)) = f_x(x_0, y_0) + f_y(x_0, y_0) \varphi'(x_0) = 0$$

ed anche

$$g_x(x_0, y_0) + g_y(x_0, y_0) \varphi'(x_0) = 0$$

e la coppia $(1, \varphi'(x_0))$ è soluzione non banale del sistema algebrico lineare omogeneo la cui matrice dei coefficienti è data da

$$\begin{pmatrix} \nabla f(x_0, y_0) \\ \nabla g(x_0, y_0) \end{pmatrix}$$

Ne segue che esistono $\alpha, \beta \in \mathbb{R}$ non entrambi nulli, tali che

$$\alpha \nabla f(x_0, y_0) + \beta \nabla g(x_0, y_0) = 0$$

e, dal momento che $\nabla g(x_0, y_0) \neq 0$, ne viene che deve essere $\alpha \neq 0$.

Possiamo pertanto affermare, a meno di dividere per α , che esiste λ tale che

$$\nabla f(x_0, y_0) + \lambda \nabla g(x_0, y_0) = 0$$

Viceversa, posto

$$h(x) = f(x, \varphi(x))$$

se $h'(x_0) = 0$ e $h''(x_0) > 0$, (x_0, y_0) è un punto di minimo relativo per f vincolato a $g = 0$.

Concludiamo osservando un semplice fatto, spesso utile quando si trattano problemi di programmazione lineare.

TEOREMA 23.10. *Sia $f : A \rightarrow \mathbb{R}$, $A \subset \mathbb{R}^2$ convesso, chiuso e limitato, f convessa e continua; allora il massimo di f in A è assunto anche in punti che sono sulla frontiera di A*

DIMOSTRAZIONE. Sia

$$f(P) = \max\{f(Q) : Q \in A\}$$

allora, se P è interno ad A , detti $Q, R \in A$ gli estremi del segmento ottenuto intersecando A con una qualunque retta passante per P , si ha

$$P = \lambda Q + (1 - \lambda)R$$

e

$$f(P) \leq \lambda f(Q) + (1 - \lambda)f(R) \leq \max\{f(Q), f(R)\}$$

□

Osservazione. Nel caso in cui A sia poliedrale, cioè se

$$A = \{P \in \mathbb{R}^2 : g_i(P) \leq 0, g_i \text{ lineare}, i = 1, \dots, m\}$$

il massimo si può cercare solo tra i vertici della frontiera.

□

INTEGRAZIONE PER LE FUNZIONI DI DUE VARIABILI

1. Definizione di integrale doppio.

Se f è una funzione di 2 variabili positiva e se $R = [a, b] \times [c, d]$ è un rettangolo contenuto nel suo dominio, possiamo considerare il problema di calcolare il volume V delimitato dal piano (x, y) dal grafico di f e dal cilindro generato da R con generatrici parallele all'asse z (si veda la figura 24.1).

Il volume può essere definito

- considerando una partizione di R ,
- definendo in corrispondenza le somme superiori e le somme inferiori di f relative alla partizione scelta,
- dichiarando una funzione integrabile se, al variare delle partizioni, l'estremo inferiore delle somme superiori e l'estremo superiore delle somme inferiori coincidono,
- in tal caso chiamiamo il loro valore comune

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

L'esistenza dell'integrale è assicurata, similmente a quanto accade per le funzioni di una variabile, dalla continuità della funzione integranda, e si può anche dimostrare che è sufficiente che f sia continua su R a meno di un insieme di area 0. Possiamo in altre parole dimostrare il seguente risultato.

Se f è una funzione di due variabili limitata su un insieme chiuso e limitato D ed è continua a meno di un sottoinsieme di misura 0, allora f è integrabile su D

(Pur non entrando nei particolari della definizione di area, possiamo ricordare che è possibile calcolare l'area di insiemi piani significativi usando la teoria dell'integrazione per le funzioni di una variabile)

Quando una funzione è integrabile possiamo approssimare il suo integrale su R anche usando le somme di Riemann; tali somme possono anche essere usate per dare la definizione di integrabilità e si calcolano come segue:

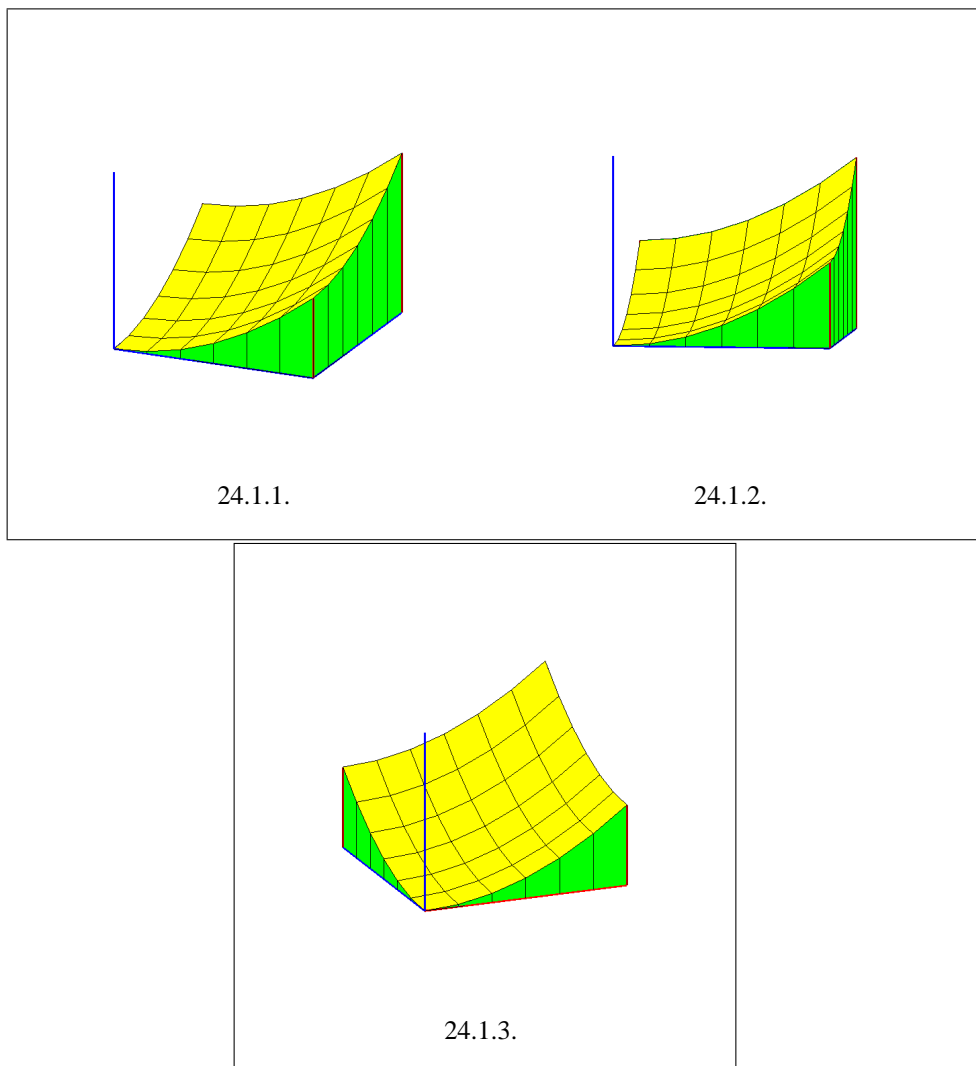


FIGURA 24.1.

- Si suddivide il rettangolo R in rettangoli più piccoli R_j , ad esempio ottenuti suddividendo i lati di R in parti uguali di ampiezza δ_x e δ_y , rispettivamente (figura 24.3);
- si sceglie in maniera arbitraria un punto (ξ_j, η_j) in ognuno dei rettangoli R_j e si calcola la quota $f(\xi_j, \eta_j)$ (figura 24.2.1)
- si sostituisce al volume delimitato su R_j dalla funzione f il volume del parallelepipedo di base R_j ed altezza $f(\xi_j, \eta_j)$ (figura 24.2.2)
- si calcola la somma di tutti i contributi così ottenuti da ciascuno dei rettangoli della partizione.

Le somme di Riemann sono così definite da

$$(24.1) \quad \mathcal{R}(f) = \sum_j f(\xi_j, \eta_j) \delta_x \delta_y$$

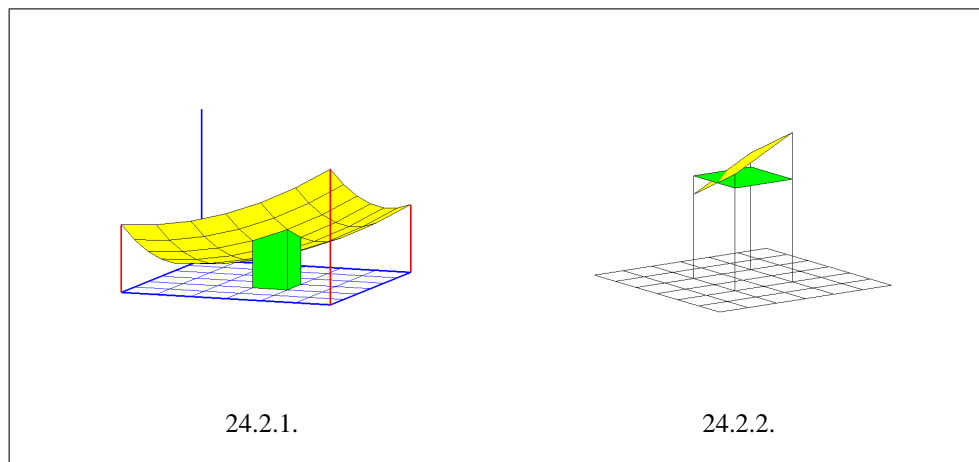


FIGURA 24.2.

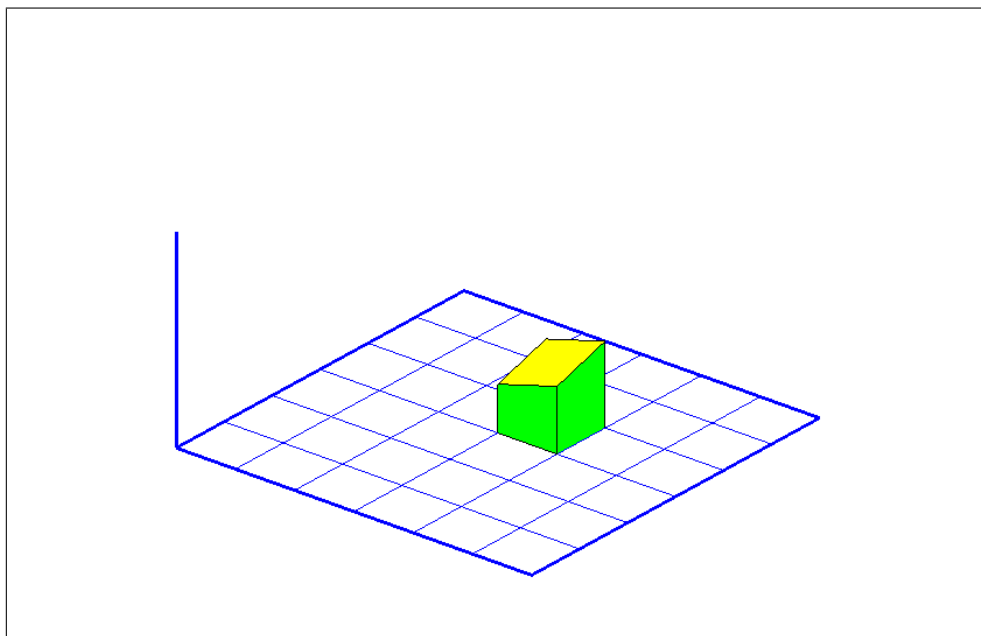


FIGURA 24.3. Singolo elemento di volume

e, quando la partizione è abbastanza fine, cioè suddivide R in rettangoli R_j abbastanza piccoli, $\mathcal{R}(f)$ approssima il valore di $\iint_R f(x, y) dx dy$

2. Formule di riduzione per gli integrali doppi.

Purtroppo non disponiamo, per il calcolo di un integrale doppio, di uno strumento tanto potente quanto il teorema fondamentale del calcolo integrale; questo risultato si può infatti estendere anche al calcolo delle funzioni di più variabili, ma si colloca in un contesto più generale: quello delle forme differenziali e del teorema di Stokes.

Occorre quindi cercare altre vie per il calcolo degli integrali doppi.

Se definiamo

$$S(x) = \{(x, y, z) \in \mathbb{R}^3 : a \leq x \leq b, 0 \leq z \leq f(x, y)\}$$

$S(x)$ rappresenta una sezione del volume V , si veda figura 24.4.

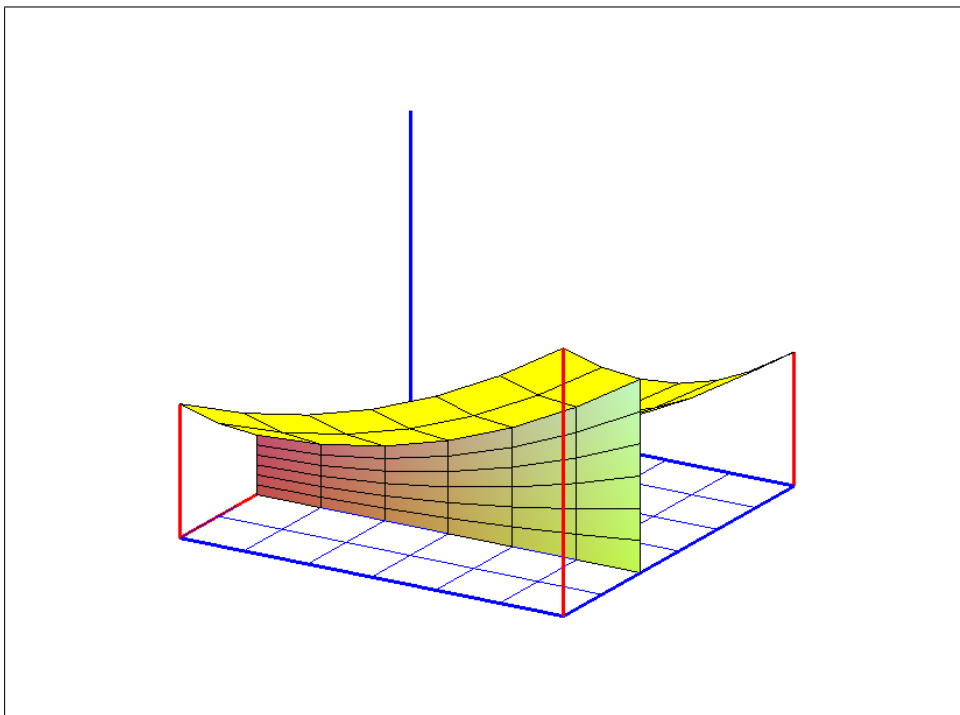


FIGURA 24.4. $S(x)$

ed il calcolo integrale per le funzioni di una variabile consente di calcolare la sua area $A(x)$ mediante la

$$A(x) = \int_a^b f(x, y) dy$$

Possiamo considerare il volume V come la somma (infinita) dei volumi elementari $A(x)$ (che sono nulli) per $x \in [a, b]$; naturalmente la somma infinita si calcola integrando $A(x)$ su $[a, b]$ e quindi

$$\iint_R f(x, y) dx dy = \int_a^b \left(\int_c^d f(x, y) dy \right) dx$$

In maniera del tutto simile possiamo calcolare

$$\iint_R f(x, y) dx dy = \int_c^d \left(\int_a^b f(x, y) dx \right) dy$$

Il calcolo di un integrale doppio può quindi essere ridotto al calcolo di due integrali semplici (formule di riduzione).

Il vincolo fin qui posto sul dominio di integrazione, (R è un rettangolo) non può tuttavia essere mantenuto e quindi è necessario definire

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

per una classe di sottoinsiemi del piano un po' più generale.

È naturale considerare per questo scopo la classe dei **domini normali**

Diciamo che un insieme D è un dominio normale rispetto all'asse x se

$$D = \{(x, y) \in \mathbb{R}^2 : a \leq x \leq b, \alpha(x) \leq y \leq \beta(x)\}$$

dove $[a, b]$ è un intervallo reale e α e β sono funzioni continue su $[a, b]$. (Si veda la figura 24.5).

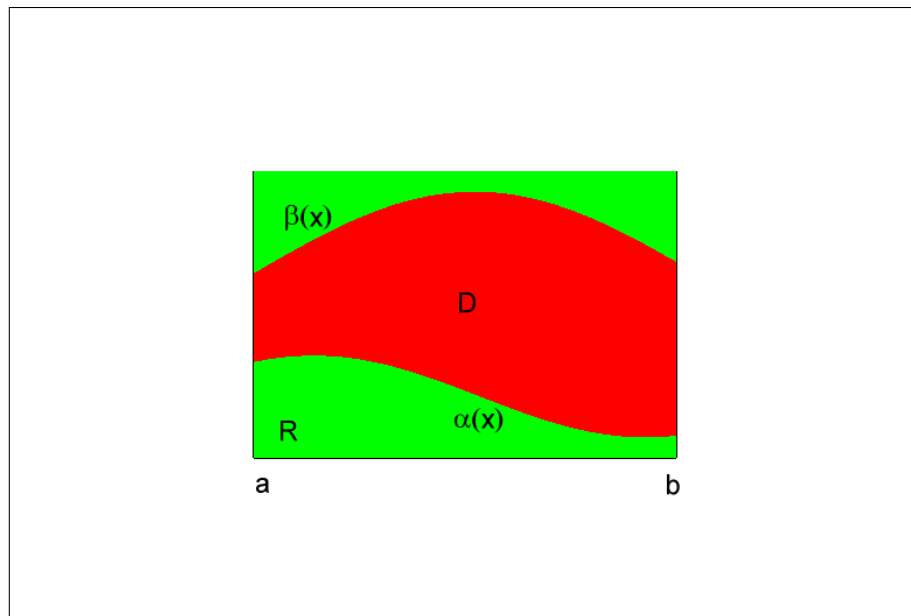
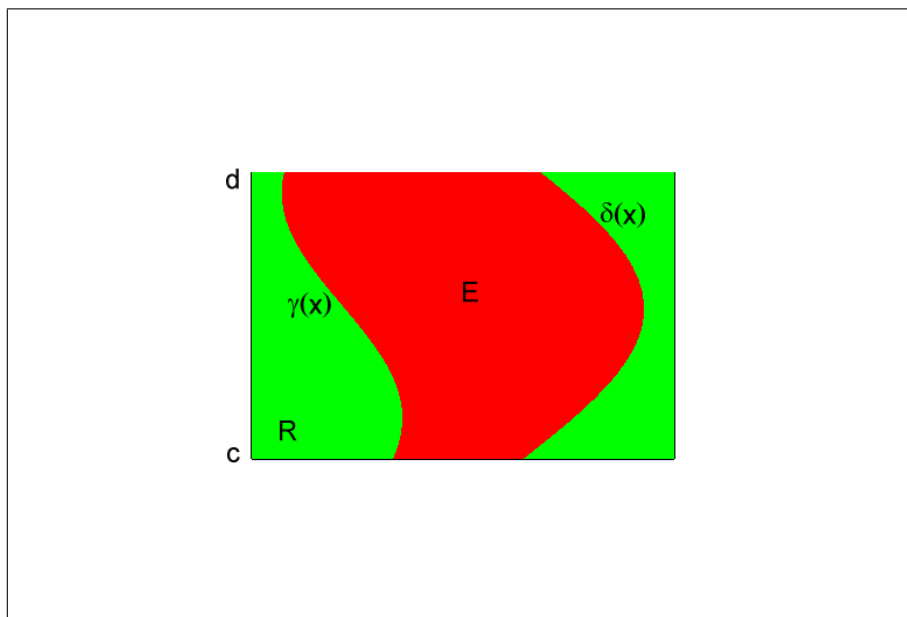


FIGURA 24.5. Dominio normale rispetto all'asse x

Diciamo che un insieme E è un dominio normale rispetto all'asse y se

$$E = \{(x, y) \in \mathbb{R}^2 : c \leq y \leq d, \gamma(y) \leq x \leq \delta(y)\}$$

dove $[c, d]$ è un intervallo reale e γ e δ sono funzioni continue su $[c, d]$. (Si veda la figura 24.6).

FIGURA 24.6. Dominio normale rispetto all'asse x

Per definire, ad esempio,

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

possiamo

- definire una funzione

$$\tilde{f}(x, y) = \begin{cases} f(x, y) & \text{se } (x, y) \in D \\ 0 & \text{se } (x, y) \notin D \end{cases}$$

- considerare un rettangolo $R \supset D$
- definire

$$\iint_D f(x, y) dx dy = \iint_R \tilde{f}(x, y) dx dy$$

Nel compiere questa procedura, possiamo osservare che, poichè si può dimostrare che il grafico di una funzione continua ha area nulla e poichè la definizione di \tilde{f} può generare discontinuità solo nei punti dei grafici di α e di β , se f è continua a meno di insiemi di area nulla tale risulta anche \tilde{f} e pertanto

Una funzione f continua a meno di un insieme di area nulla è integrabile su un dominio normale D .

A completamento occorre poi osservare che

$$\begin{aligned}\iint_D f(x, y) dx dy &= \iint_R \tilde{f}(x, y) dx dy \\ &= \int_a^b \left(\int_c^d \tilde{f}(x, y) dy \right) dx = \int_a^b \left(\int_{\alpha(x)}^{\beta(x)} f(x, y) dy \right) dx\end{aligned}$$

ed in maniera del tutto simile

$$\begin{aligned}\iint_E f(x, y) dx dy &= \iint_R \tilde{f}(x, y) dx dy \\ &= \int_c^d \left(\int_a^b \tilde{f}(x, y) dx \right) dy = \int_c^d \left(\int_{\gamma(y)}^{\delta(y)} f(x, y) dx \right) dy\end{aligned}$$

Infine osserviamo che quanto abbiamo visto è applicabile ad insiemi che siano unione finita di domini normali. Questo ci permette di considerare la maggior parte degli insiemi che si incontrano nella pratica del calcolo.

3. Cambiamento di variabili negli integrali doppi

3.1. Cambiamento di variabili lineari. Consideriamo ora il problema di calcolare l'area di un parallelogrammo A che abbia come lati i vettori (a, b) e (c, d) .

Semplici considerazioni di geometria permettono di stabilire che

$$\text{Area}(A) = ad - bc = \det \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

ed inoltre, se teniamo conto del fatto che il volume del cilindro di altezza 1 (figura 24.7) che ha per base il parallelogrammo è uguale ad $\text{Area}(A)$, possiamo anche affermare che

$$\text{Area}(A) = \iint_A 1 dx dy = \iint_B \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} du dv$$

dove

$$B = \{(u, v) : 0 \leq u \leq 1, 0 \leq v \leq 1\} = [0, 1] \times [0, 1]$$

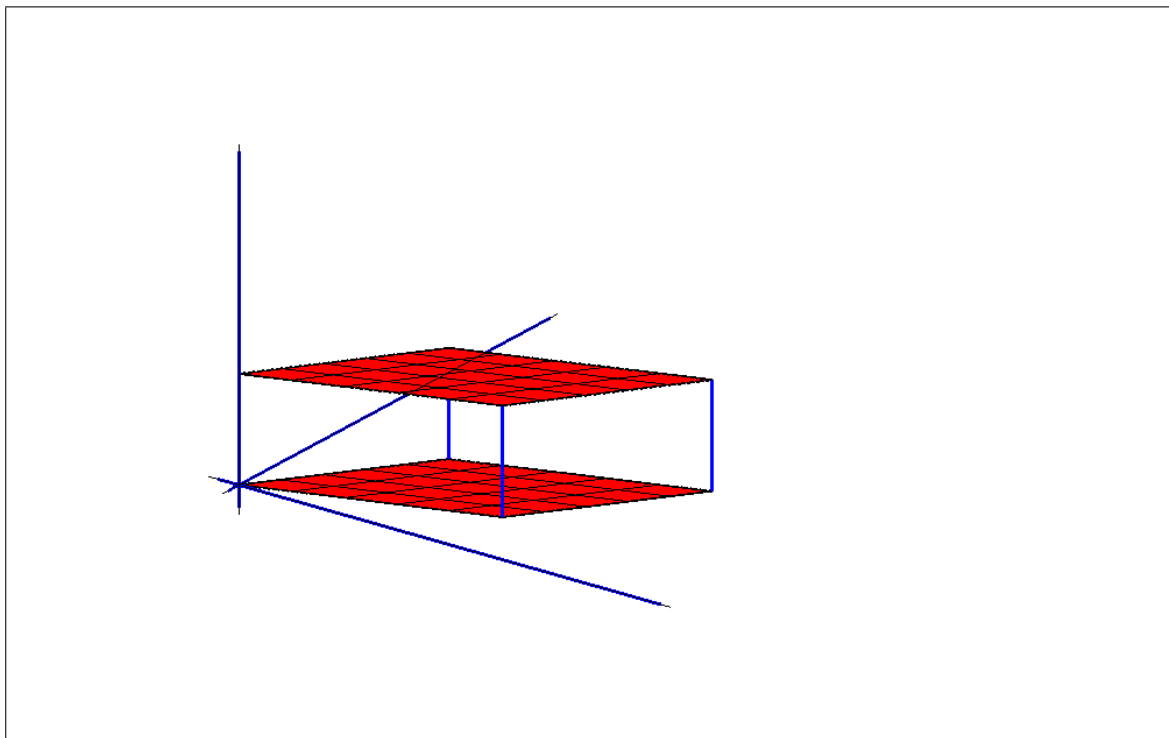


FIGURA 24.7. Volume ed area

ed osservare che il quadrato $B = [0, 1] \times [0, 1]$ si trasforma nel parallelogramma A mediante le corrispondenze

$$(24.2) \quad \begin{cases} x = au + bv \\ y = cu + dv \end{cases} \quad \text{cioè} \quad \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

Se supponiamo che

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} \neq 0$$

la corrispondenza è biunivoca e può essere invertita; sia

$$(24.3) \quad \begin{cases} u = \alpha x + \beta y \\ v = \gamma x + \delta y \end{cases} \quad \text{cioè} \quad \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

la corrispondenza inversa.

Con riferimento alla definizione di integrale possiamo anche osservare che una partizione del quadrato B in quadrati più piccoli B_j corrisponde ad una suddivisione del parallelogramma A in parallelogrammi A_j , simili, più piccoli (si veda la figura 24.8).

