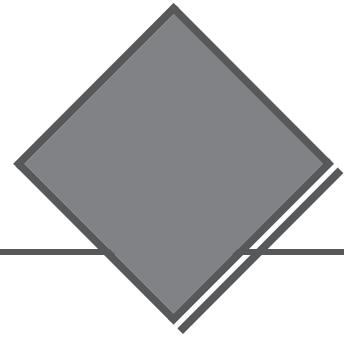


Curiosità

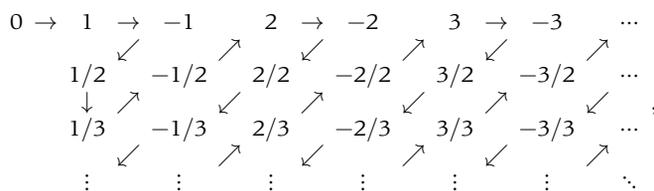


Vengono qui riportate tutte le curiosità presenti nel volume di Marco Abate *Matematica e Statistica - Le basi per le scienze della vita - Terza edizione* che sono evidenziate con l'icona .

In questa sezione le figure, le tabelle e le formule sono numerate con una "C" finale, i riferimenti senza la "C" finale rimandano al testo.

Ogni curiosità inizia a pagina nuova e vi si può facilmente accedere tramite la funzionalità dei segnalibri del PDF.

Curiosità 1.1 (Come contare i numeri razionali) È possibile elencare uno per volta, in un ordine preciso, tutti i numeri razionali. Per farlo, basta seguire le frecce nella seguente tabella:



dove la prima riga contiene tutti i numeri interi, la seconda riga tutte le frazioni con denominatore 2, la terza riga tutte le frazioni con denominatore 3, e così via. Seguendo le frecce si percorrono tutte le diagonali una alla volta, e si ottiene un elenco di tutti i numeri razionali semplicemente togliendo le ripetizioni (una volta messo 1 nell'elenco non è più necessario aggiungervi $2/2$ o $3/3$, una volta messo $-3/2$ non è necessario metterci $-6/4$ o $-9/6$, eccetera).

Questo vuol dire che è possibile *contare* i numeri razionali; i numeri razionali sono tanti quanti i numeri naturali. In termini tecnici, si dice che esiste una **corrispondenza biunivoca** (o una **bigezione**) fra l'insieme \mathbb{Q} e l'insieme \mathbb{N} . Ne ripareremo nella Curiosità 2.1.

Curiosità 1.2 ($\sqrt{2}$ è irrazionale) Come si fa a vedere che $\sqrt{2}$ è irrazionale? Bisogna verificare che nessun numero razionale elevato al quadrato faccia esattamente due. Supponiamo che non sia così; allora deve esistere un numero razionale p/q (ridotto ai minimi termini: in particolare p e q non possono essere entrambi pari) che elevato al quadrato faccia 2, cioè tale che $(p/q)^2 = 2$. Questo è equivalente a dire che

$$p^2 = 2q^2 .$$

In particolare, p^2 è pari. Ma l'unico modo in cui un quadrato può essere pari è che anche la base sia pari; quindi p è pari, e possiamo scrivere $p = 2n$. Ma allora $4n^2 = p^2 = 2q^2$, per cui

$$q^2 = 2n^2 .$$

Ma questo vuol dire che anche q è pari, per cui p/q non è ridotta ai minimi termini, contro quanto avevamo supposto all'inizio. Quindi supporre che esista un numero razionale che elevato al quadrato faccia 2 conduce necessariamente a una contraddizione, per cui un tale numero razionale non può esistere, e $\sqrt{2}$ è irrazionale.

In maniera analoga puoi provare a verificare che anche $\sqrt{3}$ è irrazionale e, se te la senti, puoi cercare di scoprire per quali numeri naturali n la radice quadrata \sqrt{n} è irrazionale (vedi l'Esercizio 1.6).

Curiosità 1.3 (Esistenza delle radici) Un dettaglio sottile ma non irrilevante da tener presente: così come abbiamo fatto vedere (nella Curiosità 1.2) che non esiste un numero razionale il cui quadrato sia 2, dovremmo far vedere che invece *esiste* un numero reale il cui quadrato sia 2. Senza questa informazione, $\sqrt{2}$ sarebbe soltanto un simbolo tipografico, una speranza fatta segno, e non un numero reale in carne ed ossa. Per fortuna, esistono delle procedure (che forse avrai studiato alle scuole medie, e che sono usate dalle calcolatrici) per costruire esplicitamente un numero reale il cui quadrato sia uguale a 2, risolvendo questo dilemma. Si ottiene

$$\sqrt{2} = 1.414\ 213\ 562\ 373\ \dots ,$$

dove i puntini indicano cifre decimali che non si ripetono mai in modo periodico, e che quindi non siamo in grado di predire conoscendo solo le precedenti.

Più in generale, esistono anche procedure per dimostrare l'esistenza (e per costruire) la radice q -esima di qualsiasi numero reale positivo.

Curiosità 1.4 (Quanti sono i numeri reali) Una delle conseguenze della costruzione dei numeri reali come numeri decimali infiniti è che mentre i numeri razionali sono tanti quanti i numeri naturali, i numeri reali sono sostanzialmente di più: non esiste alcuna corrispondenza biunivoca fra i numeri naturali (o interi o razionali) e i numeri reali. Sono talmente tanti di più che se si prende un punto a caso sulla retta si è praticamente certi di beccare un numero irrazionale (ed è possibile dare un significato matematicamente preciso a questa frase). Ne ripareremo nella Curiosità 2.1.

Curiosità 1.5 (Distanze e spazi metrici) I matematici hanno la (sana) abitudine di prendere le proprietà di un oggetto utile e trasformarle nella definizione di qualcosa di più generale che si spera sia altrettanto utile dell'oggetto di partenza. Per esempio, se X è un insieme qualsiasi, allora i matematici chiamano **distanza** sull'insieme X qualunque regola che permette di associare a ogni coppia di punti x e y di X un numero reale $d(x, y)$ non negativo in modo che siano valide le proprietà (1.1). La coppia formata dall'insieme X e dalla distanza d è chiamata **spazio metrico**, ed è uno dei concetti più importanti della matematica contemporanea.

Curiosità 1.6 (Proprietà delle potenze) A voler essere onesti, bisogna ammettere che nel Paragrafo 1.2 non abbiamo davvero fatto vedere che le proprietà (1.2) e (1.3) valgono anche per le potenze di esponente razionale. Rimediamo subito. Per far vedere che $a^{p/q} \cdot a^{r/s} = a^{(p/q)+(r/s)}$ ci basta far vedere (perché?) che $(a^{p/q} \cdot a^{r/s})^{qs} = a^{ps+rq}$. E infatti

$$\begin{aligned}(a^{p/q} \cdot a^{r/s})^{qs} &= (a^{p/q})^{qs} (a^{r/s})^{qs} = (a^{1/q})^{pqs} (a^{1/s})^{rqs} \\ &= ((a^{1/q})^q)^{ps} ((a^{1/s})^s)^{rq} = a^{ps} \cdot a^{rq} = a^{ps+rq},\end{aligned}$$

come voluto. In modo analogo (esercizio per te) si verifica la (1.3).

Curiosità 1.7 (Troncamenti e arrotondamenti) Dire che tronchiamo un numero alla k -esima cifra decimale vuol dire che scartiamo le cifre decimali dalla $(k + 1)$ -esima in poi. L'errore che introduciamo può quindi variare da 0 (se tutte le cifre decimali scartate erano uguali a 0) a 10^{-k} (se tutte le cifre decimali scartate erano uguali a 9).

Supponiamo invece di voler arrotondare alla k -esima cifra decimale. Se la prima cifra decimale scartata è fra 0 e 4, l'errore che introduciamo arrotondando varia da 0 (se tutte le cifre decimali scartate erano uguali a 0) a un massimo di $5 \cdot 10^{-(k+1)}$ (se le cifre decimali scartate erano un 4 seguito da infiniti 9). Analogamente, se la prima cifra decimale è scartata è fra 5 e 9, l'errore che introduciamo arrotondando varia da 0 (se tutte le cifre decimali scartate erano uguali a 9) a un massimo di $5 \cdot 10^{-(k+1)}$ (se le cifre decimali scartate erano un 5 seguito da infiniti 0). Quindi in ogni caso l'errore che si effettua arrotondando è minore o uguale a quello che si effettua troncando, e l'errore massimo per arrotondamento è di $5 \cdot 10^{-(k+1)} = 0.5 \cdot 10^{-k}$, la metà dell'errore massimo per troncamento.

Curiosità 1.8 (Teoria non intuitiva degli insiemi) Se vuoi farti venire un mal di testa, considera l'insieme U di tutti gli insiemi che non sono elementi di se stessi (per esempio, l'insieme dei felini non è un felino, per cui non è un elemento di se stesso), e chiediti se U è un elemento di se stesso oppure no. Questo esempio mostra come l'apparentemente semplice teoria degli insiemi può avere degli aspetti estremamente delicati, che i matematici stanno esplorando da un paio di secoli.

Curiosità 1.9 (Logica non elementare) La logica ordinaria è una logica a due valori: ogni proposizione o è vera o è falsa (*tertium non datur*, non esiste una terza possibilità). Questo è anche noto come principio del *terzo escluso*, e fu già enunciato da Aristotele nella sua *Metafisica*.

Al fine di non incorrere in paradossi, è importantissimo distinguere fra gli oggetti studiati (oggetti matematici, logici, eccetera) e il linguaggio usato per studiarli. Se i due piani si confondono, possono nascere contraddizioni che minacciano l'intera costruzione logica. Il paradosso logico più famoso è descritto dalla frase seguente:

Questa frase è falsa.

Per il principio del terzo escluso la frase è vera o falsa. Se supponiamo sia vera, allora è falsa, impossibile. Ma se supponiamo sia falsa, allora non lo è, contraddizione. Quindi questa frase non è né vera né falsa: paradosso!

Il problema sembrerebbe risiedere nel fatto che la frase parla di se stessa (autoreferenzialità). Invece possiamo ottenere lo stesso paradosso con una coppia di frasi che non parlano di se stesse:

- (1) *La frase successiva è vera.*
(2) *La frase precedente è falsa.*

Infatti, per il principio del terzo escluso, la frase (1) o è vera o è falsa. Se è vera, allora è vera anche la (2), e quindi la (1) è falsa, assurdo. Se invece la (1) è falsa, allora la frase (2) è falsa, e dunque la frase (1) è vera, contraddizione. Quindi abbiamo di nuovo un paradosso.

Usando una successione infinita di frasi possiamo costruire un esempio in cui nessuna frase della lista si riferisce a una precedente, ma nonostante ciò non è possibile decidere la verità o falsità di nessuna di esse:

- (1) *Tutte le frasi successive sono false.*
(2) *Tutte le frasi successive sono false.*
⋮
(n) *Tutte le frasi successive sono false.*
⋮

Infatti, non appena una di queste frasi è vera, la successiva dev'essere falsa, per cui tutte quelle dopo dovrebbero essere vere, contraddicendo la verità di quella precedente. Quindi dovrebbero essere tutte false, ma allora sono tutte vere: paradosso.

Questi che sembrano puri giochetti retorici sono in realtà al centro di uno dei più importanti risultati matematici del ventesimo secolo, il *teorema di incompletezza di Gödel*, la cui dimostrazione usa ragionamenti analoghi ma espressi usando soltanto numeri.

Curiosità 2.1 (Cardinalità di insiemi) Le funzioni bigettive permettono di stabilire quando due insiemi qualsiasi, anche infiniti, hanno lo stesso numero di elementi (o, con una terminologia più tecnica, quando hanno la stessa cardinalità). Infatti, si dice che due insiemi A e B hanno la stessa **cardinalità** se esiste una funzione $f: A \rightarrow B$ bigettiva. Se ci pensi un attimo, vedrai che dire che due insiemi finiti A e B hanno la stessa cardinalità vuol dire che li possiamo contare nello stesso modo, cioè che hanno lo stesso numero di elementi; e i matematici hanno scoperto che questo modo di pensare funziona bene anche per insiemi infiniti.

Per esempio, la funzione $g: \mathbb{Z} \rightarrow \mathbb{N}$ definita da

$$g(x) = \begin{cases} 2x & \text{se } x \geq 0, \\ 2|x| - 1 & \text{se } x < 0, \end{cases}$$

è bigettiva (controlla, notando che manda i numeri interi positivi nei numeri pari e i numeri interi negativi nei numeri dispari), per cui l'insieme \mathbb{Z} dei numeri interi e l'insieme \mathbb{N} dei numeri naturali hanno lo stesso numero di elementi, nonostante \mathbb{N} sia un sottoinsieme proprio di \mathbb{Z} (e nella Curiosità 1.1 abbiamo visto che anche l'insieme \mathbb{Q} dei numeri razionali ha la stessa cardinalità di \mathbb{N}). In realtà, questo non è (troppo) strano: infatti, si può far vedere che un insieme è infinito se e solo se ha la stessa cardinalità di un suo sottoinsieme proprio.

Possiamo usare le funzioni anche per dire quale fra due insiemi è più grande. Diremo che un insieme A ha **cardinalità minore o uguale** a quella di un insieme B se esiste una funzione *iniettiva* $f: A \rightarrow B$; e che ha **cardinalità maggiore o uguale** a quella di B se esiste una funzione *surgettiva* $g: A \rightarrow B$. Nota che perché queste definizioni abbiano senso bisogna verificare diverse cose: per esempio che se A ha cardinalità minore o uguale a quella di B (cioè se esiste una funzione iniettiva $f: A \rightarrow B$) e A ha anche cardinalità maggiore o uguale a quella di B (cioè esiste una funzione surgettiva $g: A \rightarrow B$), allora A ha la stessa cardinalità di B (cioè esiste una funzione $h: A \rightarrow B$ bigettiva). Tutto ciò si può fare, ma è il contenuto di un risultato tutt'altro che banale, il *Teorema di Cantor-Bernstein*.

Con questa terminologia possiamo dire, come accennato nella Curiosità 1.4, che l'insieme \mathbb{R} dei numeri reali ha cardinalità strettamente più grande di quella di \mathbb{N} : infatti, esistono (prova a costruirle per esercizio) sia funzioni iniettive da \mathbb{N} in \mathbb{R} sia funzioni surgettive da \mathbb{R} su \mathbb{N} , ma Cantor nella seconda metà del 1800 ha dimostrato che non può esistere alcuna corrispondenza biunivoca fra \mathbb{N} ed \mathbb{R} .

Curiosità 2.2 (Relazioni d'equivalenza) Fra le numerosissime relazioni studiate dai matematici uno dei tipi più comuni è quello delle relazioni d'equivalenza. Diremo che una relazione $R \subseteq A \times A$ fra gli elementi di un insieme A è una **relazione d'equivalenza** se:

1. è *riflessiva*, nel senso che ogni elemento è in relazione con se stesso: $(a, a) \in R$ per ogni $a \in A$;
2. è *simmetrica*, nel senso che se un elemento a è in relazione con b allora anche b è in relazione con a : $(a, b) \in R$ implica $(b, a) \in R$;
3. è *transitiva*, nel senso che se a è in relazione con b e b è in relazione con c allora a è in relazione anche con c : se $(a, b) \in R$ e $(b, c) \in R$ allora $(a, c) \in R$.

Per esempio, la relazione "essere parenti di" è una relazione d'equivalenza, mentre "essere genitori di" non lo è (non è né riflessiva né simmetrica né transitiva...).

Curiosità 2.3 (Cambiamento di sistema di riferimento) Scegliendo nel piano un sistema di riferimento diverso, allo stesso punto P vengono associate coordinate diverse; che relazione c'è fra loro? Indichiamo con \mathcal{A}^2 il piano euclideo, e con $\mathcal{R} = R(O, A_1, A_2)$ e $\mathcal{R}' = R(O', A'_1, A'_2)$ due sistemi di riferimento sul piano. Il sistema \mathcal{R} ci fornisce una bigezione $f: \mathcal{A}^2 \rightarrow \mathbb{R}^2$, che associa a ciascun punto $P \in \mathcal{A}^2$ la coppia di coordinate $f(P) = (x, y)$. Analogamente, il sistema \mathcal{R}' ci fornisce una seconda bigezione $f': \mathcal{A}^2 \rightarrow \mathbb{R}^2$, che associa al punto $P \in \mathcal{A}^2$ le coordinate $f'(P) = (x', y')$. La composizione $f' \circ f^{-1}: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ è quindi una bigezione di \mathbb{R}^2 con se stesso, che manda le coordinate (x, y) nelle coordinate (x', y') ; in altre parole, la funzione $f' \circ f^{-1}$ è il *cambiamento di coordinate* dal sistema \mathcal{R} al sistema \mathcal{R}' . Si può dimostrare che

$$\begin{cases} x' = b_{11}x + b_{12}y + c_1, \\ y' = b_{21}x + b_{22}y + c_2, \end{cases}$$

dove $(c_1, c_2) = f'(O)$ sono le coordinate di O nel sistema di riferimento \mathcal{R}' , e $b_{11}, b_{12}, b_{21}, b_{22}$ sono opportuni numeri reali.

Curiosità 2.4 (Teorema di Bezout) L'introduzione delle coordinate cartesiane permette di studiare insiemi geometrici molto più generali di quelli alla portata dei geometri greci, limitati a lavorare con rette, circonferenze e poco altro. Un esempio di risultato completamente fuori dalla portata dei geometri classici è il *teorema di Bezout*. Supponiamo di avere due insiemi del piano S_1 e S_2 , descritti rispettivamente dalle equazioni $F_1(x, y) = 0$ e $F_2(x, y) = 0$, dove F_1 e F_2 sono due polinomi, in due variabili, di grado rispettivamente d_1 e d_2 , e supponiamo che l'insieme $S_1 \cap S_2$ sia finito. Allora il teorema di Bezout dice che l'intersezione $S_1 \cap S_2$ contiene al più $d_1 d_2$ punti. Per esempio, una retta (che è data da un'equazione di grado 1) interseca un insieme S descritto da un polinomio di grado d in al più d punti (oppure in infiniti punti, nel qual caso si può far vedere che la retta è tutta contenuta in S).

Curiosità 3.1 (Prodotto di vettori applicati) Forse ti chiederai perché nel Paragrafo 3.1 non abbiamo tentato di definire su \mathcal{V}_0^2 e su \mathcal{V}_0^3 anche un prodotto fra vettori applicati che avesse le usuali proprietà del prodotto fra numeri reali (associatività, distributività rispetto alla somma, commutatività, eccetera). Il fatto è che mentre in \mathcal{V}_0^2 questo si può fare (sono i numeri complessi; vedi la Curiosità 9.6), in \mathcal{V}_0^3 un tale prodotto *non esiste*. La dimostrazione di questo risultato a prima vista sorprendente è, come puoi immaginare, piuttosto difficile.

Curiosità 3.2 (Spazi vettoriali) Uno **spazio vettoriale** su \mathbb{R} è un insieme V su cui sono definite una somma e un prodotto per scalari che godono delle stesse proprietà algebriche della somma e del prodotto per scalari di vettori applicati. Per la precisione, devono valere le seguenti proprietà:

- (a) $v_1 + (v_2 + v_3) = (v_1 + v_2) + v_3$ per ogni $v_1, v_2, v_3 \in V$;
- (b) esiste un elemento $O \in V$, detto **vettore nullo** tale che $v + O = O + v = v$ per ogni $v \in V$;
- (c) per ogni $v \in V$ esiste un elemento $-v \in V$, detto **opposto**, tale che $v + (-v) = (-v) + v = O$;
- (d) $v_1 + v_2 = v_2 + v_1$ per ogni $v_1, v_2 \in V$;
- (e) $\lambda(v_1 + v_2) = \lambda v_1 + \lambda v_2$ per ogni $\lambda \in \mathbb{R}$ e $v_1, v_2 \in V$;
- (f) $(\lambda + \mu)v = \lambda v + \mu v$ per ogni $\lambda, \mu \in \mathbb{R}$ e $v \in V$;
- (g) $(\lambda\mu)v = \lambda(\mu v)$ per ogni $\lambda, \mu \in \mathbb{R}$ e $v \in V$;
- (h) $1v = v$ e $0v = O$ per ogni $v \in V$.

Abbiamo visto (Osservazione 3.3) che è possibile definire sull'insieme \mathbb{R}^n delle n -uple di numeri reali una somma e un prodotto per scalari che lo rendono uno spazio vettoriale. Inoltre è facile verificare (controlla) che l'insieme di tutti i polinomi, l'insieme di tutte le funzioni continue (vedi il Paragrafo 5.7) e l'insieme di tutte le funzioni derivabili (vedi il Paragrafo 7.2), considerati con le usuali operazioni di somma e prodotto per un numero reale, sono altri esempi di spazi vettoriali.

Curiosità 3.3 (Prodotto scalare) Siamo arrivati al concetto di prodotto scalare in \mathbb{R}^2 e in \mathbb{R}^3 partendo dalla geometria del piano e dello spazio euclidei, e l'abbiamo usato per *calcolare* lunghezze e angoli di vettori. Le formule che abbiamo visto suggeriscono una naturale generalizzazione al caso di \mathbb{R}^n : definiamo il **prodotto scalare** di due elementi di \mathbb{R}^n ponendo

$$\left\langle \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \right\rangle = x_1 y_1 + \cdots + x_n y_n .$$

Puoi verificare facilmente che, se mettiamo su \mathbb{R}^n le operazioni di somma e prodotto per scalari componente per componente descritte nell'Osservazione 3.3, il prodotto scalare gode delle proprietà elencate subito dopo l'Osservazione 3.5:

$$\begin{aligned} \langle v_1 + v_2, v_3 \rangle &= \langle v_1, v_3 \rangle + \langle v_2, v_3 \rangle ; & \langle v_2, v_1 \rangle &= \langle v_1, v_2 \rangle ; & \langle \lambda v_1, v_2 \rangle &= \lambda \langle v_1, v_2 \rangle ; \\ \langle v_1, v_1 \rangle &\geq 0 & \text{e } \langle v_1, v_1 \rangle &= 0 \text{ se e solo se } v_1 = O \end{aligned} \quad (3.1C)$$

per ogni $v_1, v_2, v_3 \in \mathbb{R}^n$ e $\lambda \in \mathbb{R}$.

