Appunti per il corso di Analisi SSIS 06-07

Riccardo Ricci

Università di Firenze, Facoltà di S.M.F.N.

28 marzo 2007

Capitolo 1

Successioni

1.1 Il paradosso di Zenone

> *In this capricious world nothing is more capricious than posthumous fame. One of the most notable victims of posterity's lack of judgement is the Eleatic Zeno. Having invented four arguments all immeasurably subtle and profound, the grossness of subsequent philosophers pronounced him to be a mere ingenious juggler, and his arguments to be one and all sophisms. After two thousand years of continual refutation, these sophisms were reinstated, and made the foundation of a mathematical renaissance...*
>
> Bertrand Russell, *The Principles of Mathematics* (1903)

I paradossi di Zenone ci sono stati tramandati da Aristotele nel suo libro sulla Fisica.

Il più famoso è quello noto sotto il nome di *Achille e la Tartaruga* che dichiara l'impossibilità per il velocissimo Achille di raggiungere una lentissima tartaguga partita prima di lui nella corsa.

L'argomento di Zenone è, a grosse linee, il seguente: inizialmente la tartaruga ha un vantaggio su Achille; prima di raggiungere la tartaruga Achille dovrà quindi percorrere lo spazio che inizialmente lo separava dalla tartaruga. Ma questa nel frattempo avrà percorso un nuovo tratto di strada, che Achille dovrà sua volta percorrere, mentre la tartaruga sarà ancora in grado di spostarsi di un nuovo tratto, e così via *all'infinito*. Quindi Achille *non potrà mai raggiungere la tartaruga*.

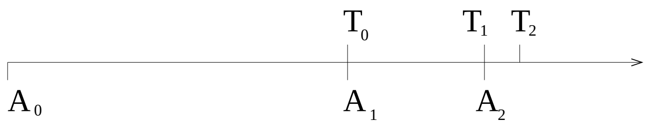

Figura 1.1: Achille e la Tartaruga

Ancora più radicale è il cosiddetto paradosso della *dicotomia* che *nega la possibilità stessa di partire*: prima di percorrere un certo tratto di strada se

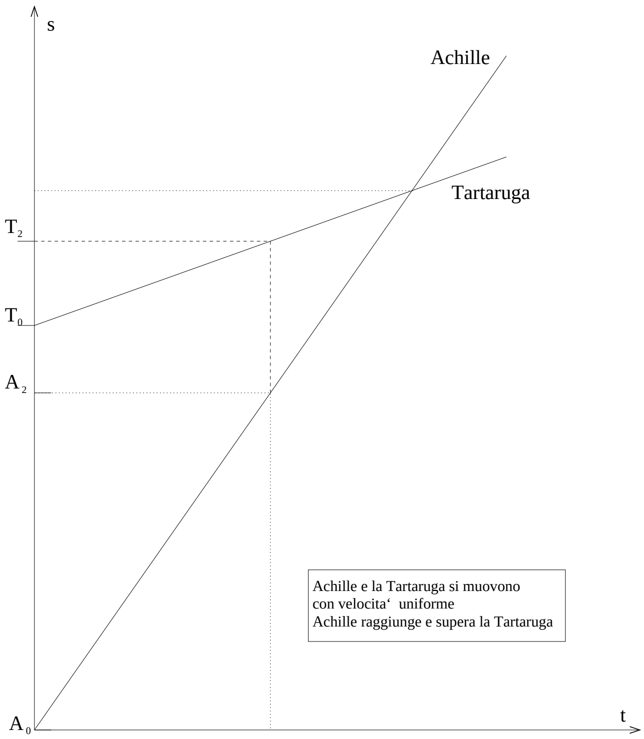

Figura 1.2: Achille e la Tartaruga, digramma tempo-spazio

ne dovrà percorrere la metà, e prima ancora la metà della metà, e così via *all'infinito*. Quindi non possiamo neppure partire!

L'aspetto paradossale è ovviamente legato al fatto che l'esperienza ci dice che Achille supera la tartaruga e per farlo è capace di muoversi (come d'altronde la stessa tartaruga).

Il paradosso non è quindi nei "fatti" ma nel nostro modello interpretativo della realtà. Esso nasce dall'idea parmenidea della indivisibilità del tutto.

Ma se assumiamo che lo spazio e il tempo siano *indefinitamente* suddivisibili possiamo, con relativa facilità, uscire da questa soluzione paradossale. Come già notava Aristotele, le successive distanze nella nostra suddivisione diventano sempre più piccole e occorre quindi un tempo sempre più piccolo per percorrerle. Si tratta quindi di mostrare che tutte queste, infinite, sempre più piccole quantità hanno *una somma finita*.

La matematica che ci permette di "dimostrare" che tali somme sono finite è nota sotto il nome di teoria delle *serie numeriche.* Essa compare nell'opera dei matematici "ellenistici" (un po' meno di due secoli dopo Zenone) in particolare nelle opere di Archimede sul metodo di esaustione (anticipato da Eudosso nel quarto secolo A.C.) e dalla teoria delle proporzioni nel V libro degli *Elementi*.

Dopo l'eclisse della matematica greca[1] si dovrà aspettare la "nascita" del calcolo differenziale per tornare a far luce su questo problema[2].

[1] Vedi il libro di Lucio Russo *La rivoluzione dimenticata* (1996).

[2] Ma la nascita della moderna fisica quantistica ha rimesso di nuovo in gioco l'assunto

1.2 La serie geometrica

Prendiamo un esempio puramente geometrico: partiamo da un quadrato e tracciamo una delle due diagonali. Il quadrato è ora diviso in due triangoli rettangoli isosceli. Lasciamo in pace uno dei due e dividiamo in due il secondo tracciando l'altezza relativa all'ipotenusa. Dividiamo ancora allo stesso modo uno dei due triangoli così ottenuti. A questo punto abbiamo una divisione del quadrato in quattro triangoli con aree rispettivamente di un mezzo, un quarto e un ottavo (due triangoli) dell'area del quadrato. A questa suddivisione corrisponde l'ovvia uguaglianza tra frazioni

$$1 = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8}\,. \tag{1.1}$$

La nostra assunzione di infinta divisibilità dello spazio ci "autorizza" a iterare il nostro processo di suddivisione ottenendo una "serie" di triangoli con aree che a ogni passo si dimezzano. La relazione algebrica corrispondente dovebbe ora avere un aspetto del genere:

$$1 = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \frac{1}{64} + \frac{1}{128} + \dots \tag{1.2}$$

dove i punti stano a indicare che la somma va avanti indefinitamente con termini analoghi (ovvero potenze negative di 2).

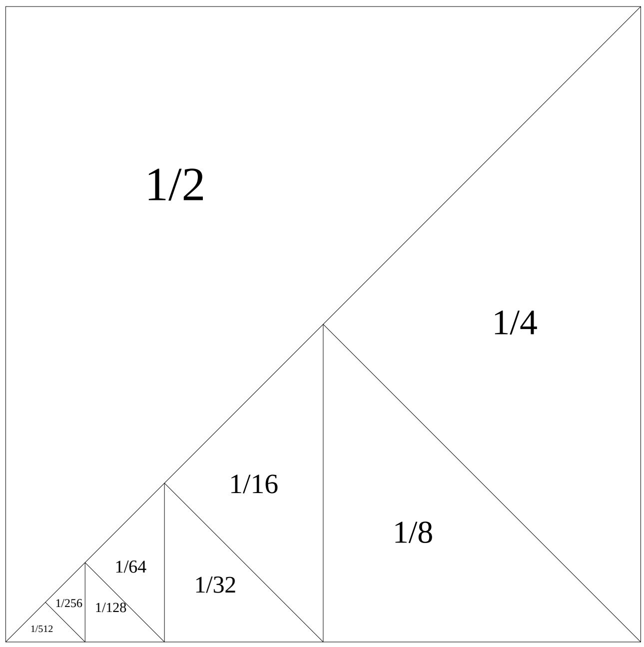

Figura 1.3: Divisione del quadrato e serie geometrica

della continuità alla base della soluzione dei paradossi. Una buona trattazione si trova in rete nelle pagine di Wikipedia dedicate agli "Zeno's Parodoxes" (purtroppo solo nella versione inglese e non nella traduzione italiana) http://en.wikipedia.org/wiki/Zeno_of_Elea e http://en.wikipedia.org/wiki/Zeno%27s_paradoxes.

Nella matematica odierna per indicare la somma a secondo membro nell'equazione (1.2) si usa una notazione introdotta da Leonhard Euler (Eulero) nel 1755:

$$\sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n , \tag{1.3}$$

che leggiamo "somma per n che va da 1 all'infinito di un mezzo elevato alla n", Il simbolo $\sum$ che deriva dalla lettere sigma maiuscola, Σ, è detto "sommatoria". Il simbolo ∞, detto "infinito", fu introdotto da John Wallis nel 1655 e sta a significare che la somma va estesa a tutti i *numeri interi* a partire a dal numero che compare sotto il segno $\sum$ (1 nel nostro caso). La lettera n è detto "indice della sommatoria" o anche "indice muto" a significare che il suo nome non ha importanza al di fuori della somma; avremmo potuto chiamarlo con qualsiasi altro nome (in genere una singola lettera) e la nostra interpretazione della somma non cambia: p.e.

$$\sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n \quad \text{e} \quad \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^k , \tag{1.4}$$

sono la stessa cosa.

Dalla nostra suddivisione del quadrato in triangoli ci aspettiamo l'uguaglianza

$$1 = \sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n , \tag{1.5}$$

che non è altro che la riscrittura in matematichese della "somma" (1.2). Anzi potremmo ritenere che la nostra suddivisione del quadrato sia una "dimostrazione" di questa uguaglianza. Ma questa argomentazione non soddisfa la necessità di rigore della matematica moderna (in verità neppure quella di Archimede, come si può vedere dalla ricostruzione del metodo di esaustione presentata nel libro di Russo): chi ci dice che il nostro processo effettivamente esaurisca tutti i punti del quadrato in termini di triangolini la cui area è presa in conto dalla somma infinita?

Vogliamo quindi costruire un contesto matematico in cui la nostra ugughlianza diventi un "teorema" cioè qualcosa di dimostrabile senza ombra di dubbio.

1.3 Serie e successioni

Facciamo un passo indietro: invece della "somma infinita" (1.2) consideriamo la somma che otteniamo "buttando via l'ultimo triangolino" in una costruzione che si fermi a un numero finito di suddivisioni. Per esempio supponiamo di aver suddiviso il quadrato in sei triangoli (facendo 5 "tagli") di cui gli ultimi due uguali. Lasciando da parte l'ultimo triangolo abbiamo quindi ricoperto una frazione di area del quadrato pari a

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} . \tag{1.6}$$

Riscriviamo questa somma alla Eulero

$$\sum_{k=1}^{5} \left(\frac{1}{2}\right)^k , \tag{1.7}$$

dove ora "l'estremo" superiore dell'indice di somma non è più l'infinito ma il numero 5. Con un ovvia generalizzazione possiamo fermarci, invece che a 5, ad un intero qualsiasi n e scrivere

$$s_n = \sum_{k=1}^{n} \left(\frac{1}{2}\right)^k , \tag{1.8}$$

(notate che ora n non è più un indice "muto" ed è "visibile" fuori dall'operazione di somma, il cui risultato dipende da n)

Il simbolo s_n, che sta a indicare la nostra somma, ci ricorda che essa è solo un pezzo della somma infinita da cui siamo partiti, e che con essa si confonde quando "n va all'infinito". Per questo motivo s_n prende il nome di *ridotta*, o *somma parziali*, n-esima della sommatoria infinita.

La somma ridotta s_n è un esempio di un oggetto matematico più generale detto *successione numerica*.

La definizione di succesione è semplice: una successione è una qualsiasi funzione numerica definita sui numeri naturali **N**, ovvero una legge che a ogni numero naturale n associa un numero a_n (a volte si usa anche la notazione $a(n)$). a_n si dice il termine n-esimo della succesione. Per indicare la successione nel suo insieme si usa la notazione, $\{a_n\}$, che consiste nel racchiudere il temine n-esimo tra parantesi graffe. In numero n è detto "indice" della successione. In genere lo si fa partire da 1 o da 0 (c'è per altro una annosa disquisizione se 0 debba essere incluso tra i naturali oppure no).

Altri esempi:

$$\{n^2\} , \quad \{2^n\} , \quad \left\{\frac{1}{n}\right\} ,$$

che rispettivamente associano al numero naturale n il suo quadrato, il numero 2 elevato alla potenza n-esima, il reciproco del numero n. Notate che nei primi due esempi il termine n-esimo è ancora un numero naturale, mentre nel terzo esempio il termine n-esimo è un numero razionale.

C'è un'altra sostanziale differenza tra le prime due successioni e la terza. Al crescere di n i termini n^2 e 2^n crescono "illimitatamente", ovvero diventano più grandi di ogni fissato numero M per n sufficientemente grande[3].

La terza successione invece ha un termine n-esimo che si avvicina a 0 quando n cresce. Come possiamo caratterizzare questo "avvicinamento"? Possiamo osservare che in questo caso il termine $\frac{1}{n}$ diventa arbitrariamente vicino a 0, ovvero per qualsiasi numero positivo ε abbiamo $|\frac{1}{n}| < \varepsilon$ se n è sufficientemente grande[4].

Possiamo ora caratterizzare questi due comportamenti in maniera formale:

Definizione 1.3.1 *Una successione $\{a_n\}$ si dice* convergente *al numero L se per ogni numero positivo ε esiste un indice N_ε per cui vale*

$$|L - a_n| < \varepsilon \quad \text{per ogni } n > N_\varepsilon . \tag{1.9}$$

[3] P.e. scegliendo $M = 10000$ abbiamo $n^2 > M$ per tutti gli $n \geq 101$ e $2^n > M$ per ogni $n \geq 14$.

[4] Se prendiamo $\varepsilon = \frac{1}{10000}$ abbiamo $|\frac{1}{n}| < \varepsilon$ per ogni $n \geq 10001$.