Pertanto se f è una funzione definita su A , per calcolare

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

possiamo calcolare le somme di Riemann usando la partizione di A in parallelogrammi, che risulta più naturale di una partizione in rettangoli; Le somme di Riemann in questo caso risultano essere

$$\mathcal{R}(f) = \sum_j f(x_j, y_j) \text{Area}(A_j) = \sum_j f(x_j, y_j) \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{Area}(B_j)$$

Ma esiste un unico punto $(u_j, v_j) \in B$ tale che

$$(x_j, y_j) = (au_j + bv_j, cu_j + dv_j)$$

per cui

$$\mathcal{R}(f) = \sum_j f(au_j + bv_j, cu_j + dv_j) \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{Area}(B_j)$$

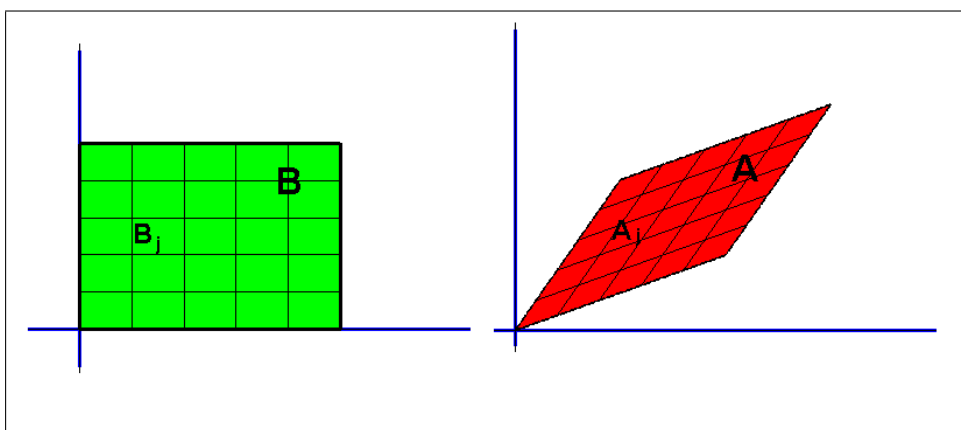


FIGURA 24.8. Cambiamento di variabili lineare

Tali somme al raffinarsi della partizione si approssimano a

$$\iint_B f(au + bv, cu + dv) \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} du dv$$

e quindi possiamo concludere che

$$\iint_A f(x, y) dx dy = \iint_B f(au + bv, cu + dv) \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} du dv$$

3.2. Coordinate Polari nel piano. Possiamo usare in luogo di 24.2 anche altre trasformazioni; ad esempio possiamo usare la trasformazione in coordinate polari che è definita da:

$$(24.4) \quad \begin{cases} x = \rho \cos \theta \\ y = \rho \sin \theta \end{cases} \quad \rho \geq 0, \quad \theta \in [0, 2\pi]$$

La 24.4 trasforma

- le rette $\rho = R$ in circonferenze centrate nell'origine di raggio R
- le rette $\theta = \alpha$ in semirette passanti per l'origine inclinate di un angolo α rispetto al semiasse positivo dell'asse x .
- i settori di corona circolare nel piano (x, y)

$$A = \{(x, y) : r \leq \sqrt{x^2 + y^2} \leq R, \alpha \leq \tan y/x \leq \beta\} = \{(x, y) : r \leq \rho \leq R, \alpha \leq \theta \leq \beta\}$$

in rettangoli

$$B = \{(\rho, \theta) : \rho \leq R, \alpha \leq \theta \leq \beta\} = [r, R] \times [\alpha, \beta]$$

nel piano (ρ, θ) . (Si veda la figura 24.9).

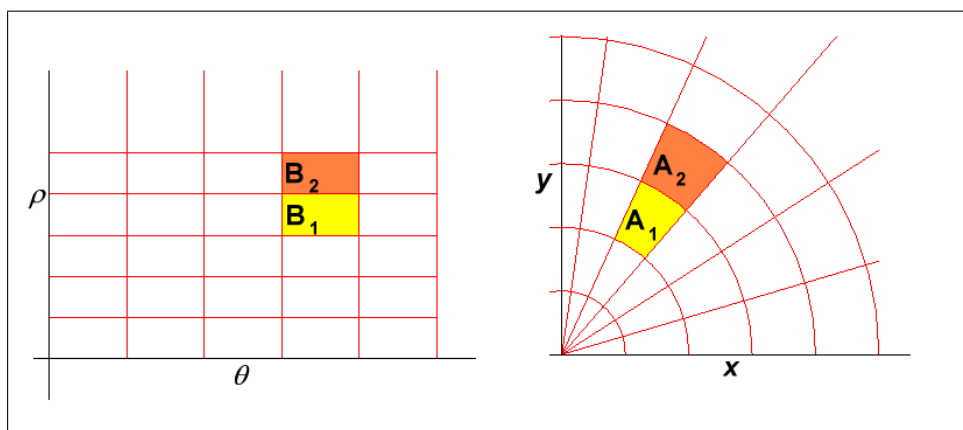


FIGURA 24.9. Cambiamento di variabili in coordinate polari

Purtroppo, la trasformazione definita dalle 24.4 non è biunivoca nè invertibile ed inoltre se

$$B_1 = [R - \delta_r, R] \times [\alpha, \beta]$$

e

$$B_2 = [R, R + \delta_r] \times [\alpha, \beta]$$

l'immagine A_1 di B_1 ed A_2 di B_2 hanno aree diverse anche se B_1 e B_2 hanno aree uguali.

Più precisamente si vede che se l'area di A_1 è più piccola di quella di A_2 poichè A_1 è più vicino all'origine di A_2 .

Possiamo calcolare che:

$$\begin{aligned} \text{Area}(B) &= (R - r)(\beta - \alpha) \\ \text{Area}(A) &= \frac{1}{2}(R^2 - r^2)(\beta - \alpha) = \frac{1}{2}(R + r)\text{Area}(B) \end{aligned}$$

Pertanto non possiamo procedere, come nel caso di 24.2 in quanto il fattore di conversione per ottenere Area (A) da Area (B) non è costante.

Possiamo tuttavia affermare che

$$(24.5) \quad \text{Area} (A) = \iint_A 1 dx dy$$

e la 24.5 si può ottenere come somma di settori circolari più piccoli A_j delimitati da circonferenze di raggio ρ e $\rho + \delta_\rho$ e aventi ampiezza δ_θ .

Se

$$B_j = [\rho, \rho + \delta_\rho] \times [\theta, \theta + \delta_\theta]$$

l'area di ciascuno dei settori A_j è data da

$$\text{Area} (A_j) = \frac{1}{2}(2\rho + \delta_\rho)\delta_\rho\delta_\theta = \frac{1}{2}(2\rho + \delta_\rho)\text{Area} (B_j)$$

ed inoltre se δ_ρ è piccolo e trascurabile avremo che

$$\text{Area} (A_j) \approx \frac{1}{2}2\rho\text{Area} (B_j)$$

Poichè

$$\text{Area} (A) = \sum_j \text{Area} (A_j) \approx \sum_j \rho \text{Area} (B_j)$$

Possiamo affermare che

$$\text{Area} (A) = \iint_B \rho d\rho d\theta$$

Se poi f è una funzione definita su A , possiamo affermare che

$$\iint_A f(x, y) dx dy = \iint_A f(\rho \cos \theta, \rho \sin \theta) \rho d\rho d\theta$$

INTEGRAZIONE DI FUNZIONI DI TRE VARIABILI

1. Definizione di integrale triplo.

Le formule di riduzione che abbiamo usato per il calcolo di un integrale doppio consentono di ricondurre il problema al calcolo di due integrali semplici.

In modo del tutto simile possiamo trovare il modo di calcolare un integrale triplo, cioè l'integrale di una funzione f di tre variabili (x, y, z) su un dominio V contenuto in \mathbb{R}^3

$$\iiint_V f(x, y, z) dx dy dz$$

Il concetto di area che è naturalmente collegato al concetto di integrale semplice e quello di volume che è caratteristico dell'integrale doppio si estende al concetto di ipervolume a quattro dimensioni per gli integrali tripli.

Inoltre, come nel caso di due variabili in cui abbiamo osservato che

$$\text{Area}(A) = \iint_A 1 dx dy = \text{Volume}(C_A)$$

se C_A è il cilindro di base A e di altezza 1, possiamo dire che

$$\text{Volume}(V) = \iiint_V 1 dx dy dz = \text{IperVolume}(C_V)$$

dove C_V è il cilindro di base V e di altezza 1.

2. Formule di riduzione per gli integrali tripli.

Per gli integrali tripli sono possibili diverse scomposizioni che danno origine a diverse formule di riduzione che riteniamo utile illustrare mediante qualche esempio.

Ci occuperemo allo scopo di calcolare

$$\iiint_V f(x, y, z) dx dy dz$$

dove

$$(25.1) \quad V = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 \leq z \leq 2 - \sqrt{x^2 + y^2}\}$$

La parte di \mathbb{R}^3 definita dalla 25.1 è quella indicata nella figura 25.1.2
25.1.1

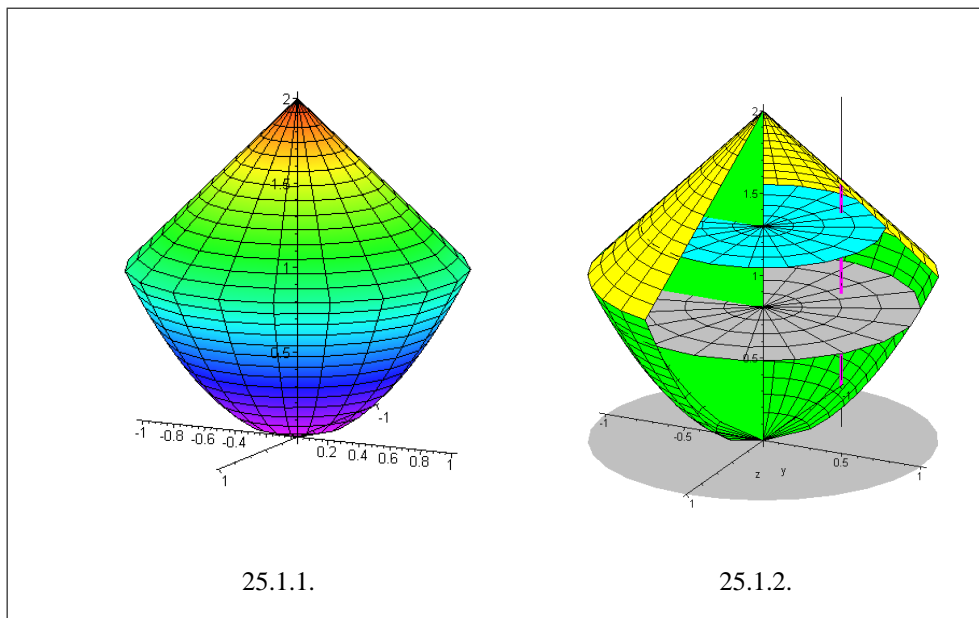


FIGURA 25.1.

Possiamo integrare su V "sommando", cioè integrando rispetto a z , i valori ottenuti mediante il calcolo dell'integrale doppio sulle sezioni di V definite da

$$S(z) = \{(x, y) : (x, y, z) \in V\}$$

Avremo pertanto che

$$\iiint_V f(x, y, z) dx dy dz = \int_0^2 \left(\iint_{S(z)} f(x, y, z) dx dy \right) dz$$

e gli integrali indicati si calcolano come già sappiamo.

Possiamo anche calcolare l'integrale triplo considerando la proiezione D del solido V e calcolando

$$\iint_D \left(\int_{F(x,y)} f(x, y, z) dx, dy \right) dz = \iint_D \left(\int_{x^2+y^2}^{2-\sqrt{x^2+y^2}} f(x, y, z) dx, dy \right) dz$$

3. Cambiamento di variabili per gli integrali tripli.

Anche per gli integrali tripli è utile considerare qualche cambiamento di variabile allo scopo di semplificare i calcoli nel caso di solidi con particolari simmetrie.

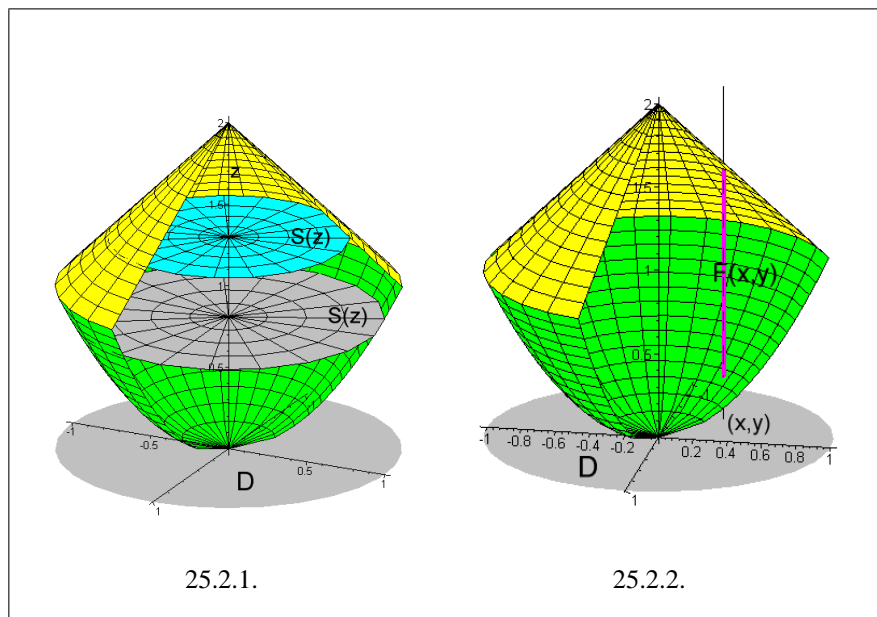


FIGURA 25.2.

I cambiamenti di variabile più comuni sono quello lineare, quello in coordinate cilindriche e quello in coordinate sferiche, che ora illustriamo brevemente.

3.1. Cambio di variabili lineare. Si tratta del cambiamento di variabili definito dalle

$$\begin{cases} x = a_1 r + b_1 s + c_1 t \\ y = a_2 r + b_2 s + c_2 t \\ z = a_3 r + b_3 s + c_3 t \end{cases}$$

Se $A \subset \mathbb{R}^3$ e se B è il trasformato di A mediante il cambiamento di variabili lineari si ha

$$\begin{aligned} \iiint_A f(x, y, z) dx dy dz = \\ \iiint_B f(a_1 r + b_1 s + c_1 t, a_2 r + b_2 s + c_2 t, a_3 r + b_3 s + c_3 t) \left| \frac{\partial(x, y, z)}{\partial(r, s, t)} \right| dr ds dt \end{aligned}$$

dove

$$\left| \frac{\partial(x, y, z)}{\partial(r, s, t)} \right| = \det \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}$$

3.2. Coordinate cilindriche. Si tratta del cambiamento di variabili definito dalle

$$\begin{cases} x = \rho \cos \theta \\ y = \rho \sin \theta \\ z = z \end{cases} \quad \rho \geq 0 \quad 0 \leq \theta \leq 2\pi \quad , z \in \mathbb{R}$$

Se $A \subset \mathbb{R}^3$ e se B è il trasformato di A mediante il cambiamento di variabili in coordinate cilindriche si ha

$$\iiint_A f(x, y, z) dx dy dz = \iiint_B f(\rho \cos \theta, \rho \sin \theta, z) \rho d\rho d\theta dz$$

3.3. Coordinate sferiche. Si tratta del cambiamento di variabili definito dalle

$$\begin{cases} x = \rho \cos \theta \cos \varphi \\ y = \rho \sin \theta \cos \varphi \\ z = \rho \sin \varphi \end{cases} \quad \rho \geq 0 \quad 0 \leq \theta \leq 2\pi \quad -\pi/2 \leq \varphi \leq \pi/2$$

Se $A \subset \mathbb{R}^3$ e se B è il trasformato di A mediante il cambiamento di variabili in coordinate sferiche si ha

$$\begin{aligned} \iiint_A f(x, y, z) dx dy dz = \\ \iiint_B f(\rho \cos \theta \cos \varphi, \rho \sin \theta \cos \varphi, \rho \sin \varphi) \rho \cos \varphi d\rho d\theta d\varphi \end{aligned}$$

INTEGRALI MULTIPLI IMPROPRI

Come nel caso degli integrali semplici, possiamo considerare il problema di calcolare l'integrale di una funzione di due o più variabili che non siano limitate o su domini di integrazione non limitati.

Qui illustriamo l'argomento con qualche esempio che è significativo anche per il seguito e che fornisce un utile strumento per affrontare, se necessario anche gli altri casi.

Consideriamo pertanto una funzione f definita su \mathbb{R}^2 limitata ed integrabile su ogni insieme limitato e chiuso di \mathbb{R}^2 (ad esempio continua) e sia $D \subset \mathbb{R}^2$ un sottoinsieme non limitato di \mathbb{R}^2 .

In tali condizioni non è lecito definire

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

in senso proprio, tuttavia possiamo procedere come segue:

Innanzitutto assicuriamoci di poter lavorare con una funzione sempre positiva; se $f \geq 0$ nulla è da fare ma se così non è basta definire

$$f_+(x, y) = \max\{f(x, y), 0\} \quad \text{e} \quad f_-(x, y) = \min\{f(x, y), 0\}$$

osservare che

$$f = f_+ + f_-$$

e calcolare

$$\iint_D f(x, y) dx dy = \iint_{D_+} f_+(x, y) dx dy + \iint_{D_-} f_-(x, y) dx dy$$

dove

$$D_+ = \{(x, y) \in D : f(x, y) \geq 0\}, \quad D_- = \{(x, y) \in D : f(x, y) \leq 0\}$$

e chiedere che entrambi gli integrali a secondo membro esistano e non diano luogo ad una forma indeterminata.

Supponiamo quindi che $f \geq 0$ e consideriamo una successione di insiemi D_n soddisfacente le seguenti condizioni:

- D_n è chiuso e limitato
- $D_{n+1} \supset D_n$
- per ogni insieme limitato e chiuso K contenuto in D si può trovare un $D_{\bar{n}}$ tale che $D_{\bar{n}} \supset K$

È evidente che le condizioni sopra elencate esprimono il concetto che la successione di domini D_n riempie, invade, l'insieme D ed infatti una successione che soddisfa tali condizioni si chiama successione di domini invadenti D .

DEFINIZIONE 26.1. Se $f \geq 0$ e se D_n è una successione di domini invadenti D allora definiamo

$$\iint_D f(x, y) dx dy = \lim_n \iint_{D_n} f(x, y) dx dy$$

Si può dimostrare che, qualora il limite esista, è indipendente dalla successione di domini invadenti usata.

1. Qualche esempio

Consideriamo il problema di calcolare

$$\iint_{\mathbb{R}^2} e^{-(x^2+y^2)/2} dx dy$$

Definiamo

$$D_n = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq n^2\}$$

allora D_n è una successione di domini invadenti \mathbb{R}^2 e quindi

$$\begin{aligned} \iint_{\mathbb{R}^2} e^{-(x^2+y^2)/2} dx dy &= \lim_n \iint_{D_n} e^{-(x^2+y^2)/2} dx dy = \\ &= \lim_n \int_0^{2\pi} \left(\int_0^n \rho e^{-(\rho^2)/2} d\rho \right) d\theta = \\ &= \lim_n 2\pi \int_0^n \rho e^{-(\rho^2)/2} d\rho = 2\pi \lim_n \left. -e^{-(\rho^2)/2} \right|_0^n = 2\pi \lim_n (1 - e^{-(n^2)/2}) = 2\pi \end{aligned}$$

Il risultato appena ricavato ha una conseguenza interessante, infatti, poiché il valore dell'integrale non dipende dalla successione di domini invadenti usata, possiamo rifare il calcolo anche usando la successione definita da

$$Q_n = [-n, n] \times [-n, n]$$

ed otterremo lo stesso risultato.

Avremo

$$\begin{aligned}
2\pi &= \iint_{\mathbb{R}^2} e^{-(x^2+y^2)/2} dx dy = \lim_n \iint_{Q_n} e^{-(x^2+y^2)/2} dx dy = \\
&= \lim_n \int_{-n}^n \left(\int_{-n}^n e^{-(x^2)/2} e^{-(y^2)/2} dx \right) dy = \\
&= \lim_n \left(\int_{-n}^n e^{-(y^2)/2} dy \right) \left(\int_{-n}^n e^{-(x^2)/2} dx \right) = \lim_n \left(\int_{-n}^n e^{-(t^2)/2} dt \right)^2 = \left(\int_{-\infty}^{+\infty} e^{-(t^2)/2} dt \right)^2
\end{aligned}$$

e possiamo affermare che

$$\int_{-\infty}^{+\infty} e^{-(t^2)/2} dt = \sqrt{2\pi}$$

e che

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-(t^2)/2} dt = 1$$

CAPITOLO 27

LA SOMMA DI INFINITI TERMINI: LE SERIE.

Il problema di sommare un numero non finito di quantità numeriche è stato per lungo tempo considerato privo di senso, ma è giustificabile facilmente, anche dal punto di vista intuitivo, non appena si consideri il seguente esempio.

Sia $I = [0, 1]$ e consideriamo una successione di intervalli così definita: poniamo

$$\begin{aligned} I_1 &= [0, 1/2] \\ I_2 &= [1/2, 1/4] \\ I_3 &= [1/4, 1/8] \\ I_4 &= [1/8, 3/4] \\ &\dots\dots\dots \end{aligned}$$

E' ovvio che

$$\cup I_k = [0, 1]$$

ed inoltre la lunghezza del segmento I_k è data da

$$\ell(I_k) = 1/2^k$$

Pertanto

$$1 = \ell([0, 1]) = \sum_{k=1}^{+\infty} \frac{1}{2^k}$$

Possiamo cercare di puntualizzare il concetto di somma infinita mediante la seguente definizione

DEFINIZIONE 27.1. *Sia a_k una successione di numeri reali e definiamo*

$$S_n = \sum_{k=1}^n a_k$$

Se $\lim S_n$ esiste finito, diciamo che

$$\sum_{k=1}^{+\infty} a_k = S = \lim S_n$$

In tal caso si dice che $\sum_{k=1}^{+\infty} a_k$ è una serie convergente che ha per somma S .

Se $\lim S_n = +\infty (-\infty)$ diciamo che $\sum_{k=1}^{+\infty} a_k$ è una serie positivamente (negativamente) divergente.

Se $\lim S_n$ non esiste diciamo che la serie non è determinata.

Consideriamo ora qualche esempio importante di serie

Sia $x \in \mathbb{R}$ possiamo considerare $a_n = x^n$ e avremo

$$S_n = \sum_{k=0}^n a_k = \sum_{k=0}^n x^k = 1 + x + x^2 + x^3 + \dots + x^n$$

Se osserviamo che

$$xS_n = x + x^2 + x^3 + x^4 + \dots + x^{n+1}$$

si ottiene

$$(1 - x)S_n = 1 - x^{n+1}$$

e, per $x \neq 1$,

$$S_n = \frac{1 - x^{n+1}}{1 - x}$$

Di qui si vede che

- se $|x| < 1$ $\lim S_n = \frac{1}{1-x}$
- se $x \geq 1$ $\lim S_n = +\infty$
- se $x \leq -1$ $\lim S_n$ non esiste.

Pertanto

$$\sum_{k=0}^{+\infty} x^k = \frac{1}{1-x} \quad \text{se } |x| < 1$$

mentre per i restanti valori di x la serie è divergente o indeterminata.

$\sum x^k$ si chiama **serie geometrica di ragione x** .

Possiamo ottenere facilmente altri esempi di serie convergenti. usando la formula di Taylor.

consideriamo lo sviluppo di McLaurin della funzione e^x

$$e^x - 1 - x - \frac{x^2}{2!} - \frac{x^3}{3!} - \dots - \frac{x^n}{n!} = R_{n+1}(x)$$

dove il resto R_{n+1} si può esprimere nella forma di Lagrange mediante la

$$|R_{n+1}(x)| \leq e^{|x|} \frac{|x|^{n+1}}{(n+1)!}$$

Pertanto, se definiamo

$$S_n = \sum_{k=0}^n \frac{x^k}{k!}$$

si ha

$$|e^x - S_n| \leq |R_{n+1}(x)| \leq e^{|x|} \frac{|x|^{n+1}}{(n+1)!}$$

e tenendo conto che

$$\lim \frac{|x|^{n+1}}{(n+1)!} = 0 \text{ per ogni } x \in \mathbb{R}$$

si ha

$$e^x = \lim S_n = \sum_{k=0}^{+\infty} \frac{x^k}{k!}$$

In maniera del tutto analoga si prova che

$$\sin(x) = \sum_{k=0}^{+\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!} \forall x \in \mathbb{R}$$

$$\cos(x) = \sum_{k=0}^{+\infty} (-1)^k \frac{x^{2k}}{(2k)!} \forall x \in \mathbb{R}$$

$$\ln(1+x) = \sum_{k=1}^{+\infty} (-1)^{k-1} \frac{x^k}{k} \forall x \in [-1/2, 1]$$

ELEMENTI DI PROBABILITÀ E STATISTICA.

Il calcolo delle probabilità nasce agli inizi del '700 per rispondere alle istanze prodotte dal gioco d'azzardo.

Il suo scopo era inizialmente introdurre una stima numerica dei rischi ad esso connessi e successivamente lo stesso metodo fu esteso per studiare i fenomeni caratterizzati da elementi di incertezza.

Vediamo innanzi tutto di stabilire quali sono gli elementi necessari per parlare di probabilità

Occorre innanzi tutto considerare una famiglia di eventi che costituiscono lo spazio su cui definiremo la probabilità.
Possiamo ad esempio associare all'idea di evento quella di sottoinsieme di un insieme dato che chiameremo evento certo.

1. Spazi di probabilità

Consideriamo, ad esempio, il più semplice tra i giochi d'azzardo, e cioè il lancio di una moneta, possiamo dire che gli eventi possibili sono

- L'uscita di Testa $\{T\}$
- L'uscita di Croce $\{C\}$

mentre l'insieme $\mathcal{U} = \{T, C\}$ costituisce l'evento certo.

In questo caso è naturale definire la probabilità degli eventi che entrano in gioco:

$$(28.1) \quad \mathcal{P}(T) = \frac{1}{2}$$

$$(28.2) \quad \mathcal{P}(C) = \frac{1}{2}$$

$$(28.3) \quad \mathcal{P}(\mathcal{U}) = \mathcal{P}(T) + \mathcal{P}(C) = 1$$

$$(28.4) \quad \mathcal{P}(\emptyset) = 0$$

giustificando la definizione con il fatto che su due possibili uscite una sola è favorevole nel caso si consideri T o C , mentre entrambe vanno bene nel caso si consideri l'unione di T e C . È ovvio che con ciò supponiamo che T e C si presentino con ugual frequenza, cioè che la moneta sia non truccata.

Un secondo esempio di spazio di probabilità si può costruire considerando il caso del lancio di due dadi.

Se le facce sono numerate, come al solito, da 1 a 6 possiamo identificare l'esito del lancio con la coppia di numeri (i, j) (punteggio) che si leggono sulla faccia superiore del primo e del secondo dado.

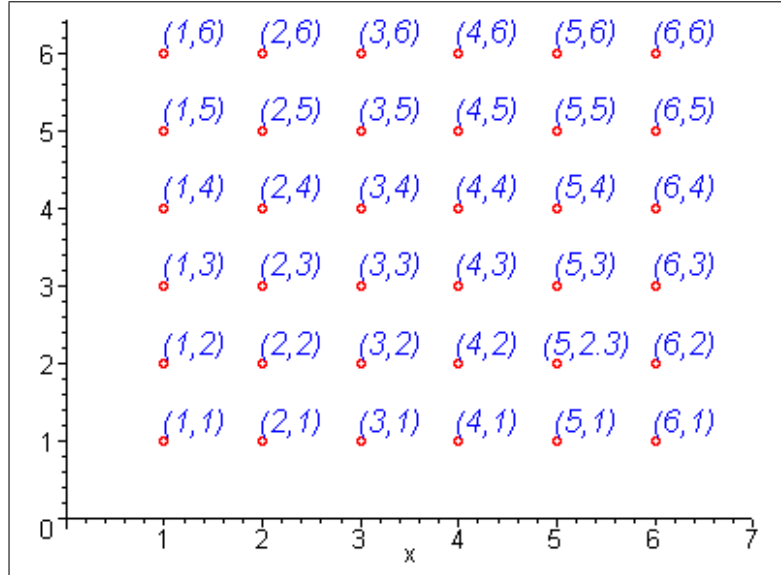


FIGURA 28.1. Lo spazio \mathcal{U} degli eventi nel caso del lancio di due dadi

In tal modo possiamo identificare ciascuna delle 36 possibili uscite (eventi) con il punto del piano cartesiano di coordinate (i, j) ; indicheremo tale evento con il simbolo $A_{i,j}$, (si veda la figura 28.1).