Più in generale, un **prodotto scalare** su uno spazio vettoriale V (vedi la Curiosità 3.2) è una qualsiasi funzione $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$ che soddisfa le (3.1C). Usando un prodotto scalare possiamo *definire* la lunghezza di un elemento di V , e il coseno dell'angolo di due elementi di V , con le stesse formule viste per il piano e per lo spazio. Nota che perché queste siano definizioni sensate occorre verificare che il prodotto scalare goda di proprietà opportune. Per esempio, siccome il coseno di un angolo è sempre un numero compreso fra -1 e 1 , possiamo usare (3.10) in V solo se

$$|\langle v_1, v_2 \rangle| \leq \|v_1\| \|v_2\|$$

per ogni coppia di elementi v_1, v_2 di V . Fortunatamente questa disuguaglianza (detta *disuguaglianza di Cauchy-Schwarz*) è sempre verificata, per cui tutto funziona. In particolare, possiamo dire che due elementi $v_1, v_2 \in V$ sono **ortogonali** se e solo se $\langle v_1, v_2 \rangle = 0$, e si verifica che v_1 e v_2 sono proporzionali se e solo se $|\langle v_1, v_2 \rangle| = \|v_1\| \|v_2\|$.

Quest'ultima osservazione ci permette di dare una nuova interpretazione del coefficiente di Pearson introdotto nel Paragrafo 11.3 per valutare la bontà dell'interpolazione fornita dalla retta di regressione; vedi la Curiosità 11.3.

Curiosità 4.1 (Sottospazi vettoriali) Un sottoinsieme $W \subseteq V$ di uno spazio vettoriale V (vedi la Curiosità 3.2) è detto **sottospazio vettoriale** se è chiuso rispetto alla somma e al prodotto per scalari, cioè se ogni volta che prendiamo $v, w \in W$ e $\lambda \in \mathbb{R}$ anche $v + w$ e λv appartengono a W .

Per esempio, l'insieme dei polinomi che si annullano in 0 è un sottospazio vettoriale dello spazio vettoriale di tutti i polinomi, in quanto la somma di due polinomi che si annullano in 0 continua ad annullarsi in zero, e un multiplo di un polinomio che si annulla in zero continua ad annullarsi in zero.

Nota che un sottospazio vettoriale W deve sempre contenere il vettore nullo, in quanto si ha $O = 0w \in W$ per ogni $w \in W$. Quindi l'insieme delle soluzioni di un sistema lineare *non* omogeneo *non* è un sottospazio vettoriale, in quanto non contiene il vettore nullo.

Curiosità 4.2 (Applicazioni lineari) Un'applicazione lineare fra due spazi vettoriali (vedi la Curiosità 3.2) V e W è una funzione $L: V \rightarrow W$ tale che

$$L(v_1 + v_2) = L(v_1) + L(v_2) \quad \text{e} \quad L(\lambda v_1) = \lambda L(v_1)$$

per ogni $v_1, v_2 \in V$ e $\lambda \in \mathbb{R}$. Per esempio, la funzione che associa a un polinomio p il suo valore in 1 è una funzione lineare; nel Capitolo 7 vedremo che l'operazione di derivata è lineare, e nel Capitolo 8 vedremo che l'operazione di integrale definito è lineare.

Si può anche dimostrare che le applicazioni lineari da \mathbb{R}^n a \mathbb{R}^m sono tutte e sole quelle della forma L_A con $A \in M_{m,n}(\mathbb{R})$.

Curiosità 4.3 (Trasposizione) Un'altra operazione utile sulle matrici è la **trasposizione**, ottenuta scambiando righe e colonne. Più precisamente, se $A = (a_{ij}) \in M_{m,n}(\mathbb{R})$, allora la **trasposta** di A è la matrice $A^T \in M_{n,m}(\mathbb{R})$ che al posto (i, j) ha l'elemento a_{ji} . Per esempio,

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \implies A^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}.$$

Una delle proprietà principali della trasposizione, di dimostrazione non semplice, è che non modifica il rango:

$$\text{rg } A^T = \text{rg } A$$

per ogni matrice $A \in M_{m,n}(\mathbb{R})$.

Invece non è difficile vedere come si comporta la trasposizione rispetto alle operazioni sulle matrici: si ha

$$(A + B)^T = A^T + B^T, \quad (\lambda A)^T = \lambda A^T \quad \text{e} \quad (AB)^T = B^T A^T.$$

Inoltre, se A è una matrice quadrata invertibile, allora anche A^T è invertibile e l'inversa della trasposta è la trasposta dell'inversa: $(A^T)^{-1} = (A^{-1})^T$.

Una delle conseguenze dell'esistenza degli sviluppi di Laplace sia lungo le righe sia lungo le colonne è il fatto che il determinante di una matrice è uguale al determinante della trasposta:

$$\det A^T = \det A.$$

Infine, supponiamo di avere due vettori $x, y \in \mathbb{R}^n$. Allora possiamo pensarli come matrici con n righe e 1 colonna, per cui i loro trasposti diventano matrici con 1 riga ed n colonne. In particolare, possiamo moltiplicare la riga y^T per la colonna x ottenendo

$$y^T x = \begin{pmatrix} y_1 & \cdots & y_n \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_1 y_1 + \cdots + x_n y_n = \langle x, y \rangle.$$

In altre parole, il prodotto scalare canonico di x e y si ottiene moltiplicando il trasposto di y per x . Nota infine che anche il prodotto xy^T ha senso, ma è una matrice quadrata di ordine n contenenti tutti i possibili prodotti fra le coordinate di x e quelle di y .

Curiosità 4.4 (Unicità del determinante) Si può dimostrare che esiste un'unica funzione $\det: M_{n,n}(\mathbb{R}) \rightarrow \mathbb{R}$ che soddisfa le proprietà (D.1)-(D.4) del Paragrafo 4.5, ed è tale che $\det I_n = 1$; inoltre questa funzione dev'essere necessariamente data dalla formula (4.23). Siccome gli sviluppi di Laplace godono di queste proprietà e assumono il valore corretto sulla matrice identica, devono coincidere con questa funzione, per cui possono essere usati per calcolare il determinante.

Curiosità 4.5 (Matrici simmetriche) Una classe significativa di matrici che ammettono sempre autovalori reali è composta dalle matrici simmetriche. Una matrice quadrata $A \in M_{n,n}(\mathbb{R})$ è **simmetrica** se $A^T = A$, dove A^T è la trasposta di A (vedi la Curiosità 4.3). Molti fenomeni fisici sono rappresentati da matrici simmetriche, il che permette di utilizzare sistematicamente autovalori e autovettori grazie all'importante **Teorema spettrale**: *se A è una matrice simmetrica di ordine n , allora esiste una base di \mathbb{R}^n composta da autovettori di A .*

Curiosità 4.6 (Molteplicità degli autovalori) La **molteplicità algebrica** di un autovalore è la sua molteplicità come radice del polinomio caratteristico (se non la conosci, vedi il Paragrafo 5.4 per la definizione di molteplicità di una radice). Siccome il polinomio caratteristico di una matrice A di ordine n ha grado n , la somma delle molteplicità algebriche degli autovalori di A è al massimo n .

Il **Teorema fondamentale dell'algebra** dice che ogni polinomio di grado n ha esattamente n radici *complesse*, (vedi la Curiosità 9.6) contate con molteplicità. Quindi se prendiamo come autovalori anche le radici complesse del polinomio caratteristico, una matrice quadrata di ordine n ha esattamente n autovalori (reali e complessi) contati con la molteplicità algebrica.

Curiosità 5.1 (Esistenza di soluzioni di equazioni) Una conseguenza del Teorema dei valori intermedi è la seguente: se $f: [a, b] \rightarrow \mathbb{R}$ è una funzione continua, allora per ogni $y_0 \in [\min f, \max f]$ l'equazione $f(x) = y_0$ ammette sempre una soluzione più piccola e una soluzione più grande. Per funzioni non continue questo risultato è falso, anche nel caso di funzioni monotone. Per esempio, se $f: [-1, 1] \rightarrow \mathbb{R}$ è la funzione crescente data da

$$f(x) = \begin{cases} x & \text{se } -1 \leq x \leq 0, \\ x + 1 & \text{se } 0 < x \leq 1, \end{cases}$$

allora l'equazione $f(x) = 1$ non ammette soluzioni (controlla, anche tracciando il grafico di f).

Curiosità 5.2 (Parabole) Una **parabola** nel piano è il luogo dei punti la cui distanza da una retta data (detta **direttrice** della parabola) è uguale alla distanza da un punto dato (detto **fuoco** della parabola). Si può dimostrare che tutte le parabole con direttrice parallela all'asse delle ascisse sono il grafico di una funzione quadratica. Quindi (perché?) una parabola qualsiasi si può ottenere ruotando il grafico di una opportuna funzione quadratica.

Curiosità 5.3 (Decomposizione in funzioni pari e dispari) Ogni funzione $f: \mathbb{R} \rightarrow \mathbb{R}$ si può scrivere (in modo unico) come somma di una funzione pari e una funzione dispari. Infatti, ponendo

$$f_+(x) = \frac{f(x) + f(-x)}{2} \quad \text{e} \quad f_-(x) = \frac{f(x) - f(-x)}{2}$$

si vede subito che f_+ è pari, f_- è dispari e che $f = f_+ + f_-$.

Curiosità 5.4 (Coniche) Nel Paragrafo 3.4 abbiamo visto che gli insiemi definiti da un'equazione lineare $ax + by + c = 0$ in due variabili erano tutte e sole le rette del piano. Gli insiemi definiti da un'equazione quadratica

$$ax^2 + bxy + cy^2 + dx + ey + f = 0$$

sono invece chiamati **coniche** (perché si può far vedere che si possono ottenere come intersezione di un cono a due falde con un piano). Le parabole (anche quelle generali di cui abbiamo parlato nella Curiosità 5.2) sono tutte coniche, ma non sono le sole; nell'Esempio 2.40 abbiamo visto che anche le circonferenze lo sono.

Le coniche si possono classificare. Quelle più interessanti (dette **irriducibili**, che non sono unione di rette o punti) sono divise in tre classi: parabole, ellissi, iperboli. Delle parabole abbiamo parlato nella Curiosità 5.2. Un **ellisse** è il luogo dei punti P del piano tali che la somma $d(P, F_1) + d(P, F_2)$ delle distanze $d(P, F_1)$ e $d(P, F_2)$ da due punti fissati F_1 e F_2 (detti **fuochi**) sia costante; se $F_1 = F_2$ otteniamo le circonferenze. Infine, un **iperbole** è il luogo dei punti P del piano tali che il modulo $|d(P, F_1) - d(P, F_2)|$ della differenza delle distanze $d(P, F_1)$ e $d(P, F_2)$ da due punti fissati F_1 e F_2 (detti nuovamente **fuochi**) sia costante.

Curiosità 5.5 (Differenza di potenze) La formula (5.13) si dimostra con il seguente conto:

$$\begin{aligned}(x-y) \sum_{i=0}^k x^i y^{k-i} &= \sum_{i=0}^k x^{i+1} y^{k-i} - \sum_{i=0}^k x^i y^{k-(i-1)} \\ &= x^{k+1} + \sum_{i=0}^{k-1} x^{i+1} y^{k-i} - \sum_{h=0}^{k-1} x^{h+1} y^{k-h} - y^{k+1} \\ &= x^{k+1} - y^{k+1},\end{aligned}$$

dove abbiamo posto $h = i - 1$ nell'ultima sommatoria (e leggi l'Osservazione 4.11 in caso tu non conosca il simbolo di sommatoria).

Curiosità 5.6 (Decomposizione dei polinomi) Se $x_0 \in \mathbb{R}$ è una radice di molteplicità $r_0 > 0$ del polinomio p , nel Paragrafo 5.4 abbiamo visto che possiamo trovare un polinomio q_0 con $q_0(x_0) \neq 0$ tale che $p(x) = (x - x_0)^{r_0} q_0(x)$. Se q_0 ha un'altra radice x_1 di molteplicità $r_1 > 0$, possiamo ripetere questa costruzione con q_0 trovando un polinomio q_1 tale che $p(x) = (x - x_0)^{r_0} (x - x_1)^{r_1} q_1(x)$. Se invece q_0 non ha altre radici reali, si può dimostrare che può essere scritto come prodotto di polinomi quadratici con discriminante negativo.

Procedendo in questo modo si riesce a dimostrare che ogni polinomio p si può scrivere in modo unico come prodotto

$$f(x) = ap_1(x)^{r_1} \cdots p_k(x)^{r_k}, \quad (5.1C)$$

dove $a \in \mathbb{R}$ è il coefficiente direttore di p , gli esponenti r_1, \dots, r_k sono numeri interi positivi e p_1, \dots, p_k sono polinomi **monici** (cioè con coefficiente direttore uguale a 1) e **irriducibili** (cioè non si possono scrivere come prodotto di due altri polinomi monici). I polinomi irriducibili sono l'equivalente per i polinomi dei numeri primi per i numeri naturali; e (5.1C) è l'equivalente della decomposizione in fattori primi di un numero naturale. Infine, si può anche dimostrare che i polinomi monici irriducibili a coefficienti reali sono o polinomi lineari della forma $x - x_0$ oppure polinomi quadratici senza radici reali (cioè con discriminante negativo).

Curiosità 5.7 (Potenze irrazionali) Un'altra proprietà cruciale dei numeri reali è che ogni numero reale può essere approssimato arbitrariamente bene da numeri razionali: dato $r \in \mathbb{R}$, per ogni $\varepsilon > 0$ arbitrariamente piccolo esiste un numero razionale $p/q \in \mathbb{Q}$ tale che $|r - p/q| < \varepsilon$. In termini tecnici, si dice che i numeri razionali sono **densi** nei numeri reali.

Si può usare la densità dei numeri razionali per definire x^r per qualsiasi $x > 0$ e $r \in \mathbb{R}$. Consideriamo prima il caso $r > 0$. L'idea è di approssimare x^r con potenze razionali $x^{p/q}$, dove $p/q \in \mathbb{Q}$ approssima arbitrariamente bene r . Formalmente, definiamo x^r come l'estremo superiore (vedi l'Osservazione 5.9) del sottoinsieme $A = \{x^{p/q} \mid p/q \in \mathbb{Q}, p/q \leq r\}$; in altre parole, approssimiamo x^r con i numeri $x^{p/q}$ dove p/q sono razionali che approssimano r dal basso. Si può anche procedere dall'alto: infatti, si può dimostrare che l'estremo superiore di A coincide con l'estremo inferiore del sottoinsieme $B = \{x^{p/q} \mid p/q \in \mathbb{Q}, p/q \geq r\}$. Un altro modo per esprimere questo concetto è dire che abbiamo scelto $x^r \in \mathbb{R}$ in modo che per ogni $\varepsilon > 0$ arbitrariamente piccolo esista un $\delta > 0$ sufficientemente piccolo tale che se $|p/q - r| < \delta$, allora $|x^{p/q} - x^r| < \varepsilon$.

Se invece $r < 0$, si procede in modo analogo, ma invertendo l'estremo superiore con l'estremo inferiore in quanto le potenze con esponente negativo sono decrescenti. Quindi se $r < 0$ si pone

$$x^r = \inf\{x^{p/q} \mid p/q \in \mathbb{Q}, p/q \leq r\} = \sup\{x^{p/q} \mid p/q \in \mathbb{Q}, p/q \geq r\}.$$

In entrambi i casi non è difficile verificare che la funzione $f: \mathbb{R}^+ \rightarrow \mathbb{R}^+$ definita da $f(x) = x^r$ soddisfa tutte le proprietà che ci si aspetta: è strettamente crescente, è continua (vedi il Paragrafo 5.7) e valgono le proprietà delle potenze

$$x^r \cdot x^s = x^{r+s} \quad \text{e} \quad x^{-r} = \frac{1}{x^r}.$$

Nelle Curiosità 6.6 e 8.4 descriveremo altri modi per definire x^r per $x > 0$ e $r \in \mathbb{R}$.

Curiosità 5.8 (Continuità delle funzioni monotone) Una funzione monotona crescente $f: I \rightarrow \mathbb{R}$ è continua in $x_0 \in I$ se e solo se

$$f(x_0) = \sup\{f(x) \mid x < x_0\} = \inf\{f(x) \mid x > x_0\},$$

dove sup e inf indicano l'estremo superiore e l'estremo inferiore introdotti nell'Osservazione 5.9. Questo segue dal fatto che se f è crescente si ha

$$\lim_{x \rightarrow x_0^-} f(x) = \sup\{f(x) \mid x < x_0\} \quad \text{e} \quad \lim_{x \rightarrow x_0^+} f(x) = \inf\{f(x) \mid x > x_0\}.$$

Infatti, posto $y_- = \sup\{f(x) \mid x < x_0\}$, la definizione di estremo superiore ci dice (perché?) che per ogni $\varepsilon > 0$ arbitrariamente piccolo esiste un $x_1 < x_0$ tale che $0 \leq y_- - f(x_1) < \varepsilon$. Posto $\delta = x_0 - x_1$, se x è tale che $0 < x_0 - x < \delta$ si deve avere $x_1 < x < x_0$; quindi, essendo f crescente, $f(x) \geq f(x_1)$ e $0 \leq y_- - f(x) < \varepsilon$. Dunque y_- soddisfa la definizione di limite da sinistra: per ogni $\varepsilon > 0$ arbitrariamente piccolo esiste un $\delta > 0$ sufficientemente piccolo tale che $0 < x_0 - x < \delta$ implica $|f(x) - y_-| < \varepsilon$. In maniera analoga si mostra che il limite da destra coincide con l'estremo inferiore, e quindi f è continua in x_0 se e solo se l'estremo inferiore coincide con l'estremo superiore e con il valore in x_0 .

Un risultato analogo vale anche per le funzioni monotone decrescenti, scambiando estremo superiore ed estremo inferiore.

Curiosità 5.9 (Operazioni sui limiti) Non è difficile verificare le (5.19). Per esempio, supponiamo che la funzione f abbia limite $a \in \mathbb{R}$ per x che tende a $x_0 \in \mathbb{R}$ e che la funzione g abbia limite $b \in \mathbb{R}$ per x che tende a $x_0 \in \mathbb{R}$. Questo vuol dire che per ogni $\varepsilon > 0$ arbitrariamente piccolo esiste $\delta > 0$ tale che $0 < |x - x_0| < \delta$ implica $|f(x) - a| < \varepsilon$ e $|g(x) - b| < \varepsilon$. Allora

$$0 < |x - x_0| < \delta \quad \Rightarrow \quad |(f(x) + g(x)) - (a + b)| \leq |f(x) - a| + |g(x) - b| < 2\varepsilon .$$

Ma anche 2ε è arbitrariamente piccolo; quindi abbiamo dimostrato che la funzione $f + g$ ha limite $a + b$ per $x \rightarrow x_0$. Ragionamenti simili (anche se un filo più complicati per il prodotto e il quoziente) funzionano anche negli altri casi.

Curiosità 5.10 (Continuità dell'inversa) Una funzione (reale di variabile reale) continua invertibile f è strettamente monotona. Infatti, supponiamo che non lo sia; allora devono esistere $x_1 < x_2 < x_3$ nel dominio di f tali che $f(x_1) \leq f(x_2) \geq f(x_3)$ oppure $f(x_1) \geq f(x_2) \leq f(x_3)$. Nel primo caso, se $f(x_1) \leq f(x_3)$, allora per il Teorema dei valori intermedi esiste un $x_4 \in [x_1, x_2]$ tale che $f(x_4) = f(x_3)$, il che contraddice (perché?) l'iniettività di f . In modo analogo si trova una contraddizione se $f(x_1) \geq f(x_3)$, o se $f(x_1) \geq f(x_2) \leq f(x_3)$. Quindi f , per essere invertibile, dev'essere strettamente monotona.

Capito questo, è facile verificare che l'inversa di una funzione continua invertibile è ancora continua. Infatti, fissiamo $y_0 = f(x_0)$ nel dominio di f^{-1} , e un $\varepsilon > 0$ arbitrariamente piccolo; supponiamo anche per semplicità che il dominio di f sia un intervallo aperto e che ε sia così piccolo che $[x_0 - \varepsilon, x_0 + \varepsilon]$ sia contenuto nel dominio di f . Allora se prendiamo come $\delta = \min\{f(x_0 - \varepsilon), f(x_0 + \varepsilon)\}$ usando la monotonia si ottiene subito (vedi la Figura 5.30) che se $y = f(x)$ dista da y_0 meno di δ allora $x = f^{-1}(y)$ dista da x_0 meno di ε , come voluto.

Curiosità 5.11 (Una funzione senza né massimo né minimo) Un esempio di funzione *non* continua che non ammette né massimo né minimo su un intervallo chiuso è la funzione $f: [-1, 1] \rightarrow \mathbb{R}$ definita da

$$f(x) = \begin{cases} |x| & \text{se } x \neq -1, 0, 1, \\ 1/2 & \text{se } x = -1, 0, 1. \end{cases}$$

Se provi a tracciarne il grafico, vedrai che la non esistenza del minimo e del massimo è dovuta a dei salti nel grafico, che (appunto) non è continuo.

Curiosità 6.1 (L'esponenziale come limite, parte prima) Vediamo come si dimostra (6.1). È piuttosto complicato, ma c'era da aspettarselo: è una delle formule fondamentali della Matematica.

Cominciamo supponendo $r > 0$. Posto

$$x_n(r) = \left(1 + \frac{r}{n}\right)^n,$$

per prima cosa dimostriamo che $\{x_n(r)\}$ è una successione *strettamente crescente*, cioè tale che $x_n(r) < x_{n+1}(r)$ per ogni $n \in \mathbb{N}$. Usando lo sviluppo del binomio possiamo scrivere

$$\begin{aligned} x_n(r) &= \sum_{k=0}^n \binom{n}{k} \frac{r^k}{n^k} = 1 + \sum_{k=1}^n \frac{r^k}{k!} \frac{n(n-1)\cdots(n-k+1)}{n \cdot n \cdots n} \\ &= 1 + \sum_{k=1}^n \frac{r^k}{k!} \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right). \end{aligned}$$

Siccome per $j = 1, \dots, n-1$ si ha

$$1 - \frac{j}{n+1} > 1 - \frac{j}{n},$$

otteniamo

$$\begin{aligned} x_{n+1}(r) &= 1 + \sum_{k=1}^{n+1} \frac{r^k}{k!} \left(1 - \frac{1}{n+1}\right) \cdots \left(1 - \frac{k-1}{n+1}\right) \\ &> 1 + \sum_{k=1}^n \frac{r^k}{k!} \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \\ &\quad + \frac{r^{n+1}}{(n+1)!} \left(1 - \frac{1}{n+1}\right) \cdots \left(1 - \frac{n}{n+1}\right) \\ &> x_n(r). \end{aligned}$$

Quindi la successione $\{x_n(r)\}$ è strettamente crescente. Adesso dimostriamo che (contrariamente a quanto pensava il funzionario della banca) è una successione *limitata* (superiormente e inferiormente; vedi il Paragrafo 6.6).

