CAPITOLO 1

La retta reale

1. I numeri naturali. Gli interi relativi.

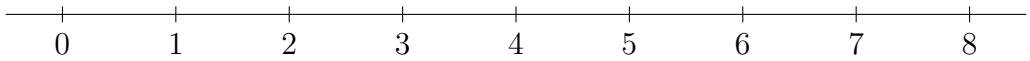
L'operazione di contare è una delle più naturali che esistano. Ognuno di noi, prima ancora di sapere che cosa vogliano dire "uno" e "due", ha già la capacità di distinguere tra (almeno) due persone differenti. Solo dopo un po' di tempo si rende conto che le persone sono in realtà "tre", ed inizia a contare sé stesso[(1)]. Per complicare un po' le cose, invece di contare da uno, decideremo di contare da zero[(2)].

Per meglio identificare i numeri che possiamo ottenere contando a partire da zero, definiamo l'insieme dei numeri **naturali**:

$$\mathbb{N} = \{0, 1, 2, 3, \ldots, n, \ldots\}\,. \tag{1.1}$$

In questo caso i secondi puntini di sospensione "..." indicano che siamo in grado di contare "fino a quando vogliamo"; in altre parole che l'insieme dei numeri naturali è "infinito", ovvero che non ammette "massimo". Non esiste, cioè, il più grande numero naturale: non appena pensiamo di averlo trovato, è sufficiente "aggiungere 1" per trovarne uno ancora più grande.

Se rappresentiamo i numeri naturali su una retta, scegliendo in maniera arbitraria la distanza tra 0 ed 1 (ovvero, l'unità di misura), avremo una immagine "discreta": per passare da un numero al successivo si deve compiere un passo lungo 1.



L'operazione di "aggiungere 1", restando ancora all'interno dello stesso insieme è una proprietà di $\mathbb{N}$, che può essere così generalizzata: dati due numeri naturali n ed m qualsiasi, la loro **somma** $n + m$ è ancora un numero naturale: quello ottenuto, appunto, partendo da n ed effettuando m passi verso destra. In linguaggio matematico, "l'insieme dei numeri naturali è chiuso rispetto alla somma". Siccome partire da n e fare m passi è la stessa cosa che partire da m e fare n passi, l'addizione è

[(1)]In alcune popolazioni aborigene dell'Amazzonia, il concetto di tre (ovvero, della paternità) non esiste.

[(2)]Storicamente, il concetto di zero compare molto dopo quello di uno, due e tre.

commutativa:

$$\forall n\,,\ m \in \mathbb{N}\,, \qquad n + m = m + n\,.$$

Inoltre, esiste un numero naturale "speciale", lo zero, che gode della proprietà di lasciare "immutato" qualsiasi numero rispetto alla somma (come è evidente: sommare zero vuol dire fare 0 passi, cioè non muoversi):

$$\forall n \in \mathbb{N}\,, \quad n + 0 = n\,.$$

Lo zero viene pertanto detto **elemento neutro** dell'operazione di somma. La chiusura rispetto alla somma permette di ordinare i numeri naturali: diremo che un numero naturale n è **maggiore** di m se esiste un numero naturale p, diverso da zero, tale che $n = m + p$. In simboli, scriveremo $n > m$. Data questa definizione, si può affermare che dati due numeri naturali **diversi** n ed m, si ha $n > m$, oppure $m > n$ (in definitiva, uno dei due sarà a destra dell'altro sulla retta). Ammettendo che il numero naturale p possa essere 0, viene definito il concetto di "maggiore od uguale"; in simboli $n \geq m$. La relazione di "$\geq$" è antisimmetrica: se $n \geq m$ e $m \geq n$, allora $n = m$[3].

Infine, dato un numero naturale n, e fissato un secondo numero naturale m che avrà la funzione di "contatore", è possibile sommare n a sé stesso ripetendo l'operazione m volte. Il risultato, ovvero "m volte n", verrà indicato con

$$m \cdot n = \overbrace{n + n + \cdots + n}^{m \text{ volte}}:$$

il **prodotto** di m con n. Il numero 1 è l'elemento neutro del prodotto, essendo $1 \cdot n = n$ per ogni n in $\mathbb{N}$. Le relazioni tra somma, prodotto e ordinamento sono quelle note:

$$\begin{aligned}
\forall n\,, m \in \mathbb{N}\,, &\qquad m \cdot n = n \cdot m\,,\\
\forall n\,, m\,, p \in \mathbb{N}\,, &\qquad p \cdot (n + m) = p \cdot n + p \cdot m\,,\\
\forall n\,, m\,, p \in \mathbb{N}\,, &\qquad n \geq m \Rightarrow p \cdot n \geq p \cdot m\,,\\
\forall n\,, m\,, p \in \mathbb{N}\,, &\qquad n > m\,,\ p \neq 0 \Rightarrow p \cdot n > p \cdot m\,.
\end{aligned}$$

Il prodotto di un numero naturale con sé stesso, ovvero $n \cdot n$, viene convenzionalmente indicato con n^2, il quadrato di n. Analogamente, il prodotto di n per sé stesso m volte viene indicato con n^m, la potenza m-sima di n. Ovviamente $n^1 = n$, mentre, per convenzione, $n^0 = 1$ per ogni n in $\mathbb{N}$. Per le potenze valgono le regole fondamentali:

$$\begin{aligned}
\forall n\,, m\,, p \in \mathbb{N}\,, &\qquad n^{m+p} = n^m \cdot n^p\,,\\
\forall n\,, m\,, p \in \mathbb{N}\,, &\qquad (n \cdot m)^p = n^p \cdot m^p\,,\\
\forall n\,, m\,, p \in \mathbb{N}\,, &\qquad (n^m)^p = n^{m \cdot p}\,.
\end{aligned}$$

---

[3] Esercizio!

Supponiamo ora di avere un sottoinsieme E di $\mathbb{N}$; ad esempio

$$E_1 = \{\text{numeri pari}\} = \{0, 2, 4, 6, 8, \ldots\},$$
$$E_2 = \{\text{numeri dispari}\} = \{1, 3, 5, 7, 9 \ldots\},$$
$$E_3 = \{\text{multipli non nulli di cinque}\} = \{5, 10, 15, 20, 25 \ldots\}.$$
$$E_4 = \{\text{potenze di due}\} = \{1, 2, 4, 8, 16, \ldots\}.$$
$$E_5 = \{\text{numeri primi}\} = \{2, 3, 5, 7, 11, \ldots\},$$
$$E_6 = \{\text{numeri di Fibonacci}\} = \{1, 1, 2, 3, 5, 8, 13, 21 \ldots\},$$

Ognuno di questi insiemi non ammette un numero più grande: dato un numero (pari o dispari) qualsiasi, è sufficiente aggiungere due per ottenere un numero più grande della stessa parità; sommando 5 ad un multiplo di 5 se ne ottiene uno più grande; raddoppiando una potenza di due se ne trova una più grande; dimostrare che esistono infiniti numeri primi (numeri, cioè, che non sono ottenibili come prodotto di naturali diversi da 1) è più complicato, ma è comunque vero; anche i numeri di Fibonacci (ognuno dei quali è somma dei due precedenti, a partire da 1, 1), perché è sufficiente sommare ad uno di essi quello che lo precede nella lista per ottenere un numero di Fibonacci maggiore. Invece, per tutti gli insiemi considerati in precedenza esiste un numero — appartenente all'insieme — più piccolo di tutti gli altri (rispettivamente, 0, 1, 5, 1, 2 e 1).

Definizione 1.1. Sia E un sottoinsieme non vuoto di $\mathbb{N}$. Diremo che E ammette **minimo** se esiste m in E tale che $n \geq m$ per ogni n in E. In altre parole, se esiste un elemento di E che è il più piccolo (il più a sinistra) di tutti. Se E ammette minimo, tale minimo è unico; se infatti ne esistessero due, si avrebbe $m_1 \geq m_2$ (perché m_2 è minimo) e $m_2 \geq m_1$ (perché m_1 è minimo); per antisimmetria, $m_1 = m_2$.

Si osservi che la richiesta di appartenenza ad E è fondamentale, dato che esiste sempre un numero naturale che è più piccolo di tutti gli elementi di E: lo zero. Inoltre, se il minimo non appartenesse ad E, l'unicità del minimo verrebbe meno: ad esempio l'insieme E_3 definito prima ammette 0, 1, 2, 3, 4 e 5 come numeri naturali "più piccoli" di tutti i numeri in E_3 (di questi, però, solo un appartiene ad E_3, ed è il maggiore di tutti).

Esercizio 1.2. Costruire un sottoinsieme non vuoto di $\mathbb{N}$ che non ammette minimo.

Svolto l'esercizio precedente? Se non lo avete fatto, fatelo o, almeno, tentate di farlo. Non ci riuscite? Non c'è da meravigliarsi: non si riesce a trovare un sottoinsieme non vuoto di $\mathbb{N}$ che non ammetta minimo, per il semplice fatto che un tale sottoinsieme non esiste. Questo è il contenuto del seguente teorema.

Teorema 1.3 (Principio del buon ordinamento)**.** *Ogni sottoinsieme non vuoto di* $\mathbb{N}$ *ammette minimo.*

Dimostrazione. Sia E un sottoinsieme non vuoto di $\mathbb{N}$. Allora, per definizione, E contiene almeno un elemento; sia esso n_0. A questo punto, esistono due possibilità: o n_0 è più piccolo di tutti gli altri elementi di E, oppure no. Se è più piccolo, n_0 è il minimo (ed allora abbiamo finito); se non è più piccolo, esiste n_1 in E, con $n_0 > n_1$ (strettamente maggiore, perché deve essere diverso). Se n_1 è più piccolo di tutti gli elementi di E, allora è il minimo; se non lo è, allora esiste n_2 in E con $n_1 > n_2$. Possiamo continuare il ragionamento, ed è chiaro che "ci troviamo nei pasticci" se la scoperta di un numero n_k in E tale che $n_{k-1} > n_k$ prosegue indefinitamente: ovvero se continuiamo a trovare elementi di E più piccoli del precedente, ma ancora maggiori di qualche altro elemento di E. La nostra fortuna è che una tale "discesa infinita" non è possibile perché, fissato n_0 (ed n_0 è fissato una volta per tutte dall'essere E non vuoto), esistono al più n_0 numeri di cui n_0 è maggiore e che possono appartenere ad E: 0, 1, 2, ..., $n_0 - 1$. Pertanto, dopo al più n_0 "scelte", non avremo più a disposizione numeri naturali; il che vuol dire che l'ultima scelta che abbiamo fatto non può essere migliorata: E ammette minimo.

Una via alternativa alla dimostrazione, dopo aver scelto n_0, è la seguente. Partiamo da 0: se 0 appartiene ad E[4], è evidente che 0 è il minimo di E; se 0 non è in E, si consideri 1 e si ripeta il ragionamento precedente. Si devono fare al più n_0 verifiche (un numero finito, quindi), alla fine delle quali si è trovato il minimo di E. ■

Fino ad ora ci siamo mossi sulla retta "andando verso destra", nella direzione cioè della somma. Che succede se, partendo da un numero n qualsiasi, ad esempio 5, ci muoviamo verso sinistra? Ovviamente, dopo un salto siamo atterrati su 4 (che è caratterizzato — rispetto alla nostra operazione di salto — dall'essere l'unico numero naturale il cui successore è 5), e se continuiamo a spostarci verso sinistra troviamo (nell'ordine), 3, 2, 1, e 0. A questo punto l'operazione di salto verso sinistra ci viene impedita dal fatto che "prima di zero" non c'è nulla. Non esiste alcun numero naturale il cui successore sia zero. In altre parole, se interpretiamo l'operazione di saltare verso sinistra come "sottrazione" (così come avevamo interpretato l'azione di saltare verso destra come addizione), l'insieme dei numeri naturali non è chiuso rispetto a tale operazione: non sempre sottraendo un numero naturale da un altro naturale si ottiene un elemento di $\mathbb{N}$. A questo punto, si aprono davanti a noi due strade: lasciare tutto così com'è, emanare un editto che vieti la sottrazione[5], e

[4] Si noti che, per quanto E possa essere complicato (si pensi ad E_6), è sempre possibile verificare se un numero vi appartiene o no.

[5] Per alcuni tipi di sottrazione tale "editto" esiste già, e viene studiato in un'altra facoltà...

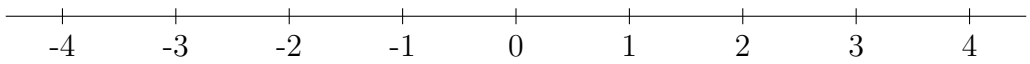
proseguire a lavorare con i naturali[6], oppure decidere che una tale restrizione non va bene, ed "aggiungere" ad $\mathbb{N}$ tutti i numeri mancanti.

Già, ma quanti sono? E come indicarli? Per convenzione, si decide di usare le stesse cifre usate per i naturali, ovvero 1, 2, 3, e di farle precedere da un segno "$-$" (a ricordare che sono numeri che si possono ottenere solo tramite sottrazioni, appunto). Inoltre, si sceglie di indicare con $-n$ il numero ottenuto facendo n passi a sinistra a partire dallo zero (così come n è il numero che si ottiene facendo n passi verso destra a partire da zero). Abbiamo allora l'insieme degli **interi relativi**:

$$\mathbb{Z} = \{\ldots, -n, \ldots, -3, -2, -1, 0, 1, 2, 3, \ldots, n, \ldots\}\,. \tag{1.2}$$

La definizione stessa di $-n$ fa poi sì che $n - n = n + (-n)$ sia uguale a zero, e che quindi $-n$ sia l'**inverso** di n rispetto all'operazione di somma (e, simmetricamente, n sia l'inverso di $-n$).

Come i naturali, anche gli interi sono rappresentabili su una retta. Siccome $\mathbb{N}$ è contenuto in $\mathbb{Z}$, conviene "prolungare" a sinistra in maniera simmetrica la rappresentazione grafica di $\mathbb{N}$, per ottenere quella di $\mathbb{Z}$:



Per costruzione, l'insieme dei numeri interi relativi è chiuso sia rispetto alla somma che alla sottrazione; tramite la sottrazione, è possibile definire la relazione di **minore**: dati n ed m in $\mathbb{Z}$, diciamo che n è minore di m, e scriviamo $n < m$, se esiste p in $\mathbb{N} \setminus \{0\}$ tale che $n = m - p$. Se p può essere anche 0, abbiamo la relazione di **minore od uguale**, che indichiamo $\leq$. Dal momento che se $n = m - p$, allora $n + p = m - p + p = m + 0 = m$, ne segue che $n < m$ ($n \leq m$) se e solo se $m > n$ ($m \geq n$). Mentre per i sottoinsiemi dei naturali non avevamo necessariamente il numero più grande, ma avevamo il minimo (per il principio di buon ordinamento), per $\mathbb{Z}$ la situazione è differente: esistono sottoinsiemi di $\mathbb{Z}$ che non posseggono minimo; ad esempio, $\mathbb{Z}$ stesso: se supponiamo che m sia il minimo, $m - 1$ è un numero intero più piccolo di m. È anche possibile definire il prodotto tra due numeri interi, e si verifica che le proprietà di somma e prodotto (come la commutatività e la distributività) si estendono anche a $\mathbb{Z}$. Notiamo, in particolare che $(-m) \cdot (-n) = m \cdot n$; infatti, essendo $0 = (m + (-m)) \cdot (-n)$, si ha che $-(m \cdot (-n)) = (-m) \cdot (-n)$. Analogamente si prova che $m \cdot (-n) = -(m \cdot n)$, da cui $m \cdot n = -(-(m \cdot n)) = (-m) \cdot (-n)$.