Definizione 1.3.2 *Una successione $\{a_n\}$ si dice* divergente *a $+\infty$ $(-\infty)$ se per ogni numero positivo M esiste un indice N_M per cui vale*

$$a_n > M \ (a_n < -M) \quad \text{per ogni } n > N_M, . \tag{1.10}$$

Esercizio Dimostrare che la successione $\{\frac{1}{n^2}\}$ converge a 0.

Possiamo ora dare una semplice caratterizzazione di una serie convergente:

Definizione 1.3.3 *Una serie numerica si dice convergente se è convergente la successione $\{s_n\}$ delle sue somme parziali.*

In modo analogo si caratterizzano le serie divergenti.

Dimostriamo ora che la serie geometrica è convergente secondo la nostra definizione, ovvero che la succesione delle sue ridotte $s_n = \sum_{k=1}^{n} \left(\frac{1}{2}\right)^k$, converge al valore 1.

Per far questo "calcoliamo" la somma, ovvero mostriamo che per ogni n si ha

$$s_n = \sum_{k=1}^{n} \left(\frac{1}{2}\right)^k = \frac{1}{2}\,\frac{1-\left(\frac{1}{2}\right)^n}{1-\frac{1}{2}}, \tag{1.11}$$

o, equivalentemente, che

$$S_n = \sum_{k=0}^{n} \left(\frac{1}{2}\right)^k = \frac{1-\left(\frac{1}{2}\right)^{n+1}}{1-\frac{1}{2}}, \tag{1.12}$$

si noti che ora l'indice di somma parte da 0 e non da 1, la differenza è che c'è un termine in più nella somma, quello con $k = 0$, che vale 1. Nota anche che per ogni n si ha $S_n = 2s_n$.

La dimostrazione si fa usando il Principio di Induzione.

La formula (1.12) è vera per $n = 0$ (nella somma c'è il solo termine $\left(\frac{1}{2}\right)^0 = 1 = \frac{1-\left(\frac{1}{2}\right)^1}{1-\frac{1}{2}}$.

Supponiamo ora cha la formula sia vera per n uguale a un qualche numero N e facciamo vedere che da ciò segue che la formula è vera anche per $N + 1$.

Per assunzione abbiamo quindi

$$\sum_{k=0}^{N} \left(\frac{1}{2}\right)^k = \frac{1-\left(\frac{1}{2}\right)^{N+1}}{1-\frac{1}{2}}, \tag{1.13}$$

e, per definizione di ridotta, $S_{N+1} = S_N + a_{N+1} = S_N + \left(\frac{1}{2}\right)^{N+1}$. Abbiamo quindi che

$$S_{N+1} = \sum_{k=0}^{N+1} \left(\frac{1}{2}\right)^k = \sum_{k=0}^{N} \left(\frac{1}{2}\right)^k + \left(\frac{1}{2}\right)^{N+1} = \frac{1-\left(\frac{1}{2}\right)^{N+1}}{1-\frac{1}{2}} + \left(\frac{1}{2}\right)^{N+1} . \tag{1.14}$$

Basta adesso calcolare l'ultima somma

$$\frac{1-\left(\frac{1}{2}\right)^{N+1}}{1-\frac{1}{2}} + \left(\frac{1}{2}\right)^{N+1} = \frac{1-\left(\frac{1}{2}\right)^{N+1} + \left(1-\frac{1}{2}\right)\left(\frac{1}{2}\right)^{N+1}}{1-\frac{1}{2}} = \frac{1-\left(\frac{1}{2}\right)^{N+2}}{1-\frac{1}{2}}, \tag{1.15}$$

che è ciò che otteniamo dalla nostra formula quando a posto di N sostituiamo $N+1$. Abbiamo così dimostrato che dalla validità della formula per un certo numero naturale (qualsiasi) segue la validità della formula per il naturale sucessivo. Poiché la formula valeva per $n=0$ (il primo intero), essa vale per tutti i numeri naturali[5].

Infine mostriamo che S_n converge a 2. Riscriviamo $S_n = 2\left(1-\left(\frac{1}{2}\right)^{n+1}\right)$. Fissiamo un numero $\varepsilon > 0$; abbiamo $0 < 2 - 2\left(1-\left(\frac{1}{2}\right)^{n+1}\right) < \varepsilon$ per ogni n che soddisfa $\left(\frac{1}{2}\right)^n < \varepsilon$, ovvero se $n > \frac{2}{\varepsilon^n}$ (per essere pignoli dobbiamo trovare, ai sensi della nostra definizione, un numero intero N_ε tale che la disuguglianza valga per $n > N_\varepsilon$; prendiamo quindi $N_\varepsilon = \left[\frac{2}{\varepsilon^n}\right]+1$, dove la parentesi quadra $[x]$ indica la parte intera di x, ovvero il più grande numero intero più piccolo di x). Ne segue che S_n converge a 2, o equivalentemente che s_n converge a 1.

Esercizio Mostrare che per ogni x vale

$$\sum_{k=0}^{n} x^n = \frac{1-x^{n+1}}{1-x}$$

e dedurne che per ogni $x<1$ sia ha

$$\sum_{k=0}^{\infty} x^n = \frac{1}{1-x}\,,$$

x si dice la "ragione" della serie geometrica.

Dire cosa succede per $x \geq 1$. Cercare inoltre di capire cosa succede per $x < 0$ (provate a "formulare un'ipotesi" sul caso $x=-1$).

Suggerimento La dimostrazione si può fare per induzione come già fatto per il caso di ragione $\frac{1}{2}$. In alternativa si può anche ragionare come segue: si moltiplica la somma per $(1-x)$

$$(1-x)\sum_{k=0}^{n} x^k = (1-x)\left(1+x+x^2+\ldots+x^n\right)$$

e si osserva che il secondo membro è la ben nota fattorizzazione di

$$\left(1-x^{n+1}\right)\,.$$

[5]**Il Principio di Induzione.** Questo principio risponde alla domanda: Come possiamo dimostrare che una proprietà è vera *per qualsiasi numero naturale*? Ci sono alcune proprietà che possiamo provare direttamente per un *generico* numero n, per esempio, la formula $(n+1)^2 = n^2+2n+1$. Ma la maggior parte delle formule "interessanti" non possono essere provate in questo modo. D'altra parte non possiamo neppure provare una proposizione per 1 e poi per 2 e poi per 3 e *così via* perché in questo modo dovremmo andare avanti indefinitamente.

Dobbiamo quindi dare un argomento che ci convinca che una certa proprietà può essere pensata vera per tutti i numeri naturali senza doverla verificare direttamente per tutti.

Questo argomento funziona nel modo seguente: supponiamo di aver dimostrato la proprietà per i primi n numeri, ovvero per $1,2,3,\ldots,n$ e inoltre supponiamo di poter dimostrare che dalla validità della proprietà per questi numeri *segua la sua validità* per il numero successivo $n+1$. Allora *asseriamo* che la proprietà *è vera per tutti i numeri naturali.*

Nella maggior parte delle applicazioni, se la proprietà coinvolge solo un numero naturale, il Principio si usa mostrando che la proprietà è vera per il numero 1 (o 0) e che *se si assume* che valga per un numero n qualsiasi allora *segue* che essa *vale per il numero sucessivo* $n+1$.

In questo modo dal fatto che la propriet è vale per 1 segue che vale per 2 e poi che vale per 3 *e così via.*

1.3.1 Rappresentazioni decimali

La serie geometrica in effetti non dovrebbe essere una "novità" anche per chi non abbia mai sentito parlare di serie o successioni.

Prendiamo il numero $\frac{1}{3}$. Sappiamo che la sua rappresentazione decimale è un "numero periodico" ovvero $\frac{1}{3} = 0,33333333\ldots$ che per convenienza si scrive anche $0,\overline{3}$ intendendo con la sopralieatura che in numero 3 si ripete indefinitamente.

Ma la rappresentazione decimale[6] di un numero, p.e. $2,53$, deve intendersi come una somma: $2,53 = 2(10)^0 + 5(10)^{-1} + 3(10)^{-2}$.

In questo modo possiamo ora scrivere

$$\frac{1}{3} = \sum_{n=1}^{\infty} \frac{3}{10^n}$$

ovvero come una serie geometrica di ragione $\frac{1}{10}$ moltiplicata per 3.

Esercizio Verificare con la formula generale per la serie geometrica che questo risultato è corretto.

Esercizio Determinare la serie che esprime il numero $\frac{1}{7}$.

Esercizio Per quali numeri n la frazione $\frac{1}{n}$ ha una rappresentazione decimale periodica?.

Un esempio un po' inquietante è il numero $0,9999999\ldots$: quanto vale? Ovviamente possiamo subito dire che $0,9999999\ldots \leq 1$, ma è un numero effettivamente minore di 1 o si tratta di "un'illusione ottica"?

Anche qui la serie geometrica ci permette di rispondere: abbiamo

$$\sum_{n=1}^{\infty} \frac{9}{10^n} = 9\,\frac{\frac{1}{10}}{1-\frac{1}{10}} = 9\,\frac{1}{9} = 1$$

ovvero $0,9999999\ldots = 1$.

1.3.2 Altre serie notevoli

1.3.3 La serie armonica

E' la serie definita da

$$\sum_{k=1}^{\infty} \frac{1}{k}\,. \tag{1.16}$$

Anche in questo caso si sommano infini termini sempre più piccoli. Ma ora la serie è divergente. Per renderse conto osserviamo che possiamo scrivere una ridotta di ordine 2^n raggruppando i termini

$$s_{2^n} = 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) + \ldots + \left(\frac{1}{2^{n-1}+1} + \ldots + \frac{1}{2^n}\right)$$

dove ogni somma dentro parentesi tonda termina con un termine del tipo $\frac{1}{2^k}$, $k = 2, 3, \ldots, n$ e contiene 2^{k-1} termini. Al totale abbiamo quindi $n-1$ parentesi.

[6] E più in genrale qualsiasi rappresentazione di un numero intero, comprese quelle non strettamente posizionali: p.e. MCMLIII (1953) è la "somma" di "Mille+(Mille-Cento)+Cinquanta+Uno+Uno+Uno"

Poiché l'ultimo addendo dentro le parentesi è più piccolo di tutti quelli che lo precedono, possiamo "minorare" le somme in parentesi sostituento tutti gli addendi con termini uguali all'ultimo, ovvero abbiamo

$$s_{2^n} > 1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) + \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}\right) + \ldots + \left(\frac{1}{2^n} + \ldots + \frac{1}{2^n}\right)$$

Ma ora ognuna delle parentesi vale $\frac{1}{2}$ (verificarlo!) e al totale quindi

$$s_{2^n} > 1 + \frac{1}{2} + (n-1)\frac{1}{2} = 1 + \frac{n}{2}\,.$$

Ma la successione $\left\{\frac{n}{2}\right\}$ diverge a $+\infty$ e così deve essere anche della serie.

La serie di Mengoli

Si tratta della serie

$$\sum_{k=1}^{\infty} \frac{1}{k(k+1)}\,. \tag{1.17}$$

Mostramo che questa serie è convergente e che la sua somma vale 1.

Osserviamo prima di tutto che

$$\frac{1}{k(k+1)} = \frac{1}{k} - \frac{1}{k+1}\,.$$

Prendiamo le ridotte della serie

$$s_n = \frac{1}{1\cdot 2} + \frac{1}{2\cdot 3} + \frac{1}{3\cdot 4} + \ldots + \frac{1}{n(n+1)}$$
$$= \left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \ldots + \left(\frac{1}{n} - \frac{1}{n+1}\right) = 1 - \frac{1}{n+1}\,,$$

e la convergenza della serie segue ora dalla convergenza a 0 della successione $\left\{\frac{1}{n+1}\right\}$.

La serie di Mengoli ha una suggestiva interpretazione geometrica: prendiamo le due iperboli di equazione $y = \frac{1}{x}$ e $y = \frac{1}{x+1}$. Nei punti di ascissa intera k (positiva) la differenza delle ordinate dei punti corrispondenti sulle due iperboli è data da

$$\frac{1}{k} - \frac{1}{k+1}\,.$$

Quindi la serie corrisponde a sommare tali differenze. E' immediato "verificare" dal grafico delle due iperboli che questa somma ci dà un segmento di lunghezza unitaria.

La serie esponenziale

Questa è forse la serie più famosa assieme alla serie geometrica:

$$\sum_{k=0}^{\infty} \frac{1}{k!}\,, \tag{1.18}$$

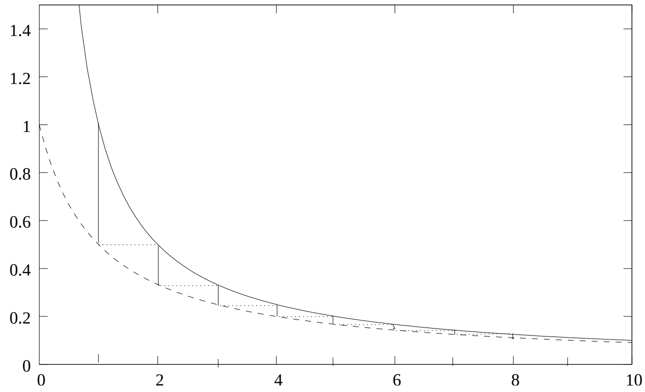

Figura 1.4: Grafico delle iperboli $y = \frac{1}{x}$ e $y = \frac{1}{x+1}$

dove il numero $k!$ è il prodotto di tutti i naturali minori o uguali a k (per definizione si pone inoltre $0! = 1$). Mostreremo che questa successione è limitata oltre ad essere ovviamente crescente (tutti i termini sono positivi). Mostreremo anche che se la somma esiste, essa non può essere un numero razionale. Si pone quindi il problema di capire di che razza di "numero" si tratti, ammesso che esista.

Per dimostare la limitatezza della somma (o meglio della successione delle ridotte s_n) basta osservare che il termine n-esimo della somma è maggiorato, per $k \geq 3$ da $\left(\frac{1}{2}\right)^{k-1}$, che segue immediatamente osservando che

$$k! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot (k-1) \cdot k > 2^{k-1}\,,$$

dove nel prodotto ci sono $k - 1$ fattori (oltre a 1) tutti maggiori o uguali a 2.