Poichè nel caso di dadi non truccati ogni evento è equiprobabile possiamo affermare che la probabilità di $A_{i,j}$ è data da

$$\mathcal{P}(A_{i,j}) = \frac{1}{36}$$

Ovviamente possiamo considerare anche altri eventi, ad esempio possiamo cercare di stimare la probabilità che si presenti l'evento $A_{1,4}$ oppure l'evento $A_{5,6}$; possiamo indicare il nuovo evento come

$$B = A_{1,4} \cup A_{5,6} = \{A_{1,4}, A_{5,6}\}$$

ed è ragionevole stimare che, poichè accettiamo 2 eventi su 36 possibili,

$$\mathcal{P}(B) = \frac{2}{36} = \frac{1}{36} + \frac{1}{36}$$

Se vogliamo stimare la probabilità che si presenti uno qualunque degli eventi $A_{i,j}$, cioè se vogliamo stimare la probabilità che si presenti l'evento

$$\mathcal{U} = \{A_{i,j} : i, j = 1..6\}$$

poichè accettiamo 36 possibilità su 36 possiamo dire che

$$\mathcal{P}(\mathcal{U}) = 1$$

e che \mathcal{U} è l'evento certo.

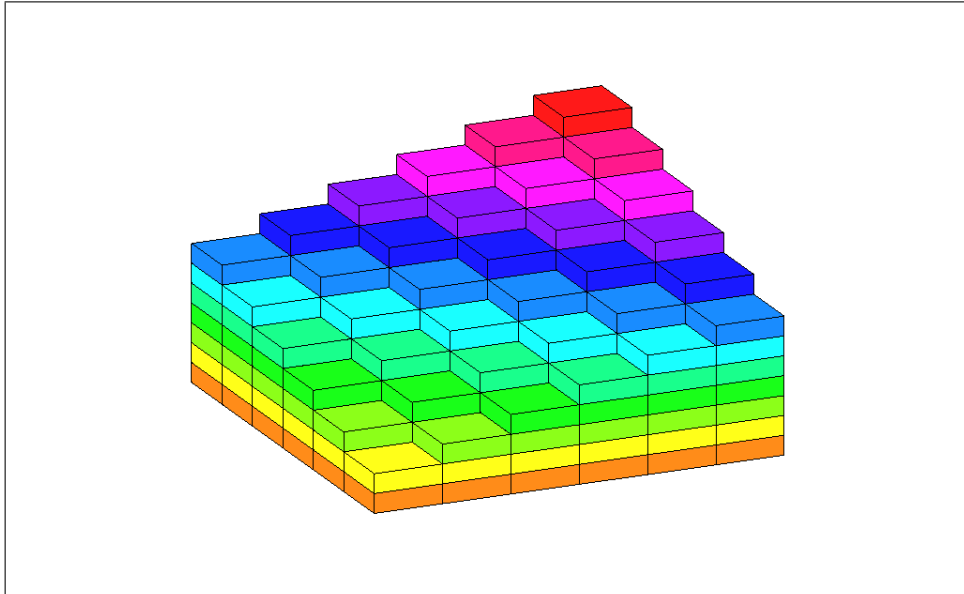


FIGURA 28.2. Istogramma relativo al lancio di due dadi, (supposti distinti)

In generale possiamo parlare di spazio di probabilità finito se è assegnata una famiglia di eventi

$$\mathcal{F} = \{A_i : i = 1..N\}$$

soddisfacente le seguenti condizioni

- $\emptyset \in \mathcal{F}$
- $A_1, A_2 \in \mathcal{F} \implies A_1 \cup A_2 \in \mathcal{F}$

ed è assegnata una funzione

$$\mathcal{P} : \mathcal{F} \rightarrow \mathbb{R}$$

che associa ad ogni $A \in \mathcal{F}$ un valore reale $\mathcal{P}(A)$, che chiamiamo probabilità che l'evento accada, soddisfacente le seguenti proprietà:

- $\mathcal{P}(A) \geq 0$
- se $\mathcal{U} = \bigcup_{A \in \mathcal{F}} A \in \mathcal{F}$, si ha $\mathcal{P}(\mathcal{U}) = 1$
- se $A_1, A_2 \in \mathcal{F}$ sono tali che $A_1 \cap A_2 = \emptyset$ allora

$$\mathcal{P}(A_1 \cup A_2) = \mathcal{P}(A_1) + \mathcal{P}(A_2)$$

Osserviamo che, dal momento che la famiglia dei possibili eventi è finita, avremo che

$$\mathcal{U} = \bigcup_{A \in \mathcal{F}} A \in \mathcal{F}$$

\mathcal{U} rappresenta l'evento certo ed anche l'ambiente in cui si individuano gli eventi.

Per assegnare uno spazio di probabilità, quindi, è necessario assegnare un insieme \mathcal{U} una famiglia \mathcal{F} di sottoinsiemi di \mathcal{U} ed una funzione P definita su \mathcal{F} a valori in \mathbb{R} .

Ci riferiremo quindi ad uno spazio di probabilità come ad un terna $(\mathcal{U}, \mathcal{F}, \mathcal{P})$

Nel caso dell'esempio precedente la famiglia \mathcal{F} è costituita da ciascuno degli eventi (i, j) che abbiamo rappresentato nel piano con i punti $A_{i,j}$ e dalle unioni finite di essi.

Ad esempio

$$B = \{A_{1,3}, A_{4,1}\} = A_{1,3} \cup A_{4,1}$$

è l'evento che accade se almeno una tra le uscite $A_{1,3}$ e $A_{4,1}$ accade.

Poichè le uscite del lancio dei dadi sono ritenute equiprobabili possiamo affermare che $\mathcal{P}(A_{i,j}) = \frac{1}{36}$ per ogni possibile uscita $A_{i,j}$ e quindi, ad esempio,

$$\mathcal{P}(B) = \mathcal{P}(A_{1,3} \cup A_{4,1}) = \mathcal{P}(A_{1,3}) + \mathcal{P}(A_{4,1}) = \frac{1}{36} + \frac{1}{36} = \frac{2}{36}$$

Ovviamente

$$\mathcal{P}(U) = \mathcal{P}\left(\bigcup_{A \in \mathcal{F}} A\right) = 36 \frac{1}{36} = 1$$

Come abbiamo già osservato, ci riferiamo ad uno spazio di probabilità come ad una terna $(\mathcal{U}, \mathcal{F}, \mathcal{P})$ in cui \mathcal{U} è lo spazio di probabilità, \mathcal{F} è la famiglia di tutti gli eventi e \mathcal{P} è la misura di probabilità definita su \mathcal{U} .

Si possono provare i seguenti risultati:

- $0 \leq \mathcal{P}(A) \leq 1$ per ogni $A \in \mathcal{F}$
- $\mathcal{P}(\emptyset) = 0$
- Se $A_1, A_2 \in \mathcal{F}$ $A_1 \subset A_2$ allora

$$\mathcal{P}(A_1) \subset \mathcal{P}(A_2)$$

ed inoltre

$$\mathcal{P}(A_2 \setminus A_1) = \mathcal{P}(A_2) - \mathcal{P}(A_1)$$

- $\mathcal{P}(A^c) = 1 - \mathcal{P}(A)$ per ogni $A \in \mathcal{F}$
- $\mathcal{P}(A \cup B) = \mathcal{P}(A) + \mathcal{P}(B) - \mathcal{P}(A \cap B)$
- Se $B = \bigcup_{i=1}^N A_i$ e se $A_i \cap A_j = \emptyset$ allora

$$\mathcal{P}(B) = \sum_{i=1}^N \mathcal{P}(A_i)$$

- $\mathcal{P}(A) = \mathcal{P}(A \cap B) + \mathcal{P}(A \cap B^c)$
- se $A_i \cap A_j = \emptyset$, ed esiste i_0 tale che $A \subset A_{j_0}$, allora

$$(28.5) \quad \boxed{\mathcal{P}(A) = \mathcal{P}(A \cap A_1) + \mathcal{P}(A \cap A_2) + \dots + \mathcal{P}(A \cap A_N)}$$

DEFINIZIONE 28.1. Se $A, B \in \mathcal{F}$ definiamo probabilità di A condizionata a B e la denotiamo con $\mathcal{P}(A|B)$ il valore

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

La probabilità di A condizionata a B , $\mathcal{P}(A|B)$, definisce la probabilità di accadimento di A nel caso sia accaduto B

Naturalmente si ha

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

Nel caso in cui

$$\mathcal{P}(A|B) = \mathcal{P}(A)$$

diciamo che A e B sono eventi indipendenti, (la probabilità di accadimento di A non è cambiata dal fatto che B è accaduto).

In tal caso si ha

$$\mathcal{P}(A \cap B) = \mathcal{P}(A)\mathcal{P}(B)$$

Si può dimostrare che

Se $A_i \cap A_j = \emptyset$, ed esiste i_0 tale che $A \subset A_{j_0}$, allora

$$\mathcal{P}(A) = \mathcal{P}(A_1)\mathcal{P}(A|A_1) + \mathcal{P}(A_2)\mathcal{P}(A|A_2) + \dots \mathcal{P}(A_N)\mathcal{P}(A|A_N)$$

e ne segue il seguente teorema di Bayes

TEOREMA 28.1. - di Bayes - Se $A_1, A_2, \dots, A_N \in \mathcal{F}$ sono eventi tali che $A_i \cap A_j = \emptyset$ e $\bigcup_{i=1}^N A_i$ allora

$$\mathcal{P}(A_k|B) = \frac{\mathcal{P}(A_k)\mathcal{P}(A|A_k)}{\sum_{i=1}^N \mathcal{P}(A_i)\mathcal{P}(A|A_i)}$$

Per capire perchè il teorema sia vero, consideriamo il caso in cui $N = 2$; allora

- $A_1 \cup A_2 \supset A$
- $A_1 \cap A_2 = \emptyset$
- $\mathcal{P}(B) = \mathcal{P}(B \cap (A_1 \cup A_2)) = \mathcal{P}(B \cap A_1) + \mathcal{P}(B \cap A_2)$

e

$$\mathcal{P}(A_1|B) = \frac{\mathcal{P}(A_1 \cap B)}{\mathcal{P}(B)}$$

$$\mathcal{P}(B|A_1) = \frac{\mathcal{P}(B \cap A_1)}{\mathcal{P}(A_1)}$$

da cui

$$\mathcal{P}(A_1|B)\mathcal{P}(B) = \mathcal{P}(B \cap A_1) = \mathcal{P}(B|A_1)\mathcal{P}(A_1)$$

Ne segue che

$$\mathcal{P}(A_1|B) = \frac{\mathcal{P}(B \cap A_1)}{\mathcal{P}(B)} = \frac{\mathcal{P}(B \cap A_1)\mathcal{P}(A_1)}{\mathcal{P}(B \cap A_1)\mathcal{P}(A_1) + \mathcal{P}(B \cap A_2)\mathcal{P}(A_2)}$$

2. Qualche applicazione della regola di Bayes.

2.1. Si consideri una popolazione in cui è diffusa una malattia e si supponga di voler sottoporre l'intera popolazione ad un test con lo scopo di determinare per ciascuna persona se è affetta o no dalla malattia.

Una situazione di questo genere si presenterebbe se si considerasse l'opportunità di procedere ad uno screening di massa per rivelare, ad esempio, la diffusione dell'AIDS.

In tal caso sarebbe ragionevole supporre che:

- Nella popolazione 1 persona su 1000 è affetta dalla malattia (**incidenza della malattia**).
- Il test usato per rivelare la malattia è positivo (e quindi indica la presenza della malattia) in 95 casi per ogni 100 persone malate esaminate (**sensività del test**).
- Il test usato per rivelare la malattia è positivo (e quindi indica la presenza della malattia) in 1 caso per ogni 100 persone sane esaminate (**specificità del test**).

Indichiamo con

- A l'evento "infetto"
- A' l'evento "sano"
- T_p l'evento "test positivo"
- T_n l'evento "test negativo"

Indichiamo inoltre con

- a l'incidenza della malattia (nel caso prima citato $a = 1/1000 = 0.001$)
- p la sensitività del test (nel caso prima citato $p = 95/100 = 0.95$)
- q la specificità del test (nel caso prima citato $q = 1/100 = 0.01$)

Possiamo facilmente calcolare che

$$\begin{aligned}
\mathcal{P}(A) &= a & \mathcal{P}(A') &= 1 - a \\
\mathcal{P}(T_p|A) &= p & \mathcal{P}(T_n|A) &= 1 - p \\
\mathcal{P}(T_p|A') &= q & \mathcal{P}(T_n|A') &= 1 - q
\end{aligned}$$

da cui

$$\begin{aligned}
\mathcal{P}(T_p) &= \mathcal{P}(T_p \cap A) + \mathcal{P}(T_p \cap A') = \\
&= \mathcal{P}(T_p|A)\mathcal{P}(A) + \mathcal{P}(T_p|A')\mathcal{P}(A') = pa + q(1 - a)
\end{aligned}$$

Pertanto, applicando il teorema di Bayes

$$\mathcal{P}(A|T_p) = \frac{\mathcal{P}(T_p|A)\mathcal{P}(A)}{\mathcal{P}(T_p|A)\mathcal{P}(A) + \mathcal{P}(T_p|A')\mathcal{P}(A')} = \frac{pa}{pa + q(1 - a)}$$

Se utilizziamo i valori di incidenza sensitività e specificità prima introdotti, possiamo ricavare che:

$$\mathcal{P}(A|T_p) = \frac{pa}{pa + q(1 - a)} = 0.0874$$

e questo significa che la probabilità che una persona in cui il test ha dato esito positivo sia effettivamente infetto è inferiore al 9%.

Se ne deduce la poca convenienza ad effettuare uno screening di massa.

Vale la pena osservare che neppure test più precisi consentono di raggiungere risultati più significativi.

Infatti se supponiamo di migliorare il test in modo da avere la certezza di individuare la malattia nelle persone infette, cioè se supponiamo la sensitività $p = 100/100 = 1$, mantenendo $a = 0.001$ e $q = 0.01$, otteniamo

$$\mathcal{P}(A|T_p) = \frac{pa}{pa + q(1 - a)} = 0.091$$

In realtà aumentando la sensitività, la specificità necessariamente diminuisce; se supponiamo che

$$a = 0.001 \quad , \quad p = 1 \quad , \quad q = 0.002$$

la situazione peggiora e si ha addirittura

$$\mathcal{P}(A|T_p) = \frac{pa}{pa + q(1 - a)} = 0.05$$

Solo migliorando la specificità senza peggiorare la sensitività si ottiene qualcosa di meglio; infatti se

$$a = 0.001 \quad , \quad p = 0.95 \quad , \quad q = 1/300 = 0.0034$$

si ha

$$\mathcal{P}(A|T_p) = \frac{pa}{pa + q(1 - a)} = 0.22$$

Un miglioramento decisivo nell'affidabilità dei risultati si ottiene invece nel caso in cui l'incidenza della malattia sia alta; infatti per

$$a = 0.01 \quad , \quad p = 0.95 \quad , \quad q = 0.01$$

si ha

$$\mathcal{P}(A|T_p) = \frac{pa}{pa + q(1-a)} = 0.5$$

e addirittura

$$\mathcal{P}(A|T_p) = \frac{pa}{pa + q(1-a)} = 0.9$$

nel caso in cui l'incidenza salga ad $a = .1$.

2.2. Un secondo interessante esempio si ha considerando la seguente situazione.

Si supponga che un capo di governo debba prendere una decisione di politica economica e disponga di tre consiglieri che identifichiamo con

$A \quad B \quad C$

Si supponga di avere una stima della loro affidabilità descritta come segue:

- la probabilità che il consigliere A abbia un'opinione corretta è $1/6$,
- la probabilità che il consigliere B abbia un'opinione corretta è $1/3$,
- la probabilità che il consigliere C abbia un'opinione corretta è $1/2$.

Indichiamo con

A_r l'evento "il consigliere A ha espresso un'opinione corretta",

B_r l'evento "il consigliere B ha espresso un'opinione corretta",

C_r l'evento "il consigliere C ha espresso un'opinione corretta".

Esprimiamo l'affidabilità dei consiglieri assegnando agli eventi A_r, B_r e C_r opportuni valori di probabilità

$$\mathcal{P}(A_r) = 1/6 \quad \mathcal{P}(B_r) = 1/3 \quad \mathcal{P}(C_r) = 1/2$$

Il capo di governo chiede ai suoi consiglieri una previsione sulle conseguenze che la sua decisione avrà sul tasso di disoccupazione, a distanza di un anno.

Al quesito i tre consiglieri rispondono secondo i valori descritti nella tabella che segue, dove nella colonna D sono riportate le probabilità che la disoccupazione diminuisca, nella colonna S che rimanga stabile e nella colonna I che il tasso di disoccupazione aumenti.

	D	S	I
A	1/10	1/10	8/10
B	6/10	2/10	2/10
C	2/10	6/10	2/10

TABELLA 28.1. Previsioni dei consiglieri sul tasso di disoccupazione.

Trascorso un anno si rileva che il tasso di disoccupazione è aumentato; alla luce di questo fatto le percentuali di affidabilità dei tre consiglieri vanno riviste e ciò può essere fatto ridefinendo la probabilità di A_r , B_r e C_r alla luce del fatto che l'evento I si è verificato.

Dovremo, in altre parole valutare

$$\mathcal{P}(A_r|I) = \frac{\mathcal{P}(I|A_r)\mathcal{P}(A_r)}{\mathcal{P}(I)}$$

$$\mathcal{P}(B_r|I) = \frac{\mathcal{P}(I|B_r)\mathcal{P}(B_r)}{\mathcal{P}(I)}$$

$$\mathcal{P}(C_r|I) = \frac{\mathcal{P}(I|C_r)\mathcal{P}(C_r)}{\mathcal{P}(I)}$$

ma

$$\begin{aligned}\mathcal{P}(I) &= \mathcal{P}(I \cap A_r) + \mathcal{P}(I \cap B_r) + \mathcal{P}(I \cap C_r) = \\ &= \mathcal{P}(I|A_r)\mathcal{P}(A_r) + \mathcal{P}(I|B_r)\mathcal{P}(B_r) + \mathcal{P}(I|C_r)\mathcal{P}(C_r)\end{aligned}$$

Ora, la probabilità che avvenga I ammesso che un consigliere abbia ragione si legge nell'ultima colonna della tabella, per cui

$$\mathcal{P}(I|A_r) = 8/10$$

$$\mathcal{P}(I|B_r) = 2/10$$

$$\mathcal{P}(I|C_r) = 2/10$$

per cui

$$P(I) = \frac{8}{10} \frac{1}{6} + \frac{2}{10} \frac{1}{3} + \frac{2}{10} \frac{1}{2} = \frac{3}{10}$$

Ne viene che le nuove percentuali di affidabilità sono

$$\mathcal{P}(A_r|I) = \frac{\mathcal{P}(I|A_r)\mathcal{P}(A_r)}{\mathcal{P}(I)} = \frac{\frac{8}{10} \frac{1}{6}}{\frac{3}{10}} = \frac{4}{9}$$

$$\mathcal{P}(B_r|I) = \frac{\mathcal{P}(I|B_r)\mathcal{P}(B_r)}{\mathcal{P}(I)} = \frac{\frac{2}{10} \frac{1}{3}}{\frac{3}{10}} = \frac{2}{9}$$

$$\mathcal{P}(C_r|I) = \frac{\mathcal{P}(I|C_r)\mathcal{P}(C_r)}{\mathcal{P}(I)} = \frac{\frac{2}{10} \frac{1}{2}}{\frac{3}{10}} = \frac{1}{3}$$

3. Qualche richiamo di calcolo combinatorio.

Per studiare un po' di probabilità discreta è utile conoscere qualche elemento di calcolo combinatorio.

Il calcolo combinatorio si occupa di stabilire il numero delle possibili uscite di semplici esperimenti; si fonda essenzialmente sul principio seguente:

Se un esperimento ha n_1 possibili esiti, un secondo esperimento ha n_2 possibili esiti, un terzo esperimento ha n_3 possibili esiti, allora il numero dei possibili esiti della sequenza dei tre esperimenti è

$$n_1 n_2 n_3$$

Le più comuni conseguenze di questo principio portano a un certo numero di definizioni che descriviamo brevemente.

3.1. Disposizioni di n elementi a k a k . Parliamo di **disposizioni** (o anche, se $k = n$, di **permutazioni**) di n elementi a k a k quando consideriamo i gruppi di k elementi scelti tra gli n dati.

Riteniamo due gruppi distinti se differiscono per un elemento o per l'ordine con cui gli elementi sono scelti.

Indichiamo con

$${}_n D_k$$

il numero delle disposizioni di n elementi a k a k .

Poichè per il primo elemento di ciascun gruppo abbiamo n scelte, per il secondo ne abbiamo $(n - 1)$ per il terzo ne abbiamo $(n - 2)$ e così via, possiamo calcolare che

$${}_n D_k = n(n - 1)(n - 2)(n - 3) \dots (n - (k - 1)) = \frac{n!}{(n - k)!}$$

Il numero delle disposizioni di n elementi ad n ad n , cioè delle permutazioni, risulta

$$P_n = {}_n D_n = n!$$

Qualora gli n elementi da cui si sceglie presentino sottogruppi di elementi uguali, siano $n_1, n_2, n_3, \dots, n_k$ le permutazioni risultano in numero di

$$\frac{n!}{n_1! n_2! n_3! \dots n_k!}$$

3.2. Combinazioni di n elementi a k a k . Parliamo di **combinazioni** di n elementi a k a k quando consideriamo gruppi di k elementi scelti tra gli n dati senza distinguerli in base all'ordine degli elementi.

Indichiamo con

$${}_nC_k$$

il numero delle combinazioni di n elementi a k a k .

Poichè un gruppo di k elementi ammette $k!$ diversi modi possiamo calcolare che

$${}_nC_k = \frac{{}_nD_k}{k!} = \frac{n(n-1)(n-2)(n-3)\dots(n-(k-1))}{k!} = \frac{n!}{k!(n-k)!}$$

Il numero ${}_nC_k$ si chiama coefficiente binomiale e si indica con

$${}_nC_k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Ricordiamo che i coefficienti binomiali possono essere ricavati dal triangolo di Tartaglia e che trovano una importante applicazione nella formula del binomio di Newton.

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}$$

3.3. Campioni ordinati. È anche utile ricordare qualche formula per stimare il numero di possibili campioni estratti da una popolazione.

Per aiutarci assimiliamo la popolazione ad un'un'urna piena di palline e l'estrazione degli elementi del campione all'estrazione delle palline dall'urna.

Possiamo operare un campionamento con ripetizione estraendo una pallina, osservandola e rimettendola nell'urna dopo aver annotato l'informazione relativa.

In tal caso, se operiamo k estrazioni, avremo

$$\overbrace{nnn\dots n}^{k \text{ volte}} = n^k$$

possibili uscite in quanto per ogni elemento estratto avremo sempre n possibili scelte.

Possiamo anche operare un campionamento senza ripetizione, estraendo, osservando e non rimettendo la pallina nell'urna; in tal caso per la prima estrazione avremo n possibilità, per la seconda $n-1$, per la terza $n-2$ e così via. Pertanto in questo caso avremo

$${}_nD_k = \frac{n!}{(n-k)!}$$

possibili uscite.

4. Le variabili aleatorie

Se è assegnato uno spazio di probabilità $(\mathcal{U}, \mathcal{F}, \mathcal{P})$ diciamo che ξ è una variabile aleatoria definita su $(\mathcal{U}, \mathcal{F}, \mathcal{P})$ se è data una funzione

$$\xi : \mathcal{U} \rightarrow \mathbb{R}$$

Nel caso del lancio di due dadi la terna $(\mathcal{U}, \mathcal{F}, \mathcal{P})$ è definita da

- \mathcal{U} è l'insieme delle 36 coppie di valori (i, j)
- \mathcal{F} è la famiglia di tutti i sottoinsiemi di \mathcal{U}
- \mathcal{P} è definita da $\mathcal{P}(A_{i,j}) = \frac{1}{36}$ ed inoltre se $A \in \mathcal{F}$ è un sottoinsieme costituito da k elementi possiamo definire $\mathcal{P}(A) = k \frac{1}{36}$

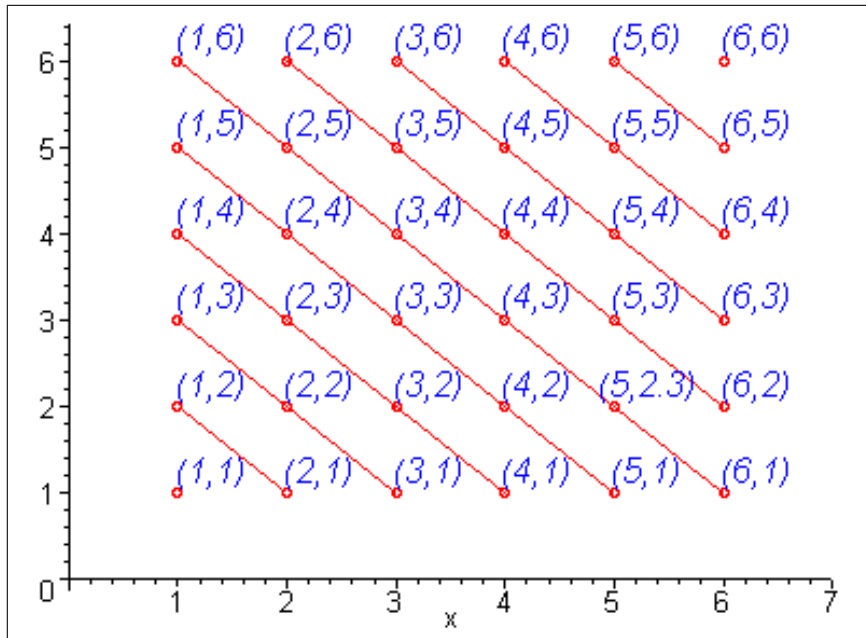


FIGURA 28.3. Eventi che forniscono lo stesso valore per ξ , (congiunti da un segmento)

Possiamo costruire un esempio di variabile aleatoria sullo spazio di probabilità associato al lancio di due dadi assegnando ad ogni uscita $(i, j) \in \mathcal{U}$ il valore ottenuto sommando i punteggi; in altre parole

$$\xi(A_{i,j}) = i + j$$

Possiamo osservare che i punti (eventi) che nella figura 28.3 sono congiunti da un segmento di retta forniscono lo stesso valore per la variabile aleatoria ξ e possiamo riassumere i valori assunti dalla variabile aleatoria ξ nella seguente tabella, dove all'incrocio della riga i -esima e della colonna j -esima possiamo leggere il valore $i + j$ assunto da ξ .