[6] In definitiva, il teorema di matematica più famoso di tutti i tempi riguarda i numeri naturali!

2. Calcolo combinatorio. Principio di induzione.

Quanti sono i numeri che si possono scrivere usando una ed una sola volta le cifre da 1 a 9? In altre parole, vogliamo contare quante "stringhe" come "135792468" si possono formare. È chiaro che se procediamo in maniera empirica, scrivendo numeri a caso su un foglio di carta e verificando che il numero che abbiamo appena scritto non compaia nella lista, corriamo il rischio da un lato di perdere molto tempo, dall'altro di non sapere quando li abbiamo scritti tutti. Per rispondere alla domanda, affrontiamo per il momento un problema più semplice: quanti sono i numeri che si possono scrivere usando una ed una sola volta le cifre 1 e 2? In questo caso la risposta è semplice: sono esattamente 2, e precisamente "12" e "21". Iniziamo con l'osservare che uno dei due numeri ha come prima cifra "1", mentre l'altro inizia con "2". Una volta scelta la prima cifra (e lo possiamo fare in due modi diversi — tanti quante sono le cifre a nostra disposizione), la seconda è obbligata (avendo sotto mano solo due cifre). E se le cifre sono 1, 2, e 3? In questo caso abbiamo tre scelte possibili per la prima cifra: "1xx", "2xx" e "3xx". Una volta scelta la prima cifra, quante scelte abbiamo per la seconda? Non più tre, perché una cifra l'abbiamo già utilizzata, ma solo due: ad esempio, "12x" e "13x". È ovvio che, scelte le prime due cifre, la terza è obbligata. In definitiva, abbiamo tre scelte per la prima cifra, due per la seconda, ed una per la terza: in tutto 6 ($= 3 \cdot 2 \cdot 1$) possibilità.

Possiamo allora rispondere alla domanda iniziale: usando una ed una sola volta le cifre da 1 a 9 possiamo scrivere $9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 326880$ numeri diversi; è chiaro che se avessimo affrontato il problema usando la forza bruta, non avremmo mai ottenuto un risultato preciso.

Siccome scrivere $9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$ è faticoso, è stata "inventata" una notazione apposita: dato un numero naturale n, indichiamo con $n!$, che si legge **n fattoriale**[7], il numero

$$n! = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 3 \cdot 2 \cdot 1\,. \tag{2.1}$$

Una delle proprietà più importanti di $n!$, che discende direttamente dalla definizione, è il fatto che $n! = n \cdot (n-1)!$; questa proprietà ci dice che, se partiamo a fare le moltiplicazioni da 1, e non da n, otteniamo via via $1!$, $2!$ e così via fino ad $n!$. I primi dieci fattoriali sono 1, 2, 6, 24, 120, 720, 5040, 40320, 362880, 3628800: come si vede, i fattoriali crescono abbastanza rapidamente all'aumentare di n.

Cosa conta esattamente $n!$? Dati n numeri (o oggetti), $n!$ conta il numero delle possibili **permutazioni** degli n numeri, dove per permutazione si intende uno dei possibili ordinamenti degli n numeri dati. Ad esempio, "12345" è una permutazione

[7] E non "enne!" con enfasi, come narrano le leggende metropolitane...

di “54321”, così come lo sono “13524” e “24135”, mentre (cane, gatto, topo) è una permutazione di (topo, cane, gatto).

Esercizio 2.1. A partire da $5! = 120$, $n!$ termina con almeno uno zero. Al crescere di n, aumenta il numero degli zeri di $n!$. Con quanti zeri termina 30!? Evidentemente, di calcolare 30! non se ne parla neppure...

Risposta 2.1: 7. Infatti, tra i fattori di 30! ci sono 2 e 5 (uno zero), 10 (uno zero), 4 e 15 (uno zero), 20 (uno zero), 16 e 25 (due zeri) e 30 (uno zero). Per i curiosi, $30! = 265252859812191058636308480000000$.

Esercizio 2.2. Si ha $20! = 243290200x176640000$. Quanto vale x?

Risposta 2.2: 8. Infatti 20!, avendo come fattore 9, deve essere divisibile per 9, e quindi la somma delle sue cifre deve essere, cone è noto, divisibile per 9. La somma delle cifre di 20! è $46 + x$ da cui $x = 8$.

Supponiamo adesso di avere n oggetti, e di sceglierne m tra questi: ad esempio, date le cifre da 1 a 9, ne secegliamo 5. Quanti numeri possiamo formare in questo modo? Ancora una volta, conviene ragionare per “numeri piccoli”, per capire cosa succede. Supponiamo di avere a disposizione 1, 2 e 3, e di dover scegliere due cifre. Ci sono, evidentemente, sei possibilità: “12”, “21”, “13”, “31”, “23” e “32”. Se le cifre a disposizione sono 1, 2, 3 e 4, e dobbiamo sempre scegliere due cifre, abbiamo dodici possibilità: “12”, “21”, “13”, “31”, “14”, “41”, “23”, “32”, “24”, “42”, “34” e “43”. Come abbiamo ottenuto questi numeri? Se, nel primo caso, consideriamo le 3! permutazioni di 1, 2 e 3, possiamo “scegliere” in tutti i modi possibili due cifre semplicemente “cancellando” l’ultima; in questa maniera da “123” si ottiene “12”, da “213” si ottiene “21”, e così via. Siccome abbiamo 6 permutazioni dalle quali cancellare l’ultima cifra, abbiamo 6 numeri che si possono ottenere scegliendo due cifre tra 1, 2 e 3. Che succede nel secondo caso? Le possibili permutazioni di 1, 2, 3 e 4 sono 4!, cioè 24, ma scegliendone una qualsiasi e cancellando le ultime due cifre, non sempre si ottiene (come resto) lo stesso numero: ad esempio, sia “1234” che “1243” generano “12” per cancellazione delle ultime due cifre; e questo perché le ultime due cifre possono essere permutate tra loro in due modi possibili (o meglio, in 2! modi possibili).

A questo punto, possiamo rispondere alla domanda: se abbiamo 9 cifre, e ne dobbiamo scegliere 5, è sufficiente considerare una qualsiasi delle 9! permutazioni di 1, 2, ..., 9; cancellare le ultime 4 cifre, ed ossservare che queste 4 cifre possono permutarsi (nelle ultime quattro posizioni) in 4! modi diversi; pertanto, il numero richiesto è $\frac{9!}{4!} = \frac{362880}{24} = 15120$.

Definiamo allora **disposizioni** di n oggetti presi a gruppi di m la quantità

$$ {}_nD_m = \frac{n!}{(n-m)!} = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-m+2) \cdot (n-m+1)\,. \tag{2.2} $$

Supponiamo ora di avere n cifre, e di chiederci (un'altra volta!) quanti numeri si possono creare scegliendo esattamente n cifre. Da un lato, sappiamo già la risposta: $n!$; dall'altro, stiamo considerando le disposizioni di n oggetti presi a gruppi di n, e quindi la risposta è ${}_nD_n$; la formula ci dà ${}_nD_n = \frac{n!}{0!}$; sappiamo che deve essere uguale ad $n!$, ma non abbiamo la più pallida idea di cosa sia $0!$. Per cavarci d'impaccio, *definiamo* $0! = 1$. In questa maniera, pur forzando l'interpretazione del fattoriale come numero delle "permutazioni", salviamo sia la correttezza della formula delle disposizioni, sia il suo significato combinatorio.

Ed ora, la prossima domanda: abbiamo sempre le cifre da 1 a 9, e vogliamo contare in quanti modi ne possiamo scegliere 5, senza tenere conto dell'ordine nel quale le scegliamo. In altre parole, non distinguiamo "12345" da "12354" e "53421". Invece di ridurci a casi semplici, questa volta ragioniamo direttamente: i numeri ottenuti scegliendo 5 cifre tra le 9 a nostra disposizione sono ${}_9D_5$. Scelto uno qualsiasi dei ${}_9D_5$ numeri, dobbiamo scartare tutti quelli che hanno le stesse cifre del nostro numero; siccome le cifre del numero che abbiamo scelto sono permutabili in $5!$ modi diversi, dovremo "buttare via" esattamente $5! - 1$ numeri. È allora evidente che il risultato voluto si ottiene dividendo ${}_9D_5$ per $5!$: $\frac{{}_9D_5}{5!} = \frac{15120}{120} = 126$.

Definiamo **combinazioni** di n oggetti presi a gruppi di m la quantità

$$ {}_nC_m = \frac{n!}{m!\,(n-m)!} = \binom{n}{m}^{(8)}\,. \tag{2.3} $$

Dalla definizione stessa di combinazione, si deduce che m deve essere compreso tra 0 e n. Tenendo conto del fatto che $0! = 1$, la formula per ${}_nC_m$ è sempre ben definita (anche per $n = 0$) e — miracolosamente — restituisce sempre un numero intero. I ${}_nC_m$, o $\binom{n}{m}$, sono anche detti **coefficienti binomiali**[9].

Una delle proprietà più notevoli di $\binom{n}{m}$ è la seguente: dati n ed m (con $0 \le m \le n$), si ha

$$ \binom{n+1}{m+1} = \binom{n}{m} + \binom{n}{m+1}\,. \tag{2.4} $$

[8] $\binom{n}{m}$ si legge "n sopra m".

[9] Per motivi che saranno chiariti tra breve.

La dimostrazione di questa formula usa in maniera cruciale il fatto che $n! = n\cdot(n-1)!$. Infatti

$$\begin{aligned}
\binom{n}{m} + \binom{n}{m+1} &= \frac{n!}{m!\,(n-m)!} + \frac{n!}{(m+1)!\,(n-m-1)!} \\
&= \frac{n!}{(m+1)!\,(n-m)!}\,[(m+1)+(n-m)] \\
&= \frac{n!}{(m+1)!\,(n-m)!}\,(n+1) \\
&= \frac{(n+1)!}{(m+1)!\,(n-m)!} = \binom{n+1}{m+1}.
\end{aligned}$$

Questa proprietà permette di scrivere i coefficienti binomiali in maniera "triangolare", formando il cosiddetto **triangolo di Tartaglia**:

$$\begin{array}{lccccccccc}
n & & & & & & & & & \\
0 & & & & & \binom{0}{0} & & & & \\
1 & & & & \binom{1}{0} & & \binom{1}{1} & & & \\
2 & & & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} & & \\
3 & & \binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3} & \\
4 & \binom{4}{0} & & \binom{4}{1} & & \binom{4}{2} & & \binom{4}{3} & & \binom{4}{4}
\end{array}$$

Ogni numero nel triangolo è somma dei due numeri che si trovano a sinistra e a destra nella riga precedente, come si vede eseguendo i calcoli:

$$\begin{array}{lccccccccc}
n & & & & & & & & & \\
0 & & & & & 1 & & & & \\
1 & & & & 1 & & 1 & & & \\
2 & & & 1 & & 2 & & 1 & & \\
3 & & 1 & & 3 & & 3 & & 1 & \\
4 & 1 & & 4 & & 6 & & 4 & & 1
\end{array}$$

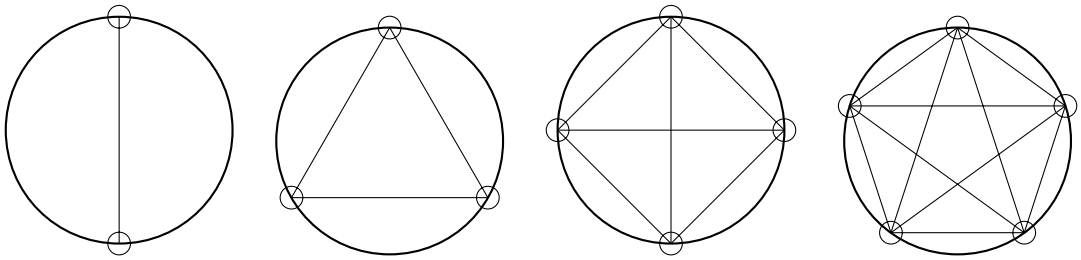
Affrontiamo ora un altro problema: dati due numeri, a e b, ed un naturale n, vogliamo calcolare la potenza n-sima del binomio $a+b$, e trovare una formula che, dato n, ci permetta di scrivere $(a+b)^n$ senza dover eseguire i calcoli tutte le volte. Alcuni casi particolari sono già noti[10]:

$$\begin{array}{lcccccccccc}
(a+b)^0 & = & & & & & 1 & & & & \\
(a+b)^1 & = & & & & a & + & b & & & \\
(a+b)^2 & = & & & a^2 & + & 2ab & + & b^2 & & \\
(a+b)^3 & = & & a^3 & + & 3a^2b & + & 3ab^2 & + & b^3 & \\
(a+b)^4 & = & a^4 & + & 4a^3b & + & 6a^2b^2 & + & 4ab^3 & + & b^4
\end{array}$$

[10] o, almeno, dovrebbero...

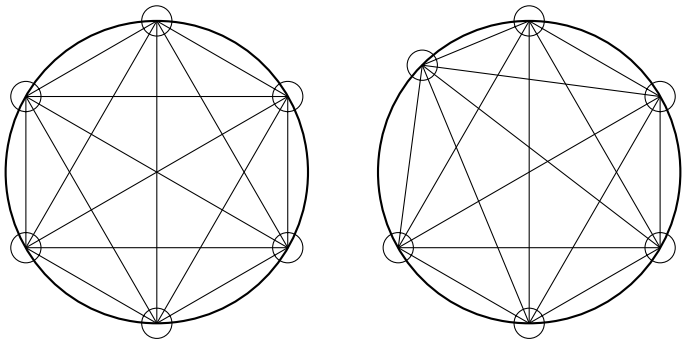
A questo punto, non v'è chi non veda la relazione strettissima che intercorre tra le potenze di $a+b$ e il triangolo di Tartaglia: i coefficienti delle potenze di a e b sono gli stessi che compaiono nel triangolo di Tartaglia.

Esercizio 2.3. Prima di andare avanti, supponiamo di avere una circonferenza, e di voler disegnare n punti su di essa in modo tale che, congiungendo i punti in tutti i modi possibili, il cerchio risulti diviso nel maggior numero di parti. Quante sono queste parti? Ovvero, in che modo dipendono da n? Come al solito, adottiamo la strategia dei "numeri piccoli"; se $n = 1$, siccome non c'è nulla da congiungere, abbiamo una sola parte (tutto il cerchio); il caso $n = 2$ è molto semplice: presi due punti sulla circonferenza, unendoli si divide in due il cerchio (e chiaramente questo è il massimo numero possibile di parti). Se $n = 3$ si ottengono 4 parti, mentre se ne hanno 8 per 4 punti. Se i punti sono 5, ne otteniamo 16.



Siamo pronti per scrivere la formula?

Risposta 2.3: No. Congetturando che il numero delle parti sia 2^{n-1}, se $n = 6$ non riusciamo in nessun modo ad ottenerne 32; se adottiamo la "strategia simmetrica" e scegliamo i sei punti come vertici di un esagono regolare, otteniamo 30 parti. Se spostiamo un po' uno dei punti arriviamo a 31, ma non di più.