Siccome $x_n(r) > x_1(r) = 1 + r$ per ogni $n \in \mathbb{N}$, la nostra successione è limitata inferiormente; facciamo vedere che è anche limitata superiormente. Scegliamo un numero naturale n_0 maggiore di r . Notando che $1 - j/n < 1$ per ogni $n, j \in \mathbb{N}^*$ troviamo che per $n > n_0$ si ha

$$\begin{aligned} x_n(r) &< \sum_{k=0}^n \frac{r^k}{k!} = \sum_{k=0}^{n_0-1} \frac{r^k}{k!} + \sum_{k=n_0}^n \frac{r^k}{k!} \\ &= \sum_{k=0}^{n_0-1} \frac{r^k}{k!} + \frac{r^{n_0}}{n_0!} \left(1 + \sum_{k=n_0+1}^n \frac{r^{k-n_0}}{(n_0+1)\cdots k}\right) \\ &< \sum_{k=0}^{n_0-1} \frac{r^k}{k!} + \frac{r^{n_0}}{n_0!} \left(1 + \sum_{k=n_0}^n \left(\frac{r}{n_0}\right)^{k-n_0}\right) = \sum_{k=0}^{n_0-1} \frac{r^k}{k!} + \frac{r^{n_0}}{n_0!} \sum_{h=0}^{n-n_0} \left(\frac{r}{n_0}\right)^h. \end{aligned}$$

Usando la (5.13) con $x = r/n_0$ e $y = 1$, e ricordando che $r/n_0 < 1$, deduciamo che

$$x_n(r) < \sum_{k=0}^{n_0-1} \frac{r^k}{k!} + \frac{r^{n_0}}{n_0!} \frac{1 - (r/n_0)^{n-n_0+1}}{1 - r/n_0} < \sum_{k=0}^{n_0-1} \frac{r^k}{k!} + \frac{r^{n_0}}{n_0!} \frac{1}{1 - r/n_0} = M_r. \quad (6.1C)$$

Abbiamo quindi ottenuto una stima $x_n(r) < M_r$ con M_r indipendente da n .

Indichiamo con $e_r \in \mathbb{R}$ l'estremo superiore (finito, perché la successione è limitata) della nostra successione $\{x_n(r)\}$; allora la Curiosità 6.12 ci dice che

$$\lim_{n \rightarrow +\infty} x_n(r) = e_r$$

per ogni $r > 0$. In particolare, $x_n(1)$ tende a uno specifico numero reale e_1 , che è proprio il numero di Nepero e .

Per concludere la dimostrazione di (6.1) per $r > 0$ dobbiamo far vedere che $e_r = (e_1)^r$. Cominciamo notando che per ogni $p > 0$ naturale la successione $\{x_p(r), x_{2p}(r), \dots, x_{mp}(r), \dots\}$ è contenuta nella successione $\{x_n(r)\}$, per cui è crescente; inoltre siccome per ogni $n \in \mathbb{N}$ esiste $m \in \mathbb{N}$ tale che $mp > n$, la successione $\{x_{mp}(r)\}$ ha lo stesso estremo superiore di $\{x_n(r)\}$, quindi lo stesso limite. Ma allora

$$\begin{aligned} e_r &= \lim_{m \rightarrow +\infty} \left(1 + \frac{r}{mp}\right)^{mp} = \lim_{m \rightarrow +\infty} \left[\left(1 + \frac{r/p}{m}\right)^{m \cdot p}\right]^p = \left[\lim_{m \rightarrow +\infty} \left(1 + \frac{r/p}{m}\right)^m\right]^p \\ &= (e_{r/p})^p. \end{aligned}$$

In particolare, prendendo $r = 1$ otteniamo $e_1 = (e_{1/p})^p$, cioè $e_{1/p} = (e_1)^{1/p}$. Quindi se $r = p/q \in \mathbb{Q}^+$ è razionale positivo otteniamo

$$e_{p/q} = (e_{1/q})^p = (e_1)^{p/q},$$

come voluto. Infine, è chiaro che se $0 < r_1 < r_2$ allora $x_n(r_1) < x_n(r_2)$ per ogni $n \in \mathbb{N}$, per cui

$$e_{r_1} = \lim_{n \rightarrow +\infty} x_n(r_1) \leq \lim_{n \rightarrow +\infty} x_n(r_2) = e_{r_2}.$$

Ricordando come abbiamo definito le potenze con esponente irrazionale nella Curiosità 5.7 ricaviamo quindi $e_r = e_1^r$ per ogni $r > 0$, come voluto.

Per trattare il caso $r < 0$ ci serve un'ulteriore stima. Usando di nuovo lo sviluppo del binomio per ogni $s > 0$ abbiamo

$$\begin{aligned} 1 &\geq \left(1 - \frac{s}{n^2}\right)^n = \sum_{k=0}^n \binom{n}{k} (-1)^k \left(\frac{s}{n^2}\right)^k = 1 + \sum_{k=1}^n \binom{n}{k} (-1)^k \frac{s^k}{n^{2k}} \\ &\geq 1 - \frac{1}{n} \sum_{k=1}^n \frac{1}{n^{k-1}} \binom{n}{k} \frac{s^k}{n^k} \geq 1 - \frac{1}{n} \sum_{k=1}^n \binom{n}{k} \frac{s^k}{n^k} \\ &\geq 1 - \frac{1}{n} (x_n(s) - 1) \geq 1 - \frac{M_s}{n}. \end{aligned}$$

Applicando il Teorema dei due carabinieri (vedi il Paragrafo 5.7) otteniamo quindi

$$\lim_{n \rightarrow +\infty} \left(1 - \frac{s}{n^2}\right)^n = 1$$

per ogni $s > 0$. Ma allora se $r > 0$ troviamo

$$\lim_{n \rightarrow +\infty} x_n(-r) = \lim_{n \rightarrow +\infty} \frac{x_n(r)x_n(-r)}{x_n(r)} = \frac{\lim_{n \rightarrow +\infty} \left(1 - \frac{r^2}{n^2}\right)^n}{\lim_{n \rightarrow +\infty} x_n(r)} = \frac{1}{e_1^r} = e_1^{-r},$$

e quindi (6.1) vale per ogni $r \in \mathbb{R}$.

Curiosità 6.2 (L'esponenziale come limite, parte seconda) Il Teorema dei due carabinieri (vedi il Paragrafo 5.7) può essere usato anche per ricavare (6.2) da (6.1). Infatti, se indichiamo con $\lfloor t \rfloor$ la *parte intera* di $t \in \mathbb{R}$ (cioè il più grande numero intero minore o uguale a t), per ogni $t > |r|$ si ha

$$\left(1 + \frac{r}{\lfloor t \rfloor + 1}\right)^{\lfloor t \rfloor} \leq \left(1 + \frac{r}{t}\right)^t \leq \left(1 + \frac{r}{\lfloor t \rfloor}\right)^{\lfloor t \rfloor + 1} \quad (6.2C)$$

se $r > 0$ e

$$\left(1 + \frac{r}{\lfloor t \rfloor + 1}\right)^{\lfloor t \rfloor + 1} \geq \left(1 + \frac{r}{t}\right)^t \geq \left(1 + \frac{r}{\lfloor t \rfloor}\right)^{\lfloor t \rfloor} \quad (6.3C)$$

se $r < 0$. Ora

$$\left(1 + \frac{r}{\lfloor t \rfloor}\right)^{\lfloor t \rfloor + 1} = \left(1 + \frac{r}{\lfloor t \rfloor}\right) \left(1 + \frac{r}{\lfloor t \rfloor}\right)^{\lfloor t \rfloor}.$$

Il primo fattore tende a 1 quando $t \rightarrow +\infty$; il secondo fattore è uguale a $x_{\lfloor t \rfloor}(r)$, quindi tende a e^r quando $t \rightarrow +\infty$. Analogamente,

$$\left(1 + \frac{r}{\lfloor t \rfloor + 1}\right)^{\lfloor t \rfloor} = \frac{x_{\lfloor t \rfloor + 1}(r)}{1 + \frac{r}{\lfloor t \rfloor + 1}} \rightarrow \frac{e^r}{1} = e^r.$$

Quindi tutte le espressioni agli estremi nelle disuguaglianze (6.2C) e (6.3C) tendono a e^r per $t \rightarrow +\infty$. Il Teorema dei due carabinieri ci dice allora che anche l'espressione in mezzo non può fare altro che tendere alla stessa quantità e^r , e abbiamo finito.

Curiosità 6.3 (Funzioni iperboliche) Altre funzioni costruibili a partire dalle funzioni esponenziali sono le cosiddette **funzioni iperboliche**: il **seno iperbolico** $\sinh: \mathbb{R} \rightarrow \mathbb{R}$, il **coseno iperbolico** $\cosh: \mathbb{R} \rightarrow [1, +\infty)$, e la **tangente iperbolica** $\tanh: \mathbb{R} \rightarrow (-1, 1)$, definite rispettivamente da

$$\sinh t = \frac{e^t - e^{-t}}{2}, \quad \cosh t = \frac{e^t + e^{-t}}{2}, \quad \tanh t = \frac{\sinh t}{\cosh t} = \frac{e^t - e^{-t}}{e^t + e^{-t}}.$$

I nomi "seno", "coseno" e "tangente" derivano da alcuni parallelismi con le funzioni trigonometriche (vedi il Paragrafo 6.4). Per esempio, seno e coseno iperbolico soddisfano la relazione

$$(\cosh t)^2 - (\sinh t)^2 = 1,$$

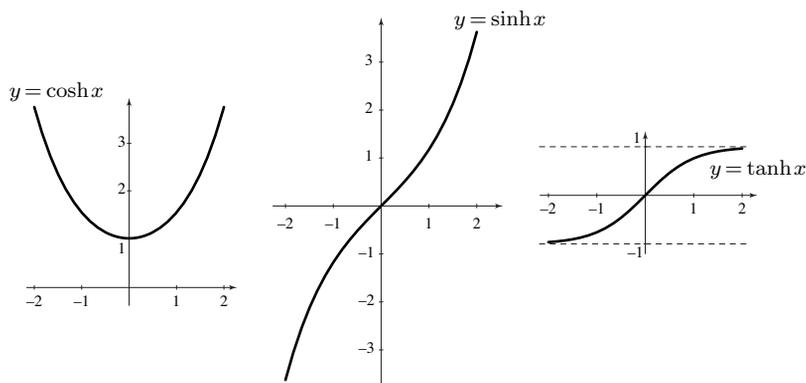
simile alla relazione fondamentale (6.15) che lega gli usuali seno e coseno. L'aggettivo "iperbolico" deriva dal fatto che l'insieme I di equazione $x^2 - y^2 = 1$ è un'iperbole equilatera avente come asintoti le bisettrici principali del piano cartesiano e $(\cosh t, \sinh t) \in I$ per ogni $t \in \mathbb{R}$.

Tra parentesi, nota che la tangente iperbolica si può scrivere come

$$\tanh t = \frac{2}{1 + e^{-2t}} - 1,$$

per cui è una funzione logistica. I grafici delle funzioni iperboliche sono riportati nella Figura 6.1C.

Figura 6.1C
Funzioni iperboliche



Curiosità 6.4 (Una funzione infinitamente piatta) Un'altra funzione che i matematici amano particolarmente è

$$f(x) = \begin{cases} e^{-1/x^2} & \text{se } x > 0, \\ 0 & \text{se } x \leq 0, \end{cases}$$

il cui grafico è riportato nella Figura 6.2C. La particolarità di questa funzione è di essere (in un senso che discuteremo nella Curiosità 7.8) infinitamente piatta nell'origine.

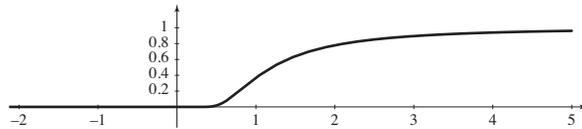


Figura 6.2C
 $f(x) = e^{-1/x^2}$

Curiosità 6.5 (Continuità delle funzioni esponenziali e logaritmiche) Le funzioni esponenziali e le funzioni logaritmiche (e quindi tutte le funzioni ottenute da loro tramite operazioni algebriche) sono continue. La continuità della funzione \exp segue dal criterio di continuità per funzioni monotone discusso nella Curiosità 5.8, che si può applicare grazie a quanto visto nella Curiosità 6.1 (e nella Curiosità 5.7). La formula (6.5) implica allora che tutte le funzioni esponenziali sono continue; allora anche tutte le funzioni logaritmiche sono continue, in quanto inverse di funzioni continue strettamente monotone (vedi la Curiosità 5.10).

Come conseguenza ulteriore, se $f, g: I \rightarrow \mathbb{R}$ sono due funzioni continue con $f(x) > 0$ per ogni $x \in I$ allora anche la funzione $h: I \rightarrow \mathbb{R}$ data da $h(x) = f(x)^{g(x)}$ è continua. Infatti, si può scrivere come

$$h(x) = \exp(g(x) \log_e f(x)) ,$$

e quindi h è composizione di funzioni continue.

Curiosità 6.6 (Potenze irrazionali) Possiamo usare il logaritmo in base e anche per dare una definizione alternativa a quella presentata nella Curiosità 5.7 delle potenze con esponente irrazionale qualsiasi. Infatti, possiamo cominciare *definendo* e^r tramite il limite (6.1). I conti fatti nella Curiosità 6.1 mostrano che la funzione $r \mapsto e^r$ è strettamente crescente e soddisfa le usuali proprietà delle potenze. Quindi possiamo definire il logaritmo in base e come inversa di questa funzione e definire q^x per $q > 0$ e $x \in \mathbb{R}$ con la formula (6.5), prendendo $a = 1$.

Nella Curiosità 8.4 vedremo un terzo modo per definire le potenze con esponente irrazionale.

Curiosità 6.7 (Funzioni curiose) Usando la funzione $\sin \frac{1}{x}$ è possibile costruire altre funzioni con andamenti curiosi. Per esempio, la funzione

$$g(x) = x \sin \frac{1}{x},$$

pur oscillando freneticamente vicino all'origine, ha limite 0 per x che tende a 0: infatti

$$|g(x)| = \left| x \sin \frac{1}{x} \right| \leq |x|,$$

per cui possiamo rendere $|g(x)|$ arbitrariamente piccolo a patto di prendere $|x|$ sufficientemente piccolo (vedi la Figura 6.3C(a)). In particolare, il rapporto $g(x)/x$ per x che tende a 0 è un caso di forma indeterminata del tipo $0/0$ in cui il limite non esiste, in quanto $g(x)/x = \sin \frac{1}{x}$.

Invece, il rapporto

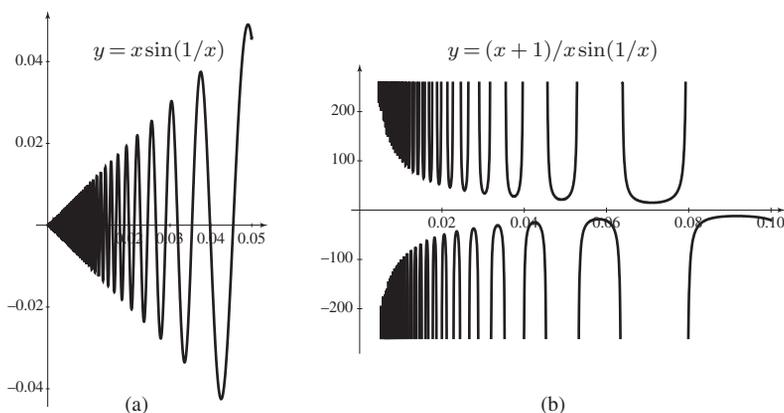
$$h(x) = \frac{x+1}{x \sin \frac{1}{x}}$$

è un esempio in cui per x che tende a 0 il numeratore tende a un limite finito non nullo (uguale a 1) e il denominatore tende a 0 ma il rapporto non ha limite $+\infty$ o $-\infty$. Infatti, per x arbitrariamente vicino a 0 la funzione $h(x)$ assume sia valori arbitrariamente grandi sia valori arbitrariamente negativi. Inoltre, h ha infinite singolarità (gli zeri di $\sin \frac{1}{x}$) che si avvicinano arbitrariamente all'origine, per cui in particolare h non è definita in alcun intervallo contenente 0; vedi la Figura 6.3C(b).

Figura 6.3C

(a) $g(x) = x \sin \frac{1}{x}$

(b) $h(x) = (x+1)/x \sin \frac{1}{x}$



Curiosità 6.8 (Un paradosso) Indichiamo con ℓ_1 la lunghezza di una semicirconferenza che ha come diametro l'intervallo $[0, 1]$. Poi indichiamo con ℓ_2 la somma delle lunghezze delle semicirconferenze di diametro $[0, 1/2]$ e $[1/2, 1]$. Più in generale, indichiamo con ℓ_n la somma delle lunghezze delle n semicirconferenze di diametro gli intervalli $[0, 1/n]$, $[1/n, 2/n]$, ..., $[(n-1)/n, 1]$ (vedi la Figura 6.4C). Tutte queste semicirconferenze si avvicinino sempre di più all'intervallo $[0, 1]$; sembrerebbe quindi naturale aspettarsi che la somma delle loro lunghezze si avvicini sempre più alla lunghezza dell'intervallo $[0, 1]$, cioè che si abbia

$$\lim_{n \rightarrow +\infty} \ell_n = 1 .$$

Invece, questo è falso. Per vederlo, basta calcolare ℓ_n . Infatti, ℓ_1 è la lunghezza di una semicirconferenza di diametro 1, cioè raggio $1/2$, per cui $\ell_1 = \pi/2$. Al passo n , ciascuna semicirconferenzina ha diametro $1/n$, per cui ha lunghezza $\pi/2n$. Ma ce ne sono n : quindi $\ell_n = n \cdot (\pi/2n) = \pi/2$. In altre parole, $\ell_n = \pi/2$ per ogni n , e il risultato esatto è quindi

$$\lim_{n \rightarrow +\infty} \ell_n = \frac{\pi}{2} > 1 .$$

La spiegazione di questo apparente paradosso è che le circonferenzine, pur avvicinandosi al segmento, diventano sempre più curve; questo aumento della curvatura controbilancia il rimpicciolimento del diametro lasciando costante la lunghezza totale.

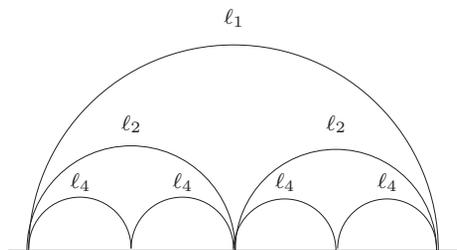


Figura 6.4C
Paradosso geometrico

Curiosità 6.9 (Coseno della differenza) Vediamo come si dimostra la formula (6.16) per il coseno della differenza. Supponiamo inizialmente che $0 \leq \phi, \psi \leq \pi/2$, e consideriamo i due triangoli rettangoli $P(\psi)AP(\phi)$ e $P(\psi)BP(\phi)$ rappresentati nella Figura 6.5C. Nel triangolo $P(\psi)AP(\phi)$, il cateto $AP(\phi)$ è lungo $|\sin \phi - \sin \psi|$, mentre il cateto $AP(\psi)$ è lungo $|\cos \phi - \cos \psi|$. Se indichiamo con d la lunghezza della corda $P(\psi)P(\phi)$, il Teorema di Pitagora ci dice che

$$d^2 = (\cos \phi - \cos \psi)^2 + (\sin \phi - \sin \psi)^2 .$$

D'altra parte, nel triangolo $P(\psi)BP(\phi)$, il cateto $BP(\phi)$ è lungo $|\sin(\phi - \psi)|$, mentre il cateto $BP(\psi)$ è lungo $|1 - \cos(\phi - \psi)|$. Una nuova applicazione del Teorema di Pitagora ci dà quindi

$$d^2 = \sin^2(\phi - \psi) + [1 - \cos(\phi - \psi)]^2 .$$

Espandendo i quadrati e confrontando le due espressioni troviamo

$$\begin{aligned} \sin^2 \phi + \cos^2 \phi + \sin^2 \psi + \cos^2 \psi - 2(\cos \phi \cos \psi + \sin \phi \sin \psi) \\ = 1 + \sin^2(\phi - \psi) + \cos^2(\phi - \psi) - 2 \cos(\phi - \psi), \end{aligned}$$

e (6.16) segue subito da (6.15).

Nel caso generale, scriviamo $\phi = \phi' + h\pi/2$ e $\psi = \psi' + k\pi/2$ per opportuni $h, k \in \mathbb{N}$ e $\phi', \psi' \in [0, \pi/2]$. Allora (6.16) segue subito dalla formula per il coseno della differenza di ϕ' e ψ' usando le relazioni di periodicità (6.13) e (6.14).