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

TABELLA 28.2. I valori assunti dalla Variabile aleatoria ξ

Possiamo ora definire alcuni elementi caratteristici di una variabile aleatoria:

DEFINIZIONE 28.2. Se ξ è una variabile aleatoria discreta definita su uno spazio di probabilità $(\mathcal{U}, \mathcal{F}, \mathcal{P})$ e se

$$\mathcal{U} = \{A_1, A_2, \dots, A_n\}$$

,

- la **media** μ di ξ è definita da

$$\mu = \sum_i \xi(A_i) \mathcal{P}(A_i)$$

- la **varianza** σ^2 di ξ è definita da

$$\sigma^2 = \text{Var}(\xi) = \sum_i (\xi(A_i) - \mu)^2 \mathcal{P}(A_i)$$

- lo **sarto quadratico medio** di ξ è definito da

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

- la **moda** M di ξ è definita da

$$M = \xi(A_i)$$

dove A_i è tale che

$$\mathcal{P}(A_i) = \max_j \mathcal{P}(A_j)$$

- la **mediana** m di ξ è definita da

$$\mathcal{P}(\xi \leq m) = \mathcal{P}(\xi \geq m)$$

- il **momento di ordine** k μ_k di ξ è definito da

$$\mu_k = \sum_i (\xi(A_i) - \mu)^k \mathcal{P}(A_i)$$

- *il momento di ordine k , rispetto all'origine μ'_k di ξ è definito da*

$$\mu'_k = \sum_i (\xi(A_i))^k \mathcal{P}(A_i)$$

DEFINIZIONE 28.3. *Chiamiamo funzione di distribuzione di probabilità della variabile aleatoria ξ la funzione*

$$\varphi : \mathcal{U} \rightarrow \mathbb{R}$$

definita da

$$\varphi(i) = \mathcal{P}(\xi = A_i)$$

Se φ è la funzione distribuzione di probabilità di una variabile aleatoria ξ avremo che

$$\begin{aligned} \mathcal{P}(n \leq \xi \leq m) &= \sum_{i=n}^m \varphi(i) \\ \mathcal{P}(\xi \leq n) &= \sum_{i \leq n} \varphi(i) \end{aligned}$$

DEFINIZIONE 28.4. *Se ξ è una variabile aleatoria discreta la cui densità di probabilità è φ , definiamo **speranza matematica di ξ***

$$E(\xi) = \sum_i \xi(A_i) \varphi(i) = \sum_i \xi(A_i) \mathcal{P}(\xi = A_i) = \mu$$

DEFINIZIONE 28.5. *Se ξ è una variabile aleatoria discreta la cui densità di probabilità è φ e se $g : \mathbb{R} \rightarrow \mathbb{R}$ è una funzione, definiamo una nuova variabile aleatoria che indichiamo con $g(\xi)$ mediante la funzione distribuzione di probabilità*

$$\psi(i) = g(\xi(A_i)) \varphi(i)$$

Definiamo inoltre

$$E(g(\xi)) = \sum_i g(\xi(A_i)) \varphi(i) =$$

Osserviamo che si ha

$$\begin{aligned} \sigma &= E((\xi - \mu)^2) \\ \mu_k &= E((\xi - \mu)^k) \\ \mu'_k &= E(\xi^k) \end{aligned}$$

DEFINIZIONE 28.6. Se ξ è una variabile aleatoria discreta la cui densità di probabilità è φ , definiamo **funzione generatrice dei momenti di ξ** la

$$M_{\xi}(t) = E(e^{t\xi}) = \sum_i e^{t\xi(A_i)} \varphi(i) =$$

4.1. Un esempio. Consideriamo il solito esempio dei dadi e la solita variabile ξ che associa ad ogni uscita in \mathcal{U} la somma dei punteggi, possiamo vedere che

$$\begin{aligned} \varphi(2) &= \mathcal{P}(\xi = 2) = \frac{1}{36} = \mathcal{P}(\xi = 12) = \varphi(12) \\ \varphi(3) &= \mathcal{P}(\xi = 3) = \frac{2}{36} = \mathcal{P}(\xi = 11) = \varphi(11) \\ \varphi(4) &= \mathcal{P}(\xi = 4) = \frac{3}{36} = \mathcal{P}(\xi = 10) = \varphi(10) \\ \varphi(5) &= \mathcal{P}(\xi = 5) = \frac{4}{36} = \mathcal{P}(\xi = 9) = \varphi(9) \\ \varphi(6) &= \mathcal{P}(\xi = 6) = \frac{5}{36} = \mathcal{P}(\xi = 8) = \varphi(8) \\ \varphi(7) &= \mathcal{P}(\xi = 7) = \frac{6}{36} \end{aligned}$$

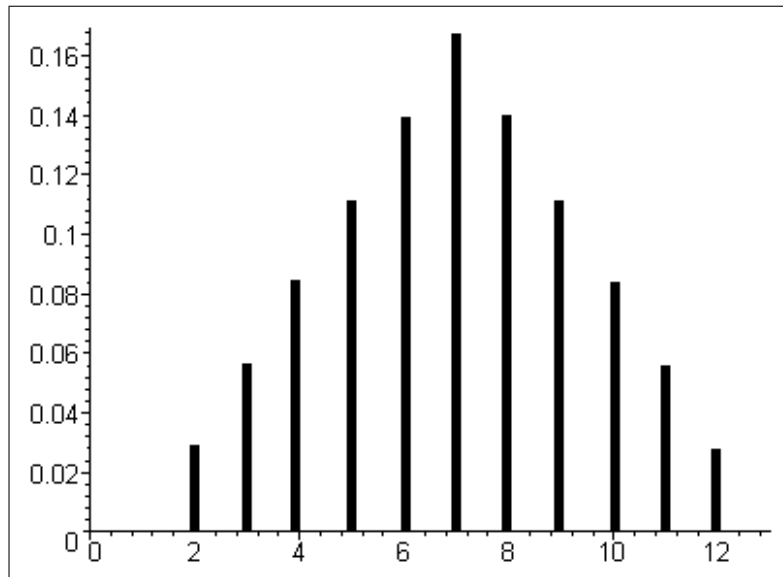


FIGURA 28.4. Distribuzione di probabilità della variabile aleatoria che restituisce il punteggio ottenuto lanciando due dadi.

Possiamo rappresentare questa distribuzioni di probabilità mediante il grafico in figura 28.4 (scatter plot) oppure mediante il grafico in figura 28.5.

Quest'ultimo grafico si chiama istogramma ed è quello che più frequentemente viene adottato per rappresentare la funzione di distribuzione di una variabile aleatoria discreta.

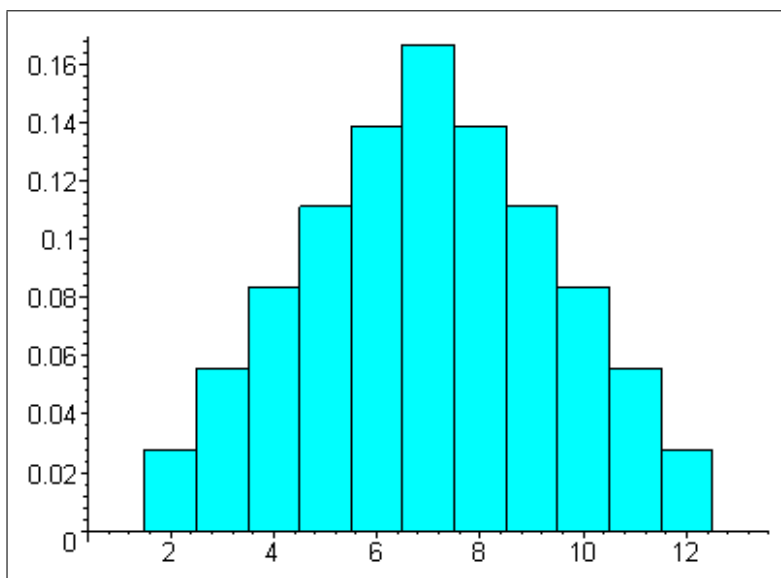


FIGURA 28.5. Istogramma relativo alla variabile aleatoria che restituisce il punteggio ottenuto lanciando due dadi.

Osservando il grafico possiamo osservare che il valore che compare con maggior frequenza è il 7, che pertanto è la moda della variabile ξ .

Possiamo calcolare la probabilità $\mathcal{P}(\xi \leq 7)$ sommando le aree dei rettangoli mostrati in figura (si osservi che la base dei rettangoli è lunga 1) ed otteniamo che $\mathcal{P}(\xi \leq 7) = \frac{21}{36}$.

In modo simile possiamo calcolare $\mathcal{P}(\xi \geq 7) = \frac{21}{36}$ e possiamo concludere che la mediana della variabile aleatoria ξ è $m = 7$.

Un semplice calcolo mostra che la varianza di ξ è $\sigma^2 = \frac{35}{6}$.

Abbiamo già notato che i rettangoli che costituiscono l'istogramma hanno base lunga 1; ciò permette di definire la funzione di distribuzione di ξ come

$$(28.6) \quad \varphi(x) = \begin{cases} \mathcal{P}(\xi = \eta) & \text{per } \eta - 0.5 \leq x \leq \eta + 0.5 \quad \eta = 2, 3, 4, \dots, 12 \\ 0 & \text{altrove} \end{cases}$$

Con tale definizione si vede che φ è definita su tutto \mathbb{R} ed inoltre si vede che

$$\int_{-\infty}^{+\infty} \varphi(t) dt = \sum_i \varphi(i) = 1$$

essendo la somma estesa a tutti i valori η che variabile discreta ξ può assumere.

5. Variabili aleatorie continue

Talvolta tuttavia non è possibile considerare uno spazio di probabilità discreto, finito o numerabile.

Ciò accade ad esempio quando si considera il problema di scegliere un numero a caso compreso tra 0 ed 1.

Infatti la probabilità di estrarre, ad esempio, il valore 0.3 non si può calcolare considerando il rapporto tra casi favorevoli, uno solo, e casi possibili, infiniti non numerabili.

Anche la definizione di media e varianza presentano qualche problema in quanto occorre definire come si intende procedere per calcolare la somma di un numero infinito, non numerabile, di addendi.

Per chiarire la questione possiamo osservare che, se è difficile definire la probabilità che la variabile aleatoria ξ il cui valore è il numero scelto a caso in $[0, 1]$ assuma il valore x , è invece naturale definire la probabilità che $\xi \in [x, x + h]$.

In tal caso infatti possiamo identificare i casi favorevoli con un segmento di lunghezza h e la totalità dei casi con l'intero intervallo $[0, 1]$ che risulta ovviamente di lunghezza 1.

Pertanto

$$\mathcal{P}(x \leq \xi \leq x + h) = \frac{h}{1}$$

Ricordando il significato di somma dell'integrale, possiamo definire la funzione distribuzione di probabilità della variabile aleatoria ξ come la funzione continua φ tale che

$$(28.7) \quad \mathcal{P}(x \leq \xi \leq x + h) = h = \int_x^{x+h} \varphi(t) dt$$

per ogni $x \in [0, 1]$ e per ogni h abbastanza piccolo.

Ne deduciamo che

$$(28.8) \quad \frac{1}{h} \int_x^{x+h} \varphi(t) dt = 1$$

e, passando al limite per $h \rightarrow 0$, poichè abbiamo supposto φ continua,

$$\varphi(x) = 1$$

Da quanto abbiamo detto appare ragionevole che, nel caso di una variabile aleatoria continua ξ , non è significativo definire

$$(28.9) \quad \mathcal{P}(\xi = x)$$

mentre è naturale definire

$$(28.10) \quad \mathcal{P}(x_0 \leq \xi \leq x_1) = \int_{x_0}^{x_1} \varphi(t) dt$$

dove φ è la funzione di distribuzione di probabilità di ξ .

Pertanto supporremo nota una variabile aleatoria continua ξ se è nota la sua funzione di distribuzione di probabilità φ .

Una funzione $\varphi : \mathbb{R} \rightarrow \mathbb{R}$, continua, è la funzione di distribuzione di probabilità di una variabile aleatoria se

- $\varphi(t) \geq 0$
- **per ogni $t \in \mathbb{R}$**
- $\int_{-\infty}^{+\infty} \varphi(t) dt = 1$

In tal caso si ha

$$\mathcal{P}(\xi \leq x) = \int_{-\infty}^x \varphi(t) dt$$

$$\mathcal{P}(x_0 \leq \xi \leq x_1) = \int_{x_0}^{x_1} \varphi(t) dt$$

La funzione

$$F(x) = \mathcal{P}(\xi \leq x) = \int_{-\infty}^x \varphi(t) dt$$

si chiama distribuzione cumulativa di probabilità della variabile aleatoria ξ di densità di probabilità φ .

Osserviamo che ad ogni variabile aleatoria discreta (finita) si può associare una variabile aleatoria continua la cui densità è una funzione costante a tratti, nulla al di fuori di un insieme limitato (nel caso in cui la variabile sia discreta e finita).

Per chiarire il concetto consideriamo la variabile aleatoria ξ che fornisce il punteggio ottenuto nel lancio di due dadi; in tal caso la funzione distribuzione di probabilità φ può essere definita come segue

$$\varphi(x) = \begin{cases} 0 & x < 1.5 \\ k/36 & k + .5 \leq x < k + 1.5 & k = 1..5 \\ 6/36 & 6.5 \leq x < 7.5 \\ (12 - k)/36 & k + .5 \leq x < k + 1.5 & k = 7..11 \\ 0 & x \geq 12.5 \end{cases}$$

Come nel caso delle variabili aleatorie discrete possiamo definire

DEFINIZIONE 28.7. Se ξ è una variabile aleatoria continua che ha densità di probabilità φ ,

- la **media** μ di ξ è definita da

$$\mu = E(\xi) = \int_{-\infty}^{+\infty} x\varphi(x)dx$$

- la **varianza** σ^2 di ξ è definita da

$$\sigma^2 = \text{Var}(\xi) = E((\xi - \mu)^2) = \int_{-\infty}^{+\infty} (x - \mu)^2 \varphi(x)dx$$

- lo **scarto quadratico medio** di ξ è definito da

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

- la **moda** M di ξ è definita da

$$M = \sup_{x \in \mathbb{R}} \varphi(x)$$

- la **mediana** m di ξ è definita da

$$\mathcal{P}(\xi \leq m) = \int_{-\infty}^m \varphi(x)dx = \int_m^{+\infty} \varphi(x)dx = \mathcal{P}(\xi \geq m)$$

- il **momento di ordine** k μ_k di ξ è definito da

$$\mu_k = E((\xi - \mu)^k) = \int_{-\infty}^{+\infty} (x - \mu)^k \varphi(x)dx$$

- il **momento di ordine** k , **rispetto all'origine** μ'_k di ξ è definito da

$$\mu_k = E(\xi^k) = \int_{-\infty}^{+\infty} x^k \varphi(x)dx$$

DEFINIZIONE 28.8. Se ξ è una variabile aleatoria continua e se $f : \mathbb{R} \rightarrow \mathbb{R}$ è una funzione possiamo definire la variabile aleatoria $f(\xi)$ assegnandone la funzione distribuzione di probabilità definita da

$$\psi(t) = f(t)\varphi(t)$$

In tal modo

$$\mathcal{P}(x_0 \leq f(\xi) \leq x_1) = \int_{x_0}^{x_1} f(t)\varphi(t)dt$$

$$E(f(\xi)) = \int_{-\infty}^{+\infty} f(t)\varphi(t)dt$$

DEFINIZIONE 28.9. Se ξ è una variabile aleatoria continua la cui densità di probabilità è φ , definiamo **funzione generatrice dei momenti** di ξ la

$$M_\xi(t) = E(e^{t\xi}) = \int_{-\infty}^{+\infty} e^{tx}\varphi(x)dx$$

5.1. Le variabili aleatorie discrete infinite. Un caso che si colloca a metà tra quello delle variabili aleatorie discrete finite ed il caso delle variabili aleatorie continue è il caso delle variabili aleatorie discrete che assumono una quantità infinita numerabile di valori.

Il caso delle variabili aleatorie infinite ci suggerisce come comportarci in questo caso.

Se lo spazio \mathcal{U} degli eventi è numerabile, allora potremo scrivere che

$$\mathcal{U} = \{A_i, i \in \mathbb{N}\}$$

e sarà sufficiente definire la probabilità di ciascun evento

$$\mathcal{P}(A_i) = p_i$$

con la condizione

$$\sum_{i=1}^{+\infty} p_i = 1$$

Media, varianza, funzione di distribuzione possono essere definiti mediante le

DEFINIZIONE 28.10. Se ξ è una variabile aleatoria discreta numerabile definiamo

- la **media** μ di ξ come

$$(28.11) \quad \mu = E(\xi) = \sum_{i=1}^{+\infty} i p_i$$

- la **varianza** σ^2 di ξ come

$$(28.12) \quad \sigma^2 = \text{Var}(\xi) = E((\xi - \mu)^2) = \sum_{i=1}^{+\infty} (i - \mu)^2 p_i$$

- lo **scarto quadratico medio** o deviazione standard di ξ

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

- il **momento k -esimo** μ_k di ξ come

$$\mu_k = \sum_{i=1}^{+\infty} (i - \mu)^k p_i$$

- il **momento k -esimo rispetto all'origine** μ'_k di ξ come

$$\mu'_k = \sum_{i=1}^{+\infty} i^k p_i$$

- la **funzione di distribuzione** φ di ξ come

$$(28.13) \quad \varphi(n) = \mathcal{P}(\xi = A_i)$$

e si ha

$$(28.14) \quad \mathcal{P}(\xi \leq n) = \sum_{i=1}^n \mathcal{P}(\xi = A_i) \quad n \in \mathbb{N}$$

Naturalmente possiamo estendere la definizione di φ ad \mathbb{R} definendola costante sugli intervalli del tipo $[k, k+1]$ e ponendola uguale a 0 prima di 1.

Inoltre

DEFINIZIONE 28.11. Se ξ è una variabile aleatoria discreta infinita la cui densità di probabilità è φ , e se $f : \mathbb{R} \rightarrow \mathbb{R}$ è una funzione possiamo definire la variabile aleatoria $f(\xi)$ assegnandone la funzione distribuzione di probabilità definita da

$$\psi(k) = f(k)\varphi(k) = f(k)p_k$$

Quindi

$$E(f(\xi)) = \sum_{i=1}^{+\infty} f(i)\varphi(i)$$

Definiamo anche definiamo **funzione generatrice dei momenti** di ξ la

$$M_\xi(t) = E(e^{t\xi}) = \sum_{k=1}^{+\infty} e^{kt}\varphi(k) =$$

6. Distribuzioni di probabilità doppie

Siano $(\mathcal{U}_1, \mathcal{F}_1, \mathcal{P}_1)$ e $(\mathcal{U}_2, \mathcal{F}_2, \mathcal{P}_2)$ due spazi di probabilità consideriamo la variabile aleatoria che indichiamo con (ξ, η) definita sullo spazio $\mathcal{U}_1 \times \mathcal{U}_2$ mediante la

$$F(x, y) = \mathcal{P}(\xi \leq x, \eta \leq y) = \int_{-\infty}^x \left(\int_{-\infty}^y f(t, s) ds \right) dt$$

F è la distribuzione cumulativa di probabilità della variabile (ξ, η) ed f è la sua funzione distribuzione di probabilità

Se f è continua possiamo affermare che

$$\frac{\partial^2 F}{\partial x \partial y} = f(x, y)$$

Naturalmente devono essere verificate le seguenti condizioni:

•

$$f(x, y) \geq 0$$

•

$$\int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(t, s) ds \right) dt = 1$$

Inoltre se

$$F_1(x) = \mathcal{P}(\xi \leq x) = \int_{-\infty}^x \left(\int_{-\infty}^{+\infty} f(t, s) ds \right) dt$$

$$F_2(y) = \mathcal{P}(\eta \leq y) = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^y f(t, s) dt \right) ds = \int_{-\infty}^y \left(\int_{-\infty}^{+\infty} f(t, s) dt \right) ds$$

F_1 ed F_2 sono le distribuzioni cumulative delle variabili aleatorie ξ e η , rispettivamente le cui funzioni di distribuzione sono date da

$$\varphi(t) = \int_{-\infty}^{+\infty} f(t, s) ds$$

$$\psi(s) = \int_{-\infty}^{+\infty} f(t, s) dt$$

Nel caso in cui le variabili aleatorie ξ e η siano indipendenti, allora (ξ, η) ha una distribuzione di probabilità

$$f(t, s) = \varphi(t)\psi(s)$$

dove φ e ψ sono le funzioni di distribuzione di ξ e η , rispettivamente.

È utile ricordare che la probabilità della variabile aleatoria ξ condizionata alla variabile aleatoria η si può definire mediante la

$$\mathcal{P}(\xi \leq x | \eta \leq y) = \int_{-\infty}^x \left(\int_{-\infty}^y \frac{f(t, s)}{\psi(s)} ds \right) dt$$

per cui $\frac{f(t, s)}{\psi(s)}$ è la sua funzione di distribuzione di probabilità.

7. Qualche proprietà di valor medio e varianza

Abbiamo già definito cosa intendiamo per

- media o valor medio
- varianza
- momento
- momento rispetto all'origine
- speranza matematica
- funzione generatrice dei momenti

di una variabile aleatoria ξ sia nel caso discreto finito sia nel caso continuo sia nel caso discreto infinito.

Tra le proprietà del valor medio ricordiamo che

- $E(\alpha\xi + \beta\eta) = \alpha E(\xi) + \beta E(\eta)$
- **se ξ e η sono variabili aleatorie indipendenti**

$$E(\xi\eta) = E(\xi)E(\eta)$$

In altre parole la speranza matematica è una applicazione lineare; La verifica di tale fatto non è banale e si può ad esempio ottenere provando il risultato nel caso discreto finito ed utilizzando un passaggio al limite per gli altri casi.

Si può provare che anche la varianza gode di proprietà di linearità, infatti

- $\text{Var}(\alpha\xi) = \alpha^2 \text{Var}(\xi)$
- **se ξ e η sono variabili aleatorie indipendenti**

$$\text{Var}(\xi \pm \eta) = \text{Var}(\xi) + \text{Var}(\eta)$$

È anche utile ricordare che

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

Vediamo ora come la funzione generatrice dei momenti si riveli molto comoda per il calcolo dei momenti di una variabile aleatoria.

Cominciamo con l'osservare che

$$(t - \mu)^k = \sum_{i=0}^k \binom{k}{i} t^i \mu^{k-i}$$

moltiplicando per $\varphi(t)$ ed integrando, otteniamo

$$(28.15) \quad \mu_k = \sum_{i=0}^k \binom{k}{i} \mu^i \mu^{k-i}$$

e se ne ricava che per trovare i momenti rispetto al valor medio μ_k è sufficiente conoscere i momenti rispetto all'origine μ'_k .

Casi particolari della 28.15 sono

$$\begin{aligned}\mu_2 &= \sigma_2 = \mu'_2 - \mu^2 \\ \mu_3 &= \mu'_3 - 3\mu'_2\mu + 2\mu^3\end{aligned}$$

(ricordiamo che $\mu_0 = \mu'_0 = 1$ e $\mu_1 = \mu'_1 = \mu$).

D'altro canto i momenti di una variabile aleatoria si possono trovare a partire dalla funzione generatrice dei momenti che è definita da:

$$M_\xi(t) = E(e^{t\xi}) = \int_{-\infty}^{+\infty} e^{ts} \varphi(s) ds$$

Infatti si può verificare che la funzione M_ξ è sviluppabile in serie di McLaurin ed il suo sviluppo è dato da

$$M_\xi(t) = \sum_{i=0}^{+\infty} \mu'_i \frac{t^i}{i!}$$

e quindi

$$\mu'_k = \frac{d^k}{dt^k} M_\xi(t)$$

8. La disuguaglianza di Tchebichev e la legge dei grandi numeri

In questa sezione ci occupiamo di due risultati fondamentali: la disuguaglianza di Tchebichev e la legge dei grandi numeri, cominciando a parlare della prima.

Sia ξ una variabile aleatoria con media μ e varianza σ^2 , allora si avrà che

$$\begin{aligned} \sigma^2 &= \int_{-\infty}^{+\infty} (t - \mu)^2 \varphi(t) dt = \\ &= \int_{\{t : |t - \mu| \leq \varepsilon\}} (t - \mu)^2 \varphi(t) dt + \int_{\{t : |t - \mu| > \varepsilon\}} (t - \mu)^2 \varphi(t) dt \geq \\ &\geq \int_{\{t : |t - \mu| \geq \varepsilon\}} (t - \mu)^2 \varphi(t) dt \geq \int_{\{t : |t - \mu| \geq \varepsilon\}} \varepsilon^2 \varphi(t) dt = \\ &= \varepsilon^2 \mathcal{P}(|\xi - \mu| \geq \varepsilon) \end{aligned}$$

se ne ricava pertanto che

$$(28.16) \quad \mathcal{P}(|\xi - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

La 28.16 è nota come disuguaglianza di Tchebichev e ne possiamo trarre una interessante conseguenza: per $\varepsilon = k\sigma$ otteniamo che

$$(28.17) \quad \mathcal{P}(|\xi - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Pertanto

$$(28.18) \quad \mathcal{P}(|\xi - \mu| < k\sigma) = 1 - \mathcal{P}(|\xi - \mu| \geq k\sigma) \geq 1 - \frac{1}{k^2}$$

Se ora consideriamo la seguente tabella

k	$1 - \frac{1}{k^2}$
1	0
2	.75
3	.88
4	.93
5	.95
6	.97

TABELLA 28.3. Valori approssimati di $1 - \frac{1}{k^2}$

Si vede pertanto che se ξ è una variabile aleatoria di media μ e di varianza σ^2 , allora la probabilità che il valore assunto da ξ sia vicino alla media μ per meno di 2 volte la varianza è del 75% e sale all'88% se ci accontentiamo di un errore inferiore a 3 volte la varianza.

Va osservato che, nonostante fornisca risultati soddisfacenti, la disuguaglianza di Tchebichev non è molto precisa.

8.1. La legge dei grandi numeri. Un'altra delle conseguenze della disuguaglianza di Tchebichev prende il nome di "Legge dei grandi numeri" e si ricava come segue.

Siano $\xi_1, \xi_2, \dots, \xi_n$ variabili aleatorie tutte con media μ e varianza σ^2 , e consideriamo la variabile aleatoria

$$S_n = \frac{\xi_1 + \xi_2 + \dots + \xi_n}{n}$$

Avremo che

$$E(S_n) = \frac{1}{n}(E(\xi_1) + E(\xi_2) + \dots + E(\xi_n)) = \mu$$

ed inoltre

$$\text{Var}(S_n) = \frac{1}{n^2}(\text{Var}(\xi_1) + \text{Var}(\xi_2) + \dots + \text{Var}(\xi_n)) = \frac{\sigma^2}{n}$$

Pertanto

$$(28.19) \quad \mathcal{P}(|S_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0$$

per $n \rightarrow +\infty$

La 28.19 è nota con il nome di **legge dei grandi numeri** ed esprime un concetto in base al quale la media delle uscite di una variabile aleatoria differisce dalla media della variabile aleatoria di una quantità infinitesima.

Va tuttavia sottolineato che la legge dei grandi numeri fornisce informazioni di carattere qualitativo e quindi non può essere usata per stime di tipo quantitativo.

9. Normalizzazione di una variabile aleatoria.

Sia ξ una variabile aleatoria di media μ e di varianza σ^2 con distribuzione di probabilità φ .

Sia cioè

$$\begin{aligned}\int_{-\infty}^{+\infty} \varphi(t) dt &= 1 \\ \mathcal{P}(a \leq \xi \leq b) &= \int_a^b \varphi(t) dt \\ E(\xi) &= \int_{-\infty}^{+\infty} t \varphi(t) dt = \mu \\ E(\xi^2) &= \int_{-\infty}^{+\infty} (t - \mu)^2 \varphi(t) dt = \sigma^2\end{aligned}$$

e consideriamo la variabile aleatoria

$$\xi^* = \frac{\xi - \mu}{\sigma}$$

Per le proprietà di media e varianza possiamo affermare che ξ^* è una variabile normalizzata (o standardizzata), intendendo con ciò che ξ^* ha media 0 e varianza 1.

Allo scopo di determinare la funzione di distribuzione di ξ^* consideriamo la variabile aleatoria η la cui funzione di distribuzione è definita da

$$\sigma \varphi(\mu + \sigma t)$$

Possiamo allora verificare che

$$\begin{aligned}\int_{-\infty}^{+\infty} \sigma \varphi(\mu + \sigma t) dt &= \int_{-\infty}^{+\infty} \varphi(s) ds = 1 \\ \int_{-\infty}^{+\infty} t \sigma \varphi(\mu + \sigma t) dt &= \int_{-\infty}^{+\infty} \left(\frac{s - \mu}{\sigma} \right) \varphi(s) ds = \\ &= \frac{1}{\sigma} \left(\int_{-\infty}^{+\infty} s \varphi(s) ds - \mu \int_{-\infty}^{+\infty} \varphi(s) ds \right) = 0 \\ \int_{-\infty}^{+\infty} t^2 \sigma \varphi(\mu + \sigma t) dt &= \int_{-\infty}^{+\infty} \left(\frac{s - \mu}{\sigma} \right)^2 \varphi(s) ds = \\ &= \frac{1}{\sigma} \int_{-\infty}^{+\infty} (s - \mu)^2 \varphi(s) ds = 1\end{aligned}$$

$$\begin{aligned}
\mathcal{P}(a \leq \eta \leq b) &= \int_a^b \sigma \varphi(\mu + \sigma t) dt = \\
&= \int_{\mu + \sigma a}^{\mu + \sigma b} \sigma \varphi(s) \frac{1}{\sigma} ds = \mathcal{P}(\mu + \sigma a \leq \xi \leq \mu + \sigma b) = \\
&= \mathcal{P}\left(a \leq \frac{\xi - \mu}{\sigma} \leq b\right)
\end{aligned}$$

Ne deduciamo che η e $\xi^* = \frac{\xi - \mu}{\sigma}$ hanno la stessa distribuzione di probabilità e quindi sono la stessa variabile aleatoria.