La formula corretta è

$$\text{Parti}(n) = \frac{n^4 - 6n^3 + 23n^2 - 18n + 24}{24} = \binom{n}{4} + \binom{n}{2} + 1\,,$$

ed è evidente che il numero delle parti è molto minore di 2^{n-1}.

Risolto l'esercizio precedente[11], appare chiaro che la frase "i coefficienti delle potenze di a e b sono gli stessi che compaiono nel triangolo di Tartaglia" è forse un po' affrettata: in definitiva, abbiamo verificato che le prime cinque righe del triangolo di Tartaglia danno i coefficienti dello sviluppo della potenza n-sima di $a+b$, ma non abbiamo in nessun modo giustificato l'affermazione che tali coefficienti siano dati dalle righe del triangolo di Tartaglia **per ogni n**. Certamente, possiamo provare a scrivere lo sviluppo di $(a+b)^5$, e controllare se le cose vanno bene, e proseguire con $(a+b)^6$, $(a+b)^7$, ma per quante "verifiche" si facciano, non saremo mai in grado di concludere che i coefficienti dello sviluppo di $(a+b)^n$ sono dati, **per ogni n**, dalla $(n+1)$-sima riga del triangolo di Tartaglia.

Prima di proseguire, però, cerchiamo di formalizzare matematicamente la "congettura" che stiamo facendo: cerchiamo cioè di scrivere in termini matematici la frase "i coefficienti dello sviluppo di $(a+b)^n$ sono dati dalla $(n+1)$-sima riga del triangolo di Tartaglia". Iniziamo con scrivere $(a+b)^3$:

$$\begin{aligned}(a+b)^3 &= a^3+3a^2b+3ab^2+b^3\\ &= \tbinom{3}{0}\,a^3\,b^0+\tbinom{3}{1}\,a^2\,b^1+\tbinom{3}{2}\,a^1\,b^2+\tbinom{3}{3}\,a^0\,b^3\,.\end{aligned}$$

Quali sono gli elementi comuni all'ultima formula? Innanzitutto, l'indice n di $\binom{n}{m}$ è sempre 3 (che, poi, è la potenza cui viene elevato il binomio $a+b$). Poi, la somma delle potenze di a e b è sempre 3, ed inoltre l'indice m di $\binom{n}{m}$ cresce da 0 a 3, ed è sempre uguale all'esponente di b. Se ripetiamo lo stesso ragionamento per $(a+b)^2$, il risultato è lo stesso, così come è lo stesso per $(a+b)^4$. A questo punto, possiamo iniziare a formalizzare la nostra "congettura": per la potenza n-sima avremo bisogno di un indice m che "conta" da 0 ad n, e che di volta in volta viene utilizzato sia come esponente di b, che in $\binom{n}{m}$; inoltre, il coefficiente sarà sempre della forma $\binom{n}{\cdot}$, e l'esponente da dare ad a sarà tale che la somma degli esponenti di a e b sia n, e quindi darà $n-m$. In definitiva,

$$(a+b)^n \stackrel{?}{=} \binom{n}{0}\,a^n\,b^0+\binom{n}{1}\,a^{n-1}\,b^1+\ldots+\binom{n}{n-1}\,a^1\,b^{n-1}+\binom{n}{n}\,a^0\,b^n\,.$$

Questa formula — sulla cui correttezza non possiamo ancora giurare — è però molto "scomoda" da manipolare; basti pensare che ha $n+1$ termini, e se n è molto grande è necessaria parecchia carta (ed altrettanto inchiostro) per poterla scrivere[12]. Fortunatamente, è stato inventato un modo più "compatto" per scrivere formule di questo tipo, basandosi sul fatto che siamo in presenza di un'operazione matematica ripetuta

[11] Non lo avete risolto? Provateci!

[12] Se la formula serve per altri calcoli, l'uso dei puntini di sospensione è impossibile...

(in questo caso l'addizione) al variare di un "contatore". Avremo così

$$(a+b)^n \stackrel{?}{=} \sum_{m=0}^{n} \binom{n}{m} a^{n-m}\, b^m \,. \tag{2.5}$$

Il simbolo

$$\sum_{m=0}^{n}$$

si legge "sommatoria per m che va da 0 ad n di...", o "somme per m da 0 ad n di...". Il suo effetto è quello di generare un contatore (m) che "corre" sui naturali muovendosi da 0 ad n (lo si può pensare come un "cronometro", ad esempio); ad ogni valore di m associamo una quantità (in questo caso il monomio $\binom{n}{m} a^{n-m}\, b^m$) che va sommata al totale ottenuto in precedenza (il totale iniziale essendo zero).

Esercizio 2.4. Calcolare

$$\sum_{m=1}^{5} m\,, \qquad \sum_{\text{Bernardo}=1}^{5} (\text{Bernardo})^2\,, \qquad \sum_{m=1}^{5} n^2\,, \qquad \sum_{m=0}^{5} \binom{5}{m}\,.$$

Risposta 2.4: 15, 55 (si noti *en passant* che l'indice di sommatoria è "muto"), $5n^2$, 32.

Esercizio 2.5. Calcolare

$$\sum_{m=1}^{n} m\,, \qquad \sum_{m=0}^{n} n^2\,, \qquad \sum_{m=0}^{n} \binom{n}{m}\,;$$

per l'ultima somma, si supponga valida la (2.5) (e si scelgano due valori opportuni per a e b).

Risposta 2.5: $\frac{n(n+1)}{2}$, $n^2(n+1)$ e 2^n, scegliendo $a = 1 = b$.

Esercizio 2.6. Scrivere sotto forma di sommatoria le seguenti quantità

$$\begin{gathered} 1+2+3+4+\ldots+n\,, \\ 1+2+4+8+\ldots+2^n\,, \\ 1\cdot 1+2\cdot 4+3\cdot 6+4\cdot 4+5\cdot 1\,, \\ 10+10^{10}+10^{10^{10}}+10^{10^{10^{10}}}+10^{10^{10^{10^{10}}}}+\ldots \end{gathered}$$

Risposta 2.6:

$$\sum_{m=1}^{n} m\,, \quad \sum_{m=0}^{n} 2^m\,, \quad \sum_{m=0}^{4} (m+1)\cdot\binom{4}{m}\,,$$

mentre l'ultima somma non può essere scritta in maniera compatta, a meno di non osservare che l'n-simo addendo è 10 elevato all'$(n-1)$-simo.

Esercizio 2.7. Sapendo che $\sum$ sta a $\prod$ (che si legge "produttoria") come la somma sta al prodotto, in quale modo possiamo anche scrivere i numeri

$$\prod_{m=0}^{n} m\,, \quad \text{e} \quad \prod_{m=1}^{n} m\,?$$

Risposta 2.7: 0 ed $n!$.

Dopo aver lavorato un po' con le sommatorie, siamo finalmente in grado di affrontare il problema di dimostrare la (2.5) per ogni n in $\mathbb{N}$. Come abbiamo detto (ammaestrati anche dall'Esercizio 2.3), il fatto di verificare che la (2.5) è vera per qualche valore di n non implica in nessun modo che lo sia per tutti. Possiamo però fare un ragionamento differente: supponiamo di aver — con lunghi e laboriosi calcoli — dimostrato che la formula è vera per un certo valore di n, e proviamo *ad usare questo fatto per dimostrare che la formula è valida per* $n+1$. In altre parole, invece di sviluppare $(a+b)^{n+1}$, e di verificare successivamente che lo sviluppo che otteniamo è proprio quello dato dal secondo membro della (2.5) con n sostituito da $n+1$, proviamo ad usare l'informazione — già nota — che $(a+b)^n$ è esattamente uguale al secondo membro della (2.5). Per utilizzare quello che già sappiamo, basta osservare che se

$$(a+b)^n = \sum_{m=0}^{n} \binom{n}{m} a^{n-m}\, b^m\,,$$

allora

$$(a+b)^{n+1} = (a+b)\,(a+b)^n = (a+b) \sum_{m=0}^{n} \binom{n}{m} a^{n-m}\, b^m\,.$$

Distribuendo la somma,

$$(a+b)^{n+1} = \sum_{m=0}^{n} \binom{n}{m} a^{n-m+1}\, b^m + \sum_{m=0}^{n} \binom{n}{m} a^{n-m}\, b^{m+1}\,.$$

Lavoriamo sulla prima sommatoria, isolando il primo termine:

$$\begin{aligned} \sum_{m=0}^{n} \binom{n}{m} a^{n-m+1}\, b^m &= \binom{n}{0} a^{n+1} + \sum_{m=1}^{n} \binom{n}{m} a^{n-m+1}\, b^m \\ &= a^{n+1} + \sum_{m=1}^{n} \binom{n}{m} a^{n-m+1}\, b^m\,. \end{aligned}$$

Nella seconda sommatoria, isoliamo l'ultimo termine,

$$\begin{aligned}\sum_{m=0}^{n}\binom{n}{m}a^{n-m}\,b^{m+1} &= \sum_{m=0}^{n-1}\binom{n}{m}a^{n-m}\,b^{m+1}+\binom{n}{n}b^{n+1}\\ &= \sum_{m=0}^{n-1}\binom{n}{m}a^{n-m}\,b^{m+1}+b^{n+1}\,,\end{aligned}$$

e successivamente osserviamo che, se chiamiamo $m+1=p$, quando m si muove da 0 a $n-1$, p si muove da 1 a n e quindi, sostituendo m con $p-1$,

$$\sum_{m=0}^{n-1}\binom{n}{m}a^{n-m}\,b^{m+1}=\sum_{p=1}^{n}\binom{n}{p-1}a^{n-p+1}\,b^{p}+b^{n+1}\,.$$

Ricordando che l'indice di sommatoria è "muto" (si veda l'Esercizio 2.4), possiamo riscrivere la formula precedente scrivendo m al posto di p[13], ottenendo

$$\sum_{m=0}^{n}\binom{n}{m}a^{n-m}\,b^{m+1}=\sum_{m=1}^{n}\binom{n}{m-1}a^{n-m+1}\,b^{m}+b^{n+1}\,.$$

In definitiva, abbiamo

$$\begin{aligned}(a+b)^{n+1} &= \sum_{m=0}^{n}\binom{n}{m}a^{n-m+1}\,b^{m}+\sum_{m=0}^{n}\binom{n}{m}a^{n-m}\,b^{m+1}\\ &= a^{n+1}+\sum_{m=1}^{n}\binom{n}{m}a^{n-m+1}\,b^{m}\\ &\quad+\sum_{m=1}^{n}\binom{n}{m-1}a^{n-m+1}\,b^{m}+b^{n+1}\\ &= a^{n+1}+\sum_{m=1}^{n}\left[\binom{n}{m}+\binom{n}{m-1}\right]a^{n-m+1}\,b^{m}+b^{n+1}\,.\end{aligned}$$

Grazie alla formula (2.4), il contenuto delle parentesi quadre non è altro che $\binom{n+1}{m}$, e quindi

$$(a+b)^{n+1}=a^{n+1}+\sum_{m=1}^{n}\binom{n+1}{m}a^{n-m+1}\,b^{m}+b^{n+1}\,.$$

Dal momento che i coefficienti di a^{n+1} e b^{n+1} si possono scrivere rispettivamente come $\binom{n+1}{0}$ e come $\binom{n+1}{n+1}$, possiamo raccogliere tutto insieme ed affermare — non senza un

[13] Questa operazione di "cambio di indici" è abbastanza frequente, per cui cercate di capire bene cosa sta succedendo!

pizzico di orgoglio — che

$$(a+b)^{n+1} = \sum_{m=0}^{n+1} \binom{n+1}{m} a^{n-m+1} b^m ,$$

e questa è esattamente la formula (2.5) con n sostituito da $n+1$.

Ricapitolando: se la (2.5) vale per un certo valore di n, allora vale anche per il successivo. Il che vuol dire che abbiamo finito: siccome la formula è vera per $n = 0$ (non si tratta di niente altro che dell'identità $(a+b)^0 = 1$), allora lo sarà per $n = 1$, e siccome è vera per $n = 1$, lo sarà per $n = 2$, e siccome è vera per $n = 2$, lo sarà anche per $n = 3$, e così via... La (2.5) è vera per ogni n in $\mathbb{N}$!

Ne siamo proprio sicuri? In altre parole, la frase "e così via..." è matematicamente accettabile? Dire che una determinata proprietà, una formula, vale "per ogni n in $\mathbb{N}$" vuol dire fare un'infinità di affermazioni: chi può controllarle tutte? Per ovviare al problema di verificare la (2.5) per ogni n, abbiamo "inventato" una via alternativa: quella di supporre la formula vera per un certo valore di n, e di dedurre da questo fatto la veridicità per $n+1$; questo fatto, unitamente al verificarsi della formula per $n = 0$, ci ha portato a concludere che la formula era vera per ogni n usando la frase "e così via...".

Ahimé, nel mondo della matematica gli "e così via..." non sono accettati[14], perché le proprietà (soprattutto quelle che coinvolgono infiniti elementi) devono essere dimostrate. Ovvero, ci serve un meccanismo matematicamente corretto che permetta di concludere che se la (2.5) è vera per $n = 0$, e se dal supporla vera per n segue che è vera per $n+1$, allora la (2.5) è vera per ogni n.

Tale strumento esiste, è chiamato **induzione matematica**, ed è il soggetto del seguente teorema.

Teorema 2.8. *Sia $\mathcal{P}(n)$ una proprietà, dipendente da un indice naturale n, tale che*

i) *$\mathcal{P}(0)$ è vera;*

ii) *se $\mathcal{P}(n)$ è vera, allora $\mathcal{P}(n+1)$ è vera.*

Allora $\mathcal{P}(n)$ è vera per ogni n in $\mathbb{N}$.

Dimostrazione. Definiamo

$$F = \{n \in \mathbb{N} : \mathcal{P}(n) \text{ è falsa}\} ,$$

e dimostriamo che F è vuoto. Supponiamo per assurdo che F non sia vuoto, e vediamo come da questo fatto si ottenga una contraddizione. Dal momento che F è un sottoinsieme non vuoto di $\mathbb{N}$, per il principio del buon ordinamento (Teorema

[14] Anche nella vita reale: il tacchino induttivista di Bertrand Russell sostiene: "Tutti i giorni mi hanno dato da mangiare; quindi anche domani mi daranno da mangiare"; ma il giorno di Natale...

1.3) esso ammette minimo. Esiste cioè $\overline{m}$ in F tale che $\overline{m}$ è più piccolo di tutti gli elementi di F. Dal momento che $\overline{m}$ appartiene ad F,

$$\mathcal{P}(\overline{m}) \text{ è } \textbf{falsa}.$$

Inoltre, $\overline{m}$ non è zero (perché, per la i), 0 non appartiene ad F). Ne segue che $\overline{m}-1$ è un numero naturale (perché $\overline{m}$ è almeno 1), e che $\mathcal{P}(\overline{m}-1)$ è **vera** (dal momento che $\overline{m}-1$ non appartiene ad F). Ma allora, per la ii),

$$\mathcal{P}((\overline{m}-1)+1) = \mathcal{P}(\overline{m}) \text{ è } \textbf{vera}.$$

Siamo così arrivati ad un assurdo, generato dall'ipotesi che F non sia vuoto.