Ne segue che

$$2 < 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \ldots + \frac{1}{k!} < 1 + 1 + \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \ldots + \frac{1}{2^{k-1}} < 3\,.$$

Torneremo poi sulla questione dell'esistenza della somma (come molti sapranno la somma della serie è il numero e).

Nota In modo analogo, dalla convergenza della serie di Melgoli è facile (provare a fare i conti) ricavare la convergenza della serie

$$\sum_{k=1}^{\infty} \frac{1}{k^2}\,. \tag{1.19}$$

Infatti per ogni termine della serie si ha $\frac{1}{k^2} \leq \frac{2}{k(k+1)}$. Ne segue che la somma degli inversi dei quadrati (se esiste) è minore di 2. (in questo caso la somma è uguale a $\frac{\pi^2}{6}$).

1.4 Successioni che vorrebbero convergere

Abbiamo finora visto esclusivamente successioni (o serie) di numeri razionali. In alcuni casi (la serie geometrica p.e.) abbiamo anche verificato che queste successioni convergevano a un numero razionale.

Poi abbiamo visto i due esempi dati dalla serie esponenziale e serie della somma dei recipoci dei quadrati. Le somme parziali di queste serie sono sucessioni che hanno un comportamento molto simile alle successioni che abbiamo mostrato essere convergenti. Infatti sono entrambe crescenti ($s_n < s_{n+1}$) e per entrambe abbiamo trovato una "maggiorazione" ovvero un numero M tale che $s_n < M$ per ogni n.

Però non siamo stati in grado di "dimostrare" che anch'esse convergono, perché non ne conosciamo la somma, e quindi non siamo in grado di applicare la definizione di successione convergente[7].

Vedremo poi che il fatto di non conoscere la loro somma non è un caso dovuto alla nostra incapacità ma nasconde una profonda questione che riguarda i "numeri". Per questi due esempi però dovremo fare dei conti un po' complicati e quindi presentiamo prima un esempio più abbordabile.

Partiamo dall'equazione di secondo grado

$$x^2 - 2x - 1 = 0 \tag{1.20}$$

e cerchiamone la soluzione positiva.

La risposta è "ben nota": abbiamo $\hat{x} = 1 + \sqrt{2}$. Ma questa "soluzione" ci pone un problema serio: la radice di 2 *non è un numero razionale*[8].

Dobbiamo quindi "definire" un oggetto a cui dare la dignità di numero e che rappresenti la soluzione di (1.20). Le sucessioni ci danno una soluzione di questo problema.

Iniziamo riscrivendo la (1.20) nella forma

$$x = 2 + \frac{1}{x} \tag{1.21}$$

ottenuta dalla nostra equazione "portando a secondo membro" i termini di grado inferiore al secondo e dividendo per x. Il "numero" $\hat{x} = 1 + \sqrt{2}$ è una soluzione anche di (1.21). Ma ora possiamo costruire una successione di numeri razionali a partire da (1.21): Partiamo da $x_0 = 2$: poichè la soluzione di (1.21) ha la forma 2 più un numero positivo abbiamo $x_0 = 2 < \hat{x} = 1 + \sqrt{2}$.

Costruiamo ora x_1 sostituendo x_0 nel secondo membro di (1.21), ovvero $x_1 = 2 + \frac{1}{x_0} = 2 + \frac{1}{2}$. Poiché sappiamo che $x_0 = 2 < \hat{x}$, ora avremo $x_1 > \hat{x}$. Possiamo proseguire a costruire la successione secondo questa regola, ovvero

$$x_0 = 2\,, \quad x_{n+1} = 2 + \frac{1}{x_n}\,,\ n = 0, 1, 2, \dots\ . \tag{1.22}$$

In accordo con quanto già osservato per i primi due termini si ha, alternativamente, che ogni termine di indice pari ($n = 2k$) è minore di $\hat{x}$ mentre ogni

[7] Si osservi che per *dimostrare*, usando la definizione di sucessione convergente, che una data successione converge, bisogna conoscerne il limite.

[8] L'irrazionalià di $\sqrt{2}$ è un risultato matematico molto antico, noto ai matematici greci nella sua forma "geometrica" di *incommensurabilità* tra il lato e la diagonale di un quadrato. La dimostrazione si può fare *per assurdo*: supponiamo che esistano due numeri interi p e q tali che $\sqrt{2} = p/q$ (possiamo assumere p e q primi tra loro); avremmo allora $2q^2 = p^2$, e quindi p^2 deve essere un numero pari. Ciò è possibile solo se, a sua volta, p è pari ovvero esiste un numero intero r tale che $p = 2r$. Ma da questo segue che $q^2 = 2r^2$ e quindi anche q deve essere un numero pari e ciò è assurdo perché avevamo assunto che p e q fossero primi tra loro. In appendice si trova un'altra dimostrazione

termine di indice dispari ($n = 2k + 1$) è maggiore di $\hat{x}$. Inoltre abbiamo che per ogni k vale

$$x_{2k} < x_{2k+2}\,, \quad x_{2k+1} > x_{2k+3}\,, \quad k = 0, 1, 2, \dots\,, \tag{1.23}$$

ovvero la sucessione dei termini di indice pari è crescente mentre quella di indice dispari è decrescente.

Infine osserviamo che

$$\begin{aligned} x_{2k+3} - x_{2k+2} &= 2 + \frac{1}{x_{2k+2}} - 2 - \frac{1}{x_{2k+1}} \\ = \frac{1}{x_{2k+2}x_{2k+1}}\left(x_{2k+1} - x_{2k+2}\right) &< \frac{1}{4}\left(x_{2k+1} - x_{2k+2}\right) \end{aligned} \tag{1.24}$$

dove abbiamo sfruttato il fatto che ogni termine della successione è maggiore di 2 (e quindi $\frac{1}{x_{2k+2}x_{2k+1}}$ è minore di $1/4$).

La diseguaglianza (1.24) ci dice che a "ogni passo" la distanza tra due termini consecutivi decresce di un fattore (almeno) un quarto rispetto alla distanza dei due termini precedenti[9].

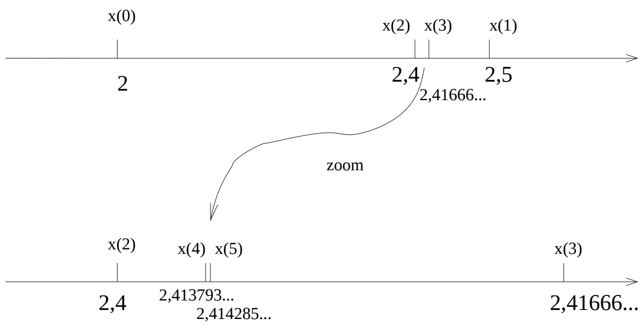

Figura 1.5: I punti della succesione approssimante di $\sqrt{2}$

La disuguaglianza (1.24) ci permette anche di provare una notevole proprietà della nostra successione. Prendiamo un numero positivo ε qualsiasi e facciamo vedere che esiste un numero naturale N_ε tale che per ogni m e n maggiori di N_ε si ha

$$|x_m - x_n| < \varepsilon\,.$$

Prima di dimostrare che questa condizione è soddisfatta dalla nostra successione, commentiamola un po'.

C'è una forte somiglianza con la definizione di successione convergente (tornare a leggere la definizione (1.3.1)) ma non si tira in ballo il valore del limite (ammesso che ci sia). Quello che ci dice questa condizione è che la differenza tra due termini della sucessione può essere resa arbitrariamente piccola purché i loro indici siano sufficientemente grandi. Ovvero i termini della successione si avvicinano sempre di più *tra loro* al crescere di n e la distanza tra due qualsiasi termini converge a zero. Questo comportamento è così fondamentale che gli si dà un nome:

[9]In realtà la decrescenza è più rapida: infatti poiché i due valori di x_{2k+2} e x_{2k+1} sono molto vicini a $1 + \sqrt{2}$, il loro prodotto è vicino al quadrato di questo numero che vale quasi $5,83$.

Definizione 1.4.1 *Una successione* $\{x_n\}$ *si dice* di Cauchy *se per ogni numero positivo* ε *esiste un numero naturale* N_ε *tale che per ogni* m *e* n *maggiori di* N_ε *si ha*

$$|x_m - x_n| < \varepsilon\,.$$

E' facile dimostrare che se una successione è convergente allora è di Cauchy[10].

Abbiamo appena visto un esempio di successione di Cauchy di numeri razionali il cui "limite" non è un numero razionale. D'altra parte la succesione che abbiamo costruito sembra individuare in modo univoco un punto sulla retta (una volta si sia scelto un segmento di lunghezza unitaria), così come si individua un "sol punto" della retta quando si riporta con il compasso la lunghezza della diagonale di un quadrato.

La nostra assunzione di "continuità" dello spazio ci "impone" quindi di ammettere come "numeri" queste successioni. Definiremo dunque i *numeri reali* come le successioni di Cauchy di numeri razionali (per precisione dovremo prendere come definizione di numero reale delle opportune *classi di equivalenza* di successioni di Cauchy). Così la radice di 2 diventa un "numero reale" identificato dalla successione di Cauchy che abbiamo costruito in questo paragrafo[11].

Ne risulta che tra numeri reali vale (*per costruzione*) anche il viceversa del criterio che abbiamo dimostrato:

Teorema 1.4.2 *Una successione di numeri reali è convergente se e solo se è di Cauchy.*

Un altro importante risultato che segue dalla nostra definizione di numeri reali è il seguente:

Teorema 1.4.3 *Ogni successione crescente superiormente limitata è convergente.*

Superiormente limitata significa che esite un numero M, fissato una volta per tutte, per cui vale $x_n \le M$ per ogni indice n.

In particolare questo risultato ci garantisce la convergenza delle successioni

$$\sum_{k=0}^{\infty} \frac{1}{k!} \quad \text{e} \quad \sum_{k=1}^{\infty} \frac{1}{k^2}\,,$$

le cui somme parziali abbiamo già visto essere limitate superiormente.

[10] Infatti se $\{x_n\}$ converge a L abbiamo che per ogni numero positivo ε esiste un numero naturale N_ε tal che $|L - x_n| < \frac{\varepsilon}{2}$ per ogni $n > N_\varepsilon$ (nota la piccola "variazione" se ε è un numero "arbitrario" altrettanto è arbitrario $\frac{\varepsilon}{2}$). Ma allora se m e n sono entrambi maggiori di N_ε abbiamo

$$|x_m - x_n| \le |L - x_n| + |L - x_m| < \varepsilon$$

che è la definizione di successione di Cauchy.

[11] Per mettere le cose a punto dovremmo far vedere che possiamo estendere le consuete operazioni tra numeri a questi nuovi "oggetti". Alcune operazioni sono defininite in modo semplice: p.e. dati due numeri individuati rispettivamente dalle sucessioni $\{x_n\}$ e $\{y_n\}$, la loro somma è il numero individuato dalla succesione $\{x_n + y_n\}$, ovvero quella il cui termine n-esimo è dato dalla somma dei termini n-esimi delle due successioni.

1.4.1 Ancora sulla serie esponenziale

Torniamo ora alla serie esponenziale. Abbiamo detto che la successione delle somme parziali è una successione convergente e quindi possiamo scrivere

$$\sum_{k=0}^{\infty} \frac{1}{k!} = e\,, \tag{1.25}$$

dove il simbolo e indica un numero reale, definito dalla somma medesima.

Dimostramo ora che e è un numero irrazionale, cioè che non possono esistere due numeri interi p e q tali che $e = p/q$.

La dimostrazione fa uso della segeunte stima dell'errore che si commette sostituendo alla somma infinita la somma parziale di ordine N

$$e - s_N = \sum_{k=0}^{\infty} \frac{1}{k!} - \sum_{k=0}^{N} \frac{1}{k!} < \frac{1}{N \cdot N!}\,, \tag{1.26}$$

che dimostremo succesivamente.

Supponiamo quindi, per assurdo, che e sia razionale ovvero $e = p/q$ con p e q interi (e primi tra loro); poiché abbiamo già visto che e è un numero strettamente compreso in $(2,3)$, estremi esclusi, q deve essere maggiore di 1 . In base alla stima (1.26) abbiamo che la ridotta s_q q dista da e per meno di $\frac{1}{q \cdot q!}$, ovvero vale

$$\frac{p}{q} - s_q < \frac{1}{q \cdot q!}\,. \tag{1.27}$$

Motiplichiamo ora la disuguaglianza (1.26) per $q!$ e otteniamo

$$p(q-1)! - q!\left(1 + \frac{1}{1!} + \frac{1}{2!} + \ldots + \frac{1}{q!}\right) < \frac{1}{q}\,. \tag{1.28}$$

Ora il secondo membro è un numero minore di 1 mentre il primo membro è un numero intero. Abbiamo quindi una contraddizione, provenuta dall'aver assunto che e sia razionale.