In altre parole la funzione

$$\sigma \varphi(\mu + \sigma t)$$

è la distribuzione di probabilità di

$$\xi^* = \frac{\xi - \mu}{\sigma}$$

QUALCHE DISTRIBUZIONE DI PROBABILITÀ

Le funzioni di distribuzione di probabilità sono fondamentali per descrivere il comportamento delle variabili aleatorie che ci interessano.

Ogni variabile aleatoria ha una sua distribuzione e per definirne la proprietà è utile fare riferimento ad alcune distribuzioni note che sono in grado di descrivere la maggior parte delle variabili aleatorie con cui normalmente si lavora.

1. La distribuzione uniforme

La più semplice funzione di distribuzione di probabilità è quella di una variabile aleatoria che restituisce un valore scelto in un intervallo $[a, b]$ con il criterio di equiprobabilità.

Abbiamo già visto che in tal caso

$$\mathcal{P}(x \leq \xi \leq x + h) = \frac{h}{b - a}$$

e che la sua distribuzione di densità è

$$\varphi(t) = \begin{cases} \frac{1}{b-a} & t \in [a, b] \\ 0 & \text{altrove} \end{cases}$$

La funzione generatrice dei momenti si calcola mediante la

$$M_{\xi}(t) = \frac{1}{b-a} \int_a^b e^{tx} dx = \frac{e^{tb} - e^{ta}}{t(b-a)}$$

Se ne ricava subito che

$$\mu = \frac{b+a}{2}$$

$$\sigma^2 = \frac{(b-a)^2}{12}$$

Descriviamo ora tre importanti distribuzioni di probabilità che sono, come vedremo, tra loro legate:

- La distribuzione binomiale di Bernoulli
- La distribuzione di Poisson
- La distribuzione normale di Gauss

Si tratta di una distribuzione discreta finita (Bernoulli), una distribuzione discreta numerabile (Poisson) ed una distribuzione continua (Gauss) che possono essere derivate a partire dal concetto elementare di prova bernoulliana.

2. La distribuzione binomiale di Bernoulli

DEFINIZIONE 29.1. Chiamiamo prova bernoulliana un esperimento che ha due soli possibili esiti:

- Successo, cui associamo il valore 1 con probabilità p
- Insuccesso, cui associamo il valore 0 con probabilità q

essendo ovviamente $p + q = 1$.

Chiamiamo variabile aleatoria bernoulliana la variabile aleatoria ξ che restituisce il numero di successi che si sono verificati su n prove ripetute (lanci) dell'esperimento.

Possiamo calcolare la probabilità che la variabile aleatoria ξ assuma il valore k mediante la

$$\varphi(k) = \mathcal{P}(\xi = k) = \binom{n}{k} p^k q^{n-k} = \binom{n}{k} p^k (1-p)^{n-k}$$

Possiamo giustificare la formula precedente se descriviamo la successione di n prove ripetute con una stringa di elementi che assumono il valore 1 oppure 0 a seconda che la corrispondente prova abbia avuto o no successo.

0	1	1	0	0	1	0	0	1	0	0	1	0	1	0	0	0	1	0	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

affinchè ci siano k successi la stringa dovrà contenere esattamente k volte il valore 1 (ed $n - k$ volte il valore 0) e quindi, poichè in ogni elemento 1 si presenta con probabilità p e 0 con probabilità q , una stringa con k successi avrà una probabilità di comparire uguale a

$$p^k q^{n-k}$$

d'altro canto, poichè siamo unicamente interessati a contare il numero di successi, e non l'ordine con cui si verificano, dovremo tener conto che si verificano, ad esempio, 2 successi in tanti modi diversi

0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

il cui numero è dato dalle combinazioni di n oggetti a k a k e cioè

$$\binom{n}{k}$$

(Infatti ciascuna stringa può essere individuata dalla sequenza dei k numeri, compresi tra 1 ed n , che indicano la posizione dei successi nella stringa.)

Possiamo calcolare la media della variabile bernoulliana ξ osservando che la media in ciascuna prova è

$$1 \cdot p + 0 \cdot q = p$$

e su n esperimenti, essendo la media lineare, avremo

$$\mu = E(\xi) = np$$

La varianza della variabile bernoulliana ξ in ciascuna prova è

$$(1-p)^2 \cdot p + 0 \cdot (0-q)^2 q = q^2 p + p^2 q = pq(p+q) = pq$$

e su n esperimenti essendo la varianza lineare avremo

$$\sigma^2 = E((\xi - \mu)^2) = npq$$

e

$$\sigma = \sqrt{npq}$$

Per calcolare la funzione generatrice dei momenti possiamo procedere come segue

$$\begin{aligned} M_\xi(t) = E(e^{t\xi}) &= \sum_{k=0}^n e^{tk} \binom{n}{k} p^k q^{n-k} = \\ &= \sum_{k=0}^n \binom{n}{k} (pe^t)^k q^{n-k} = (pe^t + q)^n \end{aligned}$$

Un esempio notevole di prove bernoulliane si costruisce considerando il lancio di una moneta ripetuto per n volte; in tal caso $p = q = \frac{1}{2}$.

3. La distribuzione di Poisson

Si consideri il caso di m prove ripetute bernoulliane in cui la probabilità di successo della singola prova sia $\frac{1}{n}$

Un esempio di tale situazione di si può realizzare considerando un pagliaio con un numero $n-1$ di pagliuzze dal quale si voglia estrarre 1 ago; la probabilità di estrarre l'ago sarà ovviamente $\frac{1}{n}$ e la prova sarà ripetuta m volte.

Sia ξ la variabile aleatoria che restituisce il numero di successi ottenuti nelle m prove fatte. Avremo che

$$\begin{aligned} \varphi(k) = \mathcal{P}(\xi = k) &= \binom{m}{k} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{m-k} = \\ &= \frac{m(m-1)(m-2) \cdots (m-(k-1))}{k!} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^m \left(1 - \frac{1}{n}\right)^{-k} = \\ &= \frac{1}{k!} \frac{m}{n} \left(\frac{m}{n} - \frac{1}{n}\right) \cdots \left(\frac{m}{n} - \frac{k-1}{n}\right) \left(1 - \frac{1}{n}\right)^m \left(1 - \frac{1}{n}\right)^{-k} \end{aligned}$$

Se ora consideriamo di far tendere n a $+\infty$ mantenendo

$$\frac{m}{n} = \lambda$$

(oppure chiedendo semplicemente che $\frac{m}{n} \rightarrow \lambda$), otteniamo che

$$\varphi(k) = \frac{1}{k!} \lambda \left(\lambda - \frac{1}{n} \right) \left(\lambda - \frac{2}{n} \right) \cdots \left(\lambda - \frac{k-1}{n} \right) \left(1 - \frac{1}{n} \right)^{n\lambda} \left(1 - \frac{1}{n} \right)^{-k}$$

e, per $n \rightarrow +\infty$

$$\varphi(k) = \frac{1}{k!} \lambda^k e^{-\lambda}$$

La funzione distribuzione di probabilità di Poisson è quindi data da

$$\boxed{\varphi(k) = \frac{1}{k!} \lambda^k e^{-\lambda}}$$

Per calcolare media, varianza ed i momenti della distribuzione di Poisson è utile calcolare la funzione generatrice dei momenti.

$$M_{\xi}(t) = \sum_{k=0}^{+\infty} e^{tk} \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=0}^{+\infty} \frac{(\lambda e^t)^k}{k!} e^{-\lambda} = e^{-\lambda} e^{\lambda(e^t-1)}$$

e le sue derivate

$$\begin{aligned} \frac{d}{dt} M_{\xi}(t) &= e^{\lambda(e^t-1)} \lambda e^t \\ \frac{d^2}{dt^2} M_{\xi}(t) &= e^{\lambda(e^t-1)} \lambda^2 e^t + e^{\lambda(e^t-1)} \lambda e^t \end{aligned}$$

e calcolarle in $t = 0$.

$$\begin{aligned} \frac{d}{dt} M_{\xi}(0) &= \lambda = \mu'_1 = \mu \\ \frac{d^2}{dt^2} M_{\xi}(0) &= \lambda^2 + \lambda = \mu'_2 \end{aligned}$$

da cui

$$\begin{aligned} \mu &= \mu'_1 = \lambda \\ \sigma^2 &= \mu'_2 - \mu^2 = \lambda^2 + \lambda - \lambda^2 = \lambda \end{aligned}$$

4. La distribuzione gaussiana

Sia $\xi_n = k$ una variabile aleatoria bernoulliana relativa ad n prove ripetute, allora

$$\mathcal{P}(\xi_n = k) = \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

Sia ha

$$\begin{aligned} E(\xi_n) &= \mu = np \\ Var(\xi_n) &= \sigma^2 = npq \end{aligned}$$

Consideriamo la variabile normalizzata (di media 0 e varianza 1)

$$\xi_n^* = \frac{\xi_n - \mu}{\sigma}$$

ξ_n^* assume i valori

$$x_k = \frac{k - \mu}{\sigma} \quad k = 1, \dots, n$$

e

$$\mathcal{P}\left(\xi_n^* = \frac{k - \mu}{\sigma}\right) = \mathcal{P}(\xi_n = k) = \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

Possiamo ricavare inoltre che

$$\begin{aligned} k &= \mu + \sigma x_k = np + \sigma x_k \\ n - k &= n - np - \sigma x_k = nq - \sigma x_k \end{aligned}$$

dove $\sigma = \sqrt{npq}$ e si vede che, per $n \rightarrow +\infty$ tanto k quanto $n - k$ tendono a $+\infty$, inoltre

$$\begin{aligned} \frac{k}{np} &= 1 + \sqrt{\frac{q}{np}} x_k \\ \frac{n-k}{nq} &= 1 + \sqrt{\frac{p}{nq}} x_k \\ x_{k+1} - x_k &= \Delta x_k = \frac{1}{\sigma} = \frac{1}{\sqrt{npq}} \end{aligned}$$

Possiamo allora scrivere che

$$\begin{aligned} \mathcal{P}(x' \leq \xi_n^* \leq x'') &= \sum_{x_k \in [x', x'']} \mathcal{P}(\xi_n^* = x_k) = \\ &= \sum_{\{k: x_k \in [x', x'']\}} \mathcal{P}(\xi_n = k) = \\ &= \sum_{\{k: x_k \in [x', x'']\}} \frac{n!}{k!(n-k)!} p^k q^{n-k} \end{aligned}$$

Poichè, per n grande, tanto k quanto $n - k$ sono grandi, possiamo usare la formula di Stirling per approssimare il fattoriale e quindi

$$\begin{aligned} \frac{n!}{k!(n-k)!} p^k q^{n-k} &\approx \frac{n^n e^{-n} \sqrt{2\pi n}}{k^k e^{-k} \sqrt{2\pi k} (n-k)^{n-k} e^{-(n-k)} \sqrt{2\pi(n-k)}} p^k q^{n-k} = \\ &= \frac{1}{\sqrt{2\pi}} \frac{n^k}{k^k} \frac{n^{n-k}}{(n-k)^{n-k}} \sqrt{\frac{n}{k(n-k)}} p^k q^{n-k} = \\ &= \frac{1}{\sqrt{2\pi}} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k} \sqrt{\frac{n}{k(n-k)}} \end{aligned}$$

Ma, sempre per n grande,

$$\sqrt{\frac{n}{k(n-k)}} = \sqrt{\frac{n}{(np + \sigma x_k)(nq - \sigma x_k)}} \approx \sqrt{\frac{n}{n^2 pq}} = \frac{1}{\sqrt{npq}}$$

D'altro canto

$$\begin{aligned}
 \ln \left(\frac{np}{k} \right)^k \left(\frac{nq}{n-k} \right)^{n-k} &= -k \ln \left(\frac{np}{k} \right) - (n-k) \ln \left(\frac{nq}{n-k} \right) = \\
 &= -(np + \sigma x_k) \ln \left(1 + \sqrt{\frac{q}{np}} \right) - \\
 &\quad - (nq - \sigma x_k) \ln \left(1 - \sqrt{\frac{p}{nq}} x_k \right) + \omega \left(\frac{1}{\sqrt{n}} \right) = \\
 &= -(np + \sigma x_k) \left(\sqrt{\frac{q}{np}} x_k - \frac{1}{2} \frac{q}{np} x_k^2 \right) - \\
 &\quad - (nq - \sigma x_k) \left(-\sqrt{\frac{p}{nq}} x_k - \frac{1}{2} \frac{p}{nq} x_k^2 \right) + \omega \left(\frac{1}{\sqrt{n}} \right) = \\
 &= -x_k \left[np \sqrt{\frac{q}{np}} - nq \sqrt{\frac{p}{nq}} \right] - x_k^2 \left[\sqrt{npq} \sqrt{\frac{q}{np}} + \sqrt{npq} \sqrt{\frac{p}{nq}} \right] - \\
 &\quad - x_k^2 \left[\frac{1}{2} np \frac{q}{np} + \frac{1}{2} nq \frac{p}{nq} \right] - x_k^3 \left[\frac{q \sqrt{npq}}{2np} + \frac{p \sqrt{npq}}{nq} \right] = \\
 &= -x_k^2 + \frac{1}{2} x_k^2 + \omega \left(\frac{1}{\sqrt{n}} \right)
 \end{aligned}$$

dove, come al solito, si indica con ω una funzione infinitesima.

Ne concludiamo pertanto che

$$\begin{aligned}
 \mathcal{P}(x' \leq \xi_n^* \leq x'') &\approx \\
 &\approx \sum_{x_k \in [x', x'']} \frac{1}{\sqrt{2\pi}} \left(\frac{np}{k} \right)^k \left(\frac{nq}{n-k} \right)^{n-k} \sqrt{\frac{n}{k(n-k)}} \approx \\
 &\approx \sum_{x_k \in [x', x'']} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \Delta x_k \rightarrow \frac{1}{\sqrt{2\pi}} \int_{x'}^{x''} e^{-\frac{1}{2}t^2} dt
 \end{aligned}$$

È pertanto naturale considerare la funzione distribuzione di probabilità definita da

$$\boxed{\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}}$$

La φ si chiama funzione di distribuzione di probabilità Gaussiana di media 0 e varianza 1. (Si ricordi che è ottenuta come limite di distribuzioni normalizzate).

Utilizzando il teorema di integrazione per sostituzione (poichè l'integrale è improprio deve essere calcolato come limite di integrali propri ed a questi si applica il teorema di integrazione per sostituzione) possiamo ricavare che la funzione

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

è una funzione di distribuzione di probabilità, che chiameremo ancora Gaussiana, di media μ e di varianza σ^2 .

Possiamo calcolare la funzione generatrice dei momenti della funzione di distribuzione gaussiana, infatti

$$M_\xi(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{tx} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

e, operando il cambio di variabili $\frac{x-\mu}{\sigma} = u$, da cui $x = \sigma u + \mu$,

$$\begin{aligned} M_\xi(t) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{t\sigma u + t\mu} e^{-\frac{u^2}{2}} \sigma du = \\ &= \sigma \frac{1}{\sigma\sqrt{2\pi}} e^{t\mu} \int_{-\infty}^{+\infty} e^{t\sigma u} e^{-\frac{u^2}{2}} du = \frac{1}{\sqrt{2\pi}} e^{t\mu} \int_{-\infty}^{+\infty} e^{\frac{1}{2}(2t\sigma u - u^2)} du = \\ &= \frac{1}{\sqrt{2\pi}} e^{t\mu} \int_{-\infty}^{+\infty} e^{\frac{1}{2}(\sigma^2 t^2 - \sigma^2 t^2 + 2t\sigma u - u^2)} du = \\ &= \frac{1}{\sqrt{2\pi}} e^{t\mu + \frac{1}{2}\sigma^2 t^2} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(\sigma t - u)^2} du = \\ &= \frac{1}{\sqrt{2\pi}} e^{t\mu + \frac{1}{2}\sigma^2 t^2} \sqrt{2\pi} = e^{t\mu + \frac{1}{2}\sigma^2 t^2} \end{aligned}$$

5. Due distribuzioni discrete

5.1. La distribuzione multinomiale. Consideriamo un esperimento che possa avere k possibili esiti, che indichiamo con

$$A_1, A_2, \dots, A_k$$

con probabilità

$$p_1, p_2, \dots, p_k, \quad p_1 + p_2 + \dots + p_k = 1$$

e supponiamo di replicarlo per n volte; consideriamo la variabile aleatoria ξ che restituisce la n -pla di valori

$$n_1, n_2, \dots, n_k, \quad n_1 + n_2 + \dots + n_k = n$$

dove n_i è il numero di volte in cui si è verificato l'evento A_i .

La funzione distribuzione di probabilità di ξ è data da

$$\begin{aligned}\varphi(n_1, n_2, \dots, n_k) &= \mathcal{P}(\xi_1 = p_1, \xi_2 = p_2, \dots, \xi_k = p_k) = \\ &= \frac{n!}{n_1! n_2! \dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}\end{aligned}$$

5.2. La distribuzione ipergeometrica. Consideriamo un'urna contenente b palline nere e w palline bianche e supponiamo di estrarre per n volte una pallina rimettendola. dopo ogni estrazione, nell'urna.

Consideriamo la variabile aleatoria ξ che restituisce il numero di volte in cui si è estratta una pallina nera; allora la densità di probabilità di ξ si può calcolare mediante la

$$\begin{aligned}\varphi(k) = \mathcal{P}(\xi = k) &= \binom{n}{k} \left(\frac{b}{b+w} \right)^k \left(\frac{w}{b+w} \right)^{n-k} = \\ &= \binom{n}{k} \frac{b^k w^{n-k}}{b+w}\end{aligned}$$

non appena si ricordi la distribuzione binomiale e si tenga presente che

$$p = \frac{b}{b+w}, \quad q = \frac{w}{b+w}$$

Qualora l'esperimento si ripeta senza rimettere la pallina estratta nell'urna, (campionamento senza ripetizione), si può vedere che la densità di probabilità della nuova variabile aleatoria ξ che conta il numero delle palline nere estratte è

$$\varphi(k) = \mathcal{P}(\xi = k) = \frac{\binom{b}{k} \binom{w}{n-k}}{\binom{b+w}{n}}$$

Si calcola anche che

$$\begin{aligned}\mu &= \frac{nb}{b+w} \\ \sigma^2 &= \frac{nbw(b+w-n)}{(b+w)^2(b+w-1)}\end{aligned}$$

6. Le distribuzioni legate ai test statistici.

6.1. La distribuzione χ^2 . Si tratta della distribuzione di probabilità di una variabile aleatoria χ^2 che restituisce la somma di ν variabili aleatorie ξ_i indipendenti, aventi distribuzione gaussiana con media 0 e varianza 1 (distribuzioni normali standardizzate).

$$\chi^2 = \xi_1^2 + \xi_2^2 + \dots + \xi_\nu^2$$

La funzione di distribuzione della variabile aleatoria χ^2 è data da

$$\varphi(t) = \begin{cases} \frac{1}{2^{\nu/2}\Gamma(\nu/2)} t^{\nu/2-1} e^{-t/2} & t \geq 0 \\ 0 & \text{altrimenti} \end{cases}$$

ν rappresenta i gradi di libertà, media, varianza e generatrice dei momenti sono date da

$$\begin{aligned} \mu &= \nu \\ \sigma^2 &= 2\nu \\ M_{\chi}^2(t) &= (1 - 2t)^{-\nu/2} \end{aligned}$$

6.2. La distribuzione T di Student. È la distribuzione di una variabile aleatoria T che restituisce il rapporto

$$T = \frac{\xi}{\sqrt{\eta/\nu}}$$

dove ξ è una variabile aleatoria con densità di probabilità gaussiana normale (media 0 e varianza 1) ed η è una variabile aleatoria con distribuzione χ^2 a ν gradi di libertà

La funzione di distribuzione della variabile aleatoria T è data da

$$\varphi(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma(\nu/2)} \left(1 + \frac{t^2}{\nu}\right)$$

ν rappresenta i gradi di libertà, media e varianza sono date da

$$\begin{aligned} \mu &= 0 \\ \sigma^2 &= \frac{\nu}{\nu - 2} \end{aligned}$$

6.3. La distribuzione F di Fisher. È la distribuzione di una variabile aleatoria F che restituisce il rapporto

$$F = \frac{\xi_1/\nu_1}{\xi_2/\nu_2}$$

dove ξ ed ξ_2 sono variabili aleatorie con distribuzione χ^2 a ν_1 e ν_2 gradi di libertà, rispettivamente.

La funzione di distribuzione della variabile aleatoria F è data da

$$\varphi(t) = \begin{cases} \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)} \nu_1^{\nu_1/2} \nu_2^{\nu_2/2} t^{\nu_2/2-1} (\nu_2 + \nu_1 t)^{-(\nu_1+\nu_2)/2} & t > 0 \\ 0 & \text{altrimenti} \end{cases}$$

media e varianza sono date da

$$\mu = \frac{\nu_2}{\nu_2 - 2}$$

$$\sigma^2 = \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 4)(\nu_2 - 2)^2}$$

7. La distribuzione γ .

La distribuzione γ è definita, per $\alpha, \beta > 0$ da

$$\varphi(t) = \begin{cases} t^{\alpha-1} e^{-t/\beta} & t > 0 \\ 0 & \text{altrimenti} \end{cases}$$

media, varianza e generatrice dei momenti sono date da

$$\mu = \alpha\beta$$

$$\sigma^2 = \alpha\beta^2$$

$$M_\gamma(t) = (1 - \beta t)^{-\alpha}$$

8. La distribuzione β .

La distribuzione β è definita da

$$\varphi(t) = \begin{cases} \frac{t^{\alpha-1}(1-t)^{\beta-1}}{B(\alpha, \beta)} & 0 < t < 1 \\ 0 & \text{altrimenti} \end{cases} = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} t^{\alpha-1}(1-t)^{\beta-1} & 0 < t < 1 \\ 0 & \text{altrimenti} \end{cases}$$

media e varianza sono date da

$$\mu = \frac{\alpha}{\alpha + \beta}$$

$$\sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)(\alpha + \beta + 1)}$$

Si vede anche la sua moda è $\frac{\alpha-1}{\alpha+\beta-2}$.

9. Variabili casuali con distribuzione assegnata.

Qualora sia necessario utilizzare dati generati casualmente con funzione distribuzione di probabilità fissata, possiamo procedere come segue.

Sia ϕ la distribuzione che si vuole considerare e sia

$$F(x) = \int_{-\infty}^x \phi(t) dt$$

la funzione di distribuzione cumulativa; allora si ha

$$\mathcal{P}(a \leq \xi \leq b) = \int_a^b \phi(t) dt = F(b) - F(a)$$

e quindi

$$\mathcal{P}(F^{-1}(a) \leq \xi \leq F^{-1}(b)) = b - a$$

e

$$\mathcal{P}(a \leq F(\xi) \leq b) = b - a$$

Pertanto $F(\xi)$ ha una distribuzione uniforme e quindi poichè

$$\xi = F^{-1}(F(\xi))$$

possiamo generare valori distribuiti con densità di probabilità ϕ , considerando valori generati con densità uniforme ed applicando a tali valori F^{-1} .

Tale procedimento non è tuttavia applicabile, ad esempio, per determinare valori distribuiti con densità gaussiana in quanto non è possibile determinare esplicitamente F^{-1} nel caso in cui

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$$

In tal caso, che peraltro è di rilevante importanza possiamo osservare che se ξ e η sono variabili aleatorie indipendenti con distribuzione gaussiana normale, allora (ξ, η) è una variabile aleatoria la cui funzione distribuzione di probabilità è

$$\frac{1}{2\pi} e^{-(t^2+s^2)/2}$$

Pertanto

$$\mathcal{P}((\xi, \eta) \in A) = \frac{1}{2\pi} \iint_A e^{-(t^2+s^2)/2} dt ds = \frac{1}{2\pi} \iint_B \rho e^{-(\rho^2)/2} d\rho d\theta$$

dove B ed A sono l'uno il trasformato dell'altro rispetto al cambio di variabili in coordinate polari.

Ne viene che possiamo identificare due nuove variabili (R, Θ) la cui densità di probabilità è data da

$$\frac{1}{2\pi} \rho e^{-(\rho^2)/2} = \left(\frac{1}{2\pi}\right) (\rho e^{-(\rho^2)/2}) = f(\theta)g(\rho)$$

dove

$$f(\theta) = \frac{1}{2\pi}$$

$$g(\rho) = \rho e^{-(\rho^2)/2}$$

Quindi per quanto visto in precedenza, per generare valori casuali di Θ e di R possiamo utilizzare valori uniformemente distribuiti θ ed r e applicare

a tali valori le funzioni F^{-1} e G^{-1} , rispettivamente, dove

$$F(t) = \int_0^t f(s)ds = 2\pi t \quad , \quad G(t) = \int_0^t g(s)ds = 1 - e^{-t^2/2}$$

Si ha allora

$$F^{-1}(s) = \frac{s}{2\pi} \quad , \quad G^{-1}(s) = \sqrt{-2 \ln(1-s)}$$

e le variabili

$$\xi = \sqrt{-2 \ln(1-t)} \cos\left(\frac{s}{2\pi}\right) \quad , \quad \eta = \sqrt{-2 \ln(1-t)} \sin\left(\frac{s}{2\pi}\right)$$

dove s e t sono distribuite uniformemente, risultano distribuite con densità gaussiana di media 0 e di varianza 1.

10. I test statistici

Consideriamo il seguente esempio.

Supponiamo di voler determinare se una moneta è truccata lanciandola un numero N di volte ed osservando il risultato dei lanci.

Possiamo definire truccata una moneta in cui la probabilità di uscita di T , che chiameremo p è diversa dalla probabilità di uscita di C che è $q = (1 - p)$. Sarà quindi truccata una moneta per la quale $p \neq 0.5$.

Osserviamo comunque che allo scopo di ottenere benefici da una scommessa potremmo anche essere interessati a conoscere se $p > 0.5$ o equivalentemente se $p < 0.5$.

I dati di cui vogliamo individuare qualche caratteristica, la popolazione (il nome deriva dalle origini dello studio statistico), sono costituiti da tutti i possibili lanci della moneta in questione.

I dati di cui disponiamo sono costituiti dalle uscite degli N lanci osservati, campione.

Del campione siamo in grado di conoscere la distribuzione di probabilità che è una distribuzione discreta binomiale la cui media è $\mu = Np$ e la cui varianza è $\sigma^2 = \sqrt{Npq}$.

Calcoliamo la probabilità che la variabile aleatoria ξ che restituisce il numero di teste uscite negli N lanci sia compresa tra n ed m .