■

Si osservi che la proprietà chiave, quella che fa "scattare" il meccanismo dell'"e così via...", è la ii), ed è dunque necessario comprendere che cosa richieda. Affinché la ii) sia vera, **non** si deve dimostrare che $\mathcal{P}(n)$ e $\mathcal{P}(n+1)$ sono vere, ma che dal **supporre** $\mathcal{P}(n)$ vera, **segue** che $\mathcal{P}(n+1)$ è vera. In altre parole, si deve trovare un modo di far "comparire" la proprietà $\mathcal{P}(n)$ all'interno della dimostrazione della proprietà $\mathcal{P}(n+1)$. Per meglio comprendere il meccanismo dell'induzione matematica, dimostriamo il seguente teorema.

Teorema 2.9 (Disuguaglianza di Bernoulli)**.** *Sia $h \geq -1$. Dimostrare che per ogni n in $\mathbb{N}$ si ha*

$$(1+h)^n \geq 1 + n\,h\,. \tag{2.6}$$

Dimostrazione. La proprietà $\mathcal{P}(n)$ è dunque la (2.6). Se $n=0$ la (2.6) diventa $(1+h)^0 = 1 \geq 1 = 1+0\,h$, ed è dunque vera; pertanto, la i) è soddisfatta. Supponiamo ora che la (2.6) sia vera per un certo valore di n, ovvero che

$$(1+h)^n \geq 1 + n\,h\,. \tag{2.7}$$

Prima di procedere, cerchiamo di capire che cosa stiamo facendo: è chiaro che si potrebbe dire che la (2.7) è "esattamente" la (2.6), e che quindi stiamo usando la tesi per dimostrare la tesi[15]. In realtà, non c'è alcuna contraddizione: non stiamo usando la (2.7) per dimostrare la (2.6), ma la stiamo supponendo vera, e la useremo per dimostrare la (2.6) per $n+1$. In altre parole, non stiamo dimostrando che "(2.6) è vera", ma che "se (2.7) è vera per un certo valore di n, allora è vera anche per il

[15] Se volete, che ci stiamo mordendo la coda!

valore successivo". Proseguendo nella dimostrazione, si ha

$$\begin{aligned}
(1+h)^{n+1} &= (1+h)\,(1+h)^n \\
\left[\begin{smallmatrix}1+h\geq 0\\ \text{per la (2.7)}\end{smallmatrix}\right] &\geq (1+h)\,(1+n\,h) \\
[\text{sviluppando il prodotto}] &= 1+(n+1)\,h+n\,h^2 \\
[n\,h^2\geq 0] &\geq 1+(n+1)\,h\,,
\end{aligned}$$

che è esattamente la (2.6) con n sostituito da $n+1$. Abbiamo così dimostrato che

$$\mathcal{P}(n) \text{ vera} \Rightarrow \mathcal{P}(n+1) \text{ vera}\,.$$

Per il principio di induzione, la (2.6) è vera per ogni n in $\mathbb{N}$. ■

Teorema 2.10 (Sviluppo del binomio)**.**

$$(a+b)^n = \sum_{m=0}^{n} \binom{n}{m} a^{n-m}\,b^m\,, \qquad \forall n\in\mathbb{N}\,. \tag{2.5}$$

Osservazione 2.11. Nello sviluppo di $(a+b)^4$, da dove viene il coefficiente di $a^2\,b^2$? Se scriviamo

$$(a+b)^4 = (a+b)\,(a+b)\,(a+b)\,(a+b)\,,$$

è evidente che, per ottenere $a^2\,b^2$ dobbiamo "scegliere" 2 volte a e 2 volte b nei quattro fattori $a+b$. In quanti modi diversi possiamo farlo? Abbiamo 4 "caselle", numerate 1, 2, 3 e 4 (i fattori del prodotto), e dobbiamo sceglierne due (dalle quali prenderemo il fattore b), indipendentemente dall'ordine, dato che la moltiplicazione è commutativa. Questo problema è lo stesso del determinare quante coppie di due numeri si possano estrarre dalle cifre 1, 2, 3 e 4, e già conosciamo la risposta: $\binom{4}{2}$. A questo punto è chiaro perché il coefficiente di $a^{n-m}\,b^m$ è $\binom{n}{m}$: per ottenere $a^{n-m}\,b^m$, dobbiamo scegliere m volte b da n fattori, indipendentemente dall'ordine...

Esercizio 2.12. Dimostrare la (2.6) nel caso $h\geq 0$ usando la (2.5).

Risposta 2.12: Usando la (2.5) si ha

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

Dal momento che il valore della somma è non negativo (perché h è maggiore o uguale a zero), si ha la (2.6).

Dell'induzione matematica si può dare una forma più debole (di dimostrazione analoga), che risponde alla necessità di verificare che una determinata proprietà valga per ogni n maggiore di un certo n_0 fissato.

Teorema 2.13. *Sia $\mathcal{P}(n)$ una proprietà, dipendente da un indice naturale n, tale che esiste n_0 in $\mathbb{N}$ per cui*

i) *$\mathcal{P}(n_0)$ è vera;*

ii) *se $n \geq n_0$, e se $\mathcal{P}(n)$ è vera, allora $\mathcal{P}(n+1)$ è vera.*

Allora $\mathcal{P}(n)$ è vera per ogni n in $\mathbb{N}$, con $n \geq n_0$.

Esercizio 2.14. Si dimostri, per induzione, che

$$\sum_{m=0}^{n} m = \frac{n^2}{2} + \frac{n}{2}\,, \qquad \forall n \in \mathbb{N}\,,$$
$$\sum_{m=0}^{n} m^2 = \frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6}\,, \qquad \forall n \in \mathbb{N}\,,$$
$$\sum_{m=0}^{n} m^3 = \frac{n^4}{4} + \frac{n^3}{2} + \frac{n^2}{4}\,, \qquad \forall n \in \mathbb{N}\,,$$
$$\sum_{m=0}^{n} m^3 = \Big(\sum_{m=0}^{n} m\Big)^2\,, \qquad \forall n \in \mathbb{N}\,,$$
$$\text{se } q \neq 1,\ \sum_{m=0}^{n} q^k = \frac{1-q^{n+1}}{1-q}\,, \qquad \forall n \in \mathbb{N}\,.$$

Risposta 2.14: Se $n = 0$ abbiamo $1 = 1$, che è vera. Se supponiamo l'uguaglianza vera per n, l'uguaglianza per $n+1$ diventa

$$\sum_{m=0}^{n+1} q^k = \sum_{m=0}^{n} q^k + q^{n+1} = \frac{1-q^{n+1}}{1-q} + q^{n+1} = \frac{1-q^{n+2}}{1-q}\,,$$

come volevasi dimostrare.

Esercizio 2.15. Si dimostri, per induzione, che

$$\binom{2n}{n} \geq 2^n\,, \qquad \forall n \in \mathbb{N}\,,$$
$$n! \geq 2^{n-1}\,, \qquad \forall n \in \mathbb{N}\,,$$
$$n^n \geq 2^{n-1}\, n!\,, \qquad \forall n \in \mathbb{N}\,,\ n \geq 1\,,$$
$$\text{determinare } n_0 \text{ tale che } n^n \geq 2^n\, n!\,, \qquad \forall n \in \mathbb{N}\,,\ n \geq n_0\,.$$

Risposta 2.15: 3) se $n = 1$, la disuguaglianza si riduce all'identità $1 = 1$, ed è pertanto vera. Supponiamo ora che sia vera per un certo n; la disuguaglianza per $n+1$ è $(n+1)^{n+1} \geq 2^n\,(n+1)!$. Dal momento che $2^n\,(n+1)! = 2(n+1)\,[2^{n-1}\,n!]$, ed essendo per ipotesi induttiva $2^{n-1}\,n! \leq n^n$, avremo dimostrato la nostra disuguaglianza se dimostreremo che

$$(n+1)^{n+1} \geq 2(n+1)\,n^n \iff (n+1)^n \geq 2n^n\,.$$

Ricordando la (2.5), abbiamo

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

Essendo $\binom{n}{0} = 1$, e $\binom{n}{1} = n$, ed essendo tutti i termini della sommatoria non negativi, si ha

$$(n+1)^n \geq n^n + n\, n^{n-1} = 2n^n\,,$$

che è quanto si voleva dimostrare. 4) $n_0 = 6$ ($6^6 = 46656$ e $2^6\, 6! = 46080$, mentre $5^5 = 3125$ e $2^5\, 5! = 3840$).

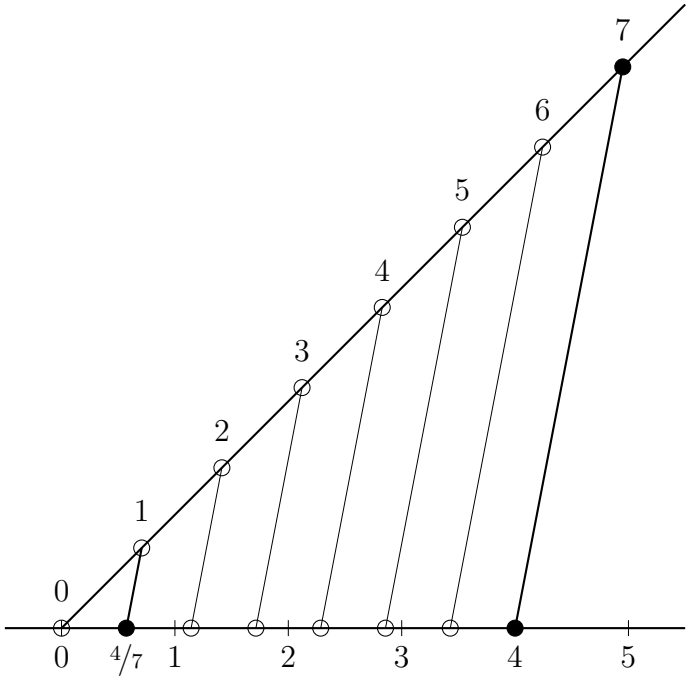
3. I razionali.

Come abbiamo visto, i numeri interi vengono introdotti per poter ben definire la sottrazione tra due numeri naturali. Una volta introdotti, ad ogni numero intero n viene associato un inverso $(-n)$, con la proprietà che $n + (-n) = 0$, l'elemento neutro dell'addizione. Ora, tra numeri interi abbiamo definito un'altra operazione, il prodotto, che ha come elemento neutro il numero 1. Dato un numero intero n, esiste un numero intero m che ne sia l'inverso rispetto al prodotto, cioè tale che $n \cdot m = 1$? A meno che il numero n non sia ± 1, la risposta è **no**. Se, ad esempio, $n = 2$, non esiste alcun numero intero m tale che $2 \cdot m = 1$. Ed infatti, se m fosse positivo, sarebbe $2 \cdot m \geq 2 > 1$, se m fosse negativo, allora $2 \cdot m \leq -2 < 1$, mentre se $m = 0$ si ha $2 \cdot m = 0 < 1$. Lo stesso ragionamento si può ripetere per n qualsiasi (diverso da ± 1). In altre parole, l'operazione di inversione rispetto al prodotto non è ben definita all'interno di $\mathbb{Z}$.

A questo punto, scartata l'idea di lasciare tutto così com'è, anche perché alle volte è necessario "dividere" (ovvero "fare le parti"), ragioniamo come abbiamo fatto nel passaggio da $\mathbb{N}$ a $\mathbb{Z}$: aggiungiamo tutte le quantità mancanti. Quello che otteniamo è l'insieme dei numeri **razionali**:

$$\mathbb{Q} = \left\{ \frac{p}{q}\,,\ p \in \mathbb{Z}\,,\ q \in \mathbb{N} \backslash \{0\} \right\}\,. \tag{3.1}$$

D'accordo, adesso abbiamo definito $\mathbb{Q}$, ma cosa vuol dire $\frac{p}{q}$? Che "numero" è? Per capirlo, torniamo alla rappresentazione di $\mathbb{N}$ sulla retta. Se, per il momento, supponiamo $p \geq 1$, possiamo rappresentare p come un punto sulla retta, a distanza p dall'origine (nel senso che per trovare p dobbiamo effettuare p salti partendo da 0). Analogamente, q è un altro punto sulla retta. Per comodità, lo rappresentiamo su una seconda retta, sghemba rispetto alla prima, con l'origine in comune. Congiungiamo ora il punto p con il punto q, e tracciamo, dal punto 1 sulla retta che contiene il punto q, la parallela alla retta che congiunge p con q. Questa retta taglierà la retta che contiene p in un punto, che per noi è $\frac{p}{q}$. D'ora in poi il numero razionale $\frac{p}{q}$ sarà il punto così costruito sulla retta.



$p = 4$ e $q = 7$.

Lo stesso ragionamento si può ripetere nel caso in cui p sia negativo.

Ricordando il Teorema di Talete[16] è facile verificare (ad esempio dalla figura) che sommando q volte $\frac{p}{q}$ (ovvero, facendo q salti sulla retta di lunghezza $\frac{p}{q}$) si ottiene p. Come conseguenza, se $n \geq 1$, moltiplicando n per $\frac{1}{n}$ si ottiene 1, e quindi $\frac{1}{n}$ è l'inverso moltiplicativo di n, mentre l'inverso moltiplicativo di $-n$ è $\frac{-1}{n}$ (per mantenere la convenzione che vuole denominatori non negativi). Un modo alternativo di indicare $\frac{1}{n}$ è scrivere n^{-1}. In questo modo si conserva la regola dei prodotti di potenze con la stessa base: $n^1 \cdot n^{-1} = n^{1-1} = n^0 = 1$. Analogamente, n^{-m} è l'inverso moltiplicativo di n^m, ovvero $\frac{1}{n^m}$ (ovviamente, deve essere $n \neq 0$).

Dati due numeri razionali $\frac{p}{q}$ e $\frac{r}{s}$ (e ricordando che i denominatori sono per convenzione positivi), si può stabilire sempre se uno dei due è maggiore dell'altro:

$$\frac{p}{q} \geq \frac{r}{s} \iff p \cdot s \geq r \cdot q\,.$$

Dati due numeri razionali, sono ben definiti sia la somma, che il prodotto:

$$\frac{p}{q} + \frac{r}{s} = \frac{p \cdot s + r \cdot q}{qs}\,, \qquad \frac{p}{q} \cdot \frac{r}{s} = \frac{p \cdot r}{q \cdot s}\,.$$

[16] Un fascio di rette parallele tagliato da due trasversali stacca su di esse segmenti a due a due proporzionali.

L'inverso additivo di $\frac{p}{q}$ è $\frac{-p}{q}$, mentre l'inverso moltiplicativo di $\frac{p}{q}$ (con $p \neq 0$) è $\frac{q}{p}$ (se $p > 0$), oppure $\frac{-q}{-p}$ (se $p < 0$).

L'insieme degli interi $\mathbb{Z}$ è un sottoinsieme di $\mathbb{Q}$; infatti, se m è un intero, allora è facile vedere che la costruzione geometrica precedente restituisce m se si calcola il numero razionale $\frac{m}{1}$.

Abbiamo dunque definito i razionali rifacendoci alla loro rappresentazione sulla retta; una tale definizione, anche se ci permette di "misurare" i razionali, è però molto poco pratica per lavorare: non è immaginabile dover prendere una riga ed un compasso tutte le volte che è necessario fare un'addizione. In altre parole, vogliamo un modo più comodo, possibilmente svincolato dalla rappresentazione geometrica dei razionali, per "scrivere" $\frac{p}{q}$. Tale modo è — ovviamente — la rappresentazione decimale.