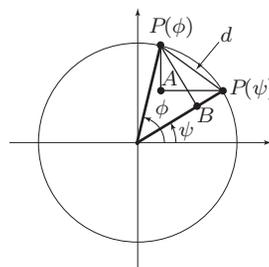


Figura 6.5C
Coseno della differenza

Curiosità 6.10 (Un limite notevole) Vogliamo dimostrare che $\lim_{\phi \rightarrow 0} \frac{\sin \phi}{\phi} = 1$. Ci basta dimostrarlo per $\phi \rightarrow 0^+$; il caso $\phi \rightarrow 0^-$ segue subito (perché?) dall'uguaglianza

$$\frac{\sin(-\phi)}{-\phi} = \frac{\sin \phi}{\phi} .$$

Facendo riferimento alla Figura 6.17 (pag. 243), confrontiamo le aree del triangolo $OQ(\phi)P(\phi)$, del settore circolare $OQP(\phi)$, e del triangolo OQS . Con le ovvie notazioni, si ha

$$\text{Area}(OQ(\phi)P(\phi)) < \text{Area}(OQP(\phi)) < \text{Area}(OQS) ,$$

in quanto i tre insiemi sono contenuti uno dentro l'altro. Ora, i due cateti di $OQ(\phi)P(\phi)$ misurano rispettivamente $\cos \phi$ e $\sin \phi$; quindi $\text{Area}(OQ(\phi)P(\phi)) = \frac{1}{2} \cos \phi \sin \phi$. Analogamente, i due cateti di OQS misurano rispettivamente 1 e $\tan \phi$; quindi $\text{Area}(OQS) = \frac{1}{2} \sin \phi / \cos \phi$. Infine, l'area di un settore circolare è proporzionale alla lunghezza dell'arco. Siccome l'area del cerchio di raggio 1, sotteso da un arco di lunghezza 2π , è uguale a π , l'area del settore circolare $OQP(\phi)$, sotteso da un arco di lunghezza ϕ , è $\text{Area}(OQP(\phi)) = \phi\pi/2\pi = \phi/2$. Riassumendo abbiamo ottenuto

$$\frac{1}{2} \sin \phi \cos \phi < \frac{\phi}{2} < \frac{1}{2} \frac{\sin \phi}{\cos \phi} .$$

Dividendo per $\sin \phi$ (che è strettamente positivo per $\phi > 0$ piccolo e non nullo), moltiplicando per 2 e prendendo il reciproco otteniamo

$$\cos \phi < \frac{\sin \phi}{\phi} < \frac{1}{\cos \phi} .$$

A questo punto, facendo tendere ϕ a 0, ricordando che $\cos 0 = 1$ e usando il Teorema dei due carabinieri (vedi il Paragrafo 5.7), otteniamo

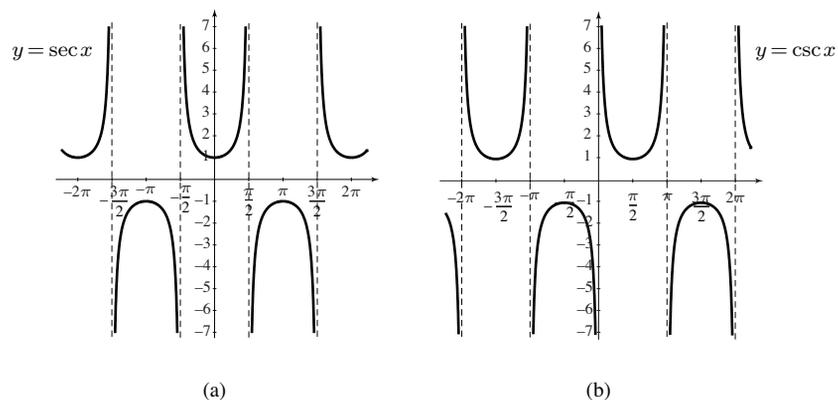
$$\lim_{\phi \rightarrow 0} \frac{\sin \phi}{\phi} = 1 .$$

Curiosità 6.11 (Altre funzioni trigonometriche) Per completezza citiamo l'esistenza di altre due funzioni trigonometriche classiche: la **secante** $\sec \phi$ e la **cosecante** $\csc \phi$ definite da

$$\sec \phi = \frac{1}{\cos \phi}, \quad \csc \phi = \frac{1}{\sin \phi}.$$

Il segmento OS della Figura 6.17 (pag. 243) è lungo $\sec \phi$, giustificando il nome "secante" per questa funzione. La Figura 6.6C(a) contiene il grafico della secante, mentre la Figura 6.6C(b) il grafico della cosecante.

Figura 6.6C
(a) $f(x) = \sec x$
(b) $f(x) = \csc x$



Curiosità 6.12 (Limite di una successione monotona) Un fatto generale piuttosto importante è che tutte le successioni monotone hanno limite. Per esempio, sia $\{x_n\}$ una successione crescente. Se è illimitata, è naturale aspettarsi che abbia limite $+\infty$; se è limitata da $M \in \mathbb{R}$, continua a crescere senza mai poter superare la barriera imposta da M , per cui è naturale pensare che vada ad accumularsi da qualche parte.

Questa intuizione è corretta, in quanto *il limite di una successione $\{x_n\}$ crescente è uguale al suo estremo superiore* (vedi l'Osservazione 5.9):

$$\text{se } x_n \text{ è crescente allora } \lim_{n \rightarrow +\infty} x_n = \sup\{x_n \mid n \in \mathbb{N}\}.$$

Infatti, indichiamo con s l'estremo superiore della successione crescente $\{x_n\}$; vogliamo far vedere che s è limite di $\{x_n\}$. Distinguiamo i due casi $s = +\infty$ e $s \in \mathbb{R}$.

Dire che $s = +\infty$ vuol dire che la successione $\{x_n\}$ è illimitata superiormente, cioè per ogni $M > 0$ esiste $N \in \mathbb{N}$ tale che $x_N > M$. Ma la successione è crescente: quindi per ogni $M > 0$ abbiamo $x_n \geq x_N > M$ per ogni $n \geq N$, quindi x_n tende a $+\infty$.

Supponiamo allora che $s \in \mathbb{R}$, cioè che la successione $\{x_n\}$ sia limitata superiormente, e scegliamo $\varepsilon > 0$ arbitrariamente piccolo. Sicuramente $s - x_n \geq 0$ per ogni $n \in \mathbb{N}$, in quanto $s \geq x_n$ per definizione. D'altra parte, deve esistere $N \in \mathbb{N}$ tale che $s - \varepsilon < x_N$; infatti altrimenti s non sarebbe il più piccolo numero maggiore di tutti gli elementi della successione. Ma la successione $\{x_n\}$ è crescente; quindi $s - \varepsilon < x_n$, e dunque $s - x_n < \varepsilon$, per ogni $n \geq N$. Riassumendo, abbiamo mostrato (ricordando che $s - x_n$ è sempre positivo) che per ogni $\varepsilon > 0$ esiste $N \in \mathbb{N}$ tale che $n \geq N$ implica $|s - x_n| < \varepsilon$, per cui $x_n \rightarrow s$, come voluto.

In modo analogo si verifica che *il limite di una successione $\{x_n\}$ decrescente è uguale al suo estremo inferiore*:

$$\text{se } x_n \text{ è decrescente, allora } \lim_{n \rightarrow +\infty} x_n = \inf\{x_n \mid n \in \mathbb{N}\}.$$

Curiosità 6.13 (I numeri di Fibonacci e la sezione aurea) La successione di Fibonacci è strettamente legata a un numero reale piuttosto noto: la **sezione aurea**

$$\phi = \frac{1 + \sqrt{5}}{2}.$$

Per esempio, la sezione aurea fornisce una formula inaspettata per il termine generale della successione di Fibonacci:

$$F_{n+2} = \phi^n + (1 - \phi)^n. \quad (6.4C)$$

Puoi verificare facilmente che F_2 e F_3 sono effettivamente dati dal secondo membro di (6.4C) per $n = 0, 1$ rispettivamente. Per vedere che (6.4C) vale sempre basta notare che ϕ e $1 - \phi$ sono le radici dell'equazione

$$x^2 - x - 1 = 0;$$

quindi

$$\begin{aligned} \phi^{n+1} + (1 - \phi)^{n+1} &= \phi^{n-1}\phi^2 + (1 - \phi)^{n-1}(1 - \phi)^2 \\ &= \phi^{n-1}(\phi + 1) + (1 - \phi)^{n-1}(1 - \phi + 1) \\ &= [\phi^n + (1 - \phi)^n] + [\phi^{n-1} + (1 - \phi)^{n-1}]. \end{aligned}$$

In altre parole, il secondo membro di (6.4C) definisce una successione che comincia come la successione di Fibonacci e soddisfa le stesse regole della successione di Fibonacci – e quindi è la successione di Fibonacci. Nota che il secondo membro contiene potenze di numeri irrazionali, mentre il primo membro è sempre un numero naturale!

Curiosità 6.14 (La serie armonica) Un modo per vedere che la serie armonica diverge consiste nel raggrupparne i termini in maniera furba. Poniamo

$$q_0 = 1, q_1 = \frac{1}{2}, q_2 = \frac{1}{3} + \frac{1}{4}, q_3 = \frac{1}{5} + \dots + \frac{1}{8}, \dots, q_n = \frac{1}{2^{n-1} + 1} + \dots + \frac{1}{2^n}.$$

Abbiamo

$$q_n = \frac{1}{2^{n-1} + 1} + \dots + \frac{1}{2^n} > \frac{1}{2^n} + \dots + \frac{1}{2^n} = 2^{n-1} \cdot \frac{1}{2^n} = \frac{1}{2};$$

quindi

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots = q_0 + q_1 + q_2 + q_3 + \dots > 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \dots,$$

e siccome l'ultimo membro diverge, la serie armonica non può far altro che divergere.

Curiosità 6.15 (Il criterio del rapporto) Il confronto con una serie geometrica fornisce un utile criterio di convergenza per serie a termini positivi: il **criterio del rapporto**. Sia $\{x_n\}$ una successione a termini positivi per cui esiste un $0 < p < 1$ tale che $x_{n+1}/x_n \leq p$ per ogni $n \in \mathbb{N}$: allora la serie $\sum_{j=0}^{+\infty} x_j$ è convergente. Infatti $x_1 \leq x_0 p$, $x_2 \leq x_1 p \leq x_0 p^2$, e in generale $x_n \leq x_0 p^n$, per cui

$$\sum_{j=0}^n x_j \leq x_0 \sum_{j=0}^n p^j = x_0 \frac{1-p^{n+1}}{1-p} < x_0 \frac{1}{1-p} .$$

Quindi la successione delle somme parziali è limitata, per cui la serie è convergente.

Curiosità 7.1 (Teorema del valor medio di Lagrange) Una formula che lega la variazione media alla variazione istantanea esiste, ed è contenuta nel **Teorema del valor medio di Lagrange**. Questo teorema dice che se $f: [a, b] \rightarrow \mathbb{R}$ è una funzione continua, derivabile in tutti i punti dell'intervallo aperto (a, b) , allora per ogni coppia di punti $x_0 < x_1$ in $[a, b]$ esiste (almeno) un punto $x_0 < \bar{x} < x_1$ tale che la variazione media di f da x_0 a x_1 sia uguale alla variazione istantanea di f in \bar{x} :

$$\frac{f(x_1) - f(x_0)}{x_1 - x_0} = f'(\bar{x}).$$

In altre parole, la retta secante per $(x_0, f(x_0))$ e $(x_1, f(x_1))$ è parallela ad almeno una retta tangente al grafico nell'intervallo (x_0, x_1) (vedi la Figura 7.1C). È importante notare che questo teorema non ci dice né come trovare \bar{x} né se ce n'è più d'uno; ma fornisce comunque un importante legame fra variazione media e variazione istantanea.

Possiamo ora dimostrare che una funzione con derivata identicamente nulla in un intervallo $[a, b]$ è costante. Infatti, per ogni $x \in (a, b)$ il Teorema del valor medio di Lagrange ci fornisce un \bar{x} fra a e x tale che $f(x) - f(a) = (x - a)f'(\bar{x})$. Ma $f' \equiv 0$; quindi $f(x) = f(a)$ per ogni $x \in [a, b]$, cioè f è costante.

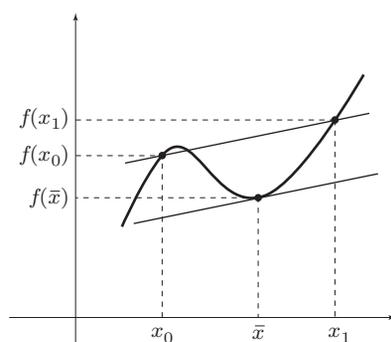


Figura 7.1C
Teorema del valor medio di Lagrange

Curiosità 7.2 (Principio d'induzione) La dimostrazione della formula (7.7) usa una procedura particolarmente utile in matematica, nota come **principio di induzione**. L'idea è la seguente: per dimostrare che una certa affermazione P_k , dipendente da un numero naturale k , è vera per ogni $k \geq 0$, basta dimostrare che:

- (I1) P_0 è vera; e che
- (I2) se P_k è vera allora anche P_{k+1} lo è.

Infatti, (I1) dice che P_0 è vera; allora usando (I2) per $k = 0$ otteniamo che anche P_1 è vera; ma allora usando (I2) con $k = 1$ otteniamo che anche P_2 è vera; ma allora usando (I2) con $k = 2$ otteniamo che anche P_3 è vera e così via, fino a dimostrare in un numero finito di passi qualsiasi P_k . Nota che siamo partiti con P_0 solo per semplicità d'esposizione; potevamo partire da P_1 , o da un qualsiasi P_{k_0} , e avremmo dimostrato P_k per tutti i $k \geq 1$, o per tutti i $k \geq k_0$.

Come esempio di applicazione del principio di induzione dimostriamo (7.7). Indichiamo con P_k l'eguaglianza (7.7) per un dato valore di k . Allora P_0 è banalmente vera (in quanto $f^0 \equiv 1$ per ogni funzione f), per cui (I1) è verificata. Supponiamo che P_k sia vera; applicando la regola di Leibniz a $f^{k+1} = f^k \cdot f$ e utilizzando P_k otteniamo

$$(f^{k+1})' = (f^k)'f + f^k f' = kf^{k-1} f' f + f^k f' = (k+1)f^k f' ,$$

per cui P_{k+1} è vera. Quindi abbiamo verificato anche (I2), e il principio di induzione ci assicura che (7.7) è vera per ogni $k \geq 0$.

Curiosità 7.3 (Derivazione di funzioni composte) Per ricavare la formula (7.11) della derivata di funzioni composte anche quando $f(x+h) - f(x) = 0$ per qualche $h \neq 0$ cominciamo notando che il fatto che g sia derivabile in $y = f(x)$ implica che se poniamo

$$s(h_1) = \frac{g(y+h_1) - g(y)}{h_1} - g'(y)$$

allora $s(h_1) \rightarrow 0$ per $h_1 \rightarrow 0$. Moltiplicando per h_1 otteniamo

$$g(y+h_1) - g(y) = g'(y)h_1 + s(h_1)h_1$$

che è vera anche quando $h_1 = 0$. Sostituendo $y = f(x)$ e $h_1 = f(x+h) - f(x)$ e dividendo per $h \neq 0$ troviamo

$$\begin{aligned} & \frac{g(f(x+h)) - g(f(x))}{h} \\ &= g'(f(x)) \cdot \frac{f(x+h) - f(x)}{h} + s(f(x+h) - f(x)) \frac{f(x+h) - f(x)}{h}. \end{aligned}$$

Facendo tendere h a 0 il rapporto incrementale di f tende a $f'(x)$, mentre $s(f(x+h) - f(x))$ tende a 0 perché $f(x+h) \rightarrow f(x)$, per cui otteniamo (7.11).

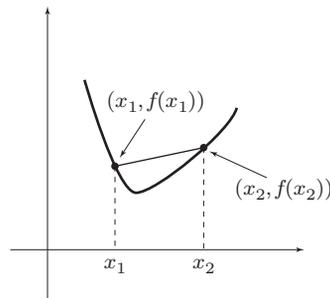
Curiosità 7.4 (Segno della derivata e monotonia della funzione) Lo strumento che permette di passare dall'informazione puntuale data dal segno della derivata all'informazione in tutto un intervallo data dalla monotonia della funzione è il teorema del valor medio di Lagrange (Curiosità 7.1). Infatti, supponiamo che $f'(x) > 0$ per ogni $x \in (a, b)$ e scegliamo due punti qualsiasi $x_1, x_2 \in (a, b)$ con $x_1 < x_2$. Allora il teorema del valor medio di Lagrange ci dice che esiste $\bar{x} \in (x_1, x_2)$ tale che $f(x_2) - f(x_1) = (x_2 - x_1)f'(\bar{x})$, quindi $f(x_1) < f(x_2)$ perché $f'(\bar{x}) > 0$ per ipotesi. Analogamente, se $f'(x) < 0$ per ogni $x \in (a, b)$ deduciamo che f è strettamente decrescente in (a, b) .

Curiosità 7.5 (Derivate di ordine superiore e massimi e minimi locali) Supponiamo che f si possa derivare n volte in x_0 e che si abbia $f'(x_0) = f''(x_0) = \dots = f^{(n-1)}(x_0) = 0$ ma $f^{(n)}(x_0) \neq 0$. Allora si può far vedere che

- se n è dispari, allora x_0 è un punto di flesso;
- se n è pari e $f^{(n)}(x_0) > 0$, allora x_0 è un minimo locale;
- se n è pari e $f^{(n)}(x_0) < 0$, allora x_0 è un massimo locale.

Curiosità 7.6 (Funzioni convesse) Si può definire cos'è una funzione convessa anche senza tirare in ballo le derivate: una funzione $f: I \rightarrow \mathbb{R}$ è detta **convessa** se per ogni $x_1 < x_2$ in I il grafico di f in $[x_1, x_2]$ si trova sotto il segmento secante congiungente $(x_1, f(x_1))$ con $(x_2, f(x_2))$ (vedi la Figura 7.2C). Analogamente, f è detta **concava** se per ogni $x_1 < x_2$ in I il grafico di f in $[x_1, x_2]$ si trova sopra il segmento secante congiungente $(x_1, f(x_1))$ con $(x_2, f(x_2))$. Si può dimostrare che ogni funzione che ammette derivata seconda sempre positiva (rispettivamente, negativa) è convessa (rispettivamente, concava) anche in questo senso.

Figura 7.2C
Convessità senza derivate



Curiosità 7.7 (Teorema del valor medio di Cauchy e regola di de l'Hôpital)

La regola di de l'Hôpital per $x_0 \in \mathbb{R}$ è conseguenza di una piccola generalizzazione, dovuta a Cauchy, del teorema del valor medio di Lagrange (vedi la Curiosità 7.1). Supponiamo di avere due funzioni f e g derivabili in un intervallo $[a, b]$. Allora per ogni $x_1, x_2 \in [a, b]$ con $x_1 < x_2$ esiste $\bar{x} \in (x_1, x_2)$ tale che

$$f'(\bar{x})[g(x_2) - g(x_1)] = g'(\bar{x})[f(x_2) - f(x_1)]$$

(teorema del valor medio di Cauchy).

Questo risultato segue subito dal teorema del valore medio di Lagrange applicato alla funzione $h(x) = f(x)[g(x_2) - g(x_1)] - g(x)[f(x_2) - f(x_1)]$.

In particolare, nelle ipotesi della regola di de l'Hôpital per la forma indeterminata 0/0 possiamo applicare questo risultato con $x_1 = x_0$ e $x_2 = x$, supponendo senza perdita di generalità che $f(x_0) = g(x_0) = 0$. Quindi non appena x è abbastanza vicino a x_0 esiste un \bar{x} fra x e x_0 tale che

$$\frac{f(x)}{g(x)} = \frac{f'(\bar{x})}{g'(\bar{x})}.$$

Facendo tendere x a x_0 anche \bar{x} tende a x_0 , per cui il limite di f/g esiste ed è uguale al limite di f'/g' .

Il ragionamento per la forma indeterminata ∞/∞ è un poco più elaborato; lo presentiamo nel caso di limite finito. Se x_1 e x sono sufficientemente vicini a x_0 possiamo scrivere

$$\frac{f'(\bar{x})}{g'(\bar{x})}[g(x) - g(x_1)] = f(x) - f(x_1),$$

dove \bar{x} è fra x e x_1 ; da questo segue che

$$\frac{1}{g(x)} \left(f(x_1) - \frac{f'(\bar{x})}{g'(\bar{x})} g(x_1) \right) = \frac{1}{g(x)} \left(f(x) - \frac{f'(\bar{x})}{g'(\bar{x})} g(x) \right) = \frac{f(x)}{g(x)} - \frac{f'(\bar{x})}{g'(\bar{x})}.$$

Quindi se indichiamo con ℓ il limite di f'/g' otteniamo

$$\frac{f(x)}{g(x)} - \ell = \frac{f'(\bar{x})}{g'(\bar{x})} - \ell + \frac{1}{g(x)} \left(f(x_1) - g(x_1) \frac{f'(\bar{x})}{g'(\bar{x})} \right),$$

per cui

$$\left| \frac{f(x)}{g(x)} - \ell \right| \leq \left| \frac{f'(\bar{x})}{g'(\bar{x})} - \ell \right| + \frac{1}{|g(x)|} \left(|f(x_1)| + |g(x_1)| \left| \frac{f'(\bar{x})}{g'(\bar{x})} \right| \right).$$

Scegliendo x_1 sufficientemente vicino a x_0 possiamo supporre che $f'(\bar{x})/g'(\bar{x})$ sia arbitrariamente vicino a ℓ . In particolare, il termine fra parentesi tonde a secondo membro è limitato; quindi scegliendo x sufficientemente vicino a x_0 possiamo supporre che il secondo addendo sia arbitrariamente piccolo, in quanto $|g(x)|$ diventa arbitrariamente grande per x sufficientemente vicino a x_0 . In conclusione, scegliendo x sufficientemente vicino a x_0 otteniamo $f(x)/g(x)$ arbitrariamente vicino a ℓ , come voluto.

Curiosità 7.8 (Serie di Taylor) Se f ammette derivate di ordine qualsiasi in x_0 (come accade per esempio per le funzioni esponenziali o trigonometriche) possiamo approssimare f con polinomi di Taylor di ordine arbitrariamente alto; quindi potremmo essere tentati di scrivere una formula del tipo

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2 + \frac{1}{3!}f'''(x_0)(x - x_0)^3 + \dots, \quad (7.1C)$$

dove i puntini sottintendono la "somma" di un'infinità di termini. Questa è una buona idea, ma non si può fare sempre. Vediamo come si può procedere e quali sono i problemi che si possono incontrare.

Per ogni numero x fissato, possiamo considerare la successione di numeri

$$a_j = \frac{1}{j!}f^{(j)}(x_0)(x - x_0)^j.$$

Allora il polinomio di Taylor $P_j(x)$ valutato in x coincide proprio con la somma parziale degli a_j , cioè

$$P_n(x) = \sum_{j=0}^n a_j.$$

Quindi la somma infinita a secondo membro della (7.1C) rappresenta la serie (vedi il Paragrafo 6.6) associata alla successione $\{a_j\}$ data da

$$T(x) = \sum_{j=0}^{+\infty} a_j = \sum_{j=0}^{+\infty} \frac{f^{(j)}(x_0)}{j!} (x - x_0)^j,$$

detta **serie di Taylor** per f in x_0 . Quando $x = x_0$ tutti gli addendi di questa serie tranne il primo sono nulli, per cui $T(x_0) = f(x_0)$. Ma per $x \neq x_0$ possono capitare tre cose:

- (a) la somma non esiste, cioè la serie non converge;
- (b) la serie converge a un valore $T(x) \in \mathbb{R}$, ma $T(x) \neq f(x)$;
- (c) la serie converge a un valore $T(x)$ uguale a $f(x)$.

La situazione migliore è chiaramente la (c): vuol dire che possiamo approssimare la funzione arbitrariamente bene con i polinomi di Taylor, in quanto (c) è equivalente ad avere

$$\lim_{n \rightarrow +\infty} E_n(x) = 0.$$

Quando questo accade per tutti gli x sufficientemente vicini a x_0 si dice che f è **analitica** in x_0 .