Avremo che

$$\mathcal{P}(n \leq \xi \leq m) = \sum_{k=n}^m \binom{N}{k} p^k q^{N-k}$$

e quindi, se ad esempio $N = 100$, $n = 45, 40, 30$, $m = 55, 60, 70$,

$$\mathcal{P}(45 \leq \xi \leq 55) \approx 0.73$$

$$\mathcal{P}(40 \leq \xi \leq 60) \approx 0.96$$

$$\mathcal{P}(30 \leq \xi \leq 70) \approx 0.99$$

Possiamo quindi affermare che

- nel caso il numero di teste uscite sia superiore a 55 o inferiore a 45, si è verificato un evento che, supponendo la moneta non truccata, ha la probabilità di verificarsi del $100 - 73 = 27\%$
- nel caso il numero di teste uscite sia superiore a 60 o inferiore a 40, si è verificato un evento che, supponendo la moneta non truccata, ha la probabilità di verificarsi del $100 - 96 = 4\%$
- nel caso il numero di teste uscite sia superiore a 70 o inferiore a 30, si è verificato un evento che, supponendo la moneta non truccata, ha la probabilità di verificarsi del $100 - 99 = 1\%$

e potremmo in base a queste considerazioni concludere che

- nel caso il numero di teste uscite sia superiore a 55 o inferiore a 45, siamo fiduciosi che la moneta sia truccata con un livello di significatività del 73%
- nel caso il numero di teste uscite sia sia superiore a 60 o inferiore a 40, siamo fiduciosi che la moneta sia truccata con un livello di significatività del 96%
- nel caso il numero di teste uscite sia superiore a 70 o inferiore a 30, siamo fiduciosi che la moneta sia truccata con un livello di significatività del 99%

Abbiamo in altre parole usato la variabile ξ che restituisce il numero di teste uscite per stimare la veridicità dell'affermazione:

La moneta non è truccata

Chiamiamo la variabile ξ stimatore e l'affermazione "La moneta non è truccata" Ipotesi H_0

Osserviamo che accettare che la moneta sia truccata, cioè rigettare l'ipotesi H_0 che "La moneta non è truccata" con un livello di significatività del 99% non vuol dire essere certi che la moneta è truccata con il 99% di possibilità in quanto non è tra l'altro affatto chiaro rispetto a quale spazio tale probabilità sia calcolata.

Potremmo anche calcolare la probabilità che il numero di teste uscito sia compreso tra n ed m utilizzando il fatto che la distribuzione binomiale si può approssimare con la distribuzione normale.

In tal caso dobbiamo normalizzare i dati osservando che

$$\begin{aligned}\mathcal{P}(n \leq \xi \leq m) &= \mathcal{P}\left(\frac{n - Np}{\sqrt{Npq}} \leq \frac{\xi - Np}{\sqrt{Npq}} \leq \frac{m - Np}{\sqrt{Npq}}\right) = \\ &= G\left(\frac{m - Np}{\sqrt{Npq}}\right) - G\left(\frac{n - Np}{\sqrt{Npq}}\right)\end{aligned}$$

dove

$$G(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$$

è la funzione distribuzione cumulativa normale standardizzata (gaussiana).

Otterremo i seguenti risultati che differiscono di poco dai precedenti

$$\mathcal{P}(45 \leq \xi \leq 55) \approx 0.68$$

$$\mathcal{P}(40 \leq \xi \leq 60) \approx 0.95$$

$$\mathcal{P}(30 \leq \xi \leq 70) \approx 0.99$$

e potremmo trarre conclusioni simili alle precedenti.

Qualora fossimo interessati a verificare l'ipotesi H_0 a fronte dell'ipotesi H_1 che la probabilità di successo p sia inferiore a 0.5, potremmo calcolare

$$\mathcal{P}(\xi \leq 55) \approx 0.86$$

$$\mathcal{P}(\xi \leq 60) \approx 0.98$$

$$\mathcal{P}(\xi \leq 70) \approx 0.99$$

e concludere

- nel caso il numero di teste uscite sia superiore a 55, siamo fiduciosi che la moneta sia truccata con $p > .05$ con un livello di significatività del 86%
- nel caso il numero di teste uscite sia superiore a 60, siamo fiduciosi che la moneta sia truccata con $p > .05$ con un livello di significatività del 98%
- nel caso il numero di teste uscite sia superiore a 70, siamo fiduciosi che la moneta sia truccata con $p > .05$ con un livello di significatività del 99%

Questo esempio mostra come sia possibile trarre conclusioni sulle caratteristiche di una popolazione (l'esito di tutti i possibili lanci di una moneta) esaminandone soltanto un campione (l'esito di un numero finito di lanci della moneta) e stimando le caratteristiche mediante la somma degli esiti (1 nel caso di T , successo, 0 in caso contrario).

Per poter procedere occorre precisare meglio i concetti di popolazione, campione e stimatore.

10.1. Popolazioni. Diciamo che è assegnata una popolazione se è assegnato un insieme \mathcal{U} ed una variabile aleatoria ξ definita su \mathcal{U} .

Chiamiamo popolazione l'insieme $\xi(\mathcal{U})$.

Se ad esempio siamo interessati a stimare il diametro di 9 sferette di acciaio contenute in un cuscinetto, possiamo indicare ciascuna sferetta con un indice e considerare la popolazione i cui elementi sono

$$\{s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, s_9\}$$

In questo caso \mathcal{U} è costituito dalle 9 sferette

$$\mathcal{U} = \{s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, s_9\}$$

e possiamo considerare la variabile aleatoria ξ che associa ad ogni sferetta il suo diametro $\xi(k) = d_k$ in modo che

$$\xi(\mathcal{U}) = \{d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, d_9\}$$

La variabile ξ descrive la popolazione che stiamo esaminando.

Se fossimo interessati a studiare il diametro delle sferette che costituiscono la produzione mensile di una macchina, sarebbe molto scomodo elencarle tutte e quindi potremmo descrivere la popolazione assegnando per la variabile ξ una funzione di distribuzione, ad esempio gaussiana di medio μ e varianza σ^2 .

Nel caso del lancio della moneta la popolazione è costituita da tutti i lanci.

10.2. Campioni. Data una popolazione definita dalla variabile aleatoria ξ sullo spazio \mathcal{U} diciamo che è assegnato un campione di taglia n se sono assegnate n variabili aleatorie X_1, X_2, \dots, X_n sullo spazio \mathcal{U} . Indichiamo con X la n -pla delle variabili aleatorie X_1, X_2, \dots, X_n che costituiscono il campione.

Ad esempio possiamo considerare

- X_1 come la variabile aleatoria che restituisce il diametro di una sferetta scelta in \mathcal{U}
- X_2 come la variabile aleatoria che restituisce il diametro di una seconda sferetta scelta in \mathcal{U}
- X_k come la variabile aleatoria che restituisce il diametro di una k -esima sferetta scelta in \mathcal{U}

Nel caso del lancio di una moneta il campione è dato dalle N variabili aleatorie che restituiscono gli esiti degli N lanci effettuati.

10.3. Stimatore. Diciamo che è assegnato uno stimatore, o riassunto campionario, se è assegnata una funzione $\phi = \phi(X_1, X_2, \dots, X_n)$ delle n variabili aleatorie X_1, X_2, \dots, X_n che costituiscono il campione.

Possiamo ad esempio considerare uno stimatore del diametro delle sferette considerando la media dei diametri delle n sferette di cui abbiamo esaminato il diametro per costruire il campione.

In tal caso

$$\bar{X} = \phi(X_1, X_2, \dots, X_n) = \frac{1}{n}(X_1 + X_2 + \dots + X_n) = \frac{1}{n}(d_1 + d_2 + \dots + d_n)$$

Nel caso del lancio di una moneta lo stimatore è costituito dalla somma dei risultati di ciascun lancio, per cui ϕ è il numero di successi ottenuti su n lanci.

Uno stimatore è a sua volta una variabile aleatoria.

Uno stimatore spesso usato è costituito dalla media dei campioni, come visto nell'esempio delle sferette); indicheremo tale stimatore con

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$$

11. Risultati sulle distribuzioni campionarie

Si consideri una popolazione individuata da una variabile aleatoria ξ su un insieme \mathcal{U}

un campione $X = (X_1, X_2, \dots, X_n)$ di taglia n estratto dalla popolazione individuata da ξ ed \mathcal{U}

Si possono provare alcuni fatti:

11.1. Se $\mu = E(\xi)$ è la media della popolazione ed \bar{X} è la media campionaria

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$$

allora

- $E(\bar{X}) = \mu_{\bar{X}} = \mu$
- $E((X - \bar{X})^2) = \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$

11.2. Se ξ è distribuita normalmente con media μ e varianza σ^2 allora \bar{X} è distribuito normalmente con media μ e varianza σ^2/n .

11.3. Se ξ è distribuita con media μ e varianza σ^2 (anche non normalmente) allora

$$\frac{\bar{X} - \mu}{\sigma\sqrt{n}}$$

è asintoticamente ($n \geq 30$) distribuito normalmente con media μ e varianza σ^2/n .

11.4. Se ξ è distribuita normalmente con media μ e se

$$S = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{n}$$

è la varianza campionaria, allora la variabile aleatoria

$$\frac{\bar{X} - \mu}{S\sqrt{n-1}}$$

ha una distribuzione di probabilità di Student con $n - 1$ gradi di libertà.

11.5. Se ξ_1, ξ_2 sono due popolazioni normalmente distribuite con varianza σ_1^2, σ_2^2 e se X_1, X_2 sono due campioni di taglia m ed n rispettivamente estratti dalle due popolazioni, allora la variabile aleatoria

$$\frac{(mS_1^2)/((m-1)\sigma_1^2)}{nS_2^2/((n-1)\sigma_2^2)}$$

ha una distribuzione di probabilità di Fisher con $n-1, m-1$ gradi di libertà.

11.6. Supponiamo di disporre di un certo numero di dati in base ai quali la probabilità di accadimento di un evento A_k è x_k e supponiamo di voler verificare se la distribuzione di probabilità di tali eventi ha una funzione densità di probabilità ξ in base alla quale la probabilità attesa degli eventi A_k è ξ_k .

Possiamo considerare che la variabile aleatoria ξ descrive la popolazione degli eventi possibili di cui i valori x_k costituiscono un campione e possiamo considerare come stimatore la variabile aleatoria definita da

$$\chi^2 = \frac{(\xi_1 - x_1)^2}{x_1} + \frac{(\xi_2 - x_2)^2}{x_2} + \dots + \frac{(\xi_n - x_n)^2}{x_n}$$

a proposito della quale possiamo affermare che ha una distribuzione di probabilità χ^2

- $n-1$ gradi di libertà, nel caso in cui della distribuzione ξ siano noti i parametri
- $n-1-m$ gradi di libertà, nel caso in cui della distribuzione ξ si debbano ricavare i parametri per mezzo di m dati dei riassunti campionari.

Per avere informazioni su una variabile aleatoria è necessario conoscere la distribuzione di probabilità.

Possiamo, per analogia con quanto visto in precedenti casi, congetturare che la variabile aleatoria in esame abbia una particolare distribuzione di probabilità, tuttavia avremo successivamente la necessità di verificare se tale congettura è ragionevole.

In altri casi può accadere di aver bisogno di verificare la ragionevolezza di una ipotesi H_0 mediante affermazioni del tipo "se H_0 è vera allora la variabile aleatoria ξ ha una distribuzione di probabilità di un certo tipo (Gaussiana, χ^2 , T , F)"

In ogni caso, stabilita la funzione di distribuzione di probabilità che si vuole sottoporre a test, si procede come segue:

- si individua la precisione ϵ con cui si vuole procedere: se ad esempio si fissa $\epsilon = .05$ significa che si vuole essere certi al 95%
- si individuano α, β in modo che

$$\mathcal{P}(\alpha \leq \xi \leq \beta) = .95$$

- si estrae un valore per ξ e se tale valore cade in $[\alpha, \beta]$ si accetta l'ipotesi, in caso contrario si rifiuta.

11.7. Il Test χ^2 . supponiamo di disporre di una serie di dati e di voler verificare se essi sono distribuiti in accordo con una certa ipotesi che chiamiamo H_1 .

Come esempio possiamo adottare il famoso esperimento di Mendel:

Mendel incrociò tra di loro due tipi di piselli ciascuno dei quali aveva due caratteri

- la forma (Rotondo o angoloso, R oppure a)
- il colore (Giallo o verde, G oppure v)

Egli supponeva che

- Rotondo e Giallo fossero caratteri dominanti (distinti dall'iniziale Maiuscola)
- angoloso e verde fossero caratteri recessivi (iniziale minuscola)

egli supponeva cioè che dall'incrocio di due piante si ottenesse una terza pianta con i caratteri scelti tra i dominanti delle due che l'avevano generata.

Mendel ottenne 556 piselli che classificò secondo i caratteri appena descritti nella seguente tabella

Caratteri	Frequenza osservata
Rotondo Giallo	315
Rotondo verde	108
angoloso Giallo	101
angoloso verde	32

TABELLA 29.1. Risultati dell'esperimento di Mendel.

I tipi di piselli possibili sono pertanto 4 ed i possibili incroci sono 16 e possono essere elencati nella tabella 29.1 dove accanto alle caratteristiche delle piante A e B incrociate sono riportate le caratteristiche dell'incrocio.

A	+	B	→	C
RG	+	RG	→	RG
RG	+	Rv	→	RG
RG	+	aG	→	RG
RG	+	av	→	RG
Rv	+	RG	→	RG
Rv	+	Rv	→	Rv
Rv	+	aG	→	RG
Rv	+	av	→	Rv
aG	+	RG	→	RG
aG	+	Rv	→	RG

aG	+	aG	→	aG
aG	+	av	→	aG
av	+	RG	→	RG
av	+	Rv	→	Rv
av	+	aG	→	aG
av	+	av	→	av

TABELLA 29.2. Possibili incroci.

Dalla tabella si evince che

- un pisello Rotondo e Giallo si presenta in 9 casi su 16
- un pisello Rotondo e verde si presenta in 3 casi su 16
- un pisello angoloso e Giallo si presenta in 3 casi su 16
- un pisello angoloso e verde si presenta in 1 caso su 16

per cui possiamo dire che

$$\mathcal{P}(RG) = \frac{9}{16}556 = 321.75$$

$$\mathcal{P}(Rv) = \frac{3}{16}556 = 104.75$$

$$\mathcal{P}(RG) = \frac{3}{16}556 = 104.75$$

$$\mathcal{P}(RG) = \frac{1}{16}556 = 34.75$$

e possiamo affiancare alle frequenze osservate nella tabella 29.1 le frequenze previste per ciascun caso:

Caratteri	Frequenza osservata	frequenza prevista	Differenza
	f_o	f_p	$f_o - f_p$
Rotondo Giallo	315	312.75	2.25
Rotondo verde	108	104.25	3.75
angoloso Giallo	101	104.25	-3.25
angoloso verde	32	34.75	-2.75

TABELLA 29.3. Risultati dell'esperimento di Mendel e frequenze previste.

Possiamo ora chiederci se i dati sperimentali confermano l'ipotesi che abbiamo fatto sui caratteri delle due piante:

Rotondo dominante e angoloso recessivo,

Giallo dominante e verde recessivo.

Si può vedere che è lecito supporre che la variabile aleatoria

$$\xi_k = \frac{f_o - f_p}{\sqrt{f_p}}, \quad k = 1, 2, 3, 4$$

segue una distribuzione gaussiana con media 0 e varianza 1.

Pertanto la somma dei quadrati delle 4 variabili aleatorie indicate in tabella 29.3 segue una distribuzione χ^2 con $3 = 4 - 1$ gradi di libertà.

Calcoliamo quindi

$$\chi^2 = \frac{(315 - 312.75)^2}{312.75} + \frac{(108 - 104.25)^2}{104.25} + \frac{(101 - 104.25)^2}{104.25} + \frac{(32 - 34.75)^2}{34.75} \approx 0.47$$

e consideriamo la tavola che fornisce i valori di χ^2 con 3 gradi di libertà.

Possiamo osservare che la tavola fornisce l'indicazione che

$$\mathcal{P}(\chi^2 < .352) = .05$$

$$\mathcal{P}(\chi^2 < .584) = .10$$

per cui

$$\mathcal{P}(\chi^2 < .352) < \mathcal{P}(\chi^2 < .47) < \mathcal{P}(\chi^2 < .584)$$

e pertanto possiamo concludere che se gli eventi fossero casuali avremmo trovato un evento che accade con probabilità maggiore del 5% ma minore del 10%, per cui possiamo affermare che l'ipotesi fatta può essere accettata al livello del 90% ma deve essere rifiutata al livello del 95%

Un secondo esempio di applicazione del test χ^2 si trova considerando il seguente problema:

Si supponga di lanciare un dado a forma di tetraedro per 100 volte e di ottenere le seguenti uscite:

1	2	3	4
23	26	24	27

TABELLA 29.4

Ci si pone il problema di stabilire se il dado è non truccato.

La frequenza attesa per ciascuno dei punteggi è $\frac{100}{4} = 25$ e possiamo quindi calcolare

$$\chi^2 = \frac{(23 - 25)^2}{25} + \frac{(26 - 25)^2}{25} + \frac{(24 - 25)^2}{25} + \frac{(27 - 25)^2}{25} \approx .4$$

dalla tabella dei valori di χ^2 con $4 - 1 = 3$ gradi di libertà si ottiene che

$$\mathcal{P}(\chi^2 < .352) = .05$$

$$\mathcal{P}(\chi^2 < .584) = .10$$

per cui

$$\mathcal{P}(\chi^2 < .352) < \mathcal{P}(\chi^2 < .4) < \mathcal{P}(\chi^2 < .584)$$

per cui, come prima, se il dado fosse truccato, avremmo sperimentato un evento la cui probabilità di accadimento è compresa tra il 5% ed il 10%, per cui possiamo affermare che l'ipotesi fatta può essere accettata al livello del 90% ma deve essere rifiutata al livello del 95%

CAPITOLO 30

IL TEOREMA DEL LIMITE CENTRALE

Il teorema del limite centrale è un risultato di grande importanza in quanto sancisce il fatto che la sovrapposizione di un gran numero di variabili aleatorie aventi media e varianza comune conduce ad una variabile con distribuzione normale (gaussiana).

Più precisamente possiamo dire che

Siano

$$\xi_1, \xi_2, \dots, \xi_n$$

variabili aleatorie indipendenti aventi la stessa distribuzione di probabilità con media μ e varianza σ^2 .

Siano

$$\eta_1, \eta_2, \dots, \eta_n$$

le corrispondenti variabili normalizzate

$$\eta_k = \frac{\xi_k - \mu}{\sigma}$$

e consideriamo la variabile aleatoria ζ definita da

$$\zeta_n = \frac{\eta_1 + \eta_2 + \dots + \eta_n}{\sqrt{n}}$$

(Poichè $E(\eta_i) = 0$ avremo che $E(\zeta_n) = nE(\eta_1 + \eta_2 + \dots + \eta_n) = 0$ e inoltre, poichè $\text{Var}(\eta_i) = 1$ avremo che $\text{Var}(\eta_1 + \eta_2 + \dots + \eta_n) = n$ e quindi $\text{Var}(\zeta_n) = 1$)

Si ha

$$\zeta_n = \frac{\xi_1 + \xi_2 + \dots + \xi_n - n\mu}{\sqrt{n}\sigma}$$

e si può dimostrare che

$$\lim_{n \rightarrow +\infty} \mathcal{P}(\alpha \leq \zeta_n \leq \beta) = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-x^2/2} dx$$

o equivalentemente che

$$\lim_{n \rightarrow +\infty} \mathcal{P}(a \leq \xi_1 + \xi_2 + \dots + \xi_n \leq b) = \frac{1}{\sqrt{2\pi}} \int_{\frac{a-n\mu}{\sqrt{n}\sigma}}^{\frac{b-n\mu}{\sqrt{n}\sigma}} e^{-x^2/2} dx$$

1. Un esempio di applicazione

Dal teorema del limite centrale risulta che moltissime variabili aleatorie possono essere approssimate mediante una variabile che abbia distribuzione, gaussiana.

Se ad esempio consideriamo la variabile aleatoria ξ che restituisce il numero di successi in una serie di N prove bernoulliane con probabilità di successo p e probabilità di insuccesso $q = 1 - p$, possiamo calcolare la probabilità di ottenere un numero di successi compreso tra n ed m utilizzando la distribuzione di Bernoulli mediante la seguente formula

$$\mathcal{P}(n \leq \xi \leq m) = \sum_{k=n}^m \binom{N}{k} p^k q^{N-k}$$

È subito evidente che, se $m - n$ è grande non è facile portare a termine il calcolo; per ovviare a questo inconveniente possiamo allora utilizzare il teorema del limite centrale e sostituire alla distribuzione di Bernoulli una distribuzione Gaussiana.

Procederemo considerando ξ come la somma di N variabili Bernoulliane ξ_i

$$\xi = \xi_1 + \xi_2 + \xi_3 + \dots + \xi_N$$

che assumono ciascuna valori 0 con probabilità q ed 1 con probabilità p .

La media di ciascuna delle ξ_i sarà $\mu = p$ e la loro varianza sarà $\sigma = \sqrt{pq}$, pertanto

$$\begin{aligned} \mathcal{P}(n \leq \xi \leq m) &= \mathcal{P}(n \leq \xi_1 + \xi_2 + \xi_3 + \dots + \xi_N \leq m) \approx \\ &\approx \frac{1}{\sqrt{2\pi}} \int_{\frac{n-Np}{\sqrt{N}\sqrt{pq}}}^{\frac{m-Np}{\sqrt{N}\sqrt{pq}}} e^{-x^2/2} dx = G\left(\frac{m-Np}{\sqrt{N}\sqrt{pq}}\right) - G\left(\frac{n-Np}{\sqrt{N}\sqrt{pq}}\right) \end{aligned}$$

dove

$$G(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-x^2/2} dx$$

è la funzione di distribuzione cumulativa Gaussiana standardizzata i cui valori si trovano tabulati.

Se ad esempio supponiamo di effettuare $N = 100$ lanci di una moneta non truccata per cui $p = q = 1/2$, possiamo valutare la probabilità che si abbiano un numero di Teste compreso tra $n = 45$ e $m = 55$ mediante la

$$\begin{aligned} \mathcal{P}(45 \leq \xi \leq 55) &\approx G\left(\frac{55 - 50}{\sqrt{100}\sqrt{1/4}}\right) - G\left(\frac{45 - 50}{\sqrt{100}\sqrt{1/4}}\right) = \\ &G(1) - G(-1) \approx 0.68 \end{aligned}$$

Abbiamo visto che, se la moneta non è truccata, su 100 lanci possiamo aspettarci un numero di successi compreso tra 45 e 55 con probabilità

del 68% e quindi potremo dire che la probabilità che i successi non siano compresi tra 45 e 55 è del 32%.

Per verificare se la moneta è truccata con un livello di significatività del 32% potremmo allora effettuare una serie di lanci e rifiutare l'ipotesi della moneta non truccata se otteniamo un numero di successi maggiore di 55 o minore di 45.

Qualora volessimo una maggiore precisione, ad esempio del 60% dovremmo avere

$$G\left(\frac{m - Np}{\sqrt{N}\sqrt{pq}}\right) - G\left(\frac{n - Np}{\sqrt{N}\sqrt{pq}}\right) = .40$$

e quindi dovrebbe risultare

$$G\left(\frac{m - Np}{\sqrt{N}\sqrt{pq}}\right) = .5 + .2 = .7 \quad \text{e} \quad G\left(\frac{n - Np}{\sqrt{N}\sqrt{pq}}\right) = .5 - .2 = .3$$

da cui

$$\frac{m - Np}{\sqrt{N}\sqrt{pq}} = .53 \quad \text{e} \quad \frac{n - Np}{\sqrt{N}\sqrt{pq}} = -.53$$

Se ne ricava che

$$m = 50 + 5 * (.53) \approx 52 \quad \quad n = 50 - 5 * (.53) \approx 48$$

REGRESSIONE LINEARE: LA RETTA DEI MINIMI QUADRATI

Siano assegnate n coppie di dati (punti di \mathbb{R}^2)

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

e si consideri il problema di determinare l'equazione di una retta

$$y = ax + b$$

in corrispondenza della quale risulti minima la quantità

$$\epsilon(a, b) = \sum_{i=1}^n (y_i - ax_i - b)^2$$

$\epsilon(a, b)$ è una funzione convessa delle variabili (a, b) che tende a $+\infty$ per $(a, b) \rightarrow \infty$ e pertanto ammette uno ed un solo punto di minimo assoluto che si può trovare annullando $\nabla \epsilon$.

Per risolvere il problema dovremo pertanto risolvere il sistema definito dalle equazioni

$$\begin{cases} \frac{\partial \epsilon}{\partial a} = \sum_{i=1}^n -2(y_i - ax_i - b)x_i = 0 \\ \frac{\partial \epsilon}{\partial b} = \sum_{i=1}^n -2(y_i - ax_i - b) = 0 \end{cases}$$

Ne viene che

$$\begin{cases} \sum_{i=1}^n x_i y_i - a \sum_{i=1}^n x_i^2 - b \sum_{i=1}^n x_i = 0 \\ \sum_{i=1}^n y_i - a \sum_{i=1}^n x_i - b \sum_{i=1}^n 1 = 0 \end{cases}$$

ovvero

$$(31.1) \quad \begin{cases} \sum_{i=1}^n x_i y_i = a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i \\ \sum_{i=1}^n y_i = a \sum_{i=1}^n x_i + nb \end{cases}$$

Dalla seconda delle 31.1 si può vedere che

$$nb = \sum_{i=1}^n y_i - a \sum_{i=1}^n x_i$$

ed anche

$$b = \frac{\sum_{i=1}^n y_i}{n} - a \frac{\sum_{i=1}^n x_i}{n} = \bar{y} - a\bar{x}$$

dove

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}, \quad \bar{y} = \frac{\sum_{i=1}^n y_i}{n}$$

indicano la media dei valori x_i ed y_i , rispettivamente.

Dalla prima delle 31.1 si può invece ottenere che

$$\begin{aligned} \sum_{i=1}^n x_i y_i &= a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i = \\ &= a \sum_{i=1}^n x_i^2 + \sum_{i=1}^n x_i \left(\frac{\sum_{i=1}^n y_i}{n} - a \frac{\sum_{i=1}^n x_i}{n} \right) \end{aligned}$$

e

$$\begin{aligned} \sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n} &= a \left(\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n} \right) \\ n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i &= a \left(n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right) \end{aligned}$$

ed infine

$$a = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

Inoltre

$$\begin{aligned} nb &= \sum_{i=1}^n y_i - \sum_{i=1}^n x_i \left(\frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \right) = \\ &= \frac{1}{n (\sum_{i=1}^n x_i)^2 - (\sum_{i=1}^n x_i)^2} \left[n \sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i - \left(\sum_{i=1}^n x_i \right)^2 \sum_{i=1}^n y_i \right. \\ &\quad \left. - n \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i + \left(\sum_{i=1}^n x_i \right)^2 \sum_{i=1}^n y_i \right] \end{aligned}$$

e se ne conclude che

$$b = \frac{n \sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i - n \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

Ora, tenendo conto che

$$\begin{aligned}
 \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) &= \\
 &= \sum_{i=1}^n x_i y_i - \sum_{i=1}^n \bar{x} y_i - \sum_{i=1}^n x_i \bar{y} + \sum_{i=1}^n \bar{x} \bar{y} = \\
 &= \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} - n\bar{x}\bar{y} + n\bar{x}\bar{y} = \\
 &= \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} = \sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}
 \end{aligned}$$

e che

$$\begin{aligned}
 \sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n x_i^2 - 2 \sum_{i=1}^n \bar{x} x_i + \sum_{i=1}^n \bar{x}^2 = \\
 &= \sum_{i=1}^n x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2
 \end{aligned}$$

si ricava che

$$\begin{aligned}
 \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} &= \\
 &= \frac{1}{n} \frac{\sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} = \\
 &= \frac{\sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - n^2 \left(\frac{\sum_{i=1}^n x_i}{n} \right)^2} = a
 \end{aligned}$$

Pertanto possiamo esprimere a e b mediante le seguenti formule

$$(31.2) \quad \begin{cases} a = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ \bar{y} = a\bar{x} + b \end{cases}$$

La prima delle due uguaglianze premette di concludere che a è invariante rispetto alla traslazione degli assi: cioè usando $x - x_0$ ed $y - y_0$ in luogo di x ed y il valore di a non cambia.