Come possiamo scrivere la rappresentazione decimale di $\frac{p}{q}$, partendo da p e q? Iniziamo a trattare il caso $p \geq 0$. Innanzitutto, è sempre possibile scrivere p come $p = m \cdot q + r_1$ con m in $\mathbb{N}$, r_1 in $\mathbb{N}$ e $0 \leq r_1 \leq q - 1$. Ad esempio,

$$10 = 3 \cdot 3 + 1\,, \quad 23 = 0 \cdot 43 + 23\,, \quad 31 = 5 \cdot 6 + 1\,.$$

Il numero intero m viene detto **parte intera** di $\frac{p}{q}$, e viene di solito indicato nel seguente modo:

$$m = \left[\frac{p}{q}\right]\,.$$

Dividendo la relazione $p = m \cdot q + r_1$ per q si ottiene $\frac{p}{q} = m + \frac{r_1}{q}$. Essendo $0 \leq \frac{r_1}{q} < 1$, ne segue che m è il più grande intero minore di $\frac{p}{q}$. Esaminiamo ora la frazione $\frac{r_1}{q}$; se $r_1 = 0$, abbiamo finito: $\frac{p}{q} = m$ è già un numero intero, e con gli interi sappiamo lavorare. Se $r_1 > 0$, consideriamo il numero intero $10 \cdot r_1$[17], e scriviamo anch'esso nella forma $10 \cdot r_1 = d_1 \cdot q + r_2$, con d_1 e r_2 in $\mathbb{N}$, e $0 \leq r_2 \leq q - 1$. Essendo $0 < 10 \cdot r_1 \leq 10 \cdot (q - 1) < 10 \cdot q$, d_1 sarà un numero compreso tra 0 e 9: ovvero, una cifra decimale. Se $r_2 = 0$, ci fermiamo, mentre se $r_2 > 0$ andiamo avanti, scriviamo $10 \cdot r_2 = d_2 \cdot q + r_3$, ed otteniamo una seconda cifra compresa tra 0 e 9, d_2. Proseguendo, a meno di non trovare uno degli $r_i = 0$, otteniamo d_3, d_4, d_5 e così via.

E così via... Se così fosse, avremmo peggiorato la situazione: se nessuno degli r_i è zero, staremmo associando ad un numero razionale $\frac{p}{q}$ un numero intero m, e poi "infinite" cifre d_1, d_2, d_3, ..., tutte comprese tra 0 e 9. Il che vuol dire che, invece di portarci appresso una riga ed un compasso, dovremmo portarci appresso molta, molta carta per scrivere tutte le cifre d_i[18]. Fortunatamente, non c'è bisogno di calcolare

[17] La scelta di $10 \cdot r_1$ ci porta ai numeri decimali. Se avessimo scelto $2 \cdot r_1$ avremmo avuto la rappresentazione binaria, eccetera.

[18] Per non parlare del tempo necessario a calcolare la somma di due razionali...

infinite cifre d_i per "conoscere" il numero razionale $\frac{p}{q}$, perché le cifre d_i si ripetono da un certo punto in poi. Vale a dire che — se nessuno degli r_i è zero — esistono due indici j e k (dipendenti da p e q) tali che $d_i = d_{i+j}$ per ogni $i \geq k$. Perché? Perché le cifre d_i sono determinate dagli r_i: ricordiamo che d_i non è altro che la parte intera di $\frac{10 \cdot r_i}{q}$; pertanto, se esistono j e k tali che $r_{j+k} = r_k$, si avrà sia $d_{j+k} = d_k$ che $r_{j+k+1} = r_{k+1}$, e quindi le cifre decimali inizieranno a ripetersi nello stesso ordine. Siccome, qualsiasi sia i, si ha $0 < r_i \leq q-1$, ne segue che i "resti possibili" sono al massimo $q-1$; pertanto, dopo aver calcolato al più $q-1$ cifre decimali, uno dei resti "dovrà" ripetersi, e con loro le cifre decimali di $\frac{p}{q}$.

Possiamo allora affermare che, dato un qualsiasi numero razionale $\frac{p}{q}$ positivo, esistono un intero m ed un numero finito d_1, d_2, ..., d_j di cifre comprese tra 0 e 9 tali che

$$\frac{p}{q} = m, d_1 d_2 d_3 \dots d_{j-1} d_j 0$$

oppure

$$\frac{p}{q} = m, d_1 d_2 d_3 \dots d_{k-1} d_k d_{k+1} \dots d_{j-1} d_j d_k \, d_{k+1} \dots \text{ per qualche } k \geq 1\,.$$

Nel caso in cui le cifre d_k, ..., d_j si ripetono, scriviamo

$$\frac{p}{q} = m, d_1 d_2 d_3 \dots d_{k-1} \overline{d_k d_{k+1} \dots d_{j-1} d_j}\,.$$

Le cifre $d_1 \dots d_{k-1}$ si dicono **antiperiodo**, mentre le cifre $d_k \dots d_j$ si dicono **periodo** del numero razionale $\frac{p}{q}$. Se lo sviluppo è della forma $m, d_1 \dots d_j 0$, possiamo considerarlo periodico con periodo $\overline{0}$:

$$m, d_1 \dots d_j 0 = m, d_1 \dots d_j \overline{0}\,,$$

anche se in genere il periodo $\overline{0}$ si omette.

Ad esempio,

$$\frac{1}{4} = 0,25\overline{0} = 0,25\,, \quad \frac{1}{3} = 0,\overline{3}\,, \quad \frac{1}{7} = 0,\overline{142857}\,,$$

$$\frac{230}{170} = 1,3\overline{58823529411764 70}\,, \quad \frac{1}{19} = 0,\overline{526315789473684210}\,.$$

Se il numero razionale considerato è negativo, ovvero se $p < 0$, la rappresentazione decimale di $\frac{p}{q}$ si ottiene scrivendo la rappresentazione decimale di $-\frac{p}{q}$ (o, meglio, di $\frac{-p}{q}$), e "cambiando segno". Ad esempio, $-\frac{1}{3} = -0,\overline{3}$.

A questo punto le operazioni con i razionali (somma, prodotto, confronto) si possono eseguire "manipolando" opportunamente le cifre decimali; ad esempio, si può stabilire quale tra due numeri razionali sia il più grande.

Definizione 3.1. Dati due numeri razionali non negativi $\frac{p}{q}$ e $\frac{r}{s}$ con rappresentazione decimale

$$\frac{p}{q} = m, d_1 d_2 d_3 \dots d_n \dots , \quad \frac{r}{s} = m', d'_1 d'_2 d'_3 \dots d'_n \dots ,$$

diremo che $\frac{p}{q} > \frac{r}{s}$ se $m > m'$, oppure se $m = m'$ ed esiste $j \geq 1$ tale che $d_i = d'_i$ per ogni $i < j$ ma $d_j > d'_j$. Se $\frac{p}{q} > 0$ e $\frac{r}{s} < 0$, allora sarà $\frac{p}{q} > \frac{r}{s}$ (indipendentemente dalla rappresentazione decimale: è il "segno" a decidere l'ordine). Se sia $\frac{p}{q}$ che $\frac{r}{s}$ sono negativi, avremo

$$\frac{p}{q} = -m, d_1 d_2 d_3 \dots d_n \dots , \quad \frac{r}{s} = -m', d'_1 d'_2 d'_3 \dots d'_n \dots ,$$

e diremo che $\frac{p}{q} > \frac{r}{s}$ se $m < m'$, oppure se $m = m'$ ed esiste $j \geq 1$ tale che $d_i = d'_i$ per ogni $i < j$ ma $d_j < d'_j$.

Ad esempio, $1,2345\overline{6} > 1,234\overline{5}$ e $-1,234\overline{5} > -1,2345\overline{6}$. L'ordinamento così definito è compatibile sia con l'ordinamento definito precedentemente, sia con l'ordinamento stabilito dalla rappresentazione dei razionali come punti sulla retta.

Supponiamo ora di aver preso un razionale $\frac{p}{q}$, di aver eseguito il procedimento precedente, e di aver ottenuto la rappresentazione decimale $\frac{p}{q} = 0,\overline{9}$. È possibile arrivare ad una rappresentazione siffatta, o abbiamo commesso un errore? Se abbiamo ottenuto 0 come parte intera, questo vuol dire che $p = 0 \cdot q + r_1$, e quindi $r_1 = p$. Siccome la prima cifra decimale è 9, allora $10 \cdot p = 9 \cdot q + r_2$; essendo anche la seconda cifra decimale 9, abbiamo $10 \cdot r_2 = 9 \cdot q + r_3$, e quindi

$$100 \cdot p = 90 \cdot q + 10 \cdot r_2 = 90 \cdot q + 9 \cdot q + r_3 = 99 \cdot q + r_3 \,.$$

Continuando, troviamo

$$1000 \cdot p = 999 \cdot q + r_4 \,, \quad 10000 \cdot p = 9999 \cdot q + r_5 \,,$$

e, in generale,

$$10^n \cdot p = (10^n - 1) \cdot q + r_{n+1} \,, \qquad \forall n \geq 1 \,. \tag{3.2}$$

Esercizio 3.2. Sapendo che, per ogni $n \geq 1$, $10 \cdot r_n = 9 \cdot q + r_{n+1}$, dimostrare per induzione la (3.2).

Risposta 3.2: Se $n = 1$, la (3.2) è vera. Supponiamo che sia vera per un certo valore di n. Essendo $10 \cdot r_{n+1} = 9 \cdot q + r_{n+2}$ per ipotesi, si ha

$$\begin{aligned} 10^{n+1} \cdot p &= 10 \cdot (10^n - 1) \cdot q + 10 \cdot r_{n+1} \\ &= 10^{n+1} \cdot q - 10 \cdot q + 9 \cdot q + r_{n+2} \\ &= (10^{n+1} - 1) \cdot q + r_{n+2} \,, \end{aligned}$$

come volevasi dimostrare.

Dividiamo ora la (3.2) per 10^n; otteniamo

$$p = q - \frac{q}{10^n} + \frac{r_{n+1}}{10^n} \quad \Rightarrow \quad p - q = \frac{r_{n+1}}{10^n} - \frac{q}{10^n}\,.$$

Prendiamo ora n tale che $10^n > q$ (ricordiamo che q è fissato una volta per tutte); una tale scelta di n è semplice: basta infatti contare le cifre di q — siano esse m — e scegliere $n = m$. Siccome $10^n > q$, allora $\frac{q}{10^n} < 1$; d'altra parte, essendo $0 < r_{n+1} \leq q - 1$, si ha $\frac{r_{n+1}}{10^n} < 1$. Di conseguenza

$$-1 < \frac{r_{n+1}}{10^n} - \frac{q}{10^n} < 1\,.$$

Essendo però $p - q = \frac{r_{n+1}}{10^n} - \frac{q}{10^n}$ un numero intero, l'unica possibilità è $p - q = 0$, ovvero $p = q$. Ma se $p = q$, il procedimento per il calcolo dell'espansione decimale di $\frac{p}{q}$ dà 1, non $0,\overline{9}$! In altre parole, scrivendo lo sviluppo decimale di un qualsiasi razionale $\frac{p}{q}$, non troveremo mai $0,\overline{9}$; in realtà, non troveremo mai uno sviluppo della forma $m, d_1 \ldots d_j\overline{9}$. Alternativamente, non tutte le "stringhe" della forma $m, d_1 d_2 d_3 \ldots d_{k-1} d_k d_{k+1} \ldots d_{j-1} d_j d_k\, d_{k+1} \ldots$ sono ottenibili tramite lo sviluppo di un numero razionale: mancano quelle di periodo "9".

Siano ora $\frac{p}{q} = \frac{1}{3} = 0,\overline{3}$ e $\frac{r}{s} = \frac{2}{3} = 0,\overline{6}$. Se calcoliamo $\frac{p}{q} + \frac{r}{s} = \frac{1}{3} + \frac{2}{3}$, troviamo ovviamente 1, ma se sommiamo gli sviluppi decimali troviamo $0,\overline{9}$, ed abbiamo appena visto che questo non è uno sviluppo "ammissibile". Come fare? Ci serve un meccanismo che ci permetta, dato lo sviluppo decimale (dato **anche** uno sviluppo decimale "scorretto" come $0,\overline{9}$), di risalire al razionale che lo ha generato (o ad un razionale "compatibile" con le operazioni che hanno generato lo sviluppo scorretto). Il metodo è noto, ed è il seguente: sia $m, d_1 d_2 d_3 \ldots d_{k-1}\overline{d_k d_{k+1} \ldots d_{j-1} d_j}$ uno sviluppo decimale periodico. Il numeratore della frazione $\frac{p}{q}$ che "corrisponde" allo sviluppo dato si ottiene considerando il numero intero

$$p = m d_1 d_2 d_3 \ldots d_{k-1} d_k d_{k+1} \ldots d_{j-1} d_j - m d_1 d_2 d_3 \ldots d_{k-1}\,,$$

mentre il denominatore q è il numero intero

$$q = \overbrace{99\ldots99}^{j-k \text{ volte}} \quad \overbrace{00\ldots00}^{k-1 \text{ volte}}\,.$$

Ad esempio,

$$0,\overline{3} = \frac{3-0}{9} = \frac{3}{9} = \frac{1}{3}\,, \quad 0,\overline{142857} = \frac{142857-0}{999999} = \frac{142857}{999999} = \frac{1}{7}\,,$$

e

$$1,1\overline{6} = \frac{116-11}{90} = \frac{105}{90} = \frac{7}{6}\,, \quad 0,\overline{9} = \frac{9-0}{9} = \frac{9}{9} = 1\,.$$

In generale, se lo sviluppo decimale è $m, d_1 \dots d_k\overline{9}$, con $d_k < 9$, allora la procedura appena descritta costruisce il numero razionale

$$\frac{p}{q} = \frac{md_1 \dots (d_k+1)}{10\dots 0} = m, d_1 \dots (d_k+1)\,,$$

dove al denominatore sono presenti esattamente k zeri. Infatti,

$$\begin{aligned} m, d_1 \dots d_k\overline{9} &= \frac{md_1 \dots d_k 9 - md_1 \dots d_k}{90\dots 0} \\ &= \frac{10 \cdot md_1 \dots d_k - md_1 \dots d_k + 9}{90\dots 0} \\ &= \frac{9 \cdot (md_1 \dots d_k + 1)}{90\dots 0} = \frac{md_1 \dots (d_k+1)}{10\dots 0}\,. \end{aligned}$$

Ad esempio, $1,234\overline{9} = 1,235$.