Resta da dimostrare la maggiorazione (1.26). Abbiamo

$$e = \sum_{k=0}^{N} \frac{1}{k!} + \sum_{k=N+1}^{\infty} \frac{1}{k!}$$

e quindi (1.26) equivale a far vedere che seconda somma $\sum_{k=N+1}^{\infty} \frac{1}{k!}$ è minore di $\frac{1}{N\,N!}$. Mostramo che vale

$$\sum_{k=N+1}^{\infty} \frac{1}{k!} < \frac{N+2}{(N+1)(N+1)!}\,, \tag{1.29}$$

dove a secondo membro si ha $\frac{N+2}{(N+1)(N+1)!} = \frac{N(N+2)}{(N+1)^2\,N!}\frac{1}{N} < \frac{1}{N\,N!}$ Possiamo riscrivere la somma in (1.29) riscalando l'idice di sommatoria

$$\sum_{k=N+1}^{\infty} \frac{1}{k!} = \sum_{j=0}^{\infty} \frac{1}{(N+1+j)!}\,, \tag{1.30}$$

(ottenuta ponendo $j = k-N-1$). Ricordando che $(k+h)! = k!\cdot(k+1)\cdot\ldots\cdot(k+h)$, possiamo mettere in evidenza in ogni addendo il fattore $\frac{1}{(N+1)!}$ e otteniamo

$$\sum_{k=N+1}^{\infty} \frac{1}{k!} = \frac{1}{(N+1)!}\left(1 + \frac{1}{N+2} + \frac{1}{(N+2)(N+3)} + \ldots\right). \tag{1.31}$$

Ora il termine nella posizione k-esima della somma può essere maggiorato con $\left(\frac{1}{N+2}\right)^k$ (per $k = 1, 2$ si ha un'uguaglianza), e quindi tutta la somma è maggiorata dalla serie geometrica di ragione $\frac{1}{N+2}$. Quindi

$$\sum_{k=N+1}^{\infty} \frac{1}{k!} < \frac{1}{(N+1)!}\sum_{k=0}^{\infty}\left(\frac{1}{N+2}\right)^k = \frac{1}{(N+1)!}\frac{N+2}{N+1}. \tag{1.32}$$

1.4.2 Ancora sul numero e

L'origine del numero e però è legata ad un altra successione, che nasce da un problema di matematica finanziaria. Se depositiamo una somma presso un conto di risparmio, alla fine di un anno ci vengono accreditati gli interessi ovvero una percentuale della somma depositata: se abbiamo messo nel conto la somma S alla fine dell'anno avremo quindi una nuova somma $S(1+h)$ dove $h > 0$ è il tasso di interesse *annuo* corrisposto e hS l'interesse. Se lasciamo la nuova somma nel conto per un altro anno, allo scadere del secondo anno avremo un capitale composto pari al capitale all'inizio del secondo anno $S(1+h)$ moltiplicato a sua volta per $(1+h)$.

Se gli interessi vengono calcolati su un periodo più breve la formula usata per calcolarli tiene conto della frazione di tempo (rispetto all'unit'a "anno") dopo la quale vengono accreditati in conto gli interessi, *rapportando il tasso di interesse alla lunghezza del periodo.* In genere si usa un tasso di interesse direttamente proporzionale alla frazione di tempo considerata. Per esempio se calcoliamo gli interessi dopo sei mesi l'interesse sarà $S\frac{h}{2}$, dove h rappresenta ancora il tasso di interesse su base annua e quindi $\frac{h}{2}$ il tasso su base semestrale.

Vediamo che è conveniente farsi accredidditare gli interessi dopo sei mesi. Infatti nei successivi sei mesi l'interesse sarà calcolato non sulla somma iniziale S ma sulla somma "capitalizzata", data da S più l'interesse corrisposto, allo scadere dei primi sei mesi, $S(1+\frac{h}{2})$. Allo scadere dell'anno avremo quindi $S(1+\frac{h}{2})(1+\frac{h}{2}) = S\left(1+h+\frac{h^2}{4}\right)$. La differenza è piccola se h è un numero piccolo (dell'ordine diciamo del 3%) ma la differenza tra gli interessi così calcolati può comunque essere non trascurabile se la somma S è ingente. Inoltre la differenza comincia a essere sensibile in tempi di forte inflazione, quando i tassi d'interesse corrisposti possono essere dell'ordine dell'unità.

Una volta capito cosa succede ci si chiede cosa cambia se diminuiamo ancora il periodo minimo per il calcolo dell'interesse. Se dividiamo l'anno in n periodi (p.e. 12 mesi) al termine di ognune dei quali capitalizziamo gli interessi, avremo, alla fine dell'anno un "guadagno" tanto maggiore quanto numerosi sono i periodi in cui abbiamo diviso l'anno. Con n periodi (di ugual durata) avremo un capitale finale pari a

$$S\left(1+\frac{h}{n}\right)^n. \tag{1.33}$$

Una domanda naturale a questo punto è: cosa succede se prendiamo intervalli temporali sempre più piccoli per la capitalizzazione degli interessi? ovvero cosa succede se facciamo crescere indefinitamente n nella formula (1.33).

La risposta è che la quantità in (1.33) cresce con n ma non supera il valore e^h, al quale converge per n che va all'infinito.

Limite della successione

$$\left(1+\frac{1}{n}\right)^n = \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array}\right) \frac{1}{n^k} = \sum_{k=0}^{n} \frac{n!}{(n-k)!k!}\frac{1}{n^k}$$

Si ha

$$\frac{n!}{(n-k)!}\frac{1}{n^k} < 1$$

quindi

$$\left(1+\frac{1}{n}\right)^n = \sum_{k=0}^{n} \frac{n!}{(n-k)!k!}\frac{1}{n^k} < \sum_{k=0}^{n} \frac{1}{k!} < \sum_{k=0}^{\infty} \frac{1}{k!}$$

inoltre si ha

$$\frac{n!}{(n-k)!}\frac{1}{n^k} = \frac{(n-k+1)\cdot \ldots \cdot n}{n^k} = \frac{n-k+1}{n} \cdot \ldots \cdot \frac{n}{n}$$

Fissiamo $m < \sqrt{n}$

$$\left(1+\frac{1}{n}\right)^n = \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array}\right) \frac{1}{n^k} > \sum_{k=0}^{m} \left(\begin{array}{c} n \\ k \end{array}\right) \frac{1}{n^k} = \sum_{k=0}^{m} \left(\frac{n-k+1}{n} \cdot \ldots \cdot \frac{n}{n}\right)\frac{1}{k!}$$

Fissiamo $\varepsilon > 0$ (piccolo), allora esiste N_ε tale che $\left(\frac{n-k+1}{n} \cdot \ldots \cdot \frac{n}{n}\right) > 1-\varepsilon$ per ogni $n > N_\varepsilon$ e per ogni $m < \sqrt{n}$ (questo dipende dal fatto che il più piccolo dei fattori del prodotto vale almeno $\frac{n-\sqrt{n}}{n}$ che converge a 1 quando n cresce).

Ora possiamo far crescere m (sempre soggetto alla limitazione $m < \sqrt{n}$ che fa crescere n) e otteniamo

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n > (1-\varepsilon)\sum_{k=0}^{\infty}\frac{1}{k!}$$

per ogni $\varepsilon > 0$ e quindi

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n \geq \sum_{k=0}^{\infty}\frac{1}{k!}$$

1.5 Appendice

1.5.1 I numeri razionali sono "numberabili"

Per quanto possa sembrare anti intiutivo è piuttosto facile dimostrare che i numeri razionali sono *numerabili* ovvero che esiste una corrispondenza biunivoca tra numeri razionali e numeri naturali.

Questo "paradosso" dell'infinito segue quelli molto più "accettabili" che consistono nel mettere in corrispondenza biunivoca l'insieme dei naturali con un suo sottoinsieme proprio (p.e. con i numeri pari, come osservato già da Galileo). Questa caratteristica è specifica degli insiemi "infiniti" tanto che viene presa come definizione di insieme infinito.

Vediamo come funziona l'esempio di Galileo: i numeri pari sono tanti quanti i numeri naturali. Basta porre la seguente corrispondenza tra $\mathbf{N}$ (i numeri naturali) e $2\mathbf{N}$ (i numeri pari): $n \to 2n$ che è ovviamente iniettiva e suriettiva, con inversa $m \to m/2$.

Abbiamo quindi una stessa numerosità per due insiemi dei quali uno è contenuto nell'altro (e non coincide con esso). Una volta imparato il trucco è facile costruire esempi di corrispondenze biunivoche tra $\mathbf{N}$ e sui sottoinsiemi sempre più "radi" (p.e. tra i naturali e le potenze di 10).

Un esempio meno banale, e dove non esiste una funzione esplicita che ci dà la corrispondenza, è quello dei numeri primi. *I numeri primi sono infiniti, ovvero sono in corrispondenza biunivoca con i numeri naturali.* Per dimostrarlo si procede per assurdo: supponiamo che ci siano soltanto un numero finito di numeri primi $2, 3, 5, 7, 11, ..., p$ dove p è più grande numero primo. Prendiamo ora il numero formato dal prodotto di tutti questi numeri primi, addizionato al numero 1, ovvero $q = 2 \cdot 3 \cdot 5 \cdot 7 \cdot \ldots \cdot p + 1$. Per ipotesi q non è primo in quanto maggiore del più grande numero primo p. Quindi deve ammettere almeno un divisore, e quindi almeno un divisore che sia numero primo, chiamimolo r. Ma r non può essere un numero della lista dei numeri primi minori o uguali a p, infatti dividendo q per uno qualsisi di questi numeri otteniamo un resto di 1. Quindi r deve essere più grande di p, in contraddizione con l'assunzione che p fosse il più grande numero primo.

Vediamo ora coma si possono contare i numeri razionali. Ci limitiamo a contare i numeri razionali positivi, lasciando per esercizio di modificare la strategia per contare anche quelli negativi.

Ogni razionale positivo può essere scritto come frazione p/q dove p e q sono numeri naturali. e possiamo disporre tutti questi numeri in un tabellone (infinito) dove la prima riga è fatti da tutti i numeri naturali divisi per 1, la seconda da tutti i numeri naturali divisi per 2 e così via. Ora possiamo numerare tutti gli elemti della tabella seguendo un percorso come indicato dalle frecce nel grafico

$$\begin{array}{cccccccccc}
1/1 & & 2/1 & \to & 3/1 & & 4/1 & \to & 5/1 & \ldots \\
\downarrow & \nearrow & & \swarrow & & \nearrow & & \swarrow & & \\
1/2 & & 2/2 & & 3/2 & & 4/2 & \ldots & & \\
& \swarrow & & \nearrow & & \swarrow & & & & \\
1/3 & & 2/3 & & 3/3 & \ldots & & & & \\
\downarrow & \nearrow & & \swarrow & & & & & & \\
1/4 & & 2/4 & \ldots & & & & & & \\
& \swarrow & & & & & & & & \\
1/5 & \ldots & & & & & & & &
\end{array}$$

Si noti che così facendo i numeri vengono contati più di una volta (p.e. 1/1 , 2/2 e 3/3 sono sempre lo stesso numero razionale) ma questo difetto è facilmente eliminabile "saltando" i numeri già contati.

1.5.2 Radice di 2 è irrazionale

Se la diagonale e il lato di un quadrato fossero commensurabili (ovvero il loro rapporto fosse un numero razionale p/q) potremmo scegliere l'unità di misura in modo da porre la lunghezza del lato uguale a q volte questa misura e quella della diagonale uguale p volte l'unità. Notiamo che p e q soddisfano alle disuguaglianze $p > q$ e $p < 2q$.

Ora riportiamo il lato sulla diagonale e prendiamo il segmento differenza per lato di un nuovo quadrato, $p - q$. Per semplici considerazioni sui segmenti di tangente alla circonferenza, questo nuovo quadrato avrà diagonale di lunghezza $2q - p$.

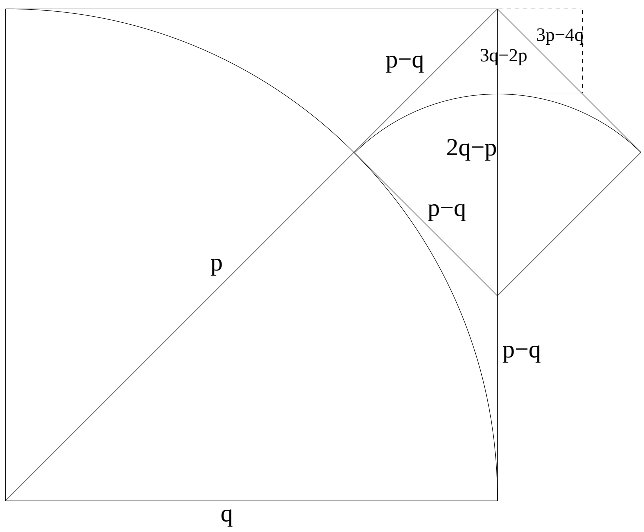

Figura 1.6: Dimostrazione dell'irrazionalità di $\sqrt{2}$

Possiamo procedere in questo modo generando una successione di quadrati i cui lati e diagonali sono segmenti di lunghezza decrescente e legati alle lunghezze del lato e diagonale del quadrato precedente dalla relazione

$$p_{k+1} = 2q_k - p_k \, , \quad q_{k+1} = p_k - q_k \, .$$

Ma tutti questi numeri p_k e q_k sono numeri interi, quindi non possiamo discendere all'infinito in quanto i numeri interi tra p e zero sono un numero finito. D'altra parte la costruzione geometrica può essere ripetuta indefinitamente per similitudine.

1.5.3 Il metodo di Erone per il calcolo della radice quadrata

Si tratta di un medoto per il calcolo approssimato della radice quadrata che conduce a un algoritmo molto simile a quello che abbiamo considerato, ma che ha una interessante interpretazione.

Vogliamo quindi determinare dei valori approssimati per $\sqrt{2}$. Partiamo da un valore "di prova" x_0. Possiamo scegliere $x_0 = 2$ che è il più piccolo intero che soddisfa la diseguaglianza $x_0^2 > 2$. Assieme a x_0 scegliamo y_0 definito dall'equazione $x_0 y_0 = 2$, ovvero $y_0 = \frac{2}{x_0}$. Con la scelta $x_0 = 2$ abbiamo quindi $y_0 = 1$, ma quello che conta è che per ogni x_0 che soddifi la condizione $x_0^2 > 2$, avremo $y_0^2 < 2$, e quindi la coppia (x_0, y_0) fornisce una approssimazione per eccesso e per difetto del "numero" cercato.

Per migliorare la nostra stima, costruiamo ora una seconda approssimazione prendendo $x_1 = \frac{x_0+y_0}{2}$, ovvero la media aritmetica delle due approssimazioni precedenti. E' facile mostare che $x_1^2 > 2$. Infatti si ha

$$x_1^2 = \frac{x_0^2 + 2x_0y_0 + y_0^2}{4} > \frac{4x_0y_0}{4} = 2$$

dove abbiamo sfruttato la disuguaglianza $x^2 + y^2 \geq 2xy$ che discende immediatamente osservando che $(x-y)^2 \geq 0$ (il segno di uguaglianza vale se e solo se $x = y$). Inoltre, ovviamente, si ha $y_0 < x_1 < x_0$, quindi il nuovo valore x_1 è una un'approssimazione per ecesso migliore della precedente. Definendo $y_1 = \frac{2}{x_1}$ abbiamo una nuova approssimazione per difetto (verificarlo!) anch'essa migliore della precedente, cioè vale $y_0 < y_1 < x_1 < x_0$ (verificare anche questo!).