Ovviamente, a questo punto si pone il problema di stabilire per quali x la serie di Taylor converge a $f(x)$. Per esempio, si può dimostrare che la serie di Taylor dell'esponenziale converge all'esponenziale per *tutti* gli $x \in \mathbb{R}$; lo stesso accade per il seno o il coseno. Invece, la serie di Taylor di $1/(1-x)$ trovata nell'Esempio 7.38 non converge per tutti gli $x \in \mathbb{R}$; per esempio, è chiaro che diverge per ogni $x > 1$ (in quanto somma di termini sempre più grandi). Però si può dimostrare che converge a $1/(1-x)$ per $|x| < 1$, per cui la funzione $1/(1-x)$ è comunque analitica in 0. Lo studio di queste problematiche e in generale delle funzioni analitiche è un campo della matematica molto sviluppato e pieno di applicazioni.

Infine, voglio citare almeno un esempio di funzione f con derivate di ogni ordine la cui serie di Taylor converge in ogni punto ma $T(x) \neq f(x)$ per ogni $x > 0$. Si tratta della funzione f definita nella Curiosità 6.4. Infatti si può verificare (vedi l'Esercizio 7.73) che questa funzione ha derivate di ogni ordine, ma $f^{(j)}(0) = 0$ per ogni $j \in \mathbb{N}$; in questo senso la funzione f è infinitamente piatta in 0. La serie di Taylor di f in 0 è dunque identicamente nulla (e quindi in particolare convergente) ma $f(x) \neq 0$ per ogni $x > 0$.

Curiosità 7.9 (Stima di Lagrange) La stima di Lagrange dell'errore è conseguenza del teorema del valor medio di Cauchy introdotto nella Curiosità 7.7. Sia $f: I \rightarrow \mathbb{R}$ una funzione con $n+1$ derivate e $x_0 \in I$. Sia

$$P_n(x) = f(x_0) + f'(x_0)(x - x_0) + \dots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n$$

il polinomio di Taylor di grado n di f in x_0 . Fissiamo ora un punto $x_1 \in I$ e definiamo le funzioni

$$F(t) = f(t) + f'(t)(x_1 - t) + \dots + \frac{f^{(n)}(t)}{n!}(x_1 - t)^n \quad \text{e} \quad G(t) = (x_1 - t)^{n+1}.$$

Si verifica subito che $F(x_1) = f(x_1)$ e $F(x_0) = P_n(x_1)$, per cui si ha $F(x_1) - F(x_0) = E_n(x_1)$; inoltre,

$$\frac{dF}{dt}(t) = \frac{(x_1 - t)^n}{n!} f^{(n+1)}(t).$$

Applicando il teorema del valor medio di Cauchy alle funzioni F e G (usando le derivate rispetto a t , non rispetto a x_1) troviamo un punto \bar{x} fra x_0 e x_1 tale che

$$-\frac{(x_1 - \bar{x})^n}{n!} f^{(n+1)}(\bar{x})(x_1 - x_0)^{n+1} = -(n+1)(x_1 - \bar{x})^n E_n(x_1),$$

per cui otteniamo la **formula di Lagrange del resto**

$$E_n(x_1) = \frac{(x_1 - x_0)^{n+1}}{(n+1)!} f^{(n+1)}(\bar{x}).$$

Quindi se indichiamo con $M_n(x_1)$ il massimo del valore assoluto di $f^{(n+1)}$ fra x_0 e x_1 abbiamo chiaramente $|f^{(n+1)}(\bar{x})| \leq M_n(x_1)$ e quindi

$$|E_n(x_1)| \leq M_n(x_1) \frac{|x_1 - x_0|^{n+1}}{(n+1)!},$$

come voluto.

Curiosità 7.10 (Serie di Taylor dell'esponenziale) Vogliamo usare la stima dell'errore di Lagrange per dimostrare che la serie di Taylor di $f(x) = e^x$ (vedi la Curiosità 7.8) converge a e^x per ogni x reale, cioè che

$$\forall x \in \mathbb{R} \quad e^x = \sum_{j=0}^{+\infty} \frac{1}{j!} x^j.$$

Dobbiamo dimostrare che $\lim_{n \rightarrow +\infty} E_n(x) = 0$ per ogni $x \in \mathbb{R}$. La stima di Lagrange dell'errore ci dice (perché?) che

$$|E_n(x)| \leq C_x \frac{|x|^{n+1}}{(n+1)!},$$

dove $C_x = e^x$ se $x > 0$, e $C_x = 1$ se $x < 0$. Quindi ci basta dimostrare che

$$\forall a > 0 \quad \lim_{n \rightarrow +\infty} \frac{a^n}{n!} = 0.$$

Ma infatti, scegliamo $n_0 \in \mathbb{N}$ tale che $n_0 + 1 > 2a$. Allora per ogni $n > n_0$ abbiamo

$$\frac{a^n}{n!} = \frac{a}{1} \cdot \frac{a}{2} \cdots \frac{a}{n_0} \cdot \frac{a}{n_0+1} \cdots \frac{a}{n} \leq \frac{a}{1} \cdot \frac{a}{2} \cdots \frac{a}{n_0} \left(\frac{a}{n_0+1} \right)^{n-n_0} < \frac{(2a)^{n_0}}{n_0!} \frac{1}{2^n}$$

che tende a 0 per $n \rightarrow +\infty$, ed è fatta. In modo analogo si può dimostrare che

$$\sin x = \sum_{n=0}^{+\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!} \quad \text{e} \quad \cos x = \sum_{n=0}^{+\infty} (-1)^n \frac{x^{2n}}{(2n)!}.$$

Curiosità 7.11 (Il metodo di Newton-Raphson-Wallis) Vediamo perché il metodo di Newton-Raphson-Wallis funziona. Supponiamo per semplicità che la funzione f sia derivabile tre volte (capirai fra un attimo a cosa ci serve) e che la derivata f' non si annulli vicino alla radice \bar{x} che stiamo cercando. Poniamo

$$g(x) = x - \frac{f(x)}{f'(x)} ;$$

allora la funzione g è definita vicino a \bar{x} , è derivabile due volte, e il metodo di Newton-Raphson-Wallis consiste esattamente nell'iterare questa funzione: $x_1 = g(x_0)$, $x_2 = g(x_1) = g(g(x_0))$ e così via.

Ora, il punto cruciale è che \bar{x} è un punto critico di g . Infatti,

$$g' = 1 - \frac{(f')^2 - f''f}{(f')^2} = \frac{f''f}{(f')^2} ,$$

per cui $f(\bar{x}) = 0$ implica $g'(\bar{x}) = 0$. Inoltre, $g(\bar{x}) = \bar{x}$, sempre perché $f(\bar{x}) = 0$; quindi lo sviluppo di Taylor al primo ordine in \bar{x} di g è

$$g(x) = \bar{x} + o(x - \bar{x}) .$$

Siccome f è derivabile tre volte e f' non si annulla vicino a \bar{x} , la funzione g è derivabile due volte vicino a \bar{x} , per cui possiamo applicare la stima di Lagrange dell'errore (Curiosità 7.9) ottenendo

$$|g(x) - \bar{x}| \leq \frac{M_1(x)|x - \bar{x}|}{2} |x - \bar{x}| ,$$

dove $M_1(x)$ è il massimo di $|g''|$ fra \bar{x} e x . Ora, se prendiamo il punto iniziale x_0 sufficientemente vicino a \bar{x} , possiamo supporre (perché?) che si abbia $M_1(x_0)|x_0 - \bar{x}| < 1$; quindi

$$|x_1 - \bar{x}| = |g(x_0) - \bar{x}| < \frac{1}{2} |x_0 - \bar{x}| ,$$

cioè la distanza di x_1 da \bar{x} è meno della metà della distanza fra x_0 e \bar{x} . Ripetendo il procedimento, a ogni passaggio la distanza si dimezza, quindi otteniamo velocemente delle ottime approssimazioni della radice cercata.

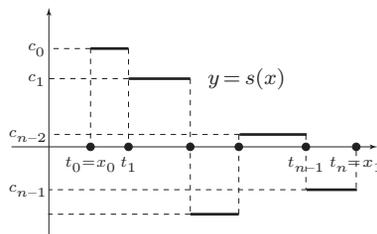
Per onestà va però detto che questo metodo ha due problemi. Il primo è che non si applica quando la derivata di f si annulla nella radice \bar{x} , cioè quando il grafico di f è tangente all'asse delle ascisse. Questo però non è un grosso problema: sommando o sottraendo a f una quantità piccola, trasliamo verticalmente il grafico in modo da fargli intersecare l'asse delle ascisse trasversalmente, per cui applicando il metodo di Newton-Raphson-Wallis alla funzione traslata otteniamo una buona approssimazione della radice originale \bar{x} .

Il secondo problema è più serio: come facciamo a scegliere un punto iniziale vicino alla radice se non sappiamo dove si trova la radice? Non basta scegliere un punto iniziale a caso; esistono esempi in cui il metodo di Newton-Raphson-Wallis applicato a certi punti iniziali non funziona: la successione $\{x_n\}$ non converge a una radice di f (prova con $f(x) = x^3 - 2x + 2$ e $x_0 = 0$). Ci sono vari metodi per superare questo problema; in particolare, molto recentemente è stata trovata una tecnica (valida quando f è un polinomio) per individuare un insieme finito e non troppo grosso di punti iniziali da cui si può partire con la certezza che per almeno uno di loro il metodo di Newton-Raphson-Wallis funziona.

Curiosità 8.1 (Integrale definito) Si può ottenere una definizione più generale di integrale definito ampliando un poco il procedimento usato per le funzioni continue.

Prima di tutto, una **funzione a scala** è una funzione costante a tratti. In altre parole, $s: [x_0, x_1] \rightarrow \mathbb{R}$ è una funzione a scala se esistono punti $t_0 < t_1 < \dots < t_n$ (non necessariamente equidistanti) con $t_0 = x_0$ e $t_n = x_1$, e numeri reali $c_0, \dots, c_{n-1} \in \mathbb{R}$ tali che $s(x)$ valga costantemente c_j sull'intervallo aperto (t_j, t_{j+1}) , per $j = 0, \dots, n-1$ (e fa quello che le pare negli estremi); vedi la Figura 8.1C.

Figura 8.1C
Funzione a scala

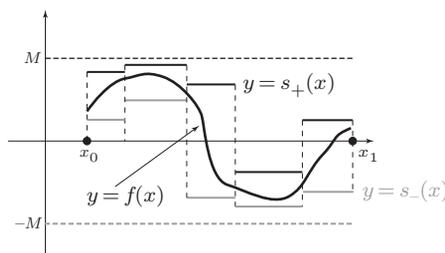


Siccome il sottografico di una funzione a scala è formato da un numero finito di rettangoli, l'area del sottografico, e quindi l'integrale definito, di una funzione a scala si calcola semplicemente con la formula

$$\int_{x_0}^{x_1} s(t) dt = \sum_{j=0}^{n-1} c_j (t_{j+1} - t_j).$$

Supponiamo ora che $f: [x_0, x_1] \rightarrow \mathbb{R}$ sia una funzione *limitata*, cioè per cui esiste un $M > 0$ tale che $|f(x)| < M$ per ogni $x \in [x_0, x_1]$; in altre parole, il grafico di f è contenuto in una porzione finita di piano. Indichiamo con \mathcal{F}_- la famiglia di tutte le funzioni a scala $s: [x_0, x_1] \rightarrow \mathbb{R}$ minori o uguali a f , cioè tali che $s(x) \leq f(x)$ per ogni $x \in [x_0, x_1]$; mentre sia \mathcal{F}_+ la famiglia di tutte le funzioni a scala $s: [x_0, x_1] \rightarrow \mathbb{R}$ maggiori o uguali a f , cioè tali che $s(x) \geq f(x)$ per ogni $x \in [x_0, x_1]$. Per esempio, la funzione costantemente uguale a $-M$ appartiene a \mathcal{F}_- , e la funzione costantemente uguale a $+M$ appartiene a \mathcal{F}_+ ; vedi la Figura 8.2C.

Figura 8.2C
 $-M < s_-(x) < f(x)$
 $< s_+(x) < M$ per ogni
 $x \in [x_0, x_1]$



Ora, l'insieme degli integrali definiti delle funzioni a scala appartenenti a \mathcal{F}_- è limitato superiormente (vedi l'Osservazione 5.9) da $M(x_1 - x_0)$ (perché); indichiamo con $I_-(f)$ l'estremo superiore (vedi di nuovo l'Osservazione 5.9) di questi integrali. Analogamente, l'insieme degli integrali delle funzioni a scala appartenenti a \mathcal{F}_+ è limitato inferiormente da $-M(x_1 - x_0)$ (perché); indichiamo con $I_+(f)$ l'estremo inferiore di questi integrali. Chiaramente (perché?) $I_-(f) \leq I_+(f)$ sempre; diremo che f è **integrabile** su $[x_0, x_1]$ se $I_-(f) = I_+(f)$, in tal caso l'**integrale definito** di f su $[x_0, x_1]$ è dato da questo valore comune:

$$\int_{x_0}^{x_1} f(t) dt = I_-(f) = I_+(f).$$

Non è difficile usare la definizione appena data per dimostrare che ogni funzione monotona è integrabile. Sia $f: [x_0, x_1] \rightarrow \mathbb{R}$ monotona; per semplicità, supponiamo sia

crescente (il ragionamento nel caso decrescente è analogo). Usando i simboli introdotti nel Paragrafo 8.1, per ogni $n \geq 1$ abbiamo (perché?)

$$\text{Area}(R_n(x_0, x_1)) \leq I_-(f) \leq I_+(f) \leq \text{Area}(R^n(x_0, x_1)) .$$

Quindi

$$\begin{aligned} 0 \leq I_+(f) - I_-(f) &\leq \text{Area}(R^n(x_0, x_1)) - \text{Area}(R_n(x_0, x_1)) \\ &\leq \frac{(f(x_1) - f(x_0))(x_1 - x_0)}{n} , \end{aligned}$$

grazie a (8.2). Siccome questo deve valere per ogni $n \geq 1$, il numero $I_+(f) - I_-(f)$ dev'essere non negativo e più piccolo di un qualsiasi numero positivo; quindi l'unica possibilità è che si abbia $I_+(f) - I_-(f) = 0$, come voluto.

Curiosità 8.2 (Integrale delle funzioni continue) Per dimostrare che le funzioni continue sono integrabili nel senso della Curiosità 8.1 ci serve un altro risultato. Abbiamo visto (Paragrafo 5.7) che una funzione $f: [a, b] \rightarrow \mathbb{R}$ è continua se per ogni $x_0 \in [a, b]$ e per ogni $\varepsilon > 0$ esiste un $\delta > 0$ (che a priori dipende sia da ε sia da x_0) tale che $|x - x_0| < \delta$ implica $|f(x) - f(x_0)| < \varepsilon$. In realtà si può dimostrare che siccome $[a, b]$ è un intervallo chiuso e limitato, allora per ogni $\varepsilon > 0$ esiste un $\delta > 0$ tale che $|x - x_0| < \delta$ implica $|f(x) - f(x_0)| < \varepsilon$ per ogni $x, x_0 \in [a, b]$; in altre parole, δ dipende solo da ε e non da x_0 . Si dice che f è **uniformemente continua** sull'intervallo $[a, b]$.

Sia allora $f: [x_0, x_1] \rightarrow \mathbb{R}$ una funzione continua. Fissiamo $N \geq 1$, e poniamo $\varepsilon = 1/N$. L'uniforme continuità ci dice (perché?) che possiamo trovare $n \geq 1$ tale che

$$|t - s| \leq (x_1 - x_0)/n \implies |f(t) - f(s)| < \varepsilon$$

per ogni $t, s \in [x_0, x_1]$. Suddividiamo allora l'intervallo $[x_0, x_1]$ nei soliti n intervallini di lunghezza $\Delta t = (x_1 - x_0)/n$. Con le notazioni introdotte sopra abbiamo

$$\sum_{j=0}^{n-1} f(t_j^-) \Delta t \leq L_-(f) \leq L_+(f) \leq \sum_{j=0}^{n-1} f(t_j^+) \Delta t.$$

Sappiamo che $t_j^-, t_j^+ \in [t_j, t_{j+1}]$, per cui $|t_j^- - t_j^+| \leq |t_{j+1} - t_j| = \Delta t$; quindi $|f(t_j^-) - f(t_j^+)| < \varepsilon$ e

$$\begin{aligned} \sum_{j=0}^{n-1} f(t_j^+) \Delta t - \sum_{j=0}^{n-1} f(t_j^-) \Delta t &= \sum_{j=0}^{n-1} (f(t_j^+) - f(t_j^-)) \Delta t \\ &< \varepsilon \sum_{j=0}^{n-1} \Delta t = \varepsilon (x_1 - x_0) = \frac{x_1 - x_0}{N}. \end{aligned}$$

Ne segue che

$$0 \leq L_+(f) - L_-(f) \leq \sum_{j=0}^{n-1} f(t_j^+) \Delta t - \sum_{j=0}^{n-1} f(t_j^-) \Delta t < \frac{x_1 - x_0}{N}.$$

Siccome questo deve valere per ogni $N \geq 1$, otteniamo di nuovo $L_+(f) - L_-(f) = 0$, come voluto.

Un ragionamento analogo (ma più complesso, e non particolarmente illuminante) mostra che per funzioni continue l'integrale definito introdotto nel Paragrafo 8.1 e quello introdotto nella Curiosità 8.1 coincidono.

Curiosità 8.3 (Integrale della somma di funzioni) Per dimostrare che l'integrale di una somma è uguale alla somma degli integrali usiamo la definizione di integrale definito data nella Curiosità 8.1. La somma di funzioni a scala è una funzione a scala (perché?) ed è facile vedere (controlla) che (8.8) è vera per le funzioni a scala. Inoltre, se s_1, s_2 sono funzioni a scala con $s_1 \leq f$ e $s_2 \leq g$, si ha $s_1 + s_2 \leq f + g$. Un risultato analogo vale se s_1 ed s_2 sono più grandi di f e g , per cui (perché?)

$$I_-(f) + I_-(g) \leq I_-(f + g) \leq I_+(f + g) \leq I_+(f) + I_+(g).$$

Ma f e g sono integrabili; quindi $I_-(f) = I_+(f)$ e $I_-(g) = I_+(g)$, per cui

$$I_-(f + g) = I_+(f + g) = I_-(f) + I_-(g) = I_+(f) + I_+(g),$$

che è quanto volevamo.

Curiosità 8.4 (Potenze irrazionali) Usando gli integrali possiamo dare una definizione precisa (equivalente a quella descritta nella Curiosità 5.7) di cosa vuol dire elevare un numero positivo a una potenza di esponente irrazionale, in un certo senso ripercorrendo nel verso opposto il percorso che abbiamo fatto fin qui.

L'idea è partire da una funzione derivabile definita su \mathbb{R}^+ , che valga 0 nel punto 1 e la cui derivata sia uguale a $1/x$; chiamiamo "log" questa funzione. Il primo teorema fondamentale del calcolo ci dice che "log"(x) è l'integrale di $1/t$ sull'intervallo $[1, x]$:

$$\text{"log"}(x) = \int_1^x \frac{1}{t} dt .$$

La prima osservazione è che per ogni $a > 0$ si ha

$$\frac{d}{dx} \text{"log"}(ax) = \frac{a}{ax} = \frac{1}{x} ;$$

cioè anche "log"(ax) risolve l'equazione differenziale $f' = 1/x$. Quindi deve esistere $c \in \mathbb{R}$ (dipendente da a) tale che "log"(ax) = "log"x + c. Ponendo $x = 1$ troviamo $c = \text{"log"}a$, e quindi abbiamo dimostrato che

$$\text{"log"}(xy) = \text{"log"}x + \text{"log"}y \quad (8.1C)$$

per ogni $x, y > 0$. Ora, la derivata di "log" è sempre positiva in \mathbb{R}^+ ; nel Paragrafo 7.4 abbiamo visto che questo implica che "log" è strettamente crescente. In particolare, è invertibile; indichiamo con "exp" la funzione inversa. Siccome la derivata di "log" non si annulla mai, "exp" è derivabile ovunque, e si verifica come nel Paragrafo 7.3 che

$$\frac{d}{dx} \text{"exp"}(x) = \text{"exp"}(x) .$$

Inoltre (8.1C) implica che

$$\text{"exp"}(x + y) = \text{"exp"}(x) \cdot \text{"exp"}(y) \quad (8.2C)$$

per ogni $x, y \in \mathbb{R}$. In particolare, se poniamo $e = \text{"exp"}(1)$, otteniamo $e^{p/q} = \text{"exp"}(p/q)$ per ogni $p/q \in \mathbb{Q}$. Questo suggerisce di *definire* e^x per $x \in \mathbb{R}$ qualsiasi ponendolo uguale a "exp"(x); in particolare, siccome "exp" è una funzione continua e derivabile, otteniamo che $x \mapsto e^x$ è una funzione continua e derivabile. Per ogni $a > 0$ definiamo allora a^x con la formula $a^x = \text{"exp"}(x \text{"log"}a)$; le formule (8.1C) e (8.2C) ci assicurano che a^x coincide con la solita definizione di potenza quando x è razionale. Infine,

$$(1 + rx)^{1/x} = \text{"exp"}\left(\frac{1}{x} \text{"log"}(1 + rx)\right) ;$$

ma

$$\begin{aligned} \lim_{x \rightarrow 0} \frac{1}{x} \text{"log"}(1 + rx) &= r \lim_{x \rightarrow 0} \frac{\text{"log"}(1 + rx) - \text{"log"}(1)}{rx} \\ &= r \lim_{h \rightarrow 0} \frac{\text{"log"}(1 + h) - \text{"log"}(1)}{h} = r \frac{d \text{"log"}}{dx}(1) = r , \end{aligned}$$

per cui

$$\begin{aligned} \lim_{x \rightarrow 0} (1 + rx)^{1/x} &= \lim_{x \rightarrow 0} \text{"exp"}\left(\frac{1}{x} \text{"log"}(1 + rx)\right) \\ &= \text{"exp"}\left(\lim_{x \rightarrow 0} \left(\frac{1}{x} \text{"log"}(1 + rx)\right)\right) = \text{"exp"}(r) . \end{aligned}$$

In particolare, "exp"(1) è il numero di Nepero, e dunque "exp" coincide con la funzione esponenziale come l'abbiamo definita noi, e "log" è il nostro logaritmo naturale.

Curiosità 8.5 (Calcolo geometrico di un integrale improprio) Un integrale improprio particolarmente importante (come vedremo nel Capitolo 12) è

$$\int_0^{+\infty} e^{-x^2} dx .$$

Descriviamo ora una tecnica che ci permette di calcolarlo anche senza bisogno di determinare gli integrali definiti

$$I(b) = \int_0^b e^{-x^2} dx$$

che, come detto nell'Esempio 8.12, non si calcolano con funzioni elementari.