Si noti che la procedura di "ricostruzione" di $\frac{p}{q}$ a partire da uno sviluppo decimale funziona anche se lo sviluppo è errato: ad esempio,

$$0,123\overline{23} = \frac{12323-123}{9900} = \frac{12200}{9900} = \frac{122}{99}\,, \quad 0,1\overline{23} = \frac{123-1}{99} = \frac{122}{99}\,.$$

In definitiva, ad ogni numero razionale della forma $\frac{p}{q}$ può essere associato uno sviluppo decimale periodico "ben formato" (ovvero, con un periodo diverso da $\overline{9}$), mentre ad un qualsiasi sviluppo decimale periodico può essere associato un numero razionale, che però non è unico: a $1 = 1,\overline{0}$ e a $0,\overline{9}$ corrisponde lo stesso numero, che è però compatibile con le operazioni algebriche eseguibili con i razionali: ad esempio $1 = 3 \cdot \frac{1}{3} = 3 \cdot 0,\overline{3} = 0,\overline{9}$. Che cosa è la rapresentazione decimale di un razionale della forma $\frac{p}{q}$? È un oggetto che racchiude in sé le proprietà di $\frac{p}{q}$ e che "misura" la distanza del punto $\frac{p}{q}$, costruito come spiegato all'inizio, dall'origine (così come un qualsiasi numero naturale misura la distanza dall'origine). E la misura in modo tale che se $\frac{r}{s}$ è un altro punto sulla retta, a sinistra di $\frac{p}{q}$, allora il numero ottenuto rappresentando $\frac{p}{q}$ in forma decimale è minore del numero ottenuto rappresentando $\frac{r}{s}$.

Osserviamo infine che mentre esiste il più piccolo naturale maggiore di zero (ed è 1), non esiste il primo numero razionale maggiore di zero. Se, infatti, supponiamo che esista, sia esso $\frac{p}{q}$; ebbene, considerando $\frac{p}{q+1}$ si ottiene un numero razionale positivo (perché p è non negativo), ma minore strettamente di $\frac{p}{q}$ dal momento che $p \cdot (q+1) = p \cdot q + p > p \cdot q$. Se pensassimo che questo nuovo numero razionale fosse il più piccolo maggiore di zero, potremmo sempre ottenerne uno più piccolo prendendo $\frac{p}{q+2}$, e così via...[19]. Non solo: dati comunque due razionali diversi $\frac{p}{q} < \frac{r}{s}$, esiste sempre un

[19] Sì, questa volta è lecito dire "e così via ..."!

razionale compreso tra i due:

$$\frac{p}{q} < \frac{\frac{p}{q} + \frac{r}{s}}{2} = \frac{p \cdot s + r \cdot q}{2qs} < \frac{r}{s}\,,$$

come si verifica facilmente. Questo vuol dire che se tentassimo una rappresentazione grafica dei razionali come abbiamo fatto con gli interi, dovremmo avere a disposizione un computer (o un pittore) in grado di disegnare linee estremamente sottili: tali che, tra due di esse, se ne deve poter mettere sempre un'altra. In altre parole, continuando ad ingrandire la nostra rappresentazione, non arriveremo mai ad una situazione simile a quella dei naturali, con dei "salti" discreti.

A questo punto, ci poniamo una domanda: prendiamo, su una retta sulla quale siano stati fissati 0 ed 1, un punto qualsiasi: esistono p e q tali che la procedura geometrica spiegata in precedenza "costruisca" questo punto? Altrimenti detto: presa una lunghezza qualsiasi sulla retta, questa lunghezza è rappresentata da un razionale?

Esercizio 3.3. Prima di andare avanti, provate a rispondere alla domanda precedente, ragionando non sui razionali come "punti sulla retta", ma come sviluppi decimali: quale numero razionale (se ne esiste uno) corrisponde allo sviluppo

$$0,10100100010000\ldots1\,\overbrace{0\ldots0}^{n \text{ zeri}}\,1\quad\overbrace{0\ldots0}^{n+1 \text{ zeri}}\quad1\ldots\,,$$

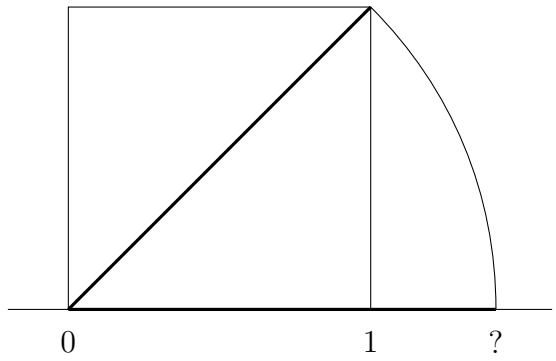
o allo sviluppo

$$0,12345678910111213141516\ldots n(n+1)(n+2)\ldots?$$

Risposta 3.3: Dato che ogni razionale dà luogo ad uno sviluppo decimale periodico, e non essendo i due sviluppi precedenti periodici, se ne deduce che nessun razionale li genera. Cosa sono, allora?

4. Non tutto ciò che è reale è razionale.

Costruiamo un quadrato di lato 1, con uno dei lati coincidente con il segmento unitario dell'asse dei numeri razionali, e disegnamone la diagonale. Riportiamo la diagonale sull'asse dei numeri, e chiediamoci se l'estremo destro del segmento cada o meno su di un razionale.



Quanto è lunga la diagonale? Usando il teorema di Pitagora, sappiamo che il quadrato costruito sulla diagonale ha come area la somma delle aree dei quadrati costruiti sui due lati del quadrato. Essendo il quadrato unitario, il quadrato costruito sulla diagonale ha area 2. Pertanto, la lunghezza della diagonale è tale che, elevata al quadrato, vale 2. Supponiamo ora che tale lunghezza sia un numero razionale $\frac{p}{q}$. Pertanto, $\frac{p^2}{q^2} = 2$, da cui segue che $p^2 = 2q^2$, ovvero che p^2 è pari. Dal momento che il prodotto tra due numeri dispari è sempre un numero dispari, dall'essere p^2 pari segue che anche p è pari, e quindi $p = 2p_1$, con p_1 un numero intero strettamente minore di p. Sostituendo nella relazione $p^2 = 2q^2$, si ottiene $4p_1^2 = 2q^2$ e, dividendo per 2, $q^2 = 2p_1^2$. Ripetendo il ragionamento precedente, se ne deduce che q è pari, e quindi $q = 2q_1$, con q_1 numero intero strettamente minore di q. Ri-sostituendo, otteniamo (dopo aver diviso per due) che $p_1^2 = 2q_1^2$, ovvero che $\frac{p_1^2}{q_1^2} = 2$. Ricapitolando, dati p e q interi e tali che $\frac{p^2}{q^2} = 2$, abbiamo trovato $1 \leq p_1 < p$ e $q_1 < q$ tali che $\frac{p_1^2}{q_1^2} = 2$. È evidente che possiamo allora ripetere il ragionamento precedente, e determinare altri due numeri interi $1 \leq p_2 < p_1$ e $q_2 < q_1$ tali che $\frac{p_2^2}{q_2^2} = 2$, e poi $1 \leq p_3 < p_2$ e $q_3 < q_2$ tali che $\frac{p_3^2}{q_3^2} = 2$ e così via. Il difetto del ragionamento consiste nel fatto che di numeri naturali p_i che verificano la disuguaglianza $1 \leq p_i < p$ ne esistono esattamente $p - 1$, e che quindi non è possibile continuare a far decrescere indefinitamente i numeratori p_i: ad un certo punto uno dei p_i non può esistere. Il fatto che non esista discende dall'aver supposto che esistano p e q tali che $\frac{p^2}{q^2} = 2$, ovvero che la lunghezza della diagonale del quadrato di lato 1 sia un numero razionale.

Esercizio 4.1. Sia m un numero naturale che non sia un quadrato perfetto (ovvero tale che non esista n in $\mathbb{N}$ per il quale $n^2 = m$). Dimostrare che non esistono p e q in $\mathbb{N}$ tali che $\frac{p^2}{q^2} = m$ (se $m = n^2$ è un quadrato perfetto, è sufficiente scegliere $p = n$ e $q = 1$). Suggerimento: fattorizzare m, ed osservare che esiste almeno uno dei fattori primi di m che compare con un esponente dispari; dimostrare successivamente che p è multiplo di questo fattore, da cui...

Risposta 4.1: Scriviamo $m = m_0 \cdot m_0^{r_0} \cdot m_1^{r_1} \cdot \ldots \cdot m_k^{r_k}$, con m_i primo per ogni i, $m_i \neq m_j$ per $i \neq j$, r_0 pari e gli r_i interi qualsiasi. Dimostriamo che p è divisibile per m_0. Scriviamo $p = km_0 + r$, con $0 \leq r \leq m_0 - 1$. Allora $p^2 = k^2 m_0^2 + 2km_0 r + r^2 = m_0 \cdot m_0^{r_0} \cdot m_1^{r_1} \cdot \ldots \cdot m_k^{r_k} q^2$, da cui segue che

$$m_0(k^2 m_0 + 2kr - m_0^{r_0} \cdot m_1^{r_1} \cdot \ldots \cdot m_k^{r_k} q^2) + r^2 = 0\,.$$

Dividendo per m_0, e cambiando segno si ottiene

$$\frac{r^2}{m_0} = m_0^{r_0} \cdot m_1^{r_1} \cdot \ldots \cdot m_k^{r_k} q^2 - k^2 m_0 - 2kr\,,$$

e quindi $\frac{r^2}{m_0}$ è un intero. Ne segue che m_0 divide r^2, ed essendo m_0 primo, m_0 divide r, il che è possibile, essendo $m_0 > r$, se e solo se $r = 0$. Ma allora $p = km_0$. Sostituendo, si trova

$$k^2 m_0^2 = m_0 \cdot m_0^{r_0} \cdot m_1^{r_1} \cdot \ldots \cdot m_k^{r_k} q^2\,.$$

Osservando che k^2 ha solo potenze pari di m_0 (o non ne ha nessuna), e a destra essendo presente una potenza dispari di m_0, si ottiene

$$m_1^{r_1} \cdot \ldots \cdot m_k^{r_k} q^2 = m_0\, m_0^{s_0}\, k_1^2\,,$$

da cui segue (analogamente a prima, ed usando il fatto che gli m_i sono a due a due distinti), che anche q deve essere un multiplo di m_0.

Pertanto, non tutto quello che possiamo "costruire" può essere misurato con i numeri razionali: alcune quantità, come la lunghezza della diagonale di un quadrato di lato 1, non possono essere rappresentate mediante un numero razionale.

Dal momento, poi, che ogni razionale della forma $\frac{p}{q}$ genera una rappresentazione decimale periodica, è chiaro che se vogliamo rappresentare come numero la lunghezza della diagonale del quadrato di lato 1, dovremo ricorrere ad una lista di decimali che "non si ripetono"; ad esempio, a qualcosa del tipo

$$1,4142135623730950488016887242096980785696718753769 4\ldots$$

Vedremo nel prossimo paragrafo come risolvere il problema.

Esercizio 4.2. Sapendo che la diagonale del cubo è un numero il cui quadrato è 3, quale diagonale ha una lunghezza il cui quadrato è n?

Risposta 4.2: Quella del cubo n dimensionale, qualsiasi cosa esso sia...

5. I numeri reali.

Come abbiamo visto nel paragrafo precedente, la lunghezza "fisica" della diagonale del quadrato di lato 1 non è rappresentabile sulla retta con un numero razionale. In altre parole, ci troviamo nella stessa situazione nella quale eravamo capitati quando avevamo provato a definire la sottrazione nei naturali, e la divisione negli interi. Però, mentre se prima potevamo "mettere da parte" il problema, e continuare a lavorare

con i naturali (o con gli interi), e l'introduzione di $\mathbb{Z}$ e $\mathbb{Q}$ poteva sembrare un'inutile complicazione, adesso non possiamo più permetterci di "chiudere gli occhi". Se vogliamo misurare distanze, ed oggetti reali, i numeri razionali non bastano più: non tutti i punti "costruibili" sulla retta — se preferite, non tutti i punti sulla retta — hanno un'etichetta rappresentabile nella forma $\frac{p}{q}$, ovvero sono costruibili mediante il teorema di Talete partendo dai numeri naturali. Qualcosa ci sfugge, nonostante i razionali siano molto fitti. Per risolvere il problema, facciamo come abbiamo sempre fatto: aggiungiamo quello che manca: siccome già sappiamo che non tutti i punti della retta sono misurabili tramite i razionali, consideriamo la retta nella sua totalità, e definiamo i **numeri reali**:

$$\mathbb{R} = \{\text{punti sulla retta}\}\,. \tag{5.1}$$

Ancora una volta, però, il rimedio rischia di essere peggiore del male: abbiamo i numeri reali, ma come ci lavoriamo? Come possiamo "manipolarli" in maniera comoda? Fra di essi ce ne sono alcuni che sono interi, altri razionali, altri (come il punto "finale" della diagonale del quadrato di lato 1) che non sono né interi, né razionali: come distinguere gli uni dagli altri? La soluzione al problema è data — una volta di più — dalla rappresentazione decimale. Così come ad ogni numero razionale possiamo associare in modo unico una rappresentazione decimale periodica, ad ogni numero reale è possibile associare una rappresentazione decimale, ovvero ogni numero reale x si può scrivere nella forma

$$x = a, b_1 b_2 b_3 \dots b_n \dots\,,$$

con a un intero, e le cifre decimali b_i numeri naturali compresi tra 0 e 9. Inoltre, a patto di escludere le rappresentazioni decimali "mal formate" (ovvero quelle nelle quali i b_i sono uguali a 9 da un certo punto in poi), ad ogni rappresentazione decimale $a, b_1 b_2 b_3 \dots b_n \dots$ corrisponde un unico numero reale x. Inoltre, se x e y sono due numeri reali, possiamo, data la loro rappresentazione decimale, decidere se x è maggiore di y seguendo lo stesso criterio dato dalla Definizione 3.1: se, ad esempio, x e y sono entrambi positivi, $x > y$ se e solo se la rappresentazione decimale di x e y "coincide" fino ad una certa cifra, e la prima cifra differente di x è più grande della corrispondente cifra di y.

È però chiaro che stiamo girando attorno al problema, che è quello della maneggevolezza: adesso sappiamo cosa è un numero reale, ma entambe le definizioni (i punti sulla retta, o gli sviluppi decimali qualsiasi[20]) sono di scarsissima utilità pratica; ad esempio, la seconda richiede molta carta per fare i calcoli... Ed infatti per poter lavorare con i numeri reali è necessario, in un certo senso, un atto di fede: i numeri reali

[20] Purché non 9 da un certo punto in poi...

esistono, punto e basta. Li si possono vedere come punti su una retta o, in maniera equivalente, come rappresentazioni decimali "ben formate"; si può lavorare con essi con operazioni algebriche come la somma, il prodotto, ed interpretare il risultato di conseguenza: la somma di due numeri reali (due punti) è il punto sulla retta che si ottiene mettendosi sul primo e spostandosi (in un verso opportuno) di una lunghezza pari alla distanza del secondo punto dall'origine, oppure la rappresentazione decimale che si ottiene sommando le due rappresentazioni decimali; il prodotto di due numeri reali è l'area (presa con il segno giusto) del rettangolo che ha per lati i due segmenti di estremi i due numeri reali, oppure la rappresentazione decimale ottenuta moltiplicando (in maniera complicatissima) le due rappresentazioni decimali.

L'unica cosa importante è che qualsiasi operazione — somme, moltiplicazioni, divisioni — si faccia con questi numeri, il risultato sarà sempre un altro numero reale (la cui interpretazione è libera); inoltre, è possibile leggere le proprietà del numero ottenuto in maniera indipendente, ma coerente, sia dalla sua rappresentazione "grafica" che dalla rappresentazione decimale.