La stessa trasformazione cambia invece il valore di b , come si vede dalla seconda uguaglianza da cui si vede anche che la retta di regressione passa per il punto di coordinate (\bar{x}, \bar{y}) che è il baricentro dei dati.

Possiamo anche osservare che a meno di operare una traslazione dei dati riportando l'origine degli assi nel baricentro (\bar{x}, \bar{y}) , possiamo supporre che

$$(31.3) \quad \begin{cases} a = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} \\ b = 0 \end{cases}$$

Ora, siano

- s_x^2 la varianza dei dati x_i

$$s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

- s_y^2 la varianza dei dati y_i

$$s_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}$$

- s_{xy} la covarianza dei dati (x_i, y_i)

$$s_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n}$$

Possiamo scrivere la retta di regressione nella forma

$$y - \bar{y} = \frac{s_{xy}}{s_x^2} (x - \bar{x})$$

e se invertiamo il ruolo di x e di y l'equazione diventa

$$x - \bar{x} = \frac{s_{xy}}{s_y^2} (y - \bar{y})$$

Possiamo misurare la correlazione tra i dati utilizzando il coefficiente definito da

$$(31.4) \quad r = \frac{s_{xy}}{s_x s_y}$$

mediante il quale le equazioni delle due rette prima introdotte diventano

$$\frac{y - \bar{y}}{s_y} = r \frac{x - \bar{x}}{s_x}$$

e

$$r \frac{y - \bar{y}}{s_y} = \frac{x - \bar{x}}{s_x}$$

Chiaramente le due rette coincidono soltanto nel caso in cui

$$r^2 = 1 \quad \text{cioè} \quad r = \pm 1$$

e il fatto che questo accada è indice della correlazione dei dati cioè del fatto che i dati si trovano su una retta.

È ragionevole quindi stimare la maggiore o minore correlazione tra i dati confrontando r^2 con 1: più r^2 è vicino ad 1 e più i dati sono da considerarsi linearmente correlati.

Possiamo inoltre misurare la dispersione dei dati attorno alla retta di regressione mediante la

$$S_{y,x} = \sqrt{\frac{\sum_{i=1}^n (y_i - y_i^s)^2}{n}}$$

dove

$$y_i^s = ax_i + b$$

Pertanto

$$\begin{aligned} S_{y,x}^2 &= \frac{\sum_{i=1}^n (y_i - ax_i - b)^2}{n} = \sum_{i=1}^n (y_i^2 + a^2 x_i^2 + b^2 - 2ax_i y_i - 2by_i + 2abx_i) = \\ &= \sum_{i=1}^n y_i^2 + a^2 \sum_{i=1}^n x_i^2 + nb^2 - 2a \sum_{i=1}^n x_i y_i - 2b \sum_{i=1}^n y_i + 2ab \sum_{i=1}^n x_i = \end{aligned}$$

e per la definizione di a e b

$$\begin{aligned} &= \sum_{i=1}^n y_i^2 + a^2 \left[\frac{1}{a} \left(\sum_{i=1}^n x_i y_i - b \sum_{i=1}^n x_i \right) \right] + nb^2 - \\ &\quad - 2a \sum_{i=1}^n x_i y_i - 2b \sum_{i=1}^n y_i + 2ab \left[\frac{1}{a} \left(\sum_{i=1}^n y_i - nb \right) \right] = \\ &= \sum_{i=1}^n y_i^2 + a \sum_{i=1}^n x_i y_i - ab \sum_{i=1}^n x_i + nb^2 - \\ &\quad - 2a \sum_{i=1}^n x_i y_i - 2b \sum_{i=1}^n y_i + 2b \sum_{i=1}^n y_i - 2nb^2 = \\ &= \sum_{i=1}^n y_i^2 - a \sum_{i=1}^n x_i y_i - ab \sum_{i=1}^n x_i - nb^2 = \\ &= \sum_{i=1}^n y_i^2 - a \sum_{i=1}^n x_i y_i - b \left(a \sum_{i=1}^n x_i + nb \right) = \end{aligned}$$

e ancora per la definizione di a e b

$$= \sum_{i=1}^n y_i^2 - a \sum_{i=1}^n x_i y_i - b \sum_{i=1}^n y_i$$

Quindi

$$S_{y,x}^2 = \sum_{i=1}^n y_i^2 - a \sum_{i=1}^n x_i y_i - b \sum_{i=1}^n y_i$$

Si può d'altro canto verificare che

$$\sum_{i=1}^n y_i^2 - a \sum_{i=1}^n x_i y_i - b \sum_{i=1}^n y_i = \sum_{i=1}^n (y_i - \bar{y})^2 - a \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

infatti

$$\begin{aligned} & \sum_{i=1}^n (y_i - \bar{y})^2 - a \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \\ &= \sum_{i=1}^n y_i^2 - 2\bar{y} \sum_{i=1}^n y_i + n\bar{y}^2 - a \sum_{i=1}^n x_i y_i + a \sum_{i=1}^n x_i \bar{y} + \\ & \quad + a \sum_{i=1}^n \bar{x} y_i - a \sum_{i=1}^n \bar{x} \bar{y} = \\ &= \sum_{i=1}^n y_i^2 - 2n\bar{y}^2 + n\bar{y}^2 - a \sum_{i=1}^n x_i y_i + an\bar{x}\bar{y} + an\bar{x}\bar{y} - an\bar{x}\bar{y} = \\ &= \sum_{i=1}^n y_i^2 - n\bar{y}^2 - a \sum_{i=1}^n x_i y_i + an\bar{x}\bar{y} = \\ &= \sum_{i=1}^n y_i^2 - a \sum_{i=1}^n x_i y_i + n\bar{y}(\bar{y} - a\bar{x}) = \\ &= \sum_{i=1}^n y_i^2 - a \sum_{i=1}^n x_i y_i + n\bar{y}b = \\ &= \sum_{i=1}^n y_i^2 - a \sum_{i=1}^n x_i y_i + b \sum_{i=1}^n y_i \end{aligned}$$

Le precedenti considerazioni permettono quindi di affermare che

$$\begin{aligned} S_{y,x}^2 &= \sum_{i=1}^n y_i^2 - a \sum_{i=1}^n x_i y_i - b \sum_{i=1}^n y_i = \\ &= \sum_{i=1}^n (y_i - \bar{y})^2 - a \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \\ &= \frac{ns_y^2 - ans_{xy}}{n} = s_y^2 - as_{xy} \end{aligned}$$

e dal momento che $a = \frac{s_{xy}}{s_x}$

$$= s_y^2 \left(1 - \frac{s_{xy}}{s_y^2} \right) = s_y^2 \left(1 - \frac{s_{xy}^2}{s_x^2 s_y^2} \right) = s_y^2 (1 - r^2)$$

Ne viene quindi che

$$\frac{S_{y,x}}{s_y^2} = (1 - r^2)$$

e

$$r^2 = 1 - \frac{\sum_{i=1}^n (y_i - y_i^s)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

D'altro canto

$$\begin{aligned} \sum_{i=1}^n (y_i - \bar{y})^2 &= \sum_{i=1}^n (y_i - y_i^s + y_i^s - \bar{y})^2 = \\ &= \sum_{i=1}^n (y_i - y_i^s)^2 + \sum_{i=1}^n (y_i^s - \bar{y})^2 + 2 \sum_{i=1}^n (y_i - y_i^s)(y_i^s - \bar{y}) \end{aligned}$$

Poichè valgono le equazioni normali **31.1** che definiscono a e b

$$\begin{aligned} \sum_{i=1}^n (y_i - y_i^s)(y_i^s - \bar{y}) &= \sum_{i=1}^n (y_i - ax_i - b)(ax_i + b - \bar{y}) = \\ &= (b - \bar{y}) \sum_{i=1}^n (y_i - ax_i - b) + a \sum_{i=1}^n x_i (y_i - ax_i - b) \\ &= (b - \bar{y}) \left(\sum_{i=1}^n y_i - a \sum_{i=1}^n x_i - nb \right) + a \left[\sum_{i=1}^n x_i y_i - a \sum_{i=1}^n x_i^2 - b \sum_{i=1}^n x_i \right] = 0 \end{aligned}$$

avremo

$$\begin{aligned} r^2 &= 1 - \frac{\sum_{i=1}^n (y_i - y_i^s)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\sum_{i=1}^n (y_i - \bar{y})^2 - \sum_{i=1}^n (y_i - y_i^s)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \\ &= \frac{\sum_{i=1}^n (y_i^s - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\text{Variazione spiegata}}{\text{Variazione totale}} \end{aligned}$$

Possiamo anche calcolare, dalla 31.4, che

$$\begin{aligned}
 r &= \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{\left(n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2\right) \left(n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i\right)^2\right)}} = \\
 &= \frac{\overline{xy} - \bar{x}\bar{y}}{\sqrt{(\overline{x^2} - \bar{x}^2)(\overline{y^2} - \bar{y}^2)}} = \\
 &= \frac{n \sum_{i=1}^n x'_i y'_i - \sum_{i=1}^n x'_i \sum_{i=1}^n y'_i}{\sqrt{\left(n \sum_{i=1}^n x'^2_i - \left(\sum_{i=1}^n x'_i\right)^2\right) \left(n \sum_{i=1}^n y'^2_i - \left(\sum_{i=1}^n y'_i\right)^2\right)}} = \\
 &= \frac{\sum_{i=1}^n x'_i y'_i}{\sqrt{(\sum_{i=1}^n x'^2_i)(\sum_{i=1}^n y'^2_i)}}
 \end{aligned}$$

CAPITOLO 32

ANALISI DEI COMPONENTI PRINCIPALI.

Quando si esaminano i dati statistici relativi ad un elevato numero di caratteristiche di una popolazione è opportuno considerare i dati in un sistema di riferimento che ne metta in evidenza le caratteristiche salienti.

Supponiamo ad esempio di campionare una funzione f su un intervallo $[a, b]$; dividiamo cioè l'intervallo in $2n - 1$ parti uguali ed indichiamo con

$$a = t_1, t_2, \dots, t_{2n} = b$$

i punti mediante i quali tale suddivisione è operata.

Consideriamo poi n coppie di dati ottenute come segue

$$P_1 = (f(t_1), f(t_2)), P_2 = (f(t_3), f(t_4)), \dots, P_n = (f(t_{2n-1}), f(t_{2n}))$$

È ragionevole che, se la funzione è continua, ci siano piccole differenze tra ascissa e ordinata di ognuno dei punti P_k e pertanto essi tenderanno a disporsi lungo la bisettrice primo-terzo quadrante $y = x$.

Quindi potremo ottenere una rappresentazione più significativa dei dati se useremo come assi coordinati le bisettrici $y = x$ ed $y = -x$, cioè se opereremo una rotazione di 45° degli assi.

Ricordiamo che se $P = (x, y)$ sono le coordinate originali, si opera una rotazione degli assi applicando una trasformazione lineare la cui matrice di rappresentazione è

$$A = \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}$$

Pertanto

$$AP_1 = \left(\frac{\sqrt{2}}{2} f(t_1) + \frac{\sqrt{2}}{2} f(t_2), -\frac{\sqrt{2}}{2} f(t_1) + \frac{\sqrt{2}}{2} f(t_2) \right)$$

ed è evidente che delle due coordinate è rilevante solo la prima, mentre la seconda è piccola.

Ne viene che si possono ricostruire i valori originali con sufficiente precisione anche se le seconde coordinate dei punti trasformati sono trascurate.

Per illustrare il metodo occorre premettere qualche considerazione sulle forme quadratiche,

1. Forme quadratiche ed autovalori.

Sia A una matrice $n \times n$ e consideriamo la funzione

$$f(u) = \langle Au, u \rangle \quad , u \in \mathbb{R}^n$$

f si chiama forma quadratica su \mathbb{R}^n e si può vedere che è sempre possibile supporre che la matrice A che la individua sia simmetrica.

Infatti, ad esempio per $n = 2$, se

$$A_1 = \begin{pmatrix} a & d_1 \\ d_2 & c \end{pmatrix}$$

ed $u = \begin{pmatrix} x \\ y \end{pmatrix}$ avremo che

$$\begin{aligned} f(u) = \langle A_1 u, u \rangle &= ax^2 + d_1 xy + d_2 yx + cy^2 = ax^2 + (d_1 + d_2)xy + cy^2 = \\ &= ax^2 + 2bxy + cy^2 = \langle Au, u \rangle \end{aligned}$$

per $b = (d_1 + d_2)/2$, da cui

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

D'altro canto, se A è una matrice simmetrica possiamo verificare che

$$\langle Au, v \rangle = \langle u, Av \rangle$$

ed inoltre

$$f(u+h) = \langle A(u+h), (u+h) \rangle = \langle Au, u \rangle + 2\langle Au, h \rangle - \langle Ah, h \rangle$$

per cui

$$f(u+h) - f(u) = 2\langle Au, h \rangle + \langle Ah, h \rangle = 2\langle Au, u \rangle + \|h\|\omega(h)$$

con ω funzione infinitesima per $h \rightarrow 0$ (si ricordi che $|\langle Ah, h \rangle| \leq \|Ah\|\|h\|$ ed $Ah \rightarrow 0$ se $h \rightarrow 0$).

Dalla definizione di differenziale si ottiene allora che f è differenziabile e che

$$\nabla f(u) = 2Au$$

Come caso particolare, se $A = I$ (la matrice identica) si ha

$$g(u) = \langle u, u \rangle = \|u\|^2 \quad , \quad \nabla g(u) = 2u$$

Consideriamo ora il problema di trovare

$$\max_{g(u)-1=0} f(u) = \max_{\|u\|^2-1=0} \langle Au, u \rangle$$

Per il teorema di Weierstraß, dal momento che f è continua e che $\|u\|^2 - 1 = 0$ definisce la superficie della sfera di centro l'origine e raggio 1, che è

chiusa e limitata, possiamo affermare che il massimo esiste. Sia u_1 il punto in cui tale massimo è assunto

$$\langle Au_1, u_1 \rangle = \max_{\|u\|^2=1} \langle Au, u \rangle$$

$$\|u_1\|^2 = 1$$

D'altro canto, il teorema dei moltiplicatori di Lagrange consente di affermare che esiste λ_1 tale che

$$\nabla f(u_1) = \lambda_1 \nabla g(u_1)$$

per cui deve essere

$$Au_1 = \lambda_1 u_1$$

Dal momento che la precedente equazione è soddisfatta

- λ_1 è un autovalore di A
- u_1 è un autovettore di A corrispondente all'autovalore λ_1

Possiamo inoltre osservare che

$$\max_{\|u\|^2=1} \langle Au, u \rangle = \langle Au_1, u_1 \rangle = \langle \lambda_1 u_1, u_1 \rangle = \lambda_1 \|u_1\|^2 = \lambda_1$$

per cui λ_1 è il valore massimo assunto da $\langle Au, u \rangle$ sulla sfera $\|u\|^2 = 1$

Consideriamo ora lo spazio vettoriale V_1 generato da u_1

$$V_1 = \{ \lambda u_1 : \lambda \in \mathbb{R} \}$$

e lo spazio V_1^\perp ortogonale a V_1

$$V_1^\perp = \{ v \in \mathbb{R}^n : \langle v, u_1 \rangle = 0 \}$$

V_1^\perp è uno spazio vettoriale $(n-1)$ -dimensionale e la matrice A definisce una forma quadratica f_1 su V_1^\perp in quanto per ogni $v \in V_1^\perp$ si ha

$$\langle Av, u_1 \rangle = \langle v, Au_1 \rangle = \langle v, \lambda_1 u_1 \rangle = 0$$

e $Av \in V_1^\perp$

Se ora ripetiamo il procedimento per f_1 su V_1^\perp possiamo provare un nuovo autovalore λ_2 ed un nuovo autovettore u_2 con le stesse caratteristiche di λ_1 e u_1 .

Chiaramente si può iterare fino a trovare:

- n autovalori $\lambda_1, \lambda_2, \dots, \lambda_n$ decrescenti in valore
- n autovettori u_1, u_2, \dots, u_n , uno per ogni autovalore, che risultano ortogonali tra loro.

Gli autovettori u_1, u_2, \dots, u_n formano quindi una base ortonormale con la caratteristica che lungo il primo asse si trova il punto di massimo della forma quadratica $\langle Au, u \rangle$ sulla sfera unitaria in \mathbb{R}^2 , lungo il secondo asse si trova il massimo della forma quadratica $\langle Au, u \rangle$ sulla sfera unitaria in V_1^\perp e così via fino all' n -esimo asse.

2. L'applicazione all'analisi delle componenti principali.

Torniamo ora ai nostri dati $P_k = (u_k^1, u_k^2, \dots, u_k^n) \in \mathbb{R}^n$ e cerchiamo di individuare una combinazione lineare delle componenti

$$Q_k = \alpha_1 u_k^1 + \alpha_2 u_k^2 + \dots + \alpha_n u_k^n$$

in modo che la varianza di Q_k sia massima (massima significatività della variabile).

Definiamo

$$v_i = \text{Var}(u_k^i)$$

$$c_{ij} = \text{Cov}(u_k^i, u_k^j)$$

la varianza delle singole componenti e la covarianza delle componenti a due a due e sia R la **matrice di covarianza** dei dati definita mediante la

$$R = \begin{pmatrix} v_1 & c_{12} & c_{13} & \cdots & c_{1n} \\ c_{21} & v_2 & c_{23} & \cdots & c_{2n} \\ \cdots & \cdots & \cdots & \ddots & \cdots \\ c_{n1} & c_{n2} & c_{n3} & \cdots & v_n \end{pmatrix}$$

Possiamo allora verificare che

$$\text{Var}(Q_k) = \langle Ra, a \rangle$$

dove

$$a = (\alpha_1, \alpha_2, \dots, \alpha_n)$$

$\text{Var}(Q_k)$ è quindi una forma quadratica cui possiamo applicare il metodo visto nella precedente sezione e mediante tale metodo possiamo individuare in ordine decrescente di significatività le componenti dei dati.

2.1. Un esempio. Per illustrare gli effetti del metodo consideriamo i punti del grafico di $\sin(t)$ nell'intervallo $[0, 2\pi]$; suddividiamo l'intervallo in 199 parti uguali, in modo da individuare 200 punti in $[0, 2\pi]$ e calcoliamo i valori assunti da $\sin(t)$ in tali punti.

Le seguenti istruzioni di Matlab producono come risultato un vettore t che contiene i 200 punti in $[0, 2\pi]$, ed i vettori x ed y che contengono i valori assunti da $\sin(6t)$ nei punti pari e dispari rispettivamente

```
clear all;
step=2*pi/199;
t=0:step:2*pi;
xx=0:2*step: 2*pi;
yy=step:2*step: 2*pi;
x=sin(6*xx);
y=sin(6*yy);
```

Possiamo esaminare nel piano la distribuzione dei punti di coordinate $(x(k), y(k))$ con $k = 1, \dots, 100$ mediante le seguenti istruzioni

```
figure(1)
plot(x,y, '+' );
```

`axis([-2 2 -2 2]);`
 che producono la seguente figura 32.1

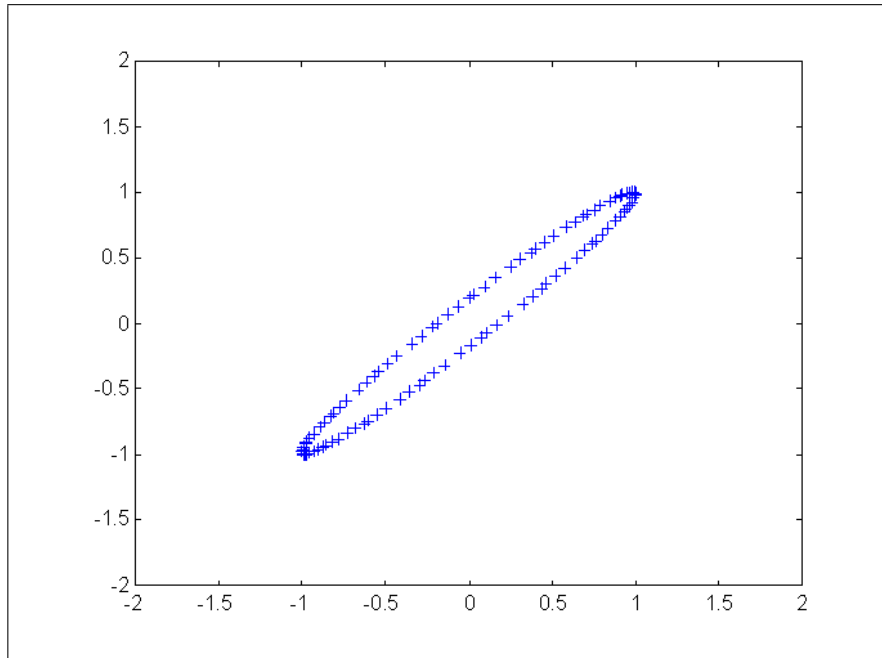


FIGURA 32.1.

Le istruzioni

```
R=cov(x,y);
[V,D]=eig(R)
```

calcolano la matrice di covarianza R e le matrici V e D , dove V è la matrice le cui colonne sono gli autovettori di R e D è una matrice diagonale con gli autovalori sulla diagonale principale; in altre parole V è la matrice tale che

$$VR = RD \quad \text{cioè tale che} \quad V^{-1}RV = D$$

La matrice V pertanto è la matrice di passaggio dal sistema di coordinate originale a quello individuato dagli autovalori di R .

Poichè ci interessa tener conto della componente relativa al massimo autovalore, di assicuriamo anche che l'autovalore più grande sia in posizione (1,1) mediante le istruzioni

```
if D(1,1)>D(2,2)
vv=V(:,1);
V(:,1)=V(:,2);
V(:,2)=vv;
end
```

La matrice V , quindi, può essere usata per effettuare un cambio di base che metta in evidenza le componenti principali.

Le seguenti istruzioni

```
tr=[x(:),y(:)]*V;
```

```

tr1=tr(:,1);
tr2=tr(:,2);
figure(2)
plot(tr1,tr2,'r+')
axis([-2 2 -2 2]);

```

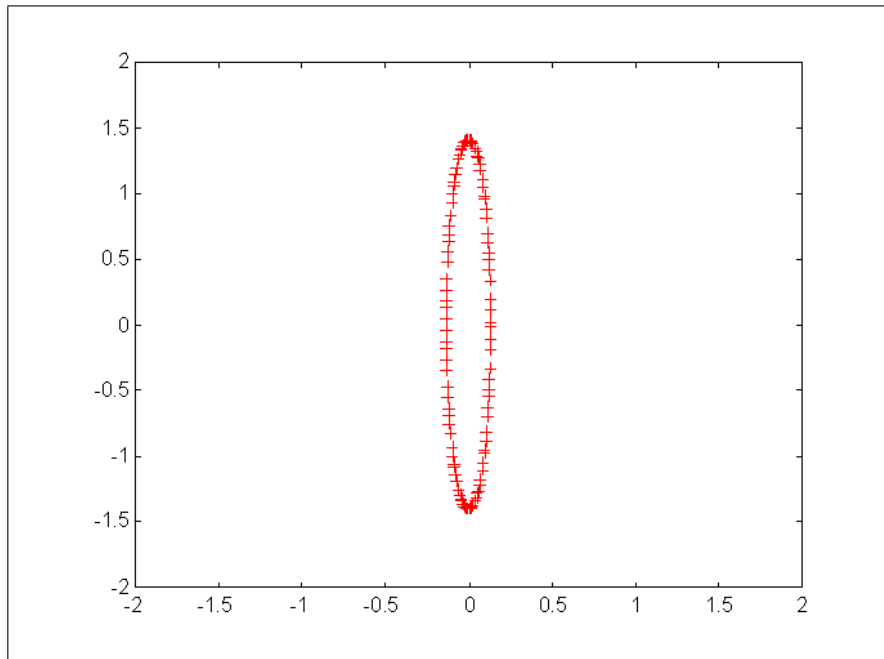


FIGURA 32.2.

Calcolano i trasformati $tr1$, $tr2$ dei punti (x, y) e li mostrano rispetto ad una coppia di assi ortogonali coincidenti con gli autovettori di R (si veda la figura 32.2).

Le successive istruzioni:

```

rtr=[tr1(:),tr2(:)]*inv(V);
rtr1=rtr(:,1);
rtr2=rtr(:,2);
figure(3)
plot(rtr1,rtr2,'g+')
axis([-2 2 -2 2]);

```

mostrano come applicando la trasformazione inversa i dati possano essere recuperati (si veda la figura 32.4.1).

Ora, possiamo osservare che la variazione della seconda componente dei dati trasformati è trascurabile rispetto alla prima, per cui, se la trascuriamo e applichiamo la trasformazione inversa ai dati privati di tale componente, otteniamo nuovi punti che differiscono di poco da quelli originali; possiamo operare usando le seguenti istruzioni:

```

nu=zeros(size(tr2));
ntr=[nu(:),tr2(:)]*inv(V);

```

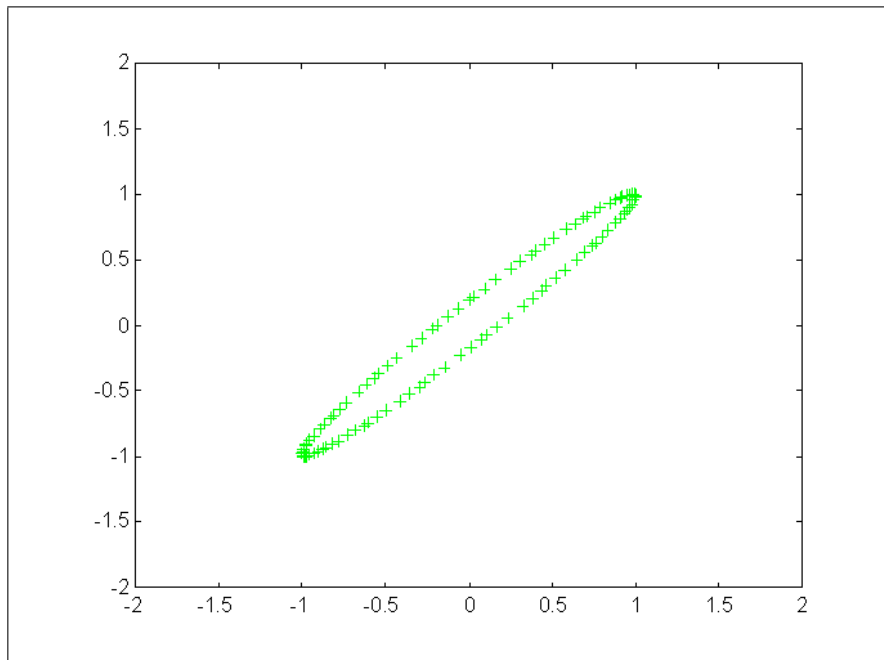


FIGURA 32.3.

```

ntr1=ntr(:,1);
ntr2=ntr(:,2);
figure(4)
plot(ntr1,ntr2,'rx',rtr1,rtr2,'g+')
axis([-2 2 -2 2]);

```

che forniscono anche una immagine dei nuovi punti come indicato in figura 32.4.1 ed un confronto con i punti originali 32.4.2.

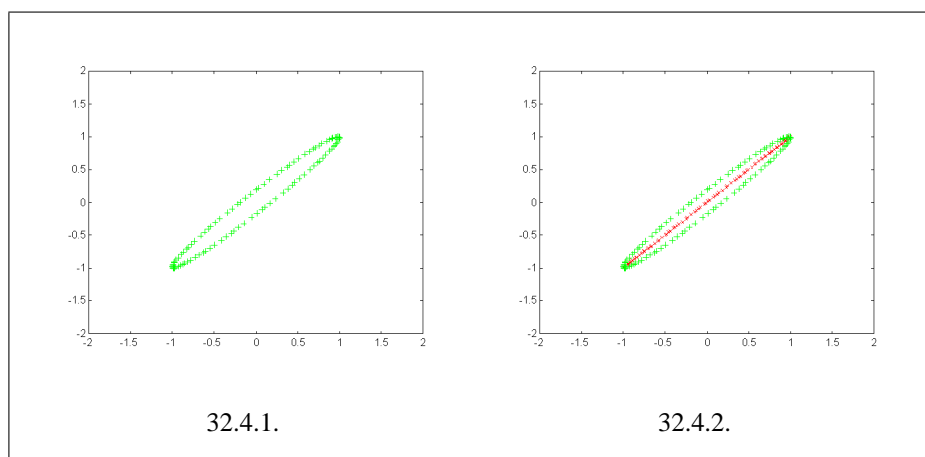


FIGURA 32.4.