Cominciamo introducendo la funzione di due variabili $F(x, y) = e^{-x^2-y^2}$ (vedi la Paragrafo 7.10 per una velocissima presentazione delle funzioni di due variabili); il grafico di $F: \mathbb{R}^2 \rightarrow \mathbb{R}$ è l'insieme $\Gamma_F = \{(x, y, F(x, y)) \mid (x, y) \in \mathbb{R}^2\} \subset \mathbb{R}^3$. Indichiamo con $Q(b) \subset \mathbb{R}^2$ il quadrato di lato $2b$ centrato nell'origine del piano; in altre parole,

$$Q(b) = \{(x, y) \in \mathbb{R}^2 \mid |x| \leq b, |y| \leq b\} .$$

Indichiamo poi con $\Gamma_F(b)$ il sottografico di F sopra $Q(b)$, cioè l'insieme

$$\Gamma_F(b) = \{(x, y, z) \mid (x, y) \in Q(b), 0 \leq z \leq F(x, y)\}$$

compreso fra il grafico di F e il quadrato $Q(b)$; vedi la Figura 8.3C(a). Se intersechiamo $\Gamma_F(b)$ con il piano di equazione $x = x_0$ (cioè il piano costituito dai punti di ascissa x_0), otteniamo (vedi la Figura 8.4C) un insieme congruente al sottografico della funzione $e^{-x_0^2}e^{-y^2}$ sopra l'intervallo $[-b, b]$.

L'area di questo insieme è

$$A_b(x_0) = \int_{-b}^b e^{-x_0^2}e^{-y^2} dy = e^{-x_0^2} \left[\int_{-b}^0 e^{-y^2} dy + I(b) \right] = 2e^{-x_0^2}I(b) ,$$

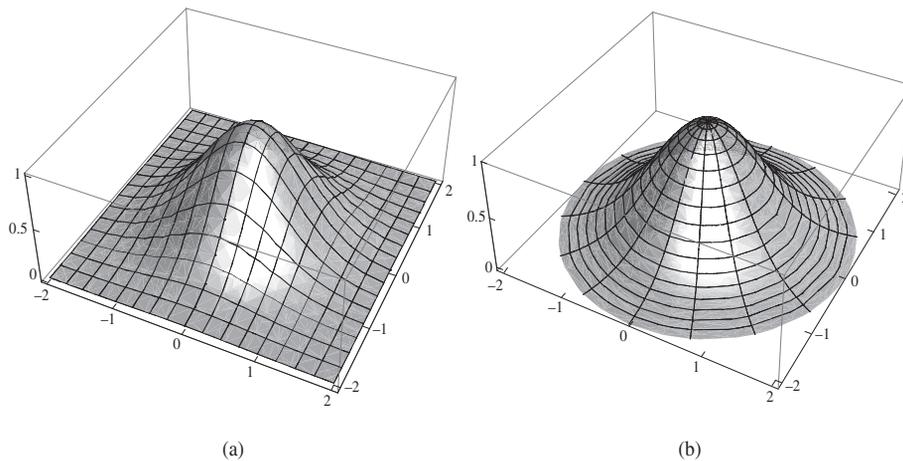
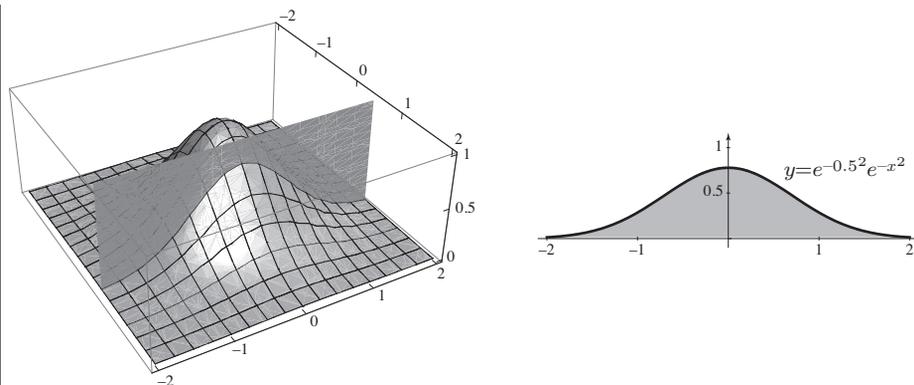


Figura 8.3C
 $F(x, y) = e^{-x^2-y^2}$ (a) su $Q(2)$
 (b) su $C(2)$

Figura 8.4C
 $\Gamma_F(2) \cap \{x = 0.5\}$



dove abbiamo usato il fatto che $e^{-(-y)^2} = e^{-y^2}$, per cui la sostituzione $g(y) = -y$ dà

$$\int_{-b}^0 e^{-y^2} dy = - \int_b^0 e^{-g^2} dg = I(b).$$

Ora, come l'area del sottografico di una funzione di una variabile è l'integrale delle lunghezze dei segmenti ottenuti intersecando il sottografico con le rette $x = x_0$, anche il volume del sottografico di una funzione di due variabili si può calcolare come l'integrale delle aree delle figure ottenute intersecando il sottografico con i piani $x = x_0$. Quindi il volume $V(b)$ di $\Gamma_F(b)$ è dato da

$$V(b) = \int_{-b}^b A_b(x) dx = 2I(b) \int_{-b}^b e^{-x^2} dx = 4I(b)^2.$$

Adesso consideriamo invece un cerchio $C(r)$ di centro l'origine e raggio $r > 0$ e indichiamo con $W(r)$ il volume del sottografico di F sopra il cerchio $C(r)$; vedi la Figura 8.3C(b). Siccome $C(b) \subset Q(b) \subset C(\sqrt{2}b)$, abbiamo

$$W(b) < V(b) < W(\sqrt{2}b).$$

Il motivo per cui abbiamo introdotto $W(b)$ è che possiamo calcolarlo esplicitamente. Infatti possiamo vedere il sottografico di F sopra $C(b)$ come l'unione, al variare di $r \in [0, b]$, dei cilindri di altezza e^{-r^2} e base la circonferenza di centro l'origine e raggio r . Allora il volume $W(b)$ è l'integrale delle aree delle superfici laterali di questi cilindri, per cui

$$W(b) = \int_0^b 2\pi r e^{-r^2} dr = \pi(1 - e^{-b^2}),$$

grazie a (8.19). Analogamente troviamo $W(\sqrt{2}b) = \pi(1 - e^{-2b^2})$; quindi

$$\pi(1 - e^{-b^2}) < 4I(b)^2 < \pi(1 - e^{-2b^2}).$$

Mandando $b \rightarrow +\infty$, sia il primo sia il terzo termine tendono a π , quindi, grazie al teorema dei due carabinieri (Paragrafo 5.7), il termine centrale è forzato ad avere lo stesso limite. Dunque $4I(+\infty)^2 = \pi$ e

$$\int_0^{+\infty} e^{-x^2} dx = \frac{\sqrt{\pi}}{2}.$$

Nel Capitolo 12 servirà l'integrale improprio di $f(x) = e^{-x^2/2}$ su tutta la retta. Usando la sostituzione $u = x/\sqrt{2}$ possiamo scrivere

$$\int_0^b e^{-x^2/2} dx = \sqrt{2} \int_0^{b/\sqrt{2}} e^{-u^2} du;$$

quindi

$$\int_0^{+\infty} e^{-x^2/2} dx = \sqrt{2} \lim_{b \rightarrow +\infty} \int_0^{b/\sqrt{2}} e^{-u^2} du = \sqrt{2} \int_0^{+\infty} e^{-u^2} du = \frac{\sqrt{2\pi}}{2} .$$

Siccome $\int_{-\infty}^0 e^{-x^2/2} dx = \int_0^{+\infty} e^{-x^2/2} dx$, otteniamo infine (vedi la Figura 8.5C)

$$\int_{-\infty}^{+\infty} e^{-x^2/2} dx = \sqrt{2\pi} .$$

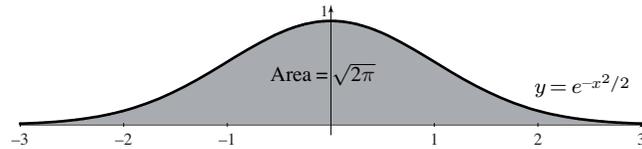


Figura 8.5C
 $\int_{-\infty}^{+\infty} e^{-x^2/2} dx$

Curiosità 9.1 (Separazione delle variabili) Vogliamo risolvere un'equazione differenziale della forma

$$y'(t) = \alpha(t)f(y(t)), \quad (9.1C)$$

per opportune funzioni α e f . Supponiamo (ipotesi da verificare a posteriori) di avere una soluzione y tale che $f(y(t))$ non si annulli mai. Allora possiamo scrivere (9.1C) nella forma

$$\frac{1}{f(y(t))} y'(t) = \alpha(t).$$

Integriamo entrambi i membri sull'intervallo di estremi t_0 e t ; otteniamo

$$\int_{t_0}^t \frac{1}{f(y(s))} y'(s) ds = \int_{t_0}^t \alpha(s) ds.$$

Ora, il metodo di integrazione per sostituzione (vedi il Paragrafo 8.5) ci dice che se poniamo

$$F(y) = \int_{y_0}^y \frac{1}{f(s)} ds,$$

allora

$$\int_{t_0}^t \frac{1}{f(y(s))} y'(s) ds = F(y(t)) + C$$

per un'opportuna costante C . Quindi (cambiando segno a C , che è arbitraria)

$$F(y(t)) = \int_{t_0}^t \alpha(s) ds + C. \quad (9.2C)$$

A volte questi conti vengono riassunti con la formula

$$\int \frac{1}{f(y)} dy = \int \alpha(s) ds + C.$$

Se la funzione F è invertibile, possiamo applicare F^{-1} a entrambi i membri di (9.2C) ottenendo la soluzione

$$y(t) = F^{-1} \left(\int_{t_0}^t \alpha(s) ds + C \right).$$

Se la y così ottenuta effettivamente è tale che $f(y(t))$ non si annulla mai, allora abbiamo trovato una soluzione dell'equazione (9.1C). Inoltre, se vogliamo una soluzione che soddisfi la condizione iniziale $y(t_0) = y_0$, basta prendere $C = 0$. Infatti,

$$y(t) = F^{-1} \left(\int_{t_0}^t \alpha(s) ds \right)$$

soddisfa $y(t_0) = F^{-1}(0) = y_0$ perché $F(y_0) = 0$.

Perché questo metodo possa funzionare occorre che si abbia $f(y_0) \neq 0$; altrimenti la funzione F è difficile sia ben definita. Viceversa, se $f(y_0) \neq 0$ allora questo metodo ci fornisce sempre una soluzione (teorica, non necessariamente esplicita) di (9.1C) vicino a t_0 . Infatti, siccome f è continua e $f(y_0) \neq 0$, abbiamo $f(y) \neq 0$ per y abbastanza vicino a y_0 ; più precisamente, $f(y)$ ha lo stesso segno (perché) di $f(y_0)$ per y abbastanza vicino a y_0 . Quindi $F(y)$ è definita per y vicino a y_0 ; inoltre, il primo teorema fondamentale del calcolo (Paragrafo 8.2) ci assicura che $F'(y) = 1/f(y)$, per cui anche $F'(y)$ ha lo stesso segno di $f(y_0)$ per y vicino a y_0 . In particolare, F è strettamente crescente vicino a y_0 se $f(y_0) > 0$, e è strettamente decrescente vicino a y_0 se $f(y_0) < 0$. In entrambi i casi, F è strettamente monotona in un intervallo I contenente y_0 ; quindi $F(I)$ è un intervallo

(per il teorema dei valori intermedi: vedi il Paragrafo 5.7) contenente $F(y_0) = 0$, e $F: I \rightarrow F(I)$ è invertibile.

Ora, per t vicino a t_0 il valore dell'integrale $\int_{t_0}^t \alpha(s) ds$ è vicino a 0, per cui appartiene al dominio $F(I)$ della funzione inversa F^{-1} di F . Quindi per t sufficientemente vicino a t_0 la funzione

$$y(t) = F^{-1} \left(\int_{t_0}^t \alpha(s) ds \right)$$

è effettivamente definita e soddisfa la condizione iniziale $y(t_0) = y_0$. Infine, usando le formule di derivazione della funzione inversa e delle funzioni composte, si verifica anche che y è una soluzione di (9.1C). Infatti,

$$\begin{aligned} y'(t) &= \frac{1}{F' \left(F^{-1} \left(\int_{t_0}^t \alpha(s) ds \right) \right)} \alpha(t) = \frac{1}{F'(y(t))} \alpha(t) \\ &= \frac{1}{1/f(y(t))} \alpha(t) = \alpha(t)f(y(t)). \end{aligned}$$

Curiosità 9.2 (Equazioni differenziali lineari non omogenee) Con un trucco classico possiamo risolvere l'equazione non omogenea

$$y'(t) = \alpha(t)y(t) + \beta(t) . \quad (9.3C)$$

Infatti, supponiamo di aver trovato una funzione mai nulla I (detta **fattore integrante**) tale che

$$Iy' - I\alpha y = (Iy)' . \quad (9.4C)$$

Allora moltiplicando (9.3C) per I possiamo riscriverla come

$$(Iy)' = I\beta ,$$

per cui ricaviamo

$$Iy = \int I(s)\beta(s) ds + C \implies y = \frac{1}{I} \int I(s)\beta(s) ds + \frac{C}{I} .$$

Quindi ci siamo ricondotti a risolvere (9.4C). Ora, $(Iy)' = Iy' + I'y$; quindi I soddisfa (9.4C) se si ha

$$I' = -\alpha I .$$

Ma abbiamo visto nel Paragrafo 9.3 che le soluzioni di questa equazione sono della forma

$$I(t) = c \exp \left(- \int_{t_0}^t \alpha(s) ds \right) .$$

Siccome ce ne basta una qualsiasi, possiamo prendere $c = 1$ e abbiamo trovato l'espressione generale delle soluzioni di (9.3C):

$$y(t) = \exp \left(\int_{t_0}^t \alpha(s) ds \right) \cdot \int_{t_0}^t e^{-\int_{s_0}^s \alpha(u) du} \beta(s) ds + C \exp \left(\int_{t_0}^t \alpha(s) ds \right) .$$

Nota che y è la somma di due pezzi: una soluzione particolare dell'equazione (9.3C) e la soluzione generale dell'equazione omogenea associata $y'(t) = \alpha(t)y(t)$.

Curiosità 9.3 (L'equazione $y' = ay^2 + by + c$) Proviamo ad applicare il metodo di separazione delle variabili all'equazione

$$y' = ay^2 + by + c, \quad (9.5C)$$

con $a \neq 0, b, c \in \mathbb{R}$.

Supponendo che $ay^2 + by + c$ non si annulli mai possiamo scrivere (9.5C) nella forma

$$\frac{y'}{ay^2 + by + c} = 1.$$

In particolare, $\alpha(s) \equiv 1$ e $f(s) = as^2 + bs + c$. Quindi dobbiamo calcolare

$$\int_{y_0}^y \frac{1}{as^2 + bs + c} ds.$$

Cominciamo supponendo che il polinomio $ay^2 + bx + c$ abbia due radici reali distinte, in modo da poter scrivere

$$ay^2 + by + c = a(y - y_1)(y - y_2),$$

dove $y_1 \neq y_2$ sono le due radici, ordinate in modo che $a(y_2 - y_1) > 0$. Nota che $ay(t)^2 + by(t) + c = 0$ se e solo se $y(t) = y_1$ o $y(t) = y_2$; quindi a posteriori dovremo verificare che la soluzione y trovata non assuma mai i valori y_1 e y_2 .

Il trucco che ci permette di calcolare l'integrale è l'identità

$$\frac{1}{a(y - y_1)(y - y_2)} = \frac{1}{a(y_2 - y_1)} \left[\frac{1}{y - y_2} - \frac{1}{y - y_1} \right].$$

Allora

$$\begin{aligned} \int \frac{1}{ay^2 + by + c} dy &= \frac{1}{a(y_2 - y_1)} \left[\int \frac{1}{y - y_2} dy - \int \frac{1}{y - y_1} dy \right] \\ &= \frac{1}{a(y_2 - y_1)} [\log |y - y_2| - \log |y - y_1|] + C \\ &= \frac{1}{a(y_2 - y_1)} \log \left| \frac{y - y_2}{y - y_1} \right| + C. \end{aligned}$$

Dunque abbiamo

$$F(y) = \int_{y_0}^y \frac{1}{as^2 + bs + c} ds = \frac{1}{a(y_2 - y_1)} \left[\log \left| \frac{y - y_2}{y - y_1} \right| - \log \left| \frac{y_0 - y_2}{y_0 - y_1} \right| \right],$$

e

$$\frac{1}{a(y_2 - y_1)} \left[\log \left| \frac{y(t) - y_2}{y(t) - y_1} \right| - \log \left| \frac{y_0 - y_2}{y_0 - y_1} \right| \right] = F(y(t)) = \int_{t_0}^t 1 ds = t - t_0,$$

da cui segue

$$\frac{y(t) - y_2}{y(t) - y_1} = -C e^{a(y_2 - y_1)t}, \quad (9.6C)$$

dove

$$C = -\frac{y_0 - y_2}{y_0 - y_1} e^{-a(y_2 - y_1)t_0}.$$

Ricavando $y(t)$ da (9.6C) e manipolando un poco il risultato troviamo

$$y(t) = \frac{y_2 + C y_1 e^{a(y_2 - y_1)t}}{1 + C e^{a(y_2 - y_1)t}} = y_2 + \frac{(y_1 - y_2) C e^{a(y_2 - y_1)t}}{1 + C e^{a(y_2 - y_1)t}}$$

e infine

$$y(t) = y_2 + \frac{y_1 - y_2}{1 + C^{-1} e^{-a(y_2 - y_1)t}}. \quad (9.7C)$$

Siccome $y(t)$ è sempre diversa da y_1 e y_2 (verifica; vedi l'Esercizio 9.34), abbiamo effettivamente trovato una soluzione della nostra equazione differenziale. Inoltre, ricordando il valore di C otteniamo

$$\begin{aligned} y(t_0) &= y_2 + \frac{y_1 - y_2}{1 + C^{-1}e^{-a(y_2 - y_1)t_0}} = y_2 + \frac{y_1 - y_2}{1 - (y_0 - y_1)/(y_0 - y_2)} \\ &= y_2 + \frac{(y_0 - y_2)(y_1 - y_2)}{(y_0 - y_2) - (y_0 - y_1)} = y_0. \end{aligned}$$

Infine, se indichiamo con $\Delta = b^2 - 4ac$ il discriminante dell'equazione $ay^2 + by + c = 0$, la formula di risoluzione delle equazioni di secondo grado ci dice che

$$a(y_2 - y_1) = \sqrt{\Delta}.$$

e quindi possiamo scrivere

$$y(t) = y_2 + \frac{-\sqrt{\Delta}/a}{1 + C_1 e^{-\sqrt{\Delta}t}},$$

dove

$$C_1 = -\frac{y_0 - y_1}{y_0 - y_2} e^{\sqrt{\Delta}t_0},$$

e y_1, y_2 sono le due radici (reali e distinte) di $ay^2 + by + c = 0$, ordinate in modo che $a(y_2 - y_1) > 0$.

Nota che se $C < 0$ può succedere che il denominatore in (9.7C) si annulli; quindi può capitare che la soluzione non sia definita su tutto l'asse reale.

Vediamo ora come risolvere (9.5C) quando $\Delta = b^2 - 4ac = 0$, per cui l'equazione $ay^2 + by + c = 0$ ha una sola radice reale $y_1 = -b/2a \in \mathbb{R}$ e $ay^2 + by + c = a(y - y_1)^2$. Allora

$$\int \frac{1}{ay^2 + by + c} dy = \frac{1}{a} \int \frac{1}{(y - y_1)^2} dy = -\frac{1}{a(y - y_1)} + C.$$

Quindi

$$F(y) = \int_{y_0}^y \frac{1}{as^2 + bs + c} ds = \frac{1}{a(y_1 - y)} - \frac{1}{a(y_1 - y_0)}$$

e

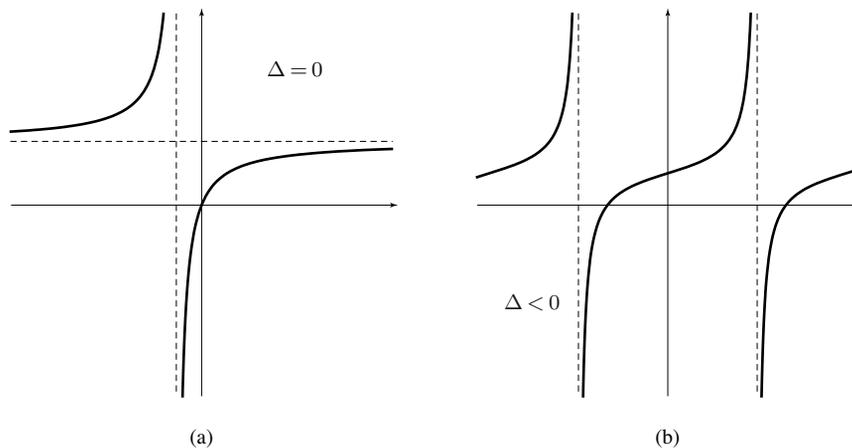
$$\frac{1}{a(y_1 - y(t))} - \frac{1}{a(y_1 - y_0)} = F(y(t)) = \int_{t_0}^t ds = t - t_0,$$

per cui

$$y(t) = y_1 - \frac{y_1 - y_0}{1 + a(y_1 - y_0)(t - t_0)}; \quad (9.8C)$$

vedi la Figura 9.1C(a). Nota che se $y_0 = y_1$, la (9.8C) si riduce a $y \equiv y_0$.