Possiamo ora andare avanti allo stesso modo costruendo due successioni di approssimazioni $\{x_n\}$ e $\{y_n\}$ con la formula di ricorrenza

$$x_{n+1} = \frac{x_n + y_n}{2}, \quad y_{n+1} = \frac{2}{x_{n+1}}$$

con le proprietà

1. per ogni n, $x_n^2 > 2$, ovvero x_n è un'approssimazione per eccesso;
2. per ogni n, $y_n^2 < 2$, ovvero y_n è un'approssimazione per difetto;
3. per ogni n, $y_n < y_{n+1} < x_{n+1} < x_n$, ovvero la successione $\{x_n\}$ è decrescente mentre la successione $\{y_n\}$ è crescente, e sono limitate rispettivamente dal basso e dall'alto.

A questo punto ci resta da verificare che le due successioni sono "contigue", ovvero che la differenza $x_n - y_n$ diventa arbitrariamente piccola al crescere di n.

Consideriamo la differenza tra

$$x_{n+1} = \frac{x_n}{2} + \frac{1}{x_n} = \frac{x_n^2 + 2}{2x_n^2}$$

e

$$y_{n+1} = \frac{2}{x_{n+1}} = \frac{4x_n}{x_n^2 + 2}\,.$$

Abbiamo quindi

$$x_{n+1} - y_{n+1} = \frac{(x_n^2+2)^2 - 8x_n^2}{2x_n(x_n^2+2)} = \frac{(x_n^2-2)^2}{2x_n(x_n^2+2)}\,.$$

Ma $x_n - y_n = x_n - \frac{2}{x_n} = \frac{x_n^2 - 2}{x_n}$ quindi possiamo riscrivere la differenza

$$x_{n+1} - y_{n+1} = \frac{x_n}{2}\frac{1}{x_n^2+2}(x_n - y_n)^2\,.$$

Osserviamo infine che $\frac{x_n}{2} \leq 1$ e $\frac{1}{x_n^2+2} < \frac{1}{4}$, e quindi

$$x_{n+1} - y_{n+1} < \frac{1}{4}(x_n - y_n)^2\,. \tag{1.34}$$

Poiché avevamo $x_0 - y_0 = 1$, applicando la (1.34) per ricorrenza otteniamo

$$x_n - y_n < \left(\frac{1}{4}\right)^{2^n - 1}, \tag{1.35}$$

che è un decrescita estremamente veloce. Per esempio dopo 4 iterazioni, la differenza tra la maggiorazione x_4 e la minorazione y_4 è minore di $0,25^{15} < 10^{-9}$, ovvero ci sono almeno nove cifre decimali coincidenti.

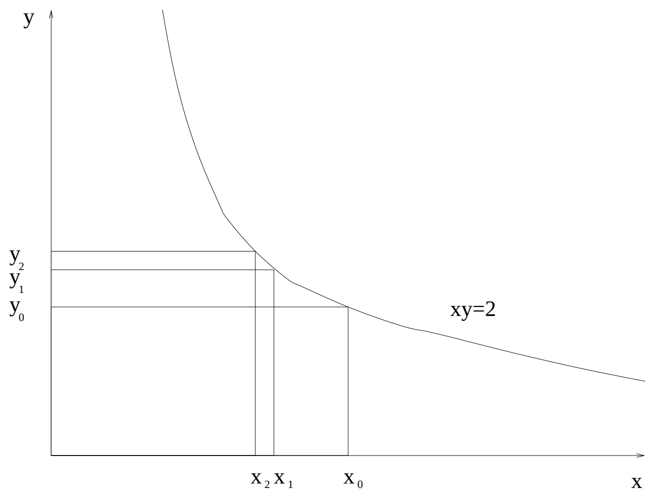

Figura 1.7: Il metodo di Erone

Il metodo di Erone e il metodo di Newton

Il metodo che abbiamo appena visto può essere usato per calcolare la radice di un numero A qualsiasi, semplicemente ponendo $y_n = A/x_n$ e mantenendo invariato $x_{n+1} = \frac{x_n - y_n}{2}$.

Ma questo metodo è comunque un metodo ad hoc per il calcolo delle radici e non si generalizza con successo al calcolo di soluzioni di altre equazioni, p.e. al calcolo di una radice cubica.

Se però guardiamo con un altra pospettiva "geometrica" al metodo di Erone ci accorgiamo che le approssimazioni x_n e y_n possono pensarsi ottenute da un altro procedimento.

Riscriviamo la formula di ricorrenza $x_{n+1} = \frac{x_n}{2} + \frac{1}{x_n}$ come

$$x_{n+1} = x_n - \frac{x_n^2 - 2}{2x_n}. \tag{1.36}$$

Ora possiamo reinterpretare la formula (1.36) osservando che il numeratore della frazione è il valore in x_n della funzione $f(x) = x^2 - 2$ di cui il numero $\sqrt{2}$ è la "radice", ovvero la soluzione dell'equazione $f(x) = 0$, mentre il denominatore è il valore della derivata $f'(x)$ in x_n. Quindi possiamo riscrivere la (1.36) nella forma

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \tag{1.37}$$

La (1.37) può interpretarsi geometricamente osservando che

$$y - f(x_n) = f'(x_n)(x - x_n) \tag{1.38}$$

è l'equazione della retta tangente nel punto di ascissa x_n al grafico della funzione $f(x)$, e quindi x_{n+1} è l'ascissa del punto in cui la retta tangente incontra l'asse delle ascisse.

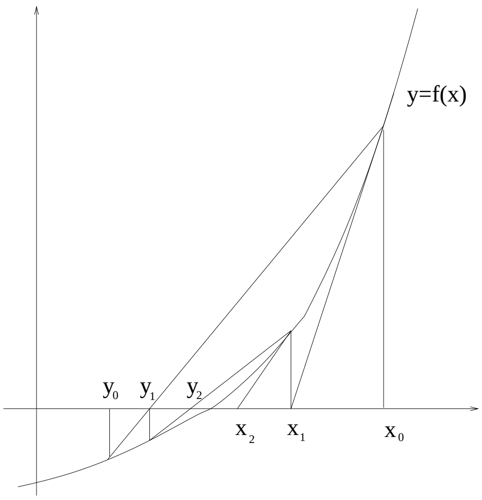

Figura 1.8: Illustrazione grafica del metodo di Newton

Analogamente possiamo trasformare la formula che ci dà la y_n

$$y_{n+1} = \frac{4x_n}{x_n^2 + 2} = \frac{2x_n + \frac{4}{x_n} + 2x_n - \frac{4}{x_n}}{x_n\left(x_n + \frac{2}{x_n}\right)}$$

che a sua volta, ricordando che $y_n = frac2x_n$, si può riscrivere

$$y_{n+1} = y_n - \frac{y_n^2 - 2}{x_n + y_n} = y_n - (y_n^2 - 2)\frac{x_n - y_n}{x_n^2 - y_n^2}$$

che possiamo riscrivere introducendo il simbolo $f(x)$ al posto di $x^2 - 2$

$$y_{n+1} = y_n - f(y_n)\frac{x_n - y_n}{f(x_n) - f(y_n)} \tag{1.39}$$

La (1.39) ha una forma molto simile alla (1.37): al posto della derivata di f compare il rapporto incrementale. La y_{n+1} rappresenta quindi l'ascissa (attenzione è un ascissa anche se si chiama "y") del punto di intersezione con l'asse x della corda (secante) del grafico $y = f(x)$ condotta dai punti di ascissa y_n e x_n.

Il metodo consistente nel cercare la radice dell'equazione $f(x)$ approssimandola successivamente con le soluzioni delle equazioni lineari ottenute cercando i punti di intersezione con l'asse delle ascisse delle tangenti e delle secanti è noto come metodo di Newton (che lo ha "inventato").

Abbiamo quindi visto che il medodo di Erone è equivalente al metodo di Newton applicato alla ricerca dell'equazione $x^2 - 2 = 0$.

1.5.4 I numeri reali non sono numerabili.

Dimostriamo ora che i numeri reali *non sono mumerabili*, ovvero che non possono essere posti in corrispondenza biunivoca con i numeri naturali.

La dimostrazione si fa per assurdo mostrando che l'assunzione che i numeri reali siano numerabili si ottiene una contraddizione.

La tecnica di dimostrazione fu proposta da Cantor nel 1891 (non era la sua prima dimostrazione della non numerabilità di $\mathbf{R}$ ma è quella che introduce una tecnica, detta *metodo di diagonalizzazione* che può essere utilizzata per altre dimostrazioni).

Supponiamo quindi che i numeri reali siano numerabili e sia data una loro enumerazione (ci limitiamo a considerare i numeri compresi tra 0 e 1)

$$\begin{aligned} r_1 &= 0, a_{1\,1}a_{1\,2}a_{1\,3}a_{1\,4}a_{1\,5}a_{1\,6}a_{1\,7}\dots \\ r_2 &= 0, a_{2\,1}a_{2\,2}a_{2\,3}a_{2\,4}a_{2\,5}a_{2\,6}a_{2\,7}\dots \\ r_3 &= 0, a_{3\,1}a_{3\,2}a_{3\,3}a_{3\,4}a_{3\,5}a_{3\,6}a_{3\,7}\dots \\ r_4 &= 0, a_{4\,1}a_{4\,2}a_{4\,3}a_{4\,4}a_{4\,5}a_{4\,6}a_{4\,7}\dots \end{aligned}$$

dove $a_{k\,h}$ è un numero tra 0 e 9, ed è la h-esima cifra decimale del k-esimo numero della lista.

Percorriamo ora la lista e costruiamo un numero

$$r_0 = 0, a_{0\,1}a_{0\,2}a_{0\,3}a_{0\,4}a_{0\,5}a_{0\,6}a_{0\,7}\dots$$

definito a partire da questa lista come segue: se $a_{k\,k}$, ovvero il termine di posto k nello sviluppo decimale del k-esimo elemento della lista, vale 0 allora poniamo uguale a 1 il k-esimo termine dello sviluppo di r_0, ovvero $a_{0\,k} = 1$. Se invece $a_{k\,k} \neq 0$, allora poniamo $a_{0\,k} = 0$.

Questo numero **non** può appartenere alla lista. Infatti se si avesse $r_0 = r_n$ per qualche n dovremmo avere $a_{0\,n} = a_{n\,n}$, ma $a_{0\,n} \neq a_{n\,n}$ per come abbiamo costruito r_0.

Ma questo contraddice la nostra assunzione di aver enumerato tutti i numeri reali (o meglio tutti quelli compresi tra 0 e 1).

1.5.5 Dimostrazione del teorema (1.4.3)

Come conseguenza del teorema (1.4.2) basta far vedere che una successione crescente, limitata superiormente, è una successione di Cauchy.

Ricordiamo che, per definizione di limitatezza, esiste (almeno) un numero M tale che $x_n \leq M$ per ogni n. Inoltre si ha che $x_{n+1} \geq x_n$ per ogni n, poiché la successione è crescente.

Prendiamo ora un numero $\varepsilon > 0$. Una delle due seguenti situazioni deve essere vera:

1. o esiste N_ε tale che $x_{N_\varepsilon} > M - \varepsilon$, e quindi, per l'ipotesi di crescenza della successione $x_k > M - \varepsilon$ per ogni $k \geq N_\varepsilon$;

2. oppure, per ogni n si ha $x_n \leq M - \varepsilon$.

Nel primo caso abbiamo verificato la condizione di Cauchy per il valore di ε, infatti per ogni $n, m > N_\varepsilon$ si ha che x_n e x_m appartengono entrambi all'intervallo $(M - \varepsilon, M)$ e quindi

$$|x_n - x_m| < \varepsilon \tag{1.40}$$

Nel secondo caso $M - \varepsilon$ è ancora un maggiorante della successione e possiamo ricominciare da capo sostituendo il valore iniziale di M con $M' = M - \varepsilon$. D'altra parte sappiamo che per i numeri (razionali) è vera la *proprietà archimedea* che dice che dati due numeri positivi a e b esiste sempre un multiplo intero di b, Kb con $K \in \mathbf{N}$, tale che $Kb > a$. Nel nostro caso sfruttiamo questa proprietà per dire che esiste sicuramente un intero K tale che $M - K\varepsilon < x_1$ (dove x_1 è il primo termine della successione). Questo ci garantisce che il processo di togliere progressivamente il numero ε a M non può generare una successione infinita di maggioranti. Quindi dovrà esistere un numero intero k tale che $M - k\varepsilon$ è un maggiorante ma $M - (k+1)\varepsilon$ no. Questo significa che esiste un x_N compreso tra questi due numeri, e di conseguenza, per la crescenza della successione, anche tutti gli x_n, con $n > N$, appartengono all'intervallo $(M - (k+1)\varepsilon\, M - k\varepsilon]$, da cui segue la (1.40).

Poiché ε è un numero arbitrario, la successione è di Cauchy, e quindi convergente.

Nota: forse qualcuno può stupirsi del fatto che in questa dimostrazione non si sia ricorsi al concetto di *estremo superiore* (ovvero il *più piccolo* dei maggioranti) ricordando una proprietà dei numeri reali che dice che *ogni insieme superiormente limitato ammette estremo superiore*. Il fatto è che questa proprietà (assunta come *assioma* nella definizione assiomatica dei numeri reali) deve essere dimostrata quando per definizione di numeri reali si sia preso l'insieme delle successioni di Cauchy. Dimostare questa proprietà è equivalente alla dimostrazione del teorema (1.4.3) (provare a fare la dimostrazione della proprietà dell'estremo superiore e dell'equivalenza).

Capitolo 2

Le funzioni

Come per molti concetti "elementari", quando parliamo di *funzioni* abbiamo abbastanza chiaro in testa di cosa stiamo parlando ma ci troviamo abbastanza a disagio a darne una "definizione" chiara e "rigorosa".