Due paragrafi fa, avevamo detto che i razionali sono molto "fitti" sulla retta: dati due razionali diversi, esiste un terzo razionale compreso tra i due. Siccome è evidente che tra due reali distinti ne esiste sempre un terzo (basta prendere il punto medio del segmento di estremi i due numeri reali), ci chiediamo ora se — per caso, o per fortuna — tra due reali distinti esista sempre un numero razionale. È chiaro che la risposta è positiva nel caso in cui i due reali siano anche razionali, ma che succede se uno dei due (o entrambi) non lo sono? La risposta è, ancora una volta, positiva.

Teorema 5.1. *Dati x e y numeri reali, con $x > y$, esiste un numero razionale q in $\mathbb{Q}$ tale che $x > q > y$.*

Dimostrazione. Suppponiamo $y \geq 0$ (se $y < 0$ il ragionamento è analogo, a patto di fare attenzione ai segni). Consideriamo le rappresentazioni decimali ben formate di x e y:

$$x = a, b_1 b_2 b_3 \dots b_n \dots , \quad y = c, d_1 d_2 d_3 \dots d_n \dots .$$

Se $a > c$, dal momento che esiste almeno un indice j tale che $d_j \neq 9$ (perché i d_i non possono essere tutti 9), sia $q = c, d_1 d_2 \dots d_{j-1}(d_j + 1)$. Evidentemente $q < x$, dato che la parte intera di q è c, mentre la parte intera di x è a; inoltre q e y hanno le stesse cifre decimali fino alla $(j-1)$-esima, mentre la j-sima di q è maggiore della corrispondente di y (per costruzione). Siccome q ha uno sviluppo decimale finito, q è un razionale.

Supponiamo ora $a = c$; siccome $x > y$, esiste un indice j tale che $b_i = d_i$ per ogni $i < j$, e $b_j > d_j$. Se esiste $h > j$ tale che $b_h \neq 0$, sia $q = a, b_1 b_2 \dots b_{j-1} b_j$; altrimenti, se $b_h = 0$ per ogni $h > j$ (vale a dire, se b_j è l'ultima cifra decimale di x), sia $k > j$

tale che $d_k \neq 9$ (un tale k esiste perché i d_i non possono essere uguali a 9 da un certo punto in poi), e sia $q = a, b_1 b_2 \dots b_{j-1} d_j \dots d_{k-1}(d_k + 1)$. In entrambi i casi abbiamo un numero razionale (dato che lo sviluppo decimale è finito) compreso tra x e y, come si verifica facilmente. ■

Ad esempio, se

$$\begin{gathered} x = 2,00012151\dots\,,\ y = 1,999234234\dots \Rightarrow q = 2\,, \\ x = 2,718281828\dots\,,\ y = 2,718281815\dots \Rightarrow q = 2,71828182\,, \\ x = 3,1415\,,\ y = 3,14149999928\dots \Rightarrow q = 3,1414999993\,. \end{gathered}$$

La proprietà enunciata dal Teorema 5.1 è detta **densità dei razionali nei reali**; un altro modo di esprimerla è il seguente: dato un numero reale qualsiasi x, ed un secondo numero reale $\varepsilon > 0$, esiste sempre un numero razionale q_ε tale che $x < q_\varepsilon < x + \varepsilon$. Se interpretiamo ε come "errore" nella stima di x, il Teorema 5.1 ci dice che ogni numero reale x può essere approssimato **a meno di un errore che possiamo rendere arbitrariamente piccolo** da numeri razionali. In definitiva, nella lunghissima rappresentazione decimale di un qualsiasi numero reale x possiamo "lasciar cadere" le cifre decimali da un certo punto in poi, e ritrovarci a lavorare con un numero razionale, sapendo sì di aver commesso un errore, ma contemporaneamente essendo in grado di misurare e valutare la portata del nostro errore. Siccome la quantità di informazione necessaria per "registrare" un numero razionale è finita (il segno, la parte intera, l'antiperiodo ed il periodo), così come ad esempio la memoria di un computer[21], la densità di $\mathbb{Q}$ in $\mathbb{R}$ è la proprietà chiave sulla quale si basano i calcoli numerici e le simulazioni al calcolatore[22].

6. Proprietà di $\mathbb{R}$.

Una delle propretà fondamentali dell'insieme dei numeri reali è l'ordinamento: dati due reali distinti x e y, si ha sempre che $x < y$, oppure $y < x$. Questo fatto è molto utile quando, invece di considerare numeri reali singoli, si lavora con insiemi di numeri reali. Le proprietà degli insiemi di numeri reali che andremo a "scoprire" sono, in un certo senso, caratterizzanti di $\mathbb{R}$, nel senso che vedremo che non sono valide — ad esempio — per $\mathbb{Q}$.

Definizione 6.1. Sia E un sottoinsieme di $\mathbb{R}$.

---

[21] Sempre maggiore man mano che passa il tempo, ma sempre finita...

[22] Si potrà obiettare che anche la frase "la lunghezza della diagonale del quadrato di lato 1" definisce in maniera precisa un unico numero reale la cui rappresentazione decimale non è periodica usando una quantità finita di informazione; ed infatti questa proprietà — che caratterizza tale numero reale — viene sfruttata da alcuni *software* "matematici" per poter lavorare in maniera "simbolica" con i numeri reali.

Diciamo che E ha **massimo** se esiste $\overline{x}$ in $\mathbb{R}$ tale che

1) per ogni y in E, si ha $y \leq \overline{x}$.
2) $\overline{x}$ appartiene ad E;

Il numero reale $\overline{x}$ si indica anche max E.

Diciamo che E ha **minimo** se esiste $\underline{x}$ in $\mathbb{R}$ tale che

3) per ogni y in E, si ha $\underline{x} \leq y$.
4) $\underline{x}$ appartiene ad E;

Il numero reale $\underline{x}$ si indica anche min E.

Di sottoinsiemi di $\mathbb{R}$ che ammettono minimo ne abbiamo già incontrati: ad esempio un sottoinsieme qualsiasi di $\mathbb{N}$ ammette minimo per il principio di buon ordinamento (Teorema 1.3). Osserviamo — e ripetiamo — che se E ammette massimo (o minimo), allora tale massimo (o minimo) è unico. La proprietà 1) afferma che $\overline{x}$ è più grande di tutti gli elementi di E (il che non implica che appartenga all'insieme); unita alla 2), si ottiene che $\overline{x}$ è "il" più grande elemento di E. Analogamente, 3) e 4) affermano che $\underline{x}$ è "il" più piccolo elemento di E.

Esercizio 6.2. Trovare massimo e minimo dei seguenti insiemi:

$$E = \{x \in \mathbb{R} : 0 \leq x \leq 1\},$$
$$E = \{x \in \mathbb{R} : x^2 \leq 1\},$$
$$E = \{x \in \mathbb{R} : (x^2 - 1)(x^2 - 4) \leq 0\},$$
$$E = \{x \in \mathbb{R} : x \in \mathbb{Q}, -3 \leq x \leq 4\},$$
$$E = \{x \in \mathbb{R} : 0 < x \leq 1\}.$$

Una volta svolto l'esercizio, è evidente che c'è qualcosa che non va nell'ultimo insieme: se è facile vedere che 1 è il massimo dell'inseme dei numeri reali x tali che $0 < x \leq 1$, è altrettanto evidente che tale insieme non ha minimo. Se lo avesse, sarebbe per definizione il più piccolo numero reale maggiore di zero, ma un tale numero non esiste: se pensassimo che $\underline{x}$ fosse il più piccolo reale maggiore di zero, potremmo considerare $\frac{\underline{x}}{2}$, che è ancora maggiore di zero, ma minore di $\underline{x}$, e così via. Pertanto, a differenza dei naturali, e come i razionali, non tutti i sottoinsiemi di $\mathbb{R}$ ammettono minimo: aggiungendo a $\mathbb{Q}$ altri "oggetti", non abbiamo migliorato la situazione.

Esercizio 6.3. Trovare un sottoinseme E di $\mathbb{R}$, contenuto in $\{x \in \mathbb{R} : 0 \leq x \leq 1\}$, che non abbia massimo, ed uno che non abbia né massimo né minimo.

Abbiamo dunque visto che non tutti i sottoinsiemi di $\mathbb{R}$ ammettono massimo e (o) minimo. Torniamo per un momento a considerare l'insieme senza minimo $E = \{x \in \mathbb{R} : 0 < x \leq 1\}$. Perché non ha minimo? In altre parole, quale tra le proprietà

3) e 4) della Definizione 6.1 non è verificata? Se prendiamo $\underline{x} = -1$, è evidente che la 3) è vera, ma la 4) è falsa; se prendiamo $\underline{x} = -\frac{1}{2}$, è sempre la 4) a saltare, e continua a non essere verificata se scegliamo $\underline{x} = 0$. Se, invece, la 4) fosse vera per un certo $\underline{x}$, che sarebbe quindi maggiore di zero, la 3) non potrebbe essere vera, perché (come già detto prima) $\frac{\underline{x}}{2}$ è ancora in E ma non è maggiore di $\underline{x}$.

Dal momento che non riusciamo a far valere contemporaneamente la 3) e la 4), accontentiamoci di una sola di esse, e vediamo se è possibile salvare qualche proprietà interessante. Dal momento che la 4) da sola non ci dice niente di più sull'insieme (è solo la condizione di appartenenza ad E), manteniamo solo la 3), e andiamo a considerare l'insieme dei numeri reali che verificano tale proprietà:

$$m(E) = \{x \in \mathbb{R} : x \leq y \text{ per ogni } y \text{ in } E\}.$$

Tale insieme è non vuoto (abbiamo già detto che contiene -1, $-\frac{1}{2}$ e 0), e può essere così caratterizzato:

$$m(E) = \{x \in \mathbb{R} : x \leq 0\}.$$

Infatti, se $x \leq 0$, e se y è in E, allora $x \leq 0 < y$ e quindi x è in $m(E)$; viceversa, se x è in $m(E)$, allora x deve essere minore od uguale a zero: se fosse maggiore di zero, $\frac{x}{2}$ sarebbe contemporaneamente minore di x e appartenente ad E. Tutti gli elementi di $m(E)$ sono "candidati minimo" per E, dato che ognuno di essi soddisfa la 3); tra tutti, ne esiste uno che è "migliore" di tutti gli altri? Uno che è più "vicino" ad E degli altri? Uno per il quale non esiste un altro numero in $m(E)$ più grande di lui? Evidentemente, sì: il numero reale 0 — il massimo di $m(E)$ — è, tra tutti i numeri in $m(E)$, il migliore; ovvero, non appena ci si sposta un po' verso destra, anche di pochissimo, si trovano elementi di E (quelli, cioè, che soddisfano la 4), e per i quali la 3) non vale).

Ricapitolando: l'insieme $E = \{x \in \mathbb{R} : 0 < x \leq 1\}$ non ha minimo; abbiamo però identificato un numero reale, 0, il massimo di $m(E)$, che è quanto di più "vicino" al minimo sia possibile ottenere. Ovviamente, siamo riusciti ad identificarlo in questo caso: dal momento che abbiamo trovato dei sottoinsiemi di $\mathbb{R}$ che non ammettono massimo, potrebbe essere senza speranza fare affermazioni del tipo: "D'accordo, l'insieme non ha minimo, ma si può considerare al suo posto il massimo di $m(E)$" perché non è detto che tale massimo esista. Prima di proseguire, definiamo in maniera precisa gli insiemi come $m(E)$.

Definizione 6.4. Sia E un sottoinsieme di $\mathbb{R}$. Definiamo $M(E)$, l'insieme dei **maggioranti** di E, come

$$M(E) = \{x \in \mathbb{R} : x \geq y \text{ per ogni } y \text{ in } E\}.$$

Definiamo $m(E)$, l'insieme dei **minoranti** di E, come

$$m(E) = \{x \in \mathbb{R} : x \leq y \text{ per ogni } y \text{ in } E\}.$$

Siamo allora interessati a cercare, se esiste, il massimo di $m(E)$ e, specularmente, il minimo di $M(E)$. Un primo caso di "esistenza" si ha quando l'insieme E ammette minimo (o massimo).

Teorema 6.5. *Sia E un sottoinsieme di $\mathbb{R}$. Se E ammette massimo $\overline{x}$, allora $M(E)$ è non vuoto, ammette minimo, e tale minimo è $\overline{x}$. Se E ammette minimo $\underline{x}$, allora $m(E)$ è non vuoto, ammette massimo, e tale massimo è $\underline{x}$.*

Dimostrazione. Essendo $\overline{x}$ il minimo di E, $\overline{x}$ è un numero reale tale che $y \leq \overline{x}$ per ogni y in E; pertanto $\overline{x}$ è un maggiorante di E, e quindi $M(E)$ è non vuoto. Se dimostriamo che, preso comunque z in $M(E)$, si ha $\overline{x} \leq z$, avremo allora dimostrato che $\overline{x}$ è il minimo di $M(E)$. Sia allora z in $M(E)$, e supponiamo per assurdo che $z < \overline{x}$; se così fosse, z non potrebbe essere un maggiorante di E perché sarebbe minore di un elemento di E, che è proprio $\overline{x}$. Quindi, $\overline{x} \leq z$. L'altra metà della dimostrazione[23] è lasciata per esercizio. ■

Esercizio 6.6. Determinare $m(E)$ e $M(E)$ per i seguenti sottoinsiemi di $\mathbb{R}$:

$$E = \{x \in \mathbb{R} : -2 < x \leq 1\} \cup \{2\},$$
$$E = \{x \in \mathbb{R} : x^2 - 1 < 0\},$$
$$E = \{x \in \mathbb{R} : x^2 - 1 \geq 0\} \cap \{x \in \mathbb{R} : x^2 < 4\},$$
$$E = \{x \in \mathbb{Q} : 3 < x < 4\}.$$

Per ognuno degli insiemi precedenti determinare (se esistono) il massimo di $m(E)$ ed il minimo di $M(E)$.

Una svolto l'esercizio precedente[24], siamo fortemente tentati[25] di affermare che, per ogni sottoinsieme E di $\mathbb{R}$, ed indipendentemente dall'esistenza o meno del massimo o del minimo di E, esistono sia il massimo di $m(E)$ che il minimo di $M(E)$. Tanto per smontare subito le nostre certezze, consideriamo il seguente insieme:

$$E = \{x \in \mathbb{R} : x^2 > 1\}.$$

Risolvendo la disequazione, vediamo subito che

$$E = \{x \in \mathbb{R} : x < -1\} \cup \{x \in \mathbb{R} : x > 1\},$$

[23] Che potrebbe essere scritta da un *word processor* capace di eseguire sostituzioni nel testo...

[24] Se l'abbiamo svolto, ovviamente!

[25] Ma già temiamo che la cosa non sarà comunque facile da dimostrare...

e che quindi sia $m(E)$ che $M(E)$ sono vuoti, come si verifica facilmente. Essendo vuoti, $m(E)$ non ha massimo e $M(E)$ non ha minimo, e quindi non possiamo più affermare che, qualsiasi sia il sottoinsieme E di $\mathbb{R}$, esistono sia il massimo dei minoranti che il minimo dei maggioranti. Quello che possiamo fare è modificare l'affermazione "$m(E)$ ammette massimo per ogni sottoinsieme E di $\mathbb{R}$" nell'affermazione (comunque tutta da provare) "$m(E)$ ammette massimo per ogni sottoinsieme E di $\mathbb{R}$ tale che $m(E)$ non sia vuoto".