Figura 9.1C
 $y' = ay^2 + by + c$
per (a) $\Delta = 0$ (b) $\Delta < 0$



Supponiamo infine che $\Delta = b^2 - 4ac < 0$, per cui l'equazione $ay^2 + by + c = 0$ non ha radici reali. Allora possiamo scrivere

$$ay^2 + by + c = a(u^2 + \rho^2),$$

dove $u = y + b/2a$ e $\rho = \frac{1}{2a}\sqrt{-\Delta}$. Quindi si ha

$$\begin{aligned} \int \frac{1}{ay^2 + by + c} dy &= \frac{1}{a} \int \frac{1}{u^2 + \rho^2} du = \frac{1}{a\rho} \int \frac{1}{(u/\rho)^2 + 1} \frac{1}{\rho} du \\ &= \frac{1}{a\rho} \arctan(u/\rho) + C = \frac{2}{\sqrt{-\Delta}} \arctan\left(\frac{2ay + b}{\sqrt{-\Delta}}\right) + C. \end{aligned}$$

Ricavando da questa formula $F(y)$ come al solito, e ponendo $F(y(t))$ uguale a $t - t_0$, otteniamo

$$\frac{2ay(t) + b}{\sqrt{-\Delta}} = \tan\left(\frac{\sqrt{-\Delta}}{2}(t - t_0) + C\right).$$

Quindi otteniamo

$$y(t) = \frac{\sqrt{-\Delta}}{2a} \tan\left(\frac{\sqrt{-\Delta}}{2}(t - t_0) + C\right) - \frac{b}{2a}, \quad (9.9C)$$

dove

$$C = \arctan\left(\frac{2ay_0 + b}{\sqrt{-\Delta}}\right);$$

vedi la Figura 9.1C(b).

Nota infine che sia (9.8C) sia (9.9C) non sono definite per tutti i valori di t . Infatti, nel primo caso la soluzione diverge per

$$t = t_0 + \frac{1}{a(y_0 - y_1)};$$

invece nel secondo caso la soluzione diverge per

$$t = t_0 - \frac{2}{\sqrt{-\Delta}} \left(\pm \frac{\pi}{2} + C\right).$$

Curiosità 9.4 (Un'altra applicazione del metodo di separazione delle variabili) Il metodo di separazione delle variabili può essere utile anche quando non si riesce a invertire esplicitamente la funzione F . Infatti, in ogni caso arriviamo alla formula

$$F(y(t)) = \int_{t_0}^t \alpha(s) ds + C$$

che esprime una relazione funzionale (ma non differenziale: non compaiono derivate) fra y e t che a volte fornisce comunque informazioni interessanti anche senza bisogno di invertire la funzione F .

Per esempio, prendiamo l'equazione differenziale

$$y' = -\frac{t}{y}.$$

Separando le variabili otteniamo

$$yy' = -t.$$

Integrando entrambi i membri rispetto a t otteniamo

$$\frac{1}{2}y^2 = \int y dy = \int yy' dt = \int -t dt = -\frac{1}{2}t^2 + C,$$

per cui troviamo

$$\frac{1}{2}y^2 = -\frac{1}{2}t^2 + C \implies y^2 + t^2 = r^2,$$

dove $r^2 = 2C$. Quindi, anche senza bisogno di risolvere esplicitamente l'equazione, possiamo dire che i punti del grafico di una soluzione dell'equazione $y' = -t/y$ appartengono sempre a una circonferenza di centro l'origine.

Curiosità 9.5 (Sistemi di equazioni differenziali lineari) Vogliamo verificare che quanto detto nel Paragrafo 9.4 riguardo le soluzioni del sistema

$$\begin{cases} y_1' = a_{11}y_1 + a_{12}y_2, \\ y_2' = a_{21}y_1 + a_{22}y_2. \end{cases} \quad (9.10C)$$

è corretto. Ricordiamo prima di tutto che l'equazione caratteristica di questo sistema è

$$k^2 - (a_{11} + a_{22})k + (a_{11}a_{22} - a_{12}a_{21}) = 0,$$

il cui discriminante è

$$\Delta = (a_{11} - a_{22})^2 + 4a_{12}a_{21}.$$

Poniamo poi $\lambda = (a_{11} + a_{22})/2$.

Cominciamo supponendo $\Delta > 0$, in modo che l'equazione caratteristica abbia due radici reali distinte

$$k_{1,2} = -\lambda \pm \frac{\sqrt{\Delta}}{2},$$

e consideriamo la coppia di funzioni

$$(y_1(t), y_2(t)) = (\alpha_1 e^{k_1 t}, \alpha_2 e^{k_1 t}).$$

Allora

$$(y_1'(t), y_2'(t)) = (k_1 \alpha_1 e^{k_1 t}, k_1 \alpha_2 e^{k_1 t}),$$

per cui (y_1, y_2) sono una soluzione di (9.10C) se e solo se (α_1, α_2) risolvono il sistema lineare

$$\begin{cases} k_1 \alpha_1 = a_{11} \alpha_1 + a_{12} \alpha_2, \\ k_1 \alpha_2 = a_{21} \alpha_1 + a_{22} \alpha_2, \end{cases}$$

che coincide con (9.19). Nota che (grazie a quanto visto nel Capitolo 4) questo sistema ammette una soluzione non nulla se e solo se la sua matrice dei coefficienti è singolare, cosa che accade se e solo se il suo determinante è nullo; ma il determinante della matrice dei coefficienti di (9.19) è

$$(a_{11} - k_1)(a_{22} - k_1) - a_{12}a_{21} = k_1^2 - (a_{11} + a_{22})k_1 + a_{11}a_{22} - a_{12}a_{21},$$

che si annulla proprio perché k_1 è una delle radici caratteristiche.

In maniera analoga si vede che $(y_1(t), y_2(t)) = (\beta_1 e^{k_2 t}, \beta_2 e^{k_2 t})$ è una soluzione di (9.10C) se e solo se (β_1, β_2) è una soluzione del sistema lineare (9.20), che di nuovo ammette soluzioni non nulle perché k_2 è una radice caratteristica.

Infine, siccome la derivata della somma è uguale alla somma delle derivate, e il secondo membro di (9.10C) è costituito da polinomi lineari omogenei in y_1 e y_2 , la somma di soluzioni di (9.10C) è ancora una soluzione; e quindi otteniamo che tutte le coppie di funzioni della forma

$$(y_1(t), y_2(t)) = (\alpha_1 e^{k_1 t} + \beta_1 e^{k_2 t}, \alpha_2 e^{k_1 t} + \beta_2 e^{k_2 t}).$$

sono soluzioni di (9.10C), come annunciato.

Supponiamo ora $\Delta = 0$, in modo che l'equazione caratteristica abbia come unica radice $k_0 = -\lambda$. Consideriamo la coppia di funzioni

$$(y_1(t), y_2(t)) = ((\alpha_1 t + \beta_1) e^{k_0 t}, (\alpha_2 t + \beta_2) e^{k_0 t}).$$

Derivando troviamo

$$(y_1'(t), y_2'(t)) = ((k_0 \alpha_1 t + \alpha_1 + k_0 \beta_1) e^{k_0 t}, (k_0 \alpha_2 t + \alpha_2 + k_0 \beta_2) e^{k_0 t}).$$

Mettendo questo in (9.10C) e considerando separatamente i coefficienti di $t e^{k_0 t}$ e quelli di $e^{k_0 t}$ vediamo che (α_1, α_2) deve risolvere (9.21) mentre (β_1, β_2) deve risolvere (9.22).

Infine, sempre derivando, uguagliando e separando i coefficienti delle varie funzioni che compaiono, si verifica che quando $\Delta < 0$ la coppia di funzioni

$$(y_1(t), y_2(t)) = ([\alpha_1 \cos(\omega t) + \alpha_2 \sin(\omega t)]e^{\lambda t}, [\beta_1 \cos(\omega t) + \beta_2 \sin(\omega t)]e^{\lambda t}),$$

dove $\omega = \sqrt{-\Delta}/2$, è una soluzione di (9.10C) non appena $(\alpha_1, \alpha_2, \beta_1, \beta_2)$ soddisfano (9.23).

Curiosità 9.6 (Numeri complessi) Nell'Osservazione 9.11 abbiamo suggerito l'esistenza di una relazione inaspettata fra funzioni esponenziali e funzioni trigonometriche; per descriverla, ci servono i numeri complessi.

Molto in breve, un **numero complesso** z è semplicemente una coppia (x, y) di numeri reali. La cosa che distingue l'insieme \mathbb{C} dei numeri complessi dal piano \mathbb{R}^2 delle coppie di numeri reali è la presenza di due operazioni, una somma e un prodotto, definiti dalle formule

$$\begin{aligned}(x_1, y_1) + (x_2, y_2) &= (x_1 + x_2, y_1 + y_2), \\ (x_1, y_1) \cdot (x_2, y_2) &= (x_1 x_2 - y_1 y_2, x_1 y_2 + x_2 y_1).\end{aligned}$$

Con un po' di pazienza puoi verificare che queste operazioni soddisfano tutte le usuali proprietà della somma e del prodotto ordinari che abbiamo elencate nel Paragrafo 1.2.

In particolare, l'elemento neutro per il prodotto è $(1, 0)$ e il reciproco di (x, y) è

$$(x, y)^{-1} = \left(\frac{x}{x^2 + y^2}, \frac{-y}{x^2 + y^2} \right).$$

Il prodotto ha una semplice interpretazione geometrica in termini delle coordinate polari (vedi il Paragrafo 6.4): la distanza dall'origine di $(x_1, y_1) \cdot (x_2, y_2)$ è il prodotto delle distanze dall'origine di (x_1, y_1) e (x_2, y_2) , mentre l'angolo di $(x_1, y_1) \cdot (x_2, y_2)$ rispetto all'asse delle ascisse è la somma degli angoli di (x_1, y_1) e (x_2, y_2) .

Se identifichiamo un numero reale x con la coppia $(x, 0)$, ritroviamo dentro \mathbb{C} una copia dei numeri reali, operazioni comprese: $(x, 0) + (y, 0) = (x + y, 0)$ e $(x, 0) \cdot (y, 0) = (xy, 0)$, per cui eseguendo le operazioni in \mathbb{R} oppure in \mathbb{C} otteniamo lo stesso risultato. Per questo motivo spesso scriveremo x per indicare il numero complesso $(x, 0)$.

L'insieme \mathbb{C} contiene un elemento molto importante: la coppia $(0, 1)$, usualmente indicata con i e detta **unità immaginaria**. La principale proprietà di i è che è una *radice quadrata di -1* : infatti,

$$i \cdot i = (0, 1) \cdot (0, 1) = (-1, 0) = -1,$$

dove nell'ultimo passaggio abbiamo identificato la coppia $(-1, 0)$ con il numero reale -1 . In particolare otteniamo anche che

$$i^{2n} = (-1)^n \quad \text{e} \quad i^{2n+1} = (-1)^n i \quad (9.11C)$$

per ogni $n \in \mathbb{N}$.

Usando l'unità immaginaria i possiamo scrivere i numeri complessi in una forma molto comoda per le operazioni: infatti si vede subito che

$$(x, y) = (x, 0) + (0, 1) \cdot (y, 0),$$

per cui viene naturale scrivere $z = x + iy$ per rappresentare il numero complesso $z = (x, y)$. Questa scrittura è molto comoda perché permette di effettuare le operazioni di somma e moltiplicazione fra numeri complessi ricorrendo semplicemente alle usuali proprietà algebriche delle operazioni e ricordando che $i^2 = -1$. Per esempio,

$$(3 + 2i)(2 - i) = 3 \cdot 2 - 3i + 2i \cdot 2 - 2i^2 = 6 + 2 + (2 - 3)i = 8 - i.$$

Se $z = x + iy$, diremo che x è la **parte reale** del numero complesso z , mentre y è la **parte immaginaria** di z .

Ora, se $z = x + iy$ è un numero complesso, si definisce l'**esponenziale complesso** e^z ponendo

$$e^z = e^{x+iy} = e^x (\cos y + i \sin y). \quad (9.12C)$$

Se ci pensi un attimo, vedrai che, grazie a (9.11C), questa definizione è perfettamente coerente con gli sviluppi in serie di Taylor delle funzioni esponenziale, seno e coseno che abbiamo visto nella Curiosità 7.10; quindi l'esponenziale reale, il seno e il coseno possono tutti essere recuperati da un'unica funzione, l'esponenziale complesso.

Derivando separatamente la parte reale e la parte immaginaria dell'esponenziale complesso, trattando i come una costante (cosa che effettivamente è), non è difficile verificare che la derivata della funzione

$$f(t) = e^{zt}$$

con t reale è data da

$$f'(t) = ze^{zt}$$

per *ogni* numero complesso z . In particolare, se $z = i\omega$ troviamo che

$$\frac{d}{dt}e^{i\omega t} = i\omega e^{i\omega t}$$

e

$$\frac{d^2}{dt^2}e^{i\omega t} = -\omega^2 e^{i\omega t} ;$$

essendo

$$e^{i\omega t} = \cos(\omega t) + i \sin(\omega t) ,$$

abbiamo recuperato le soluzioni (9.32) di (9.31) sotto forma di esponenziali.

Concludiamo notando che ponendo $z = 0 + i\pi$ in (9.12C) otteniamo la fondamentale **relazione di Eulero**

$$e^{i\pi} + 1 = 0 ,$$

che collega le cinque costanti fondamentali e , π , i , 1 e 0 (ed è una delle più belle formule di tutta la Matematica).

Curiosità 10.1 (La probabilità in insiemi infiniti) Una scoperta decisamente inaspettata della matematica contemporanea è che esistono dei sottoinsiemi del piano di cui non è possibile misurare in alcun modo l'area (ma non ti preoccupare; sono talmente orrendi che se non sei un matematico non ti capiterà mai di incontrarli). Questo vuol dire che la probabilità come rapporto di aree non può essere definita su tutti i possibili sottoinsiemi di Ω , ma solo su alcuni, più decenti, di cui si può calcolare l'area. Per superare questo problema, si è deciso che per parlare di distribuzione di probabilità basta poter associare una probabilità a una famiglia \mathcal{A} di sottoinsiemi di Ω , e non necessariamente a tutti. Questa famiglia non può essere qualsiasi, però. Deve contenere sia Ω sia l'insieme vuoto \emptyset ; se contiene un insieme E deve contenere anche il suo complementare E^c ; e se contiene una successione finita o infinita di insiemi E_1, E_2, E_3, \dots allora deve contenere anche l'unione $E_1 \cup E_2 \cup E_3 \cup \dots$. Data una famiglia \mathcal{A} di sottoinsiemi di Ω con queste proprietà (in termini tecnici \mathcal{A} si chiama **-algebra**), una **misura di probabilità** p su \mathcal{A} è un modo di associare a ciascun elemento $E \in \mathcal{A}$ un numero reale $p(E)$ in modo che siano verificate (P1)-(P3) e la seguente proprietà (P4): *data una successione infinita E_1, E_2, E_3, \dots di elementi di \mathcal{A} a due a due disgiunti, la probabilità $p(E_1 \cup E_2 \cup \dots)$ dell'unione è uguale alla somma infinita $p(E_1) + p(E_2) + \dots$* (puoi trovare informazioni sulle somme infinite nel Paragrafo 6.6). Con questo trucco si riesce a parlare in modo efficace di probabilità anche nel caso di spazi degli eventi infiniti particolarmente complicati.

Curiosità 10.2 (Indipendenza di n eventi) La definizione precisa di indipendenza per più di 2 eventi è un attimo complicata. Si dice che n eventi E_1, \dots, E_n sono **indipendenti** fra loro se e solo se per ogni $2 \leq k \leq n$ e ogni scelta di k eventi E_{i_1}, \dots, E_{i_k} fra E_1, \dots, E_n si ha

$$p(E_{i_1} \cap \dots \cap E_{i_k}) = p(E_{i_1}) \dots p(E_{i_k}) .$$

Attenzione: il fatto che tre eventi E_1, E_2, E_3 siano a due a due indipendenti non implica che i tre eventi assieme siano indipendenti. Per esempio, nel caso del lancio di due dadi a 6 facce distinti, consideriamo gli eventi "primo dado 1" $E_1 = \{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6)\}$, "secondo dado 2" $E_2 = \{(1, 2), (2, 2), (3, 2), (4, 2), (5, 2), (6, 2)\}$, e "somma dei dadi uguale a 7" $E_3 = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$. Allora si verifica subito che i tre eventi sono a due a due indipendenti (cioè $p(E_1 \cap E_2) = p(E_1)p(E_2)$ e così via), ma i tre eventi assieme non sono indipendenti:

$$p(E_1 \cap E_2 \cap E_3) = 0 \neq (1/6)^3 = p(E_1)p(E_2)p(E_3) .$$

Analogamente, sapere che $p(F_1 \cap F_2 \cap F_3) = p(F_1)p(F_2)p(F_3)$ non implica che i tre eventi siano indipendenti a due a due (e quindi non implica che siano indipendenti tutti assieme). Per esempio, prendiamo $F_1 = \Omega \setminus \{(1, 1), (2, 1), (3, 1), (4, 1), (5, 1), (1, 2), (2, 2), (3, 2), (4, 3)\}$, $F_2 = \{(1, 3), (2, 3), (3, 3), (4, 3), (5, 3), (6, 3), (6, 2), (6, 1)\}$ ed $F_3 = E_3$. Allora si vede subito che $F_1 \cap F_2 \cap F_3 = \{(6, 1)\}$ per cui

$$p(F_1 \cap F_2 \cap F_3) = 1/36 = (27/36)(8/36)(6/36) = p(F_1)p(F_2)p(F_3)$$

ma $p(F_2 \cap F_3) = 1/18 \neq (8/36)(6/36) = p(F_2)p(F_3)$.

Il succo di questi esempi è che per vedere se n eventi E_1, \dots, E_n sono indipendenti bisogna controllare che la probabilità dell'intersezione di *qualsiasi* sottoinsieme di essi è uguale al prodotto delle probabilità; considerare solo qualcuno dei sottoinsiemi non basta.

Curiosità 10.3 (Coefficienti binomiali) La formula per lo sviluppo del binomio fornisce altre due identità che a volte sono utili (e che sono un buon esercizio per imparare a usare il simbolo di sommatoria):

$$\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n} = \sum_{k=0}^n \binom{n}{k} = 2^n,$$

e

$$\binom{n}{0} - \binom{n}{1} + \cdots + (-1)^n \binom{n}{n} = \sum_{k=0}^n (-1)^k \binom{n}{k} = 0.$$

Per vedere che sono vere basta notare che (10.20) ci dà

$$\sum_{k=0}^n \binom{n}{k} = \sum_{k=0}^n \binom{n}{k} 1^{n-k} 1^k = (1+1)^n = 2^n,$$

e

$$\sum_{k=0}^n (-1)^k \binom{n}{k} = \sum_{k=0}^n \binom{n}{k} 1^{n-k} (-1)^k = (1-1)^n = 0.$$

Verificate sulle righe del triangolo di Tartaglia.

Curiosità 10.4 (Combinazioni con ripetizioni) Vediamo come si può ottenere la formula (10.21). Il numero $\tilde{C}_{n,k}$ conta in quanti modi possiamo scegliere k numeri fra 1 e n , con ripetizioni e a meno dell'ordine. Siccome l'ordine non è rilevante, possiamo metterli in ordine crescente e chiederci in quanti modi possiamo scegliere i numeri a_1, \dots, a_k in modo che si abbia

$$1 \leq a_1 \leq a_2 \leq \dots \leq a_k \leq n .$$

Ora, sommiamo 1 ad a_2 , 2 ad a_3 e così via; otteniamo

$$1 \leq a_1 < a_2 + 1 < a_3 + 2 < \dots < a_k + (k-1) \leq n + k - 1 ,$$

cioè una lista ordinata di k numeri *distinti* scelti fra 1 e $n + k - 1$. In altre parole, una combinazione con ripetizioni di n oggetti k alla volta ci ha fornito una combinazione senza ripetizioni di $n + k - 1$ oggetti k alla volta. Viceversa, supponiamo di avere una combinazione senza ripetizioni di $n + k - 1$ oggetti k alla volta, cioè di aver scelto k numeri distinti b_1, \dots, b_k fra 1 e $n + k - 1$. Siccome l'ordine non conta, possiamo scrivere

$$1 \leq b_1 < b_2 < \dots < b_k \leq n + k - 1 .$$

Sottraiamo ora 1 a b_2 , 2 a b_3 e così via; otteniamo

$$1 \leq b_1 \leq b_2 - 1 \leq b_3 - 2 \leq \dots \leq b_k - (k-1) \leq n ,$$

cioè una lista ordinata di k numeri, eventualmente ripetuti, scelti fra 1 e n . In altre parole, invertendo il processo precedente abbiamo trovato che una combinazione senza ripetizioni di $n + k - 1$ oggetti k alla volta fornisce una combinazione con ripetizioni di n oggetti k alla volta. Di conseguenza possiamo concludere che il numero delle combinazioni con ripetizioni di n oggetti k alla volta è uguale al numero delle combinazioni senza ripetizioni di $n + k - 1$ oggetti k alla volta, cioè che

$$\tilde{C}_{n,k} = C_{n+k-1,k} = \binom{n+k-1}{k} ,$$

come affermato.

Curiosità 11.1 (Media geometrica) Nel caso di dati *positivi*, invece degli errori assoluti $x - x_i$ possiamo considerare gli errori relativi x/x_i . In questo caso, l'assenza di errore corrisponde ad avere errore relativo uguale a 1; quindi possiamo chiederci se esiste un $\hat{x} \in \mathbb{R}$ per cui il *prodotto* $\prod_{i=1}^n (\hat{x}/x_i)$ sia esattamente uguale a 1. È facile rispondere: siccome

$$\prod_{i=1}^n \frac{\hat{x}}{x_i} = \frac{\hat{x}^n}{\prod_{i=1}^n x_i},$$

il prodotto degli errori relativi è uguale a 1 se e solo se

$$\hat{x} = \left(\prod_{i=1}^n x_i \right)^{1/n}.$$

Questo \hat{x} è detto *media geometrica* di x_1, \dots, x_n . Si può dimostrare che la media geometrica è sempre minore o uguale alla media aritmetica, cioè che

$$(x_1 \cdots x_n)^{1/n} \leq \frac{x_1 + \cdots + x_n}{n},$$

con uguaglianza se e solo se $x_1 = \cdots = x_n$.

Curiosità 11.2 (Interpretazione probabilistica della media ponderata) I pesi hanno una naturale interpretazione probabilistica. Se indichiamo con

$$F = f_1 + \dots + f_n = \sum_{i=1}^n f_i$$

la somma dei pesi e poniamo $p_i = f_i/F$, allora $p_1 + \dots + p_n = 1$, e possiamo interpretare p_i come la probabilità di ottenere il valore x_i facendo la nostra misura.