Per prima cosa osserviamo che la parola *funzione* è generalmente usata nel senso ristretto di "funzione reale di variabile reale a valori reali", e a tale senso faremo riferimento anche se faremo un excursus attraverso funzioni tra insiemi qualsiasi.

Il libro di Courant e Robbins[1] inizia la discussione del concetto di funzione con alcuni esempi: il primo esempio è il polinomio $x^2 + 2x - 3$ e il calcolo dei suoi valori quando la "lettera" x viene sostituita da un numero (2 nell'esempio). Poi prosegue con un esempio di funzione definita sui numeri naturali, ovvero la funzione $\pi(n)$ che indica quanti sono i numeri primi minori del numero naturale n.

Poi passa a esempi tratti dalla geometria e dalla fisica per mostrare la pervasività delle funzioni nelle applicazioni della matematica.

Segue una lunga spiegazione di cosa si debba intendere per *variabile* e poi dice:

"Può accedere che a ogni valore di una variabile X sia associato un valore ben determinato di un'altra variabile U. Allora U dice *funzione* di X"

Questa "definizione" si applica chiaramente nel caso in cui U sia il valore del polinomio $X^2 + 2X - 3$. Resta da definire il come si faccia a stabilire quando *a ogni valore di una variabile X è associato un valore ben determinato di un'altra variabile U*. Non abbiamo dubbi se si dispone di una "esplicita" procedura di calcolo del valore U a pertire da X (che può anche essere solo approssimata nella pratica ma sorretta da una teoria che ci dice che in quel caso l'associazione è ben definita: p.e. se X è il valore dei radianti di un angolo e U il seno corrispondente).

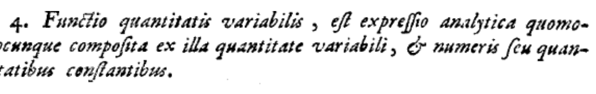
4. *Functio quantitatis variabilis, eſt expreſſio analytica quomodocunque compoſita ex illa quantitate variabili, & numeris ſeu quantitatibus conſtantibus.*

Figura 2.1: Definizione di funzione secondo Eulero

[1] R.Courant e H.Robbins, Che cos'è la matematica, ed. Boringhieri 1971

Questa era, più o meno, la visione classica della "definizione" di funzione resa espicita da Eulero nel libro *Introductio ad Analysin Infinitorum* del 1748: "Una funzione di una quantità variabile è un'espressione analitica in qualsiasi modo composta da quella quantità variabile, e da numeri o quantità costanti". Cosa intendeva Eulero per "espressione analitica"? Certamente queste comprendevano il catagolo delle funzioni "elementari" che ancora si insegnano nella scuola secondaria (polinomi, funzioni razionali, funzioni trigonometriche e funzioni esponenziali) e certamente le serie di potenze.

Questa visione mostrò i sui limiti quando si venne ad affrontare il problema della soluzione del moto di una corda vibrante. La soluzione appariva naturalmente come sovrapposizione di funzioni trigonometriche, e il problema, posto da D. Bernoulli, diventava se con somme (infinite) di tali funzioni fosse possibile rappresentare una "funzione qualsiasi" in modo analogo a quanto fatto con le serie di potenze.

Ne sorse una disputa piuttosto accanita: l'argomento principale degli oppositori della visione di Bernoulli era il fatto che "sommando funzioni periodiche si ottiene una funzione periodica" e non tutte le funzioni sono periodiche.

Questa polemica portò tuttavia a interrogarsi più a fondo sul cosa debba essere una funzione. Già sette anni dopo la pubblicazione della *Introductio ad Analysin Infinitorum*, Eulero tornava sulla definizione di funzione: "Quando certe quantità dipendono da un'altra in modo che esse subiscano dei cambiamenti quando quest'ultima cambi, allora le prime sono dette funzioni della seconda. Questo nome ha un carattere molto esteso, esso racchiude tutti i modi in cui una quantità può erre determinata usandone altre." Cosa realmente intedesse Eulero con quest'ultima frase è un mistero sepolto con lui. Tuttavia questa "definizione" riappare alla fine del secolo XVIII nel trattato di Lacroix sul calcolo differenziale: "Ogni quantità che dipende da una, o più, altre quantità è detta funzione di queste, sia che si conoscano o non si conoscano quali operazioni si debbano eseguire per ottenere la prima dalle seconde."

Questa definizione è fondamentalmente quella che viene spesso ancora insegnata. Per esempio in un libro di Aleksandrov del 1967 si può leggere una definizione di questo stile: "Se in una qualche maniera un elemento y di un insieme Y è fatto corrispondere a ogni elemento x di un insieme X, diciamo che c'è una applicazione dell'insieme X nell'insieme Y, o una funzione f il cui argomento varia nell'insieme X e il cui valore appartiene all'insieme Y."

Ma se si guarda con un po' di acribia una tale definizione ci rende conto che la "definizione" di funzione (o "applicazione") riposa sul verbo "corrispondere" il cui significato sembra demandato al "senso comune" e non a una "definizione" rigorosa. Altri esempi possono essere facilmente trovati dove invece di "corrispondenza" si usano i termini "legge" o "regola" o "operazione" senza che questi vengano espicitamente definiti. E' questa, per altro, una situazione tipica quando si debbano definire concetti "primitivi": le "definizioni" prendono la forma di circoli viziosi di sinomini che rinviano l'uno all'altro e la cui "giustificazione" è basata fondamentalmente sull'uso e il senso comune. Le *assiomatizzazioni* permettono di uscire da questi circoli viziosi ma sono del tutto incapaci di veicolare il "significato" del concetto assiomatizzato (penso che nessuno sia in grado di "capire" cosa sia un numero naturale a partire dagli assiomi di Peano, al contrario questi assiomi sono comprensibili alla luce dalla concettualizzazione informale che abbiamo dei numeri).

Nel caso delle funzioni tuttavia si può fare un "passo avanti" nel senso di *diminuire il numero di ingredienti base* della definizione e scaricare tutti i termini indefiniti sulla teoria degli insiemi[2].

Siano quindi X e Y due insiemi. Definiamo il *prodotto cartesiano* dei due insiemi, che indicheremo con $X \times Y$ come l'insieme di tutte le coppie ordinate (x, y) dove $x \in X$ e $y \in Y$. Definiamo poi una *relazione* come un sottoinsieme Γ del prodotto cartesiano $X \times Y$ e diremo che due elementi $x \in X$ e $y \in Y$ sono in relazione (e indicheremo questo scrivendo $R(x, y)$) se l'elemento $z = (x, y)$ appartiene al sottoinsieme Γ . Γ sarà anche detto *grafico della relazione R*.

Ora siamo in grado di dire quando una relazione è una funzione: basta attribuire al grafico Γ le propritá di cui gode il grafico di una funzione (secondo la nostra "definizione" informale).

Una relazione si dice *funzionale* se per ogni x in X esiste uno e un sol y in Y tale che $(x, y) \in \Gamma$ (oppure, equivalentemente, tale che vale $R(x, y)$).

Nota 2.0.1 *Possiamo anche divertirci a fare qualche variazione sul tema: p.e. possiamo definire per prima cosa la "proiezione" di $X \times Y$ su X e su Y rispettivamente tramite $p_X(x, y) = x$ e $p_Y(x, y) = y$. Allora la richiesta di "funzionalità" di una relazione la possiamo esprimere tramite le due condizioni:*

$$p_X(\Gamma) = X\,, \quad \text{e, se } z_1, z_2 \in \Gamma \text{ e } p_X(z_1) = p_X(z_2)\,, \text{ allora } p_Y(z_1) = p_Y(z_2)\,.$$

Non è inopportuna a questo punto una piccola osservazione. Nella nostra ansia di purezza definitoria abbiamo incluso nella definizione di funzione il fatto che essa deve assumere valori per tutti gli elementi dell'insieme X (quello che chiamiamo il *dominio* della funzione). E si insiste in molti testi anche sul fatto che l'assegnazione del dominio è parte integrante della definizione di una specifica funzione (ovvero la funzione definita su $\{x \in \mathbf{R} \mid x > 0\}$ da $f(x) = x^2$ è da considerarsi una funzione diversa dalla funzione definita da $f(x) = x^2$ *per ogni* numero reale)[3].

Poi si passa alla "pratica", ovvero alla parte dei programmi scolastici dedicata allo studio dei grafici di funzione, e la prima domanda si presenta immancabilmente sotto la forma: data la funzione $f(x) = \ldots$, determinarne il dominio.

A rigore una tale richiesta è mal posta: l'espressione $f(x) = \ldots$ non basta a definire di quale funzione si tratta! Si potrebbe pensare a uno studente che risponde che, non essendo completa la definizione della funzione in esame, possiamo completarne la definizione ponendo la funzione uguale al valore di $f(x)$ là dove l'espressione è calcolabile e a *ERRORE* là dove non lo è. Così avremo una funzione definita su tutto $\mathbf{R}$.

Questo modo di vedere la cosa non è poi semplicemente "provocatorio". E' così che risponde una calcolatrice elettronica quando le facciamo calcolare il valore di una funzione!

Possiamo descrivere questa situazione usando la metafora della "scatola nera", che si usa nella Teoria dei Sistemi. La funzione è una scatola nera che rende un output in risposta a un input: premete **1** (l'input) sulla vostra macchinetta,

[2] Questo approccio risale alla scuola bourbakista.

[3] Per aumentare la confusione c'è poi la teoria della funzioni di variabile complessa dove con la parola "funzione" si denotano anche funzioni "a più valori" come il logaritmo complesso definito, per $z = r\,e^{i\theta}$, da $\log(z) = \log(r) + i(\theta + 2k\pi)$ dove k è un intero *qualsiasi*

poi premete il tasto **sin** (la scatola nera) e ottene una risposta sul display (output) tipo **0,01745....** Fate lo stesso esperimento premendo prima **-1** poi il tasto log e otterrete l'output "valore non valido" o "**errore**".

Insomma l'argomento è delicato e richiede prudenza e "moderazione".

Come ultimo contributo vi invito a leggere il paragrafo dedicato alle funzioni nell'Introduzione del libro[4] di Alonzo Church (uno dei grandi logici del Novecento) che inizia con una definizione informale di funzione: "By a *function* - on, more explicitly, a *one-valued singulary* function - we shall undestand an operation which, when applied to something as *argument*, yields a certain thing as the *value* of the function *for* that argument. It is not required that the function be applicable to every possible thing as argument, but rather it lies in the nature of any given function to be applicable to certain things and, when applied to one of them as argument, to yield a certain value. The things to which the function is applicable constitute the *range of* the function (or the *range of arguments of* the function) and values constitute the *range of values of* the function. The function itself consists in the yielding or determination of a value from each argument in the range of the function."

Tutto questo è commentato in una lunga nota a piè di pagina: "Of course the words "operation," "yelding," "determination" as here used are a near-synonyms of "function" and therefore our statement, if taken as a definition, would be open to the suspicion of circularity. Throughout this Indroduction, however, we are engaged in informal explanation rather than definition, and, for this purpose, elaboration by means of synonyms may be a useful procedure. Ultimately, it seems, we must take the notion of function as primitive or undefined, or else some related notion as that of a class."

[4] A. Church, *Introduction to Mathematical Logic*, Princeton U.P., 1951

Capitolo 3

Il problema della tangente e la derivata

L'introduzione della derivata è collegata a due problemi "classici" . Da un lato, spinti dalle applicazioni del calcolo alla fisica, si ha la necessità di introdurre il concetto di velocità istantanea e di accelerazione (anch'essa "istantanea"), che sono gli oggetti matematici che compaiono nerlla meccanica classica sviluppata da Newton. Storicamente è proprio nell'opera di Newton che si trova la prima formulazione del calcolo moderno (mettiamo da parte la questione della "priorità" Newton-Leibniz sull'invenzioone del calcolo differenziale).

Ma ancora prima i matematici del diciassettesimo secolo avevano ripreso dalla matematiaca greca il problema delle tangenti, ovvero il problema che possiamo enunciare come: data una curva determinare la retta tangente ad essa in un suo punto qualsiasi.

Rispetto alla tradizione greca il problema sin era trasformato grazie dall'introduzione della geometria analitica di Cartesio. Ora metodi puramente geometrici e metodi algebrici potevano entrambi concorrere alla risoluzione del problema. Anzi il problema si poteva riformualare in termini puramente algebrici della determinazione dei coefficienti m e y_0 dell'equazione

$$y = mx + y_0 \tag{3.1}$$

della retta tangente cercata.

Ma ancor prima di "risolvere" il problema, si trattava molto spesso di definirlo con precisione.

Se abbiamo una circonferenza è chiaro cosa si debba intendere per retta tangente in un suo punto: è la retta che passa per quel punto e perpendicolare al diametro per il punto stesso. Questa definizione inoltre si presta a una costruzione geometrica "classica" (con riga e compasso).

Analogamente si può definire per via "sintetica" la retta tangente a una conica generica e tracciarla con strumenti geometrici. Si ha infatti che normale e tangente ad un ellisse sono rispettivamente la bisettrice e la perpendicolare alla bisettrice dell'angolo formato dalle congiungenti del punto con i fuochi (da cui le proprietà ottico-acustiche dell'ellisse) e lo stesso per la parabola a patto di spostare uno dei fuochi all'infinito. Per l'iperbole la costruzione è la stessa ma i ruoli della tangente e della normale si scambiano.

Ma cosa si deve intendere per tangente a una curva definita, p.e., da un polinomio di grado superiore al secondo o da una funzione non polinomiale come il logaritmo?

Per cercare una risposta torniamo indietro al problema della tangente a una conica e formuliamolo in termini algebrici. Data una conica di equazione

$$ax^2 + 2bxy + cy^2 + dx + ey + f = 0 \tag{3.2}$$

sappiamo che le ordinate dei punti di intersezione della conica con una retta generica sono date dalle soluzioni del sistema composto dalle equazioni (3.1) e (3.2).