È interessante ora osservare come dai punti originali e da quelli privati della componente meno significativa si possa ricostruire la funzione $\sin(6t)$

```

Le seguenti istruzioni
z=zeros(1,200);
zt=zeros(1,200);
for k=0:99
    zt(2*k+1)=x(k+1);
end
for k=1:100
    zt(2*k)=y(k);
end
for k=0:99
    z(2*k+1)=ntr2(k+1);
end
for k=1:100
    z(2*k)=ntr1(k);
end
figure(5)
plot(t,z)
figure(6)
plot(t,zt)

```

producono i due grafici riportati in figura 32.5, il primo dei quali riporta la funzione $\sin(t)$ ricostruita congiungendo con segmenti di retta i 200 punti originali, mentre la seconda riporta il grafico di $\sin(6t)$ ricostruito a partire dai punti ottenuti applicando la trasformazione inversa ai punti prima trasformati e poi privati della seconda componente.

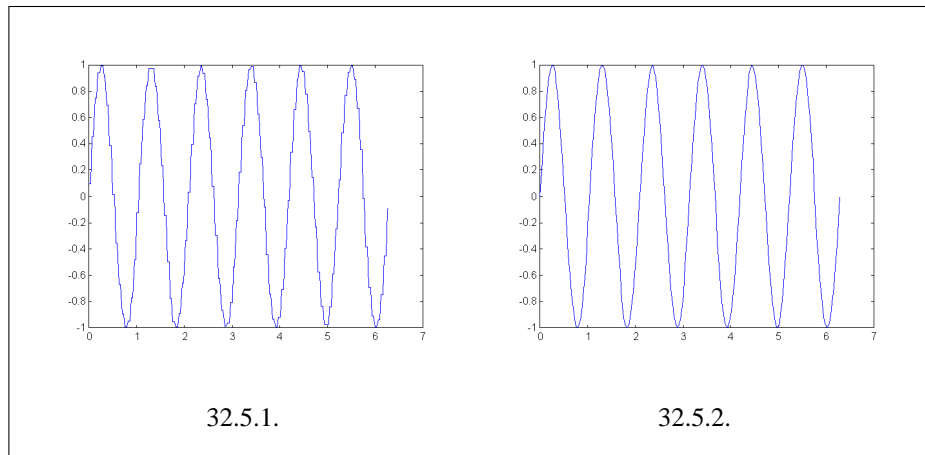


FIGURA 32.5.

Come si vede è evidente che la componente trascurata non ha peggiorato di molto il grafico, mentre la quantità di dati necessari a ricostruire l'immagine si è dimezzata.

Questo indica come può essere sviluppato un procedimento che consenta di immagazzinare dati (i punti del grafico della funzione) utilizzando al meglio le informazioni che contengono.

CAPITOLO 33

Appendice

1. LE FORMULE DI WALLIS E DI STIRLING-DE MOIVRE.

Intendiamo qui provare due formule che sono di grande utilità ed interesse.

La prima, nota come formula del prodotto di Wallis, consente di individuare una successione di razionali che ha per limite π , la seconda, dovuta a Stirling e De Moivre, consente di valutare, con notevole precisione, $n!$.

cominciamo con l'osservare che la definizione di fattoriale consente si affermare che

Si ha

$$(2n)!! = 2^n n! \quad , \quad (2n-1)!! = \frac{(2n)!}{2^n n!}$$

Possiamo provare il seguente teorema

TEOREMA 33.1. -Wallis - Posto

$$w_n = \frac{((2n)!!)^2}{((2n-1)!!)^2(2n+1)}$$

si ha

$$\lim w_n = \frac{\pi}{2}.$$

Infatti sia

$$s_n = \int_0^{\pi/2} (\sin x)^n dx$$

si ha

$$s_0 = \pi/2 \quad , \quad s_1 = 1$$

ed anche

$$\begin{aligned} s_n &= \int_0^{\pi/2} (\sin x)^{n-1} (\sin x) dx = -(\sin x)^{n-1} \cos x \Big|_0^{\pi/2} + \\ &\quad + \int_0^{\pi/2} (n-1)(\sin x)^{n-2} (\cos x)^2 dx = \\ &= (n-1) \int_0^{\pi/2} ((\sin x)^{n-2} - (\sin x)^n) dx = (n-1)(s_{n-2} - s_n) \end{aligned}$$

onde

$$s_n = \frac{n-1}{n} s_{n-2} \quad , \quad n \geq 2$$

Pertanto

$$\begin{aligned} s_{2n} &= \frac{2n-1}{2n} s_{2n-2} = \frac{2n-1}{2n} \frac{2n-3}{2n-2} s_{2n-4} = \\ &= \frac{2n-1}{2n} \frac{2n-3}{2n-2} \cdots \frac{1}{2} s_0 = \frac{(2n-1)!!}{(2n)!!} \frac{\pi}{2} \end{aligned}$$

Analogamente

$$s_{2n+1} = \frac{2n}{2n+1} \frac{2n-2}{2n-1} \cdots \frac{2}{3} s_1 = \frac{(2n)!!}{(2n+1)!!}.$$

Ora, poiché

$$0 \leq s_n = \int_0^{\pi/2} (\sin x)^n dx \leq \frac{\pi}{2},$$

mentre

$$s_{n+1} = \int_0^{\pi/2} (\sin x)^{n+1} dx \leq \int_0^{\pi/2} (\sin x)^n dx = s_n,$$

si ha che

$$\lim s_n = s \in \mathbb{R}.$$

Ma, come si verifica facilmente

$$w_n = \frac{s_{2n+1}}{s_{2n}} \frac{\pi}{2}$$

e pertanto

$$\lim w_n = \frac{s}{s} \frac{\pi}{2} = \frac{\pi}{2}$$

Se ora osserviamo che si ha

$$w_n = \frac{(2^n n!)^2 (2^n n!)^2}{((2n)!)^2 (2n+1)}$$

da cui

$$v_n = \frac{(n!)^2 2^{2n}}{(2n)! \sqrt{n}} = \sqrt{w_n \frac{(2n+1)}{n}} \longrightarrow \sqrt{\pi}$$

possiamo concludere che

Si ha

$$(33.1) \quad \lim \frac{(n!)^2 2^{2n}}{(2n)! \sqrt{n}} = \sqrt{\pi}$$

Come conseguenza della precedente formula possiamo provare la Formula di Stirling che è di grande utilità per approssimare i valori del fattoriale.

Formula di Stirling-De Moivre**Si ha**

$$\lim \frac{n^n e^{-n} \sqrt{2\pi n}}{n!} = 1$$

da cui

$$n! = n^n e^{-n} \sqrt{2\pi n} \theta_n, \quad \text{con} \quad \lim \theta_n = 1.$$

La verifica della formula di Stirling procede come di seguito illustrato. consideriamo la funzione $1/x$ sull'intervallo $[n, n+1]$; la figura 33.1 mostra come sia vera la disuguaglianza

$$\frac{2}{2n+1} = \frac{1}{n+1/2} \leq \int_n^{n+1} \frac{dx}{x} = \ln \left(1 + \frac{1}{n} \right).$$

Ne segue che

$$1 \leq (n+1/2) \ln(1+1/n)$$

ed

$$e \leq \left(1 + \frac{1}{n} \right)^{n+1/2}.$$

Tenendo conto che \ln è una funzione concava, si ottiene,

$$\begin{aligned} 1 + (\ln 2 + \ln 3 + \dots + \ln(n-1)) + \frac{1}{2} \ln n &\geq \\ &\geq \int_1^{3/2} \ln x dx + \int_{3/2}^{n-1/2} \ln x dx + \int_{n-1/2}^n \ln x dx = \\ &= \int_1^n \ln x dx = n \ln n - n + 1 = \\ &= \ln(n^n e^{-n}) + 1 \end{aligned}$$

da cui

$$\ln \left(\frac{n!}{\sqrt{n}} \right) \geq \ln(n^n e^{-n})$$

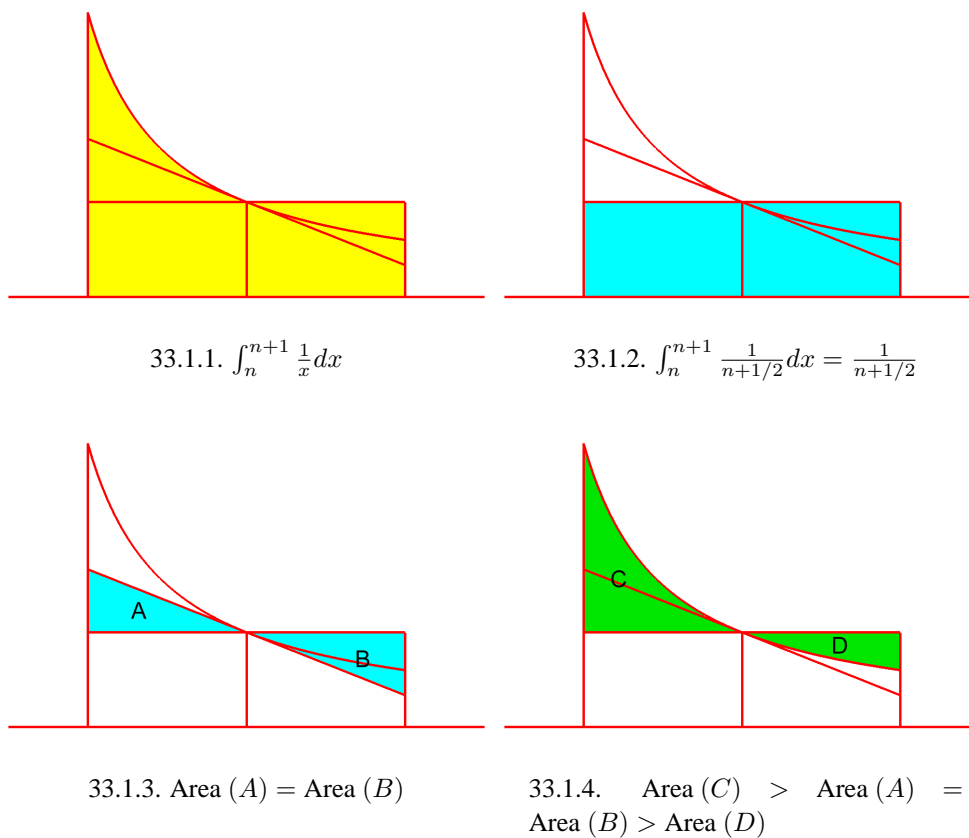


FIGURA 33.1. Grafici di funzioni di due variabili

$$0 \leq u_n = \frac{n^n e^{-n} \sqrt{n}}{n!} \leq 1.$$

Pertanto u_n è limitata e possiamo di più provare che è crescente, verificando che $u_{n+1}/u_n \geq 1$. Si ha infatti

$$\frac{u_{n+1}}{u_n} = \frac{(n+1)^{n+1} e^{-n-1} \sqrt{n+1}}{(n+1)!} \frac{n!}{n^n e^{-n} \sqrt{n}} = \frac{1}{e} \left(1 + \frac{1}{n}\right)^{n+1/2} \geq 1.$$

Possiamo a questo punto affermare che

$$u_n \rightarrow u \in \mathbb{R}$$

e ne consegue che

$$\frac{u_n^2}{u_{2n}} \rightarrow u.$$

Ma

$$\frac{u_n^2}{u_{2n}} = \frac{n^{2n} e^{-2n} n}{(n!)^2} \frac{(2n)!}{(2n)^{2n} e^{-2n} \sqrt{2n}} = \frac{(2n)! \sqrt{n}}{(n!)^2 2^{2n}} \frac{1}{\sqrt{2}}.$$

Da cui, per la **33.1**

$$u = \lim \frac{u_n^2}{u_{2n}} = \lim \frac{1}{v_n \sqrt{2}} = \frac{1}{\sqrt{2\pi}}$$

e si conclude che

$$\frac{n^n e^{-n} \sqrt{n}}{n!} \longrightarrow \frac{1}{\sqrt{2\pi}}$$

Possiamo anche osservare che,

$\forall n \in \mathbb{N}$

$$.367879 \leq \frac{1}{e} = u_1 \leq u_n \leq \lim u_n = \frac{1}{\sqrt{2\pi}} \leq .398943$$

Elenco delle figure

2.1	Costruzione della somma di due numeri reali	12
2.2	Costruzione del prodotto di due numeri reali	12
2.3	Sistema di riferimento Cartesiano	25
4.1	Potenze ad esponente naturale	35
4.2	Radici ad esponente naturale	35
4.3	Potenze ad esponente negativo	36
4.4	Potenze ad esponente positivo	38
4.5	Potenze ad esponente reale	40
4.6	Esponenziali di base $a > 1$	41
4.7	Esponenziali di base a , $0 < a < 1$	42
4.8	Esponenziali	42
4.9	Logaritmi in base maggiore di 1	43
4.10	Logaritmi in base minore di 1	44
4.11	Logaritmi	45
4.12	Una poligonale inscritta	45
4.13	Definizione di radiante	46
4.14	Definizione di sin e cos	47
4.15	Grafico della funzione sin	47
4.16	grafico della funzione cos	48
4.17	grafico della funzione tan	49
4.18	Grafico della funzione arcsin	49
4.19	Grafico della funzione arccos	50
4.20	Grafico della funzione arctan	51
4.21	Grafico della funzione $\arctan(\tan)$	52
4.22	grafico della funzione $\arcsin(\sin)$	52
4.23	Grafico della funzione $\arccos(\cos)$	53
14.1		119
14.2		120
14.3		120
14.4		122

14.5		123
14.6		124
14.7		125
14.8		126
17.1		137
17.2	Confronto tra somme superiori e somme inferiori	139
17.3	Confronto tra $U(f, P_n) - L(f, P_n)$ e $U(f, P_\epsilon) - L(f, P_\epsilon)$	143
17.4	Integrabilità delle funzioni monotone	146
18.1	Grafico di $f(x)$	159
18.2	Grafico di $g(x)$	160
19.1	Un punto materiale soggetto alla gravità	163
19.2	Il sistema di riferimento	164
21.1		176
21.2		177
21.3		178
21.4		179
21.5		181
21.6		182
21.7		183
22.1	Soluzioni del polinomio caratteristico reali distinte una positiva ed una negativa	199
22.2	Soluzioni del polinomio caratteristico reali distinte una positiva ed una negativa	199
22.3	Soluzioni del polinomio caratteristico reali distinte entrambe positive	200
22.4	Soluzioni del polinomio caratteristico complesse e coniugate con parte reale negativa	200
22.5	Soluzioni del polinomio caratteristico reali coincidenti negative	200
22.6	Grafico di A in funzione di α ed ω	201
23.1	Grafici di funzioni di due variabili	204
23.2	Curve di livello delle superfici in figura 23.1	205
23.3	Sezioni, per y fissato, dei grafici di figura 23.1	206
23.4	Sezioni, per x fissato, dei grafici di figura 23.1	207
23.5	Curve di livello e segno di f	211

23.6	Il grafico di una funzione derivabile parzialmente, non continua.	213
23.7	Definizione di derivata direzionale	215
23.8	Curve di livello e gradiente	217
23.9	Principio dei moltiplicatori di Lagrange	224
24.1		228
24.2		229
24.3	Singolo elemento di volume	229
24.4	$S(x)$	230
24.5	Dominio normale rispetto all'asse x	231
24.6	Dominio normale rispetto all'asse x	232
24.7	Volume ed area	234
24.8	Cambiamento di variabili lineare	235
24.9	Cambiamento di variabili in coordinate polari	236
25.1		240
25.2		241
28.1	Lo spazio \mathcal{U} degli eventi nel caso del lancio di due dadi	252
28.2	Istogramma relativo al lancio di due dadi, (supposti distinti)	253
28.3	Eventi che forniscono lo stesso valore per ξ , (congiunti da un segmento)	262
28.4	Distribuzione di probabilità della variabile aleatoria che restituisce il punteggio ottenuto lanciando due dadi.	265
28.5	Istogramma relativo alla variabile aleatoria che restituisce il punteggio ottenuto lanciando due dadi.	266
32.1		315
32.2		316
32.3		317
32.4		317
32.5		318
33.1	Grafici di funzioni di due variabili	322

Indice

Capitolo 1. UN PO' DI LOGICA	3
Capitolo 2. I NUMERI REALI	9
1. Rappresentazione dei numeri naturali in base b	19
2. Approssimazione dei numeri reali in base b	22
Capitolo 3. FUNZIONI REALI DI UNA VARIABILE REALE	27
Capitolo 4. LE FUNZIONI ELEMENTARI	33
1. Le funzioni Potenze	33
2. La funzione esponenziale	40
3. La funzione logaritmo	41
4. Le funzioni trigonometriche	44
Capitolo 5. DEFINIZIONE DI LIMITE E SUE CONSEGUENZE	55
Capitolo 6. LE SUCCESSIONI	63
1. Infinitesimi ed Infiniti	76
Capitolo 7. LA CONTINUITÀ	79
Capitolo 8. I TEOREMI SULLA CONTINUITÀ	81
Capitolo 9. LA DERIVABILITÀ.	87
Capitolo 10. I TEOREMI DI ROLLE, LAGRANGE E CAUCHY.	95
Capitolo 11. LA REGOLA DI DE L'HÔPITAL	99
Capitolo 12. LA FORMULA DI TAYLOR	103
Capitolo 13. QUALCHE SVILUPPO DI TAYLOR NOTEVOLE	109
1. Lo sviluppo di McLaurin di e^x	109
2. Lo sviluppo di McLaurin di $\sin x$	110
3. Lo sviluppo di McLaurin di $\cos x$	111
4. Lo sviluppo di McLaurin di $\ln(1 + x)$	112
5. Lo sviluppo di McLaurin di $\sqrt{1 + x}$	113
6. Lo sviluppo di McLaurin di $\frac{1}{1-x}$	114
7. Come ricavare altri sviluppi	116
Capitolo 14. LA CONVESSITÀ	119
Capitolo 15. ESTREMI RELATIVI E ASINTOTI.	129

Capitolo 16. RICERCA NUMERICA DI ZERI E MINIMI.	133
Capitolo 17. INTEGRAZIONE.	135
Capitolo 18. QUALCHE STUDIO DI FUNZIONE INTEGRALE	157
1. Esempio	158
2. Esempio	160
3. Esempio	161
Capitolo 19. INTRODUZIONE AI MODELLI DIFFERENZIALI	163
Capitolo 20. EQUAZIONI DIFFERENZIALI A VARIABILI SEPARABILI.	169
Capitolo 21. ESEMPI NOTEVOLI DI PROBLEMI DI CAUCHY	175
1. Esempio	175
2. Esempio	176
3. Esempio	177
4. Esempio	178
5. Esempio	180
6. Esempio	181
7. Esempio	183
Capitolo 22. SISTEMI ED EQUAZIONI DIFFERENZIALI LINEARI	185
1. L'oscillatore armonico	199
Capitolo 23. FUNZIONI DI DUE VARIABILI	203
1. La struttura di \mathbb{R}^2 .	206
2. Limiti e continuità per le funzioni di 2 variabili.	209
3. Derivabilità e differenziabilità per funzioni di 2 variabili.	212
4. Derivate del secondo ordine: forma quadratica Hessiana.	218
5. Massimi e minimi per le funzioni di 2 variabili.	220
6. Massimi e minimi vincolati.	222
Capitolo 24. INTEGRAZIONE PER LE FUNZIONI DI DUE VARIABILI	227
1. Definizione di integrale doppio.	227
2. Formule di riduzione per gli integrali doppi.	229
3. Cambiamento di variabili negli integrali doppi	233
Capitolo 25. INTEGRAZIONE DI FUNZIONI DI TRE VARIABILI	239
1. Definizione di integrale triplo.	239
2. Formule di riduzione per gli integrali tripli.	239
3. Cambiamento di variabili per gli integrali tripli.	240
Capitolo 26. INTEGRALI MULTIPLI IMPROPRI	243
1. Qualche esempio	244
Capitolo 27. LA SOMMA DI INFINITI TERMINI: LE SERIE.	247

Capitolo 28. ELEMENTI DI PROBABILITÀ E STATISTICA.	251
1. Spazi di probabilità	251
2. Qualche applicazione della regola di Bayes.	256
3. Qualche richiamo di calcolo combinatorio.	259
4. Le variabili aleatorie	262
5. Variabili aleatorie continue	267
6. Distribuzioni di probabilità doppie	271
7. Qualche proprietà di valor medio e varianza	272
8. La disuguaglianza di Tchebichev e la legge dei grandi numeri	274
9. Normalizzazione di una variabile aleatoria.	276
Capitolo 29. QUALCHE DISTRIBUZIONE DI PROBABILITÀ	279
1. La distribuzione uniforme	279
2. La distribuzione binomiale di Bernoulli	280
3. La distribuzione di Poisson	281
4. La distribuzione gaussiana	282
5. Due distribuzioni discrete	285
6. Le distribuzioni legate ai test statistici.	286
7. La distribuzione γ .	288
8. La distribuzione β .	288
9. Variabili casuali con distribuzione assegnata.	288
10. I test statistici	290
11. Risultati sulle distribuzioni campionarie	294
Capitolo 30. IL TEOREMA DEL LIMITE CENTRALE	299
1. Un esempio di applicazione	300
Capitolo 31. REGRESSIONE LINEARE: LA RETTA DEI MINIMI QUADRATI	303
Capitolo 32. ANALISI DEI COMPONENTI PRINCIPALI.	311
1. Forme quadratiche ed autovalori.	312
2. L'applicazione all'analisi delle componenti principali.	314
Capitolo 33. Appendice	319
1. LE FORMULE DI WALLIS E DI STIRLING-DE MOIVRE.	319
Elenco delle figure	325
Indice analitico	333

Indice analitico

Symbols

0, 9
1, 10
 $\cos(x)$, 46
 $\sin(x)$, 46
 $\tan(x)$, 48
è implicato, 4

Bolzano-Weierstraß , 66
complementare, 6
Criterio di Cauchy, 62
integrale indefinito, 149
integrale inferiore, 139
integrale superiore, 139
intersezione, 6
Principio di Archimede, 16
prodotto cartesiano, 6
relazione di equivalenza, 6
semifattoriale, 70
somma di Cauchy-Riemann, 138

A

and, 3
aperto, 25, 26
Approssimazione dei numeri reali in base b , 22
argomento di f , 27
asintoto, 131
assi cartesiani, 24

B

binomio di Newton, 71

C

chiuso, 26
coefficiente binomiale, 70
compatto, 26
concava, 127
condizione di integrabilità, 140
continua, 79

continua in D , 79
convessa, 122
crescente, 29, 64, 174
Criterio di Cauchy, 75
Criterio di convergenza di Cauchy, 68

D

De-Morgan, 6
decrecente, 29
dei numeri razionali, 12
derivabile, 87
derivata, 87
derivata destra, 88
derivata seconda, 92
derivata sinistra, 88
destrorsa, 24
differenziabile, 89
differenziabilità, 89
differenziale, 199
dimostrazione per assurdo, 5
divergente, 64
dominio, 27

E

elemento neutro, 9, 21
elemento neutro rispetto alla moltiplicazione, 10
elemento neutro rispetto alla somma, 9
elemento separatore, 10
equazione differenziale, 169
equivalente, 4
esponenziale, 40
estremo inferiore, 15
estremo superiore, 15

F

fattoriale, 70
flesso, 127
formula di Taylor, 103
Formula di Taylor con il resto di Lagrange , 106

Formula di Taylor con il resto di Peano ,

105

funzione, 27, 28

funzione limitata, 30

funzione composta, 28

funzione dispari, 29

funzione logaritmo, 41

funzione pari, 29

funzione potenza ad esponente razionale,

37

funzione superiormente limitata, 30

funzioni trigonometriche, 44

G

grafico, 27

Gronwall, 173

I

implica, 4

inferiormente limitato, 14

infinitesimo campione, 76

iniettiva, 28

insieme, 5

insieme compatto, 76

insieme dei numeri interi, 14

insieme dei numeri razionali, 14

insieme derivato, 55

insieme induttivo, 12

integrabile, 139

integrabile secondo Cauchy-Riemann, 140

integrale generale, 186, 195, 196

integrazione per parti, 150

integrazione per sostituzione, 151

intorno bucato di centro x e ampiezza δ ,

55

intorno di centro x e ampiezza δ , 55

inversa, 28

inverso rispetto all'addizione, 10

inverso rispetto alla moltiplicazione, 10

invertibile, 28

L

l'integrale generale, 185

legge del terzo escluso, 4

legge di cancellazione, 11

legge di non contraddizione, 4

limitato, 14, 26

limite, 55

lineare, 199

logaritmo naturale, 73

lunghezza di un arco, 44

M

maggiorante, 14

massimo, 14

matrice fondamentale, 188

metodo della 'regula falsi', 133

Metodo della regula falsi, 134

metodo delle tangenti, 133

metodo di Lagrange di variazione delle

costanti arbitrarie , 193

metodo di Newton, 133

Metodo di Newton (o delle tangenti), 133

minimo, 14

minorante, 14

modulo, 17

monometrica, 25

monotona, 29

N

norma, 17

not, 3

notazione binaria, 21

notazione decimale, 21

notazione esadecimale, 21

notazione ottale, 21

numeri decimali finiti, 24

numeri interi, 12

numeri naturali, 12

numeri reali, 9

O

or, 3

ordine di infinitesimo, 76

ordine di infinito, 76

origine, 24

ortogonale, 24

oscillatore armonico, 199

P

parte intera, 17

partizione dell'intervallo, 135

partizione di $[a, b]$, 137

partizione più fine, 138

periodica, 48

potenza, 18

potenza ad esponente frazionario, 36

potenza ad esponente reale, 38

potenza di esponente n , 33

potenze ad esponente negativo, 34

primitiva, 149

principio di induzione, 13

problema di Cauchy, 169–171, 186

prodotto cartesiano, 24

prolungamento, 28
 proposizione, 3
 punto di accumulazione, 55
 punto di massimo assoluto, 30
 punto di minimo assoluto, 30

R

radiante, 46
 radice n-esima, 34
 rango, 27
 Rappresentazione dei numeri naturali in
 base b, 19
 regola di De l'Hôpital, 99
 relazione binaria, 5
 relazione d'ordine, 6
 restrizione, 28

S

sinistrorsa, 24
 sistema cartesiano, 25
 sistema fondamentale di soluzioni, 191
 somme inferiori, 138
 somme superiori, 138
 strettamente convessa, 122
 strettamente crescente, 29
 strettamente decrescente, 29
 strettamente monotona, 29
 successione, 63
 successione convergente, 64
 successione crescente, 65
 successione estratta, 64
 successione ordinata di partizioni, 138
 superiormente limitato, 14
 surgettiva, 28
 sviluppo di McLaurin di $\cos x$, 111
 sviluppo di McLaurin di $\sin x$, 110
 sviluppo di McLaurin di e^x , 109

T

tabella di verità, 3
 teorema degli zeri, 81
 teorema dei valori intermedi, 82
 teorema della media, 148
 teorema di Weierstraß, 82
 teorema di Cauchy, 97
 teorema di Lagrange, 96
 teorema di Peano, 96
 teorema di Rolle, 95
 triangolo di Tartaglia, 70, 71

U

unione, 6

V

valore assoluto, 17
 variabili separabili, 169
 vuoto, 6

W

wronskiano, 189

X

xor, 3

Z

zero, 21