Usando i p_i possiamo calcolare la media ponderata con la formula

$$\bar{x} = \sum_{i=1}^n p_i x_i. \quad (11.1C)$$

Più in generale, in un esperimento probabilistico di spazio degli eventi $\Omega = \{x_1, \dots, x_n\}$ composto da un numero finito di numeri reali la formula (11.1C) (dove p_i ora è la probabilità dell'evento semplice x_i) fornisce il *valore atteso* (o *media*) dell'esperimento. Per esempio, il valore atteso della somma del lancio di due dadi a sei facce non truccati è

$$\frac{1}{36} \cdot 2 + \frac{2}{36} \cdot 3 + \frac{3}{36} \cdot 4 + \frac{4}{36} \cdot 5 + \frac{5}{36} \cdot 6 + \frac{6}{36} \cdot 7 + \frac{5}{36} \cdot 8 + \frac{4}{36} \cdot 9 + \frac{3}{36} \cdot 10 + \frac{2}{36} \cdot 11 + \frac{1}{36} \cdot 12 = 7.$$

Ne riparleremo nel Paragrafo 12.2 (vedi l'Esempio 12.10).

Curiosità 11.3 (Un'altra interpretazione del coefficiente di Pearson) Usando un po' di Algebra Lineare possiamo capire il significato geometrico del coefficiente di Pearson. Abbiamo visto che la retta di regressione dei dati $(x_1, y_1), \dots, (x_n, y_n)$ ha equazione

$$y = \bar{m}x + \bar{q},$$

dove \bar{x} (rispettivamente, \bar{y}) è la media di x_1, \dots, x_n (rispettivamente, di y_1, \dots, y_n). In particolare, siccome $\bar{q} = \bar{y} - \bar{m}\bar{x}$, un punto (x_i, y_i) appartiene alla retta di regressione se e solo se

$$y_i - \bar{y} = \bar{m}(x_i - \bar{x}).$$

Questo suggerisce di introdurre i vettori

$$v_x = \begin{bmatrix} x_1 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{bmatrix} \quad \text{e} \quad v_y = \begin{bmatrix} y_1 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{bmatrix}.$$

Il discorso appena fatto ci dice che le coppie di dati appartengono *tutte* alla retta di regressione (cioè l'interpolazione è perfetta) se e solo se $v_y = \bar{m}v_x$, cioè se e solo se v_y è proporzionale a v_x . Ma, per quanto detto nella Curiosità 3.3 riguardo al prodotto scalare e alla lunghezza di vettori di \mathbb{R}^n , questo accade se e solo se

$$\frac{|\langle v_x, v_y \rangle|}{\|v_x\| \|v_y\|} = 1.$$

Quindi il rapporto $\langle v_x, v_y \rangle / \|v_x\| \|v_y\|$ misura la bontà dell'interpolazione data dalla retta di regressione: più questo rapporto è vicino a ± 1 migliore è l'interpolazione. Ma

$$\|v_x\| = \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad \|v_y\| = \sqrt{\sum_{j=1}^n (y_j - \bar{y})^2}$$

e

$$\langle v_x, v_y \rangle = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}),$$

per cui (11.10) ci dice che $\langle v_x, v_y \rangle / \|v_x\| \|v_y\|$ è proprio uguale al coefficiente di Pearson.

Curiosità 12.1 (Variabili aleatorie indipendenti) Diremo che n variabili aleatorie X_1, \dots, X_n sono **indipendenti** se per ogni scelta di k variabili diverse X_{i_1}, \dots, X_{i_k} fra le X_1, \dots, X_n (con $1 \leq k \leq n$) abbiamo

$$p(\{X_{i_1} \in I_1\} \cap \dots \cap \{X_{i_k} \in I_k\}) = p(X_{i_1} \in I_1) \dots p(X_{i_k} \in I_k)$$

per ogni k -upla di intervalli o semirette $I_1, \dots, I_k \subset \mathbb{R}$; confronta con la Curiosità 10.2.

Curiosità 12.2 (Valore atteso e varianza di una variabile aleatoria bernoulliana) Sia X una variabile aleatoria bernoulliana di tipo (n, p) . Siccome i possibili valori di X sono $0, \dots, n$, otteniamo

$$\begin{aligned} E(X) &= \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} k = \sum_{k=1}^n \frac{n!}{(n-k)!k!} k p^k q^{n-k} \\ &= \sum_{k=1}^n \frac{n!}{(n-k)!(k-1)!} p^k q^{n-k} = np \sum_{k=1}^n \frac{(n-1)!}{(n-k)!(k-1)!} p^{k-1} q^{n-k} \\ &= np \sum_{h=0}^{n-1} \frac{(n-1)!}{(n-h-1)!h!} p^h q^{n-1-h} \\ &= np(p+q)^{n-1} = np \end{aligned}$$

(controlla di aver capito bene tutti i passaggi di questo conto; a un certo punto abbiamo posto $h = k - 1$ e poi usato la formula per lo sviluppo del binomio).

Il calcolo della varianza è simile ma un poco più complicato. L'idea è cercare di ricondursi a somme del tipo

$$\sum_{k=0}^{\ell} \binom{\ell}{k} p^k q^{\ell-k},$$

che valgono $(p+q)^\ell = 1$. Scrivendo

$$k - np = k - n(1-q) = nq - (n-k)$$

otteniamo

$$\begin{aligned} \text{Var}(X) &= \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} (k - np)^2 = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} (k - np)(nq - (n-k)) \\ &= n \sum_{k=0}^n \binom{n}{k} k p^k q^{n-k+1} - \sum_{k=0}^n \binom{n}{k} k(n-k) p^k q^{n-k} \\ &\quad - n^2 \sum_{k=0}^n \binom{n}{k} p^{k+1} q^{n-k+1} + n \sum_{k=0}^n \binom{n}{k} (n-k) p^{k+1} q^{n-k} \\ &= npq \left[n \sum_{k=1}^n \frac{(n-1)!}{(n-k)!(k-1)!} p^{k-1} q^{n-k} \right. \\ &\quad \left. - \sum_{k=1}^{n-1} \frac{(n-1)!}{(n-k-1)!(k-1)!} p^{k-1} q^{n-k-1} - n \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} \right. \\ &\quad \left. + n \sum_{k=0}^{n-1} \frac{(n-1)!}{(n-k-1)!k!} p^k q^{n-k-1} \right] \\ &= npq \left[n \sum_{h=0}^{n-1} \binom{n-1}{h} p^h q^{n-1-h} - (n-1) \sum_{h=0}^{n-2} \binom{n-2}{h} p^h q^{n-2-h} \right. \\ &\quad \left. - n \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} + n \sum_{k=0}^{n-1} \binom{n-1}{k} p^k q^{n-1-k} \right] \\ &= npq [n(p+q)^{n-1} - (n-1)(p+q)^{n-2} - n(p+q)^n + n(p+q)^{n-1}] \\ &= npq. \end{aligned}$$

Quindi $\sigma_X^2 = npq$, e la deviazione standard è $\sigma_X = \sqrt{npq}$.

Curiosità 12.3 (Dalla densità di probabilità alla distribuzione di probabilità)

Supponiamo di avere una funzione $f: \mathbb{R} \rightarrow \mathbb{R}^+$ sempre maggiore o uguale a 0, integrabile su ogni intervallo (o semiretta), e tale che $\int_{-\infty}^{+\infty} f(x) dx = 1$, che sono le proprietà di una densità di probabilità viste nel Paragrafo 12.4. Allora possiamo definire una distribuzione di probabilità $p: \mathcal{A} \rightarrow [0, 1]$ su \mathbb{R} ponendo

$$p(A) = \int_A f(x) dx ,$$

dove \mathcal{A} è la famiglia dei sottoinsiemi di \mathbb{R} su cui f è integrabile. Le proprietà dell'integrale permettono di verificare facilmente che p soddisfa gli assiomi (P1)-(P3) che abbiamo introdotto nel Paragrafo 10.5, per cui p è effettivamente una distribuzione di probabilità su \mathbb{R} . Inoltre, p soddisfa anche l'assioma (P4) citato nella Curiosità 10.1; ma per vederlo occorre definire l'integrale in modo più generale di quanto abbiamo fatto noi, usando la *teoria dell'integrazione di Lebesgue*.

Curiosità 12.4 (Proprietà di minimizzazione del valore atteso di variabili aleatorie continue) Il valore atteso di una variabile aleatoria continua soddisfa proprietà di minimizzazione analoghe a quelle della media di numeri. Per esempio, è l'unico $\mu \in \mathbb{R}$ per cui l'integrale degli "errori" $x - \mu$, pesati in base alla densità di probabilità f_X , è nullo. Infatti,

$$\int_{-\infty}^{+\infty} (x - \mu)f_X(x) dx = 0$$

se e solo se

$$\mu \int_{-\infty}^{+\infty} f_X(x) dx = \int_{-\infty}^{+\infty} xf_X(x) dx ,$$

per cui

$$\mu = \mu \int_{-\infty}^{+\infty} f_X(x) dx = \int_{-\infty}^{+\infty} xf_X(x) dx = \mu_X .$$

Possiamo anche far vedere che μ_X minimizza l'integrale degli scarti quadratici $(x - \mu)^2$, pesati in base alla densità di probabilità f_X . Consideriamo la funzione $s: \mathbb{R} \rightarrow \mathbb{R}$ data da

$$s(y) = \int_{-\infty}^{+\infty} (x - y)^2 f_X(x) dx .$$

Sviluppando il quadrato otteniamo

$$\begin{aligned} s(y) &= y^2 \int_{-\infty}^{+\infty} f_X(x) dx - 2y \int_{-\infty}^{+\infty} xf_X(x) dx + \int_{-\infty}^{+\infty} x^2 f_X(x) dx \\ &= y^2 - 2\mu_X y + \int_{-\infty}^{+\infty} x^2 f_X(x) dx . \end{aligned}$$

Quindi s è una funzione quadratica con coefficiente direttivo positivo, per cui ammette un unico minimo in $y_0 = -b/2a = \mu_X$, come voluto.

Curiosità 12.5 (Come calcolare $p(X = x)$ usando la funzione di distribuzione) Sia X una variabile aleatoria qualsiasi, e $x \in \mathbb{R}$. Allora da

$$\{X < x\} = \bigcup_{t < x} \{X \leq t\} \quad \text{e} \quad \{X = x\} = \{X \leq x\} \setminus \{X < x\},$$

otteniamo

$$p(X < x) = \lim_{t \rightarrow x^-} F_X(t) \quad \text{e} \quad p(X = x) = F_X(x) - \lim_{t \rightarrow x^-} F_X(t);$$

nota che il limite esiste in quanto F è crescente e limitata. In particolare, se F_X è una funzione continua a sinistra, allora $p(X = x) = 0$ per ogni $x \in \mathbb{R}$ – che equivale a dire che è impossibile effettuare la misura X con precisione infinita. L'Osservazione 12.12 dice in particolare che questo accade per le variabili aleatorie continue.

Curiosità 12.6 (Teorema del limite centrale) Siano X_1, \dots, X_n variabili aleatorie indipendenti, e poniamo

$$S_n = X_1 + \dots + X_n \quad \text{e} \quad M_n = \frac{1}{n} S_n.$$

Il Teorema del limite centrale ci dice che, sotto opportune ipotesi, le variabili aleatorie S_n e M_n tendono per $n \rightarrow +\infty$ a una variabile aleatoria normale. Siccome si può dimostrare (vedi l'Osservazione 12.9 e l'Esercizio 12.52) che

$$E(S_n) = \mu_1 + \dots + \mu_n, \quad \text{Var}(S_n) = \sigma_1^2 + \dots + \sigma_n^2,$$

e

$$E(M_n) = \frac{\mu_1 + \dots + \mu_n}{n}, \quad \text{Var}(M_n) = \frac{\sigma_1^2 + \dots + \sigma_n^2}{n^2},$$

il valore atteso e la varianza di S_n (e forse anche quelle di M_n) potrebbero divergere al tendere di n all'infinito; cosa vuol dire, allora, che S_n può tendere a una variabile aleatoria normale, la quale ha valore atteso e varianza finite?

L'idea è di normalizzare S_n e M_n portando la media a 0 e la varianza a 1, in modo da poter confrontare meglio le variabili aleatorie S_n e M_n con una variabile aleatoria normale. Vediamo come si normalizza una variabile aleatoria qualsiasi.

Sia X una variabile aleatoria qualsiasi, di valore atteso μ_X e varianza σ_X^2 . Allora la **centrata** di X è la variabile aleatoria $X^{(c)}$ data da

$$X^{(c)} = X - \mu_X,$$

e la **standardizzata** di X è la variabile aleatoria $X^{(s)}$ data da

$$X^{(s)} = \frac{1}{\sigma_X} X^{(c)} = \frac{X - \mu_X}{\sigma_X}.$$

Vediamo di calcolare valore atteso e varianza delle variabili aleatorie centrate e standardizzate. Prima di tutto abbiamo

$$F_{X^{(c)}}(t) = p(X^{(c)} \leq t) = p(X - \mu_X \leq t) = p(X \leq t + \mu_X) = F_X(t + \mu_X);$$

quindi derivando troviamo

$$f_{X^{(c)}}(x) = f_X(x + \mu_X).$$

Dunque con la sostituzione $y = x + \mu_X$ otteniamo

$$\begin{aligned} E(X^{(c)}) &= \int_{-\infty}^{+\infty} x f_{X^{(c)}}(x) dx = \int_{-\infty}^{+\infty} x f_X(x + \mu_X) dx \\ &= \int_{-\infty}^{+\infty} (y - \mu_X) f_X(y) dy \\ &= \int_{-\infty}^{+\infty} y f_X(y) dy - \mu_X \int_{-\infty}^{+\infty} f_X(y) dy = \mu_X - \mu_X = 0, \end{aligned}$$

e

$$\begin{aligned} \text{Var}(X^{(c)}) &= \int_{-\infty}^{+\infty} x^2 f_{X^{(c)}}(x) dx = \int_{-\infty}^{+\infty} x^2 f_X(x + \mu_X) dx \\ &= \int_{-\infty}^{+\infty} (y - \mu_X)^2 f_X(y) dy = \sigma_X^2. \end{aligned}$$

In altre parole, la variabile aleatoria centrata ha valore atteso 0 e varianza uguale a quella di X . Poi,

$$F_{X^{(s)}}(t) = p(X^{(s)} \leq t) = p(\sigma_X^{-1} X^{(c)} \leq t) = p(X^{(c)} \leq \sigma_X t) = F_{X^{(c)}}(\sigma_X t);$$

quindi derivando troviamo

$$f_{X^{(s)}}(x) = \sigma_X f_{X^{(c)}}(\sigma_X x) .$$

Dunque con la sostituzione $y = \sigma_X x$ otteniamo

$$\begin{aligned} E(X^{(s)}) &= \int_{-\infty}^{+\infty} x f_{X^{(s)}}(x) dx = \int_{-\infty}^{+\infty} \sigma_X x f_{X^{(c)}}(\sigma_X x) dx \\ &= \frac{1}{\sigma_X} \int_{-\infty}^{+\infty} y f_{X^{(c)}}(y) dy = 0 , \end{aligned}$$

e

$$\begin{aligned} \text{Var}(X^{(s)}) &= \int_{-\infty}^{+\infty} x^2 f_{X^{(s)}}(x) dx = \int_{-\infty}^{+\infty} \sigma_X x^2 f_{X^{(c)}}(\sigma_X x) dx \\ &= \frac{1}{\sigma_X^2} \int_{-\infty}^{+\infty} y^2 f_{X^{(c)}}(y) dy = 1 . \end{aligned}$$

In altre parole, *una variabile aleatoria standardizzata ha sempre valore atteso 0 e varianza 1.*

Allora l'enunciato esatto del Teorema del limite centrale è: *se le ipotesi (C1)-(C3) sono soddisfatte, allora le standardizzate $S_n^{(s)}$ e $M_n^{(s)}$ tendono per $n \rightarrow +\infty$ a una variabile aleatoria normale di media 0 e varianza 1 nel senso che*

$$\forall t \in \mathbb{R} \quad \lim_{n \rightarrow +\infty} F_{S_n^{(s)}}(t) = \lim_{n \rightarrow +\infty} F_{M_n^{(s)}}(t) = \Phi(t) .$$

Curiosità 12.7 (Distribuzione T di Student) La densità di probabilità della distribuzione di Student T_ν con ν gradi di libertà è data da

$$f_{T_\nu}(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{\nu}{t^2 + \nu}\right)^{(\nu+1)/2},$$

dove Γ è la **funzione Gamma di Eulero** definita da

$$\Gamma(t) = \int_0^{+\infty} x^{t-1} e^{-x} dx;$$

si può dimostrare che $\Gamma(n) = (n-1)!$ per ogni numero naturale $n \geq 1$.

Curiosità 12.8 (Un altro uso del test T) È possibile usare il test T anche per controllare l'ipotesi nulla che la media nella popolazione sia minore o uguale di un certo valore μ_0 . Si usa la quantità pivotale T senza il modulo (per cui si tratta di un test T a una coda), con nuovi valori di soglia. Per l'esattezza, si può dimostrare che il valore di soglia corrispondente al livello α per questo tipo di test è uguale al valore di soglia per il test T a due code ma con livello di affidabilità $\alpha/2$.

È possibile controllare anche l'ipotesi nulla che la media nella popolazione sia maggiore o uguale di un certo valore μ_0 . In tal caso, il test ha esito negativo con livello di affidabilità α se il valore di T (senza modulo) è *minore* del valore di soglia con livello di affidabilità $1 - (\alpha/2)$.

Curiosità 12.9 (La distribuzione F) La densità di probabilità della distribuzione F_{ν_n, ν_d} con (ν_n, ν_d) gradi di libertà è identicamente nulla per $t < 0$, mentre è data per $t \geq 0$ da

$$f_{F_{\nu_n, \nu_d}}(t) = \frac{\Gamma\left(\frac{\nu_n + \nu_d}{2}\right)}{t \Gamma\left(\frac{\nu_n}{2}\right) \Gamma\left(\frac{\nu_d}{2}\right)} \left(\frac{\nu_n t}{\nu_n t + \nu_d}\right)^{\nu_n/2} \left(1 - \frac{\nu_n t}{\nu_n t + \nu_d}\right)^{\nu_d/2},$$

dove Γ è ancora la funzione Gamma di Eulero (vedi la Curiosità 12.7).

Curiosità 12.10 (Relazione fra il test T e il test F) Nonostante l'apparenza diversa, nel caso di due campioni di uguale numerosità il test T e il test F sono del tutto equivalenti. Infatti, supponiamo di avere due campioni X e Y , entrambi di numerosità n , con media M_X e M_Y e deviazione standard campionaria s_X e s_Y rispettivamente. La varianza entro i gruppi è

$$s_{\text{entro}}^2 = \frac{s_X^2 + s_Y^2}{2},$$

mentre la varianza tra i gruppi è

$$\begin{aligned} s_{\text{tra}}^2 &= n \frac{1}{2-1} \left[\left(M_X - \frac{1}{2}(M_X + M_Y) \right)^2 + \left(M_Y - \frac{1}{2}(M_X + M_Y) \right)^2 \right] \\ &= n \left[\left(\frac{1}{2}M_X - \frac{1}{2}M_Y \right)^2 + \left(\frac{1}{2}M_Y - \frac{1}{2}M_X \right)^2 \right] = n \left[2 \frac{1}{4} (M_X - M_Y)^2 \right] \\ &= \frac{n}{2} (M_X - M_Y)^2. \end{aligned}$$

Ma allora

$$F_{1,2(n-1)} = \frac{(M_X - M_Y)^2}{s_X^2 + s_Y^2} n = (T_{2(n-1)})^2,$$

grazie alla formula (12.26). Quindi verificare che $F_{1,2(n-1)}$ sia maggiore di un valore di soglia k^* è equivalente a verificare che $|T_{2(n-1)}|$ sia maggiore del valore di soglia $\sqrt{k^*}$.

Curiosità 12.11 (Test per il confronto di più campioni) Il test F permette di decidere se almeno uno dei campioni esaminati proviene da una popolazione diversa, ma non determina quali sono i campioni spuri. Sono stati elaborati diversi test per risolvere questo problema, chiamati **procedure per confronti multipli**. In buona parte sono basati su applicazioni ripetute del test T , modificato opportunamente in modo da tenere conto del fatto che si stanno confrontando molte medie diverse. Una delle procedure più semplici è nota come **test T di Bonferroni**. L'idea è di applicare il test T a tutte le coppie di campioni ma con le seguenti modifiche:

- sostituiamo la deviazione standard campionaria combinata con la deviazione standard entro i gruppi;
- il numero di gradi di libertà per ciascun test T è dato dalla somma delle numerosità dei campioni meno il numero dei campioni;
- se vogliamo un livello di affidabilità α sul totale dei k confronti dobbiamo applicare ciascun test T con un livello di affidabilità α/k .

Altri test più sofisticati per confronti multipli sono il **test di Holm**, il **test di Student-Newman-Keuls** e il **test di Tukey**.

Curiosità 12.12 (La distribuzione χ^2) La densità di probabilità della distribuzione χ^2_ν con ν gradi di libertà è identicamente nulla per $t < 0$, ed è data per $t \geq 0$ da

$$f_{\chi^2_\nu}(t) = \frac{1}{2^{\nu/2} \Gamma(\frac{\nu}{2})} t^{(\nu/2)-1} e^{-t/2} ,$$

dove Γ è sempre la funzione Gamma di Eulero (vedi la Curiosità 12.7).

Curiosità 12.13 (Intervalli di confidenza) Le distribuzioni T di Student e χ^2 possono essere usate per determinare intervalli di confidenza in maniera non dissimile da quanto visto nel Paragrafo 12.9. Per riassumere i risultati, introduciamo la seguente notazione: indicheremo con z_α (rispettivamente, t'_α, q'_α) il valore di soglia per il test Z (rispettivamente, il test T con ν gradi di libertà, e il test χ^2 con ν gradi di libertà) con livello di affidabilità α . Allora se abbiamo un campione X di $n > 30$ individui, con media M e varianza campionaria s^2 , possiamo dire che:

- (a) se conosciamo la varianza σ^2 della popolazione, la media μ della popolazione appartiene all'intervallo

$$\left[M - \frac{\sigma}{\sqrt{n}} z_{1-\beta}, M + \frac{\sigma}{\sqrt{n}} z_{1-\beta} \right]$$

con probabilità β ;

- (b) se non conosciamo la varianza σ^2 della popolazione, la media μ della popolazione appartiene all'intervallo

$$\left[M - \frac{s}{\sqrt{n}} t'_{1-\beta}, M + \frac{s}{\sqrt{n}} t'_{1-\beta} \right]$$

con probabilità β ;

- (c) la varianza σ^2 della popolazione appartiene all'intervallo

$$\left[(n-1) \frac{s^2}{q'_{(1-\beta)/2}}, (n-1) \frac{s^2}{q'_{(1+\beta)/2}} \right]$$

con probabilità β .