Anche in questo è facile definire la tangente come quella retta per cui le intersezioni con la conica sono coincidenti, ovvero, in termini puramente algebrici, il sistema ammette una sola soluzione di molteplicità due. Questo permette di calcolare facilmente la retta tangente con metodi algebrici elementari (sono i tipici esercizi di "geometria analitica" liceale) ma ha il grave difetto di meccanizzare il procedimento con grave pregiudizio della comprensione di cosa si sta facendo (la tipica "lezione" che ne ricava lo studente è: *la retta tangente è quando il delta è uguale a zero*).

Prendiamo per semplicità la parabola di equazione

$$y = x^2 \tag{3.3}$$

e un punto su di essa di coordinate x_0 e $y_0 = x_0^2$. Tutte le rette del fascio

$$y = m(x - x_0) + y_0 \tag{3.4}$$

hanno ovviamente in comune con la parabola il punto $(x_0, y_0)2$, mentre il secondo punto è determinato dalla seconda soluzione del sistema (3.3), (3.4).

Riscriviamo ora l'equazione (3.3) *centrando il sistema di coordinate nel punto* $(x_0, y_0)2$

$$y - y_0 = (x - x_0)^2 + 2x_0(x - x_0) \tag{3.5}$$

(ricordarsi che $y_0 = x_0^2$). Riordinando i termini possiamo scrivere

$$y = x_0^2 + 2x_0(x - x_0) + (x - x_0)^2 \tag{3.6}$$

Se "buttiamo via" il termine $(x - x_0)^2$ abbiamo *l'equazione di una retta del fascio centrato in* (x_0, y_0), ovvero

$$y = x_0^2 + 2x_0(x - x_0) \tag{3.7}$$

E' facile verificare che è proprio la retta che incontra la parabola in due punti coincidenti, ovvero la *tangente* alla parabola nel punto (x_0, y_0).

Quello che però è importante notare sono i singoli termini che compaiono a secondo membro nella (3.6). Il primo, x_0^2, è *il valore di* x^2 *nel punto di ascissa* x_0. Il secondo è il prodotto di una costante, $2x_0$, per la differenza tra l'ascissa generica x e l'ascissa x_0 del punto da cui passa la retta. E' proprio il valore della costante che seleziona, tra tutte le rette del fascio, quella tangente alla parabola. Infine l'ultimo termine, $(x - x_0)^2$, è caratterizzato dal fatto di "essere trascurabile", in confronto ai primi due termini, quando x si avvicina a x_0. In maniera più formale, se prendiamo l'ultimo termine e lo dividiamo per $(x - x_0)$ otteniamo

$$\frac{(x - x_0)^2}{x - x_0} = x - x_0$$

che va a zero quando x va a x_0.

In altre parole, la "differenza" tra i valori della nostra funzione $y = x^2$ e la retta $y = x_0^2 + 2x_0(x - x_0)$ è una quantità "più piccola" (quando siamo molto vicini a x_0) della differenza tra le ascisse. Questo è vero solo per questa retta del fascio. Se prendiamo una retta $y = m(x - x_0) + y_0$ con $m \neq 2x_0$, la differenza sarà data $(2x_0 - m)(x - x_0) + (x - x_0)^2$ che è paragonabile a $x - x_0$ (ovvero se dividiamo questa differenza per $x - x_0$ otteniamo qualcosa che è, circa, una costante diversa da zero).

Se ripetiamo i conti fatti, in particolare la riscrittura in termini della differenza $x - x_0$, per la funzione $y = x^3$, otteniamo

$$y = x_0^3 + 3x_0^2(x - x_0) + 6x_0(x - x_0)^2 + (x - x_0)^3 \tag{3.8}$$

dove abbiamo di nuovo una prima parte

$$y = x_0^3 + 3x_0^2(x - x_0) \tag{3.9}$$

che è una retta del fascio per il punto (x_0, x_0^3) e un "resto" $6x_0(x-x_0)^2+(x-x_0)^3$, "più piccolo" della diffrenza $x - x_0$.

La discussione precedente ci suggerisce come definire la retta tangente al grafico di una funzione $y = f(x)$. Proviamo a scrivere la funzione "vicino" a un punto x_0 come

$$f(x) = f(x_0) + A(x - x_0) + \omega(x - x_0)\,. \tag{3.10}$$

Ovviamente questa "scomposizione"di $f(x)$ è sempre possibile qualsiasi sia la scelta della costante A, basta porre $\omega(x - x_0) = f(x) - f(x_0) - A(x - x_0)$. Ma come abbiamo già visto, il "resto" $\omega(x)$ is comporta in modo che dipende essenzialemnet dalla scelta di A.

Definizione 3.0.2 *Una funzione f si dice derivabile in un punto x_0 se esiste una costante A tale che il resto $\omega(x - x_0)$ nella (3.10) soddisfa la condizione*

$$\lim_{x \to x_0} \frac{\omega(x - x_0)}{x - x_0} = 0\,. \tag{3.11}$$

Il valore di tale costante si indica con $f'(x_0)$ e si dice la derivata *della funzione f nel punto x_0.*

La retta di equazione

$$y = f(x_0) + f'(x_0)(x - x_0) \tag{3.12}$$

è, per definizione, la *retta tangente* al grafico $y = f(x)$ nel punto di coordinate $(x_0, f(x_0)$.

Nota 3.0.3 *La definizione è (3.0.2) equivalente alla più consueta definizione data attraverso il limite del rapporto incrementale (dimostrarlo!). Questa definizione ha comunque due vantaggi. Da un lato si generalizza immediatamente a funzioni* *da* $\mathbf{R}^n$ *in* $\mathbf{R}$*, e dà luogo alla definizione di differenziale (o di gradiente). Da un altro punto di vista avremmo potuto evitare l'uso del limite nella definizione* maggiorando esplicitamente il resto ω. *Possiamo, p.e., dire che una funzione è derivabile se troviamo una costante* C *e un numero* δ *tali che*

$$|\omega(x - x_0)| \leq C(x - x_0)^2 \tag{3.13}$$

per tutti gli $x \in (x_0 - \delta, x_0 + \delta)$ *(ovvero "vicino a* x_0*"). Con un tale modifica della definizione, non tutte le funzioni derivabili nel senso della (3.0.2) risultano derivabili, in quanto la maggiorazione (3.13) è una condizione più forte della (3.11). Restano tuttavia "derivabili" la stragrande maggioranza delle funzioni che si presentano agli studenti nella scuola superiore. In altri termini è possibile dare una definizione (coerente) di derivata senza usare la definizione di limite.*

Esercizio Provare che la funzione $\exp(x)$ è "derivabile" anche nel senso che soddisfa la maggiorazione (3.13).
Far vedere che la funzione $x^{3/2}$ è derivabile in $x_0 = 0$, ma non soddisfa la maggiorazione (3.13).

Esercizio Dimostrare le proprietà delle derivate usando come definizione di funzione derivabile (e di derivata): f è derivabile nell'intervallo (a, b) se per ogni $x \in (a, b)$ esiste un numero $f'(x)$ tale che la maggiorazione (3.13) è soddisfatta con $C = f'(x)$, in un intorno di x.

Capitolo 4

Il problema della quadratura e l'integrale

La nozione di integrale sorge da due problemi apparentemente scollegati tra loro.

Da un lato il problema delle quadrature, ovvero della determinazione delle aree racchiuse da delle curve piane: la "risposta" a questo problema è quello che chiamiamo *integrale definito*.

Ma l'integrale viene anche introdotto come *l'operazione inversa* della derivazione, ovvero come quella operazione che, data una funzione, ne trova un'altra, detta *primitiva*, la cui derivata è la funzione di partenza: questa operazione è nota come *integrale indefinito*.

Il legame tra queste due operazioni è dato dal *Teorema fondamentale del calcolo integrale* che lega l'area di sottografico all'incremento della primitiva tra i due estremi di integrazione.

Nel seguito ci occuperemo prevalentemente dell'integrale definito e useremo il teorema fondamentale come legame con il problema delle primitive.

4.1 Casi elementari

Iniziamo da due casi "banali". Prendiamo per prima una funzione costante positiva. Per evitare inutili complicazioni, per il momento supponiamo che anche le ascisse a e b siano positive. In questo caso il grafico della funzione $f(x) = c$ è una retta parallela all'asse delle ascisse e il sottografico tra due estremi a e b è un rettangolo di base $b - a$ e altezza c. L'area del sottografico è quindi data dal prodotto $c\,(b - a)$.

Un po' meno elementare è il caso di una funzione lineare $f(x) = mx + c$. In questo caso il grafico è una retta e il sottografico sull'intervallo $[a, b]$, supponendo che $f(x)$ sia positiva in $[a, b]$, è un trapezio di basi $f(b) = mb+c$ e $f(a) = ma+c$ e di altezza $b-a$ (nota che ora le "basi" sono prese nella direzione parallela all'asse delle y e l'altezza lungo l'asse delle x). L'area quindi è data da

$$\frac{1}{2}(b-a)(mb+c+ma+c) = \frac{1}{2}m(b^2-a^2) + c(b-a)\,. \tag{4.1}$$

Possiamo osservare una prima, fondamentale, proprietà: la funzione $f(x) =$

$mx+c$ si può scomporre nella somma delle due funzioni $f_1(x) = mx$ e $f_2(x) = c$. Nel termine $c(b-a)$ il (4.1) riconosciamo l'area del sottografico su $[a,b]$ della funzione costante $f_2(x)$, mentre il termine $\frac{1}{2}m(b^2 - a^2)$ è l'area del trapezio "sottografico" della funzione $f_1(x)$ in accordo con la (4.1) per $c = 0$. Quindi l'area del sottografico di $f(x)$ è la somma delle aree dei sottografici di $f_1(x)$ e di $f_2(x)$. Come vedremo questa è un caso particolare di una proprietà generale dell'integrale.

4.2 Verso il caso generale

Nei due esempi precedenti abbiamo calcolato le aree sfruttando il fatto che la figura geometrica definita dal sottografico era una figura geometrica elementare di cui si sa come calcolare l'area.

Veniamo ora a un caso più complesso. Prendiamo la funzione $f(x) = x^2$ e chiediamoci "quanto vale l'area" della regione di piano compresa tra l'asse delle ascisse, il grafico $y = x^2$ e le rette $x = a$ e $x = b$, prendendo ancora a e b positivi.

In questo caso non abbiamo informazioni da parte della geometria elementare. La figura formata da sottografico non è data, come nel caso precedente, da una somma di triangoli, e quindi non possiamo calcolare l'area in modo elementare.

Dobbiamo quindi *dare una procedura* che *definisca l'area* del sottografico, rifacendoci al calcolo di aree note, e a operazioni "algebriche" ben definite.

Iniziamo osservando che il sottografico è una regione di piano compresa all'interno del rettangolo che ha per base il segmento $[a,b]$ sull'asse delle ascisse e per altezza $f(b) = b^2$ (che il più grande valore assunto da f in questo intervallo). A sua volta il sottografico di f contiene il rettangolo con la stessa base e altezza $f(a) = a^2$ (che il più piccolo valore assunto da f in questo intervallo).

Procediamo ora *suddividendo l'intervallo* $[a,b]$ in un certo numero di sottointervalli, per esempio dividendo $[a,b]$ in N parti uguali. Avremo quindi N intevallini di estremi $\frac{b-a}{N}(k-1)+a$ e $\frac{b-a}{N}k+a$, con $k = 1, 2, \ldots, N$.

In ogni intervallo $\left(\frac{b-a}{N}(k-1)+a, \frac{b-a}{N}k+a\right)$ costruiamo i due rettangoli con base questo intervallo e altezze rispettivamente $f\left(\frac{b-a}{N}k+a\right)$ e $f\left(\frac{b-a}{N}(k-1)+a\right)$ che chiemeremo rispettivamente R_k e r_k. Il sottografico della funzione f è contenuto nell'unione dei rettangoli R_k, $k = 1, 2, \ldots, N$ e contiene l'unione dei rettangoli r_k, $k = 1, 2, \ldots, N$. Ne segue che l'area A del sottografico di f deve essere compresa tra i due numeri

$$s_N = \sum_{k=1}^{N} f\left(\frac{b-a}{N}(k-1)+a\right)\frac{b-a}{N} < A < S_N = \sum_{k=1}^{N} f\left(\frac{b-a}{N}k+a\right)\frac{b-a}{N} \tag{4.2}$$

dove s_N è l'area dell'unione dei rettangoli contenuti nel sottografico e S_N quella dell'unione dei rettangoli che lo contengono.

E' abbastanza facile mostrare che le due successioni numeriche $\{s_N\}$ e $\{S_N\}$ sono rispettivamente crescenti e decrescenti al crescere di N. Possiamo in effetti calcolare esplicitamente queste due somme. Per non appesenatire il calcolo, consideriamo il caso in cui $a = 0$.

Abbiamo

$$S_N = \sum_{k=1}^{N}\left(\frac{b}{N}k\right)^2\frac{b}{N} = \left(\frac{b}{N}\right)^3\sum_{k=1}^{N}k^2 \tag{4.3}$$

e

$$s_N = \sum_{k=1}^{N} \left(\frac{b}{N}(k-1) \right)^2 \frac{b}{N} = \left(\frac{b}{N} \right)^3 \sum_{k=1}^{N} (k-1)^2 . \tag{4.4}$$

Ricordiamo che

$$\sum_{k=1}^{N} k^2 = \frac{N(N+1)(2N+1)}{6}$$

e analogamente

$$\sum_{k=1}^{N} (k-1)^2 = \sum_{k=1}^{N-1} k^2 = \frac{(N-1)N(2N-1)}{6} ,$$

da cui otteniamo

$$s_n = \frac{b^3}{3} \frac{(N-1)N(2N-1)}{2N^3} < A < S_N = \frac{b^3}{3} \frac{N(N+1)(2N+1)}{2N^3} , \tag{4.5}$$

da cui, passando al limite in N,

$$A = \frac{b^3}{3} .$$

Nel caso in cui sia $a > 0$, otteniamo, con un po' più di pazienza,

$$A = \frac{b^3}{3} - \frac{a^3}{3} .$$

4.3 La definizione di integrale

Abbiamo visto nel caso del sottografico di una parabola come sia possibile determinare un valore che assegnamo all'area del sottografico stesso. In verità si tratta di una vera *definizione* di questa area[1].

Posiamo ora generalizzare tale definizione al caso di una funzione qualsiasi $f(x)$. Supponiamo ancora che la funzione f sia positiva nell'intervallo $[a, b]$.

Iniziamo con il suddividere l'intervallo $[a, b]$ in un certo numero N di sottointervalli $[x_{k-1}, x_k]$ scegliendo $N + 1$ punti $x_k,\ k = 0, \ldots, N$ con $x_0 = a$ e $x_N = b$ e tali che per ogni $k = 1, \ldots, N$ sia $x_{k-1} < x_k$. Chiameremo una tale scelta di intervalli una *partizione* di $[a, b]$ (nota che non abbiamo supposto che i punti x_k siano equidistribuiti nell'intervallo).

Definiamo ora la somma inferiore e la somma superiore relative alla partizione scelta ponendo

$$s_N = \sum_{k=1}^{N} \inf_{x_{k-1} \le x \le x_k} f(x)\,(x_k - x_{k-1}) \tag{4.6}$$

e

$$S_N = \sum_{k=1}^{N} \sup_{x_{k-1} \le x \le x_k} f(x)\,(x_k - x_{k-1}) \tag{4.7}$$

[1] Lo stesso vale per l'area del cerchio calcolata con il metodo di Archimede dei poligoni regolari inscritti e circoscritti.

Nota: avendo assunto f positiva in $[a, b]$, la somma s_N è certamente ben definita, non altrettanto si può dire di S_N. Tuttavia anche S_N risulta ben definita se supponiamo che f sia limitata in $[a, b]$, cosa che faremo in quello che segue.

I due numeri s_N e S_N soddisfano (banalmente) la diseguaglianza

$$s_n < S_N \,. \tag{4.8}$$

Cosa succede se "cambiamo partizione" e in particolare se la "infittiamo" cioè se aggiungiamo nuovi punti come estremi degli intervallini?

Una cosa che è facile capire è che l'aggiunta di punti alla partizione fa aumentare la somma inferiore e diminuire la somma superiore. Per esempio se abbiamo inizialmente diviso $[a, b]$ in due sole parti ($N = 2$) e poi suddividiamo i due sottointevalli $[a, x_1]$ e $[x_1, b]$ ancora in due ognuno, i due termini delle s_1 e S_1 sono ora rispettivamente maggiore e minore della somma dei primi due e del terzo e quarto termine della somma inferiore e superiore con la nuova suddivisione. Questo per la (naturale) proprietà degli estremi inferiore e superiore che, se l'intervallo A è contenuto nell'intervallo B, allora $\inf_{x\in A} f(x) \geq \inf_{x\in B} f(x)$ e $\sup_{x\in A} f(x) \leq \sup_{x\in B} f(x)$.

Con questo trucco di "infittire la partizione" si può mostrare una proprietà meno intuitiva, ovvero che qualsiasi sia la partizione che si utilizza per costrire la somma inferiore e qualsiasi sia quella (in generale diversa) che si usa per costuire la somma superiore queste sono sempre ordinate, ovvero la somma inferiore è minore della somma superiore[2].

Ora abbiamo di nuovo un caso di due insiemi di numeri $\sigma =\{$tutte le possibili somme inferiori$\}$ e $\Sigma =\{$tutte le possibili somme superiori$\}$ per cui vale che qualunque siano $s \in \sigma$ e $S \in \Sigma$ si ha $s \leq S$. Se i due insiemi sono *contigui*[3] allora la funzione f si dirà *integrabile* (secondo Riemann) nell'intervallo $[a, b]$ e l'elemento separatore (un numero) si dirà l'integrale di f su $[a, b]$.

Non esiste una caratterizzazione alternativa delle funzioni integrabili. Ci sono però dei teoremi che danno delle condizioni sufficienti per l'integrabilità di una funzione. Il più famoso è quello che assicura che *ogni funzione continua su un intevallo chiuso e limitato è integrabile secondo Riemann in questo intervallo.* Purtroppo la dimostrazione di questo teorema non è facile[4] e non può essere presentata a degli studenti liceali. Tuttavia per una classe più ristretta di funzioni continue, ovvero le funzioni lipschitziane, la dimostrazione dell'integrabilità è abbastanza semplice da poter essere fatta anche alla scuola secondaria.

Ricordiamo che una funzione si dice lipschitziana in un intervallo $[a, b]$ se esiste una costante C tale che per ogni coppia di numeri $x, y \in [a, b]$ si ha

$$|f(x) - f(y)| \leq C|x - y| \,.$$

Ricordiamo in particolare che sono lipschitziane tutte le funzioni derivabili con

[2] La dimostrazione non è difficile e si trova in qualsiasi testo universitario di Analisi Matematica. Essa può tuttavia sembrare un mero "giochetto" agli studenti liceali, per cui forse è consigliabile non avventurarsi nella definizione di integrale che qui presentiamo e limitarsi a prendere delle partizioni uniformi e "inscatolate" p.e. scegliendo $N = 2^M$ in modo da passare da una partizione all'altra dimezzando gli intervalli già presenti.

[3] Ricorda che *contigui* significa che per ogni $\varepsilon > 0$ esistono due elementi s_ε e S_ε rispettivamente nella prima e nella seconda classe tali che $|S_\varepsilon - s_\varepsilon| < \varepsilon$.

[4] Si deve far uso della proprietà di *equicontinuità* della funzione, a sua volta garantita dal teorema di Cantor che dice ogni funzione continua su un intevallo chiuso e limitato è ivi equicontinua.

derivata limitata. Inoltre ogni funzione lipschitziana in un intevallo limnitato è una funzione continua e limitata (dimostarlo).

Vediamo ora di dimostrare che una tale funzione è integrabile. Per far questo basta mostare che le classi delle somme inferiori e superiori sono contigue, ovvero che, qualunque sia $K \in \mathbf{N}$ posso trovare una partizione (p.e. uniforme) tale che $S_N - s_N < 1/K$. Scriviamo la differenza $S_N - s_N$

$$S_N - s_N = \sum_{k=1}^{N} \left(\sup_{x_{k-1} \leq x \leq x_k} f(x) - \inf_{x_{k-1} \leq y \leq x_k} f(y) \right) (x_k - x_{k-1}) \tag{4.9}$$

dove abbiamo usato la lettera y nell'espressione dell'estremo inferiore per mettere in evidenza che i punti in cui si assumono i valori estremi sono diversi.

La condizione di Lipschitz tuttavia ci assicura che per tutte le coppie di punti x, y nell'intervallo $[x_k, x_{k-1}]$ si ha

$$|f(x) - f(y)| \leq C|x - y| \leq C|x_k - x_{k-1}|\,. \tag{4.10}$$

da cui segue che

$$\sup_{x_{k-1} \leq x \leq x_k} f(x) - \inf_{x_{k-1} \leq y \leq x_k} f(y) \leq C|x_k - x_{k-1}|\,. \tag{4.11}$$

Possiamo ora usare la (4.11) per maggiorare la differenza in (4.9).

$$S_N - s_N \leq C \sum_{k=1}^{N} (x_k - x_{k-1})\,(x_k - x_{k-1})\,. \tag{4.12}$$

Se prendiamo la partizione uniforme avremo

$$S_N - s_N \leq C \frac{b-a}{N} \sum_{k=1}^{N} (x_k - x_{k-1}) = C \frac{(b-a)^2}{N}\,, \tag{4.13}$$

che è un numero arbitrariamente piccolo se N è sufficientemente grande.

D'altra parte è facile costruire un esempio di funzione integrabile ma non continua, p.e. la funzione uguale a 0 in $[0, 1]$ e uguale a 1 in $[1, 2]$. Possiamo poi costruire esempi analoghi con un numero sempre più grande di discontinuità. Se poi ci si chiede "quanto discontinua" debba essere una funzione per non essere integrabile, la risposta è però assai difficile (vedi Hans Rademacher, On the condition of Riemann integrability, Amer. Math. Monthly 61, (1954). 1–8, dove si dà la condizione necessaria e sufficiente perchè una funzione da $\mathbf{R}$ in $\mathbf{R}$ sia integrabile.)

L'esempio classico di funzione non integrabile secondo Riemann è la funzione di Dirichlet che è definita uguale a 0 su i numeri *irrazionali* e uguale a 1 sui numeri *razionali.* Infatti in ogni intervallino, per quanto piccolo, ci sono sia numeri irrazionali che razionali, quindi avremo sempre l'estremo inferiore uguale a zero e quello superiore uguale a uno, e quindi le somme superiori sono tutte uguali alla lunghezza dell'intervallo mentre quelle inferiori sono sempre nulle.

Un esempio "mostruoso" di funzione integrabile è invece dato dalla funzione definita da

$$f(x) = \begin{cases} \frac{1}{n} & \text{se } x = \frac{m}{n}\,, \quad m,\, n \text{ primi tra loro} \\ 0 & \text{altrimenti} \end{cases}$$

Vediamo coma si dimostra che f è integrabile in $[0,1]$. Anche in questo caso non è difficile capire che le somme inferiori sono tutte nulle (in ogni intervallo c'è almeno un irrazionale). Resta da far vedere che l'estremo inferiore delle somme superiori è ancora zero. Dobbiamo quindi far vedere che per ogni $\varepsilon > 0$ possiamo costruire una partizione la cui somma superiore S sia minore di ε. Prendiamo $A_N = \{x \in [0,1] \,|\, x = \frac{m}{n} \text{ con } n < N\}$. In questo sottoinsieme la funzione assume valori $\frac{1}{n}$ con $n = 1, 2, 3, \ldots, N-1$, mentre nel complementare di A_N la funzione è maggiorata da $\frac{1}{N}$ (ovvero assume valori minori o uguali di $1/N$). Inoltre i punti di A_N sono al più $\sum_{k=1}^{N-1} k = \frac{N(N-1)}{2}$, poiché per ogni valore del denominatore n ci sono al più $n-1$ possibili numeratori. Attorno a ogni punto x di A_N prendiamo ora un intervallo $I_{1/N^3}(x)$, centrato in x e di ampiezza $\frac{1}{N^3}$. La partizione sarà fatta da l'unione I_N degli intervalli $I_{1/N^3}(x)$ per $x \in A_N$ e da quel che resta dell'intervallo ovvero $\overline{I}_N = [0,1] \backslash I_N$. Si osservi che $\overline{I}_N$ è contenuto nel complementare di A_N quindi abbiamo $f(x) \leq 1/N$ per ogni $x \in \overline{I}_N$.

Maggioriamo ora la somma parziale relativa e questa partizione: il contributo del sottografico sull'insieme $\overline{I}_N$ è sicuramente minore di $1/N$ ($1/N$ maggiora la funzione e 1, lunghezza dell'intervallo $[0,1]$, maggiora la somma delle lunghezze degli intervalli che compongono $\overline{I}_N$). Il contributo del sottografico sull'insieme I_N è invece maggiorato dalla somma dei sottografici sui singoli intevalli $I_{1/N^3}(x)$ (che possono parzialmente sovrapporsi) e questa, a sua volta, è maggiorata da $\frac{1}{N^3}\frac{N(N-1)}{2}$ (1 maggiora la funzione $f(x)$, $\frac{1}{N^3}$ è la lunghezza del singolo intervallo, $\frac{N(N-1)}{2}$ maggiora il numero degli intervalli). Al totale abbiamo

$$S < \frac{1}{N} + \frac{1}{N^3}\frac{N(N-1)}{2}$$

che tende a zero quando N va all'infinito.

4.4 Legame tra integrale e derivata: il teorema fondamentale

Torniamo ora a punto di vista che abbiamo menzionato all'iniio di questo capitolo sull'integrale (indefinito) come "operazione inversa" della derivazione. Poiché abbiamo chiamato con la stessa parola due cose apparentemente diverse dobbiamo in qualche modo giustificare questa scelta di nome trovando un qualche legame tra l'integrale definito e la derivata.

Questo legame è fornito dal Teorema Fondamentale del Calcolo Integrale:

Teorema 4.4.1 *Sia f una funzione continua su $[a,b]$, Allora per ogni $x \in (a,b)$ si ha*

$$\frac{\mathbf{d}}{\mathbf{d}x}\int_a^x f(s)\,ds = f(x)\,.$$

o nella versione più popolare nei testi scolastici

Teorema 4.4.2 *Sia f una funzione continua su $[a,b]$, e F una sua primitiva (ovvero $\frac{\mathbf{d}F}{\mathbf{d}x} = f$). Allora per ogni $x \in (a,b)$ si ha*

$$\int_a^x f(s)\,ds = F(x) - F(a)\,.$$

In altri termini, la funzione $\int_a^x f(s)\,ds$ *è una primitiva di* f.

Nota: l'ipotesi di continuità di f nella seconda versione non è essenziale, e basta che f sia integrabile in ogni intervallo $[a, x]$ (si noti che nella seconda versione non compare la valutazione di f in un singolo punto, ma solo il suo integrale). A parte questa maggior generalità della seconda formulazione, è facile vedere che le due versioni sono "equivalenti".

Dimostriamo la prima versione del teorema.
Scriviamo il rapporto incrementale per la funzione $F(x) = \int_a^x f(s)\,ds$

$$\frac{F(x+h) - F(x)}{h} = \frac{1}{h}\int_x^{x+h} f(s)\,ds$$

Poiché f è una funzione continua, per ogni $\varepsilon > 0$ esiste un $\delta > 0$ tale che $|f(\xi) - f(x)| < \varepsilon$ se $|\xi| < \delta$. Se prendiamo il valore di h nel rapporto incrementale tale che $|h| < \delta$, allora avremo

$$[f(x) - \varepsilon] = \frac{1}{h}\int_x^{x+h} [f(x) - \varepsilon]\,ds < \frac{1}{h}\int_x^{x+h} f(s)\,ds$$

$$< \frac{1}{h}\int_x^{x+h} [f(x) + \varepsilon]\,ds = [f(x) + \varepsilon]$$

e il teorema segue dall'arbitrarietà di ε.