Cosa vuol dire che $m(E)$ non è vuoto? Vuol dire che esiste almeno un minorante di E, ovvero un numero reale x tale che $x \leq y$ per ogni y di E. Se rappresentiamo E sulla retta, questo vuol dire che x è una sorta di "barriera", di "limitazione" per E: l'insieme E si trova tutto a destra di x; analogamente, se $M(E)$ è non vuoto, esiste z in $M(E)$ e quindi (essendo z più grande di tutti gli elementi di E), E si trova tutto a sinistra di z. Viceversa, se E ammette una "barriera" a destra (o a sinistra), $M(E)$ (o $m(E)$) non sarà vuoto.

Definizione 6.7. Sia E un sottoinsieme di $\mathbb{R}$. E si dice **limitato superiormente** se esiste M in $\mathbb{R}$ tale che $x \leq M$ per ogni x in E; equivalentemente, se $M(E)$ non è vuoto. E si dice **limitato inferiormente** se esiste m in $\mathbb{R}$ tale che $m \leq x$ per ogni x in E; equivalentemente, se $m(E)$ non è vuoto. Un sottoinsieme E di $\mathbb{R}$ che sia limitato sia inferiormente che superiormente si dice **limitato**.

Alla luce della definizione precedente, le nostre affermazioni diventano: "se E è un sottoinsieme di $\mathbb{R}$ limitato inferiormente, esiste il massimo dei minoranti" e "se E è un sottoinsieme di $\mathbb{R}$ limitato superiormente, esiste il minimo dei maggioranti". E adesso viene il difficile: ci "sembra", dopo aver fatto molti esempi, che le due affermazioni precedenti siano vere; adesso è necessario **dimostrare** che sono vere. Ed invece...

Assioma 6.8. *Sia E un sottoinsieme di $\mathbb{R}$ limitato superiormente, allora esiste il minimo dei maggioranti.*

Sia E un sottoinsieme di $\mathbb{R}$ limitato inferiormente, allora esiste il massimo dei minoranti.

In altre parole, il fatto — evidente dagli esempi — che per i sottoinsiemi inferiormente limitati esista il massimo dei minoranti non è dimostrabile: è una verita "evidente", che viene accettata senza dimostrazione[26].

Una conseguenza dell'assioma precedente è la seguente definizione.

[26] Questa è la prima volta che un assioma viene enunciato in queste note; a ben vedere, di assiomi ne abbiamo usati parecchi in precedenza. Ad esempio, che si può contare...

Definizione 6.9. Sia E un sottoinsieme di $\mathbb{R}$.

Se E è limitato superiormente, definiamo **estremo superiore** di E il minimo dei maggioranti:

$$\sup E = \min M(E)\,.$$

Se E è limitato inferiormente, definiamo **estremo inferiore** di E il massimo dei minoranti:

$$\inf E = \max m(E)\,.$$

Grazie al Teorema 6.5, possiamo affermare che se E ammette massimo, allora si ha $\sup E = \max E$, e analogamente, se E ammette minimo, allora $\inf E = \min E$. In altre parole, i concetti di estremo superiore ed inferiore estendono il concetto di massimo e minimo.

Nella Definizione 6.1 avevamo detto che $\overline{x}$ era un massimo per E se 1) $\overline{x}$ era un maggiorante di E che 2) apparteneva ad E. Il seguente teorema modifica la proprietà 2) nel caso dell'estremo superiore (con analoghe modifiche nel caso dell'estremo inferiore).

Teorema 6.10. *Sia E un sottoinsieme di $\mathbb{R}$.*

Sia E limitato superiormente. Allora $S = \sup E$, l'estremo superiore di E, è l'unico numero reale caratterizzato dalle seguenti proprietà:

a) *per ogni y in E si ha $y \leq S$;*
b) *per ogni $z < S$ esiste y in E (eventualmente dipendente da z) tale che $z < y \leq S$.*

Sia E limitato inferiormente. Allora $I = \inf E$, l'estremo inferiore di E, è l'unico numero reale caratterizzato dalle seguenti proprietà:

c) *per ogni y in E si ha $I \leq y$;*
d) *per ogni $z > I$ esiste y in E (eventualmente dipendente da z) tale che $I \leq y < z$.*

Dimostrazione. Dimostriamo la prima parte del teorema; la seconda si dimostra in modo analogo.

In primo luogo la a) è equivalente a dire che S è un maggiorante di E, che è esattamente la 4) nella definizione di minimo per $M(E)$. D'altra parte, la b) afferma che nessun numero reale minore di S è un maggiorante di E, vale a dire che tutti i maggioranti di E sono maggiori od uguali a S, e questo è esattamente ciò che afferma la 3) nella definizione di minimo per $M(E)$. Pertanto, a) e b) sono equivalenti a 3) e 4) nella definizione di minimo per $M(E)$, ovvero S verifica a) e b) se e solo se è il minimo dei maggioranti di E, cioè l'estremo superiore di E. ∎

Osservazione 6.11. Un modo alternativo di enunciare la b) e la d) è il seguente:

b') per ogni $\varepsilon > 0$ esiste y_ε in E tale che $S - \varepsilon < y_\varepsilon \leq S$;

d') per ogni $\varepsilon > 0$ esiste y_ε in E tale che $I \leq y_\varepsilon < I + \varepsilon$.

La b) e la b') esprimono entrambe la stessa proprietà: non appena ci si sposta un po' a sinistra dall'estremo superiore, si "scavalcano" elementi di E; analogamente, la d) e la d') affermano che spostandosi a destra dall'estremo inferiore si "scavalcano" elementi di E. Si noti che in entrambi i casi **non** si richiede che z appartenga ad E. Evidentemente, e siccome ci interessa solo quello che accade "vicino" ad S e ad I, per dimostrare la b') o la d') basta verificare l'esistenza di y_ε per ogni ε "sufficientemente piccolo"; anche perché se riusciamo a trovare y_ε tale che la b') è vera per un certo valore di $\varepsilon > 0$, lo stesso y_ε continua a soddisfare la b') per ogni $\varepsilon' > \varepsilon$. Osserviamo esplicitamente che non è affatto detto che l'insieme E sia "pieno" a sinistra dell'estremo superiore, o a destra dell'estremo inferiore. Ad esempio, se $E = \{-2\} \cup \{x \in \mathbb{R} : -1 < x < 1\} \cup \{2\}$, l'estremo superiore di E è 2, ma spostandosi di poco a sinistra troviamo solo il numero 2[27].

Esercizio 6.12. Calcolare estremo superiore ed inferiore (usando $m(E)$ e $M(E)$) dei seguenti insiemi (per l'ultimo si richiede un po' di immaginazione...), e dire se sono rispettivamente massimo e minimo:

$$E = \{x \in \mathbb{R} : 1 < x^2 \leq 9\} \cap \{x \in \mathbb{R} : x > -2\},$$
$$E = \{x \in \mathbb{R} : x^3 > 1\} \cap \{x \in \mathbb{R} : x^5 \leq 32\},$$
$$E = \{x \in \mathbb{Q} : x^2 \leq 4\},$$
$$E = \bigcup_{n=3}^{+\infty} \left\{ x \in \mathbb{R} : \frac{1}{n} \leq x \leq 1 - \frac{1}{n} \right\}.$$

Una volta calcolati estremo superiore ed inferiore, dimostrare che sono tali usando le a), b) (o b')), c) e d) (o d')).

Risposta 6.12: 4) Dimostriamo che $m(E) = \{x \in \mathbb{R} : x \leq 0\}$; è chiaro che se $x \leq 0$, allora x è minore di tutti gli elementi di E, dato che ogni insieme che contribuisce all'unione è composto da numeri positivi, e quindi appartiene a $m(E)$. Viceversa, se x è un minorante di E, x non può essere positivo. Se, per assurdo, x fosse positivo, consideriamo il numero $y = \frac{1}{x}$, e sia $n = [y] + 1$ (ricordiamo che $[\cdot]$ è la parte intera). Allora $x > \frac{1}{n}$, perché $n > \frac{1}{x}$; se ne deduce che x è più grande di un elemento $(\frac{1}{n})$ appartenente ad E, e quindi non può essere un minorante di E. Concludendo, l'estremo inferiore di E è 0 (e non è un minimo); con ragionamenti analoghi si verifica che l'estremo superiore di E è 1 (e non è un massimo). Per dimostrare che 0 è l'estremo inferiore di E usando c) e d'), basta dimostrare che per ogni

[27] Il che spiega per quale motivo nella b) e nella b') si richiede $y \leq S$: se la disuguaglianza fosse stretta saremmo nei guai...

$\varepsilon > 0$ esiste un numero naturale n tale che $\varepsilon > \frac{1}{n}$; come prima, si può scegliere $n = \left[\frac{1}{\varepsilon}\right] + 1$. Come esercizio aggiuntivo, verificare che $E = \{x \in \mathbb{R} : 0 < x < 1\}$.

Un'altra proprietà utile soddisfatta da estremo superiore ed inferiore è la "monotonia".

Teorema 6.13. *Sia E un sottoinsieme di $\mathbb{R}$. Se E è limitato superiormente, e se $F \subseteq E$, allora* $\sup F \leq \sup E$. *Se E è limitato inferiormente, e se $F \subseteq E$, allora* $\inf E \leq \inf F$.

Dimostrazione. Sia S l'estremo superiore di E. Allora S è un maggiorante di E, ovvero è maggiore o uguale di tutti gli elementi di E. Siccome F è contenuto in E, S è anche più grande di tutti gli elementi di F, ovvero S è in $M(F)$. Essendo in $M(F)$, è maggiore o uguale del minimo di $M(F)$, che è l'estremo superiore di F. La seconda parte della dimostrazione è lasciata per esercizio. ■

Come conseguenza del teorema precedente, se E è limitato, e $F \subseteq E$ (cosicché anche F è limitato), si ha

$$\inf E \leq \inf F \leq \sup F \leq \sup E\,.$$

L'unica disuguaglianza da dimostrare è la centrale, che discende dal fatto (di dimostrazione evidente) che se x è un maggiorante di F e y è un minorante di F, allora si ha $y \leq x$. In particolare, il massimo dei minoranti è minore del minimo dei maggioranti.

Un altro utile teorema è il seguente.

Teorema 6.14. *Siano E ed F due sottoinsiemi di $\mathbb{R}$ tali che per ogni x di E e per ogni y di F si abbia $x \leq y$. Allora*

$$\sup E \leq \inf F\,.$$

Dimostrazione. Sia y fissato in F; per ipotesi, ogni x di E è minore o uguale di y, cosicché $M(E)$ non è vuoto. Inoltre, per l'arbitrarietà di y, $M(E)$ contiene F. Per il Teorema 6.13,

$$\sup E = \min M(E) = [\text{Teorema 6.5}] = \inf M(E) \leq \inf F\,,$$

come volevasi dimostrare. ■

Osservazione 6.15. Osserviamo esplicitamente che, anche se E ed F sono tali che $x < y$ per ogni x di E e per ogni y di F, si può solo concludere che $\sup E \leq \inf F$, e non che si ha il minore stretto. Ad esempio, siano $E = \{x \in \mathbb{R} : 0 < x < 1\}$ e $F = \{x \in \mathbb{R} : 1 < x < 2\}$. Se, invece, E ha massimo e F ha minimo, allora $\max E < \min F$.

Definizione 6.16. Dati due numeri reali a e b, con $a < b$, definiamo **intervallo** di estremi a e b l'insieme dei numeri reali x compresi tra a e b. Dato che non abbiamo specificato se "compresi" voglia dire maggiori, o maggiori o uguali, di a e minori, o minori o uguali, di b, abbiamo quattro possibilità, che indichiamo come segue:

$$\begin{aligned} [a,b] &= \{x \in \mathbb{R} : a \leq x \leq b\}, \\ [a,b) &= \{x \in \mathbb{R} : a \leq x < b\}, \\ (a,b] &= \{x \in \mathbb{R} : a < x \leq b\}, \\ (a,b) &= \{x \in \mathbb{R} : a < x < b\}. \end{aligned}$$

Il primo si dice intervallo **chiuso** di estemi a e b, mentre l'ultimo si dice intervallo **aperto** di estremi a e b; il secondo ed il terzo intervallo si dicono **semiaperti** o **semichiusi**, a seconda dei "gusti".

Esercizio 6.17. Calcolare estremo superiore ed inferiore di ognuno dei quattro tipi di intervalli.

Risposta 6.17: a e b.

Abbiamo detto all'inizio del paragrafo che avremmo enunciato delle proprietà "caratterizzanti" l'insieme dei numeri reali. In effetti, il concetto di insieme dei maggioranti e dei minoranti può essere definito anche sui razionali: se E è un sottoinsieme dei razionali, allora $M_{\mathbb{Q}}(E) = \{q \in \mathbb{Q} : y \leq q \text{ per ogni } y \text{ in } E\}$ (analogamente per $m_{\mathbb{Q}}(E)$), e può avere ancora senso chiedersi se esista il minimo di $M_{\mathbb{Q}}(E)$ per ogni sottoinsieme di $\mathbb{Q}$ limitato superiormente.

Ebbene, una tale proprietà non è vera: in altre parole, pur essendo i razionali "fitti", e densi in $\mathbb{R}$, non sono abbastanza per garantire che ogni sottoinsieme dei razionali limitato superiormente ammetta estremo superiore in $\mathbb{Q}$[28]. Il controesempio è costruito a partire dall'unico (o quasi) numero che sappiamo non essere razionale, e dalla proprietà che lo definisce. Sia

$$E = \{q \in \mathbb{Q} : q \geq 0 \text{ e } q^2 \leq 2\}.$$

Innanzitutto, dal momento che se q è in E, si ha $q \leq 2$ (se fosse $q > 2$, allora $q^2 > 2$), E è limitato superiormente. Sia allora

$$M_{\mathbb{Q}}(E) = \{p \in \mathbb{Q} : q \leq p \text{ per ogni } q \text{ in } E\} = \{p \in \mathbb{Q} : p \geq 0 \text{ e } p^2 \geq 2\},$$

e dimostriamo per assurdo che $M_{\mathbb{Q}}(E)$ non ammette minimo. Supponiamo cioè che esista il minimo di $M_{\mathbb{Q}}(E)$; sia esso $\underline{p}$, un numero razionale tale che $\underline{p}^2 \geq 2$. Siccome non può essere $\underline{p}^2 = 2$, perché già sappiamo che non esistono numeri razionali (come $\underline{p}$) il cui quadrato vale 2, deve essere $\underline{p}^2 > 2$. Ma allora, se x è il numero reale positivo il cui quadrato vale 2, si ha $\underline{p} > x$. Sia ora ε un numero razionale tale che

[28] Che ammetta estremo superiore reale, già lo "sappiamo".

$0 < \varepsilon < \underline{p} - x$ (tale numero esiste perché i razionali sono densi nei reali, si veda il Teorema 5.1), ovvero tale che $\underline{p} - \varepsilon > x > 0$. Ma allora $(\underline{p} - \varepsilon)^2 > x^2 = 2$. Pertanto, $\underline{p} - \varepsilon$ appartiene a $M_{\mathbb{Q}}(E)$, ma è minore di $\underline{p}$ (che quindi non può essere il minimo di $M_{\mathbb{Q}}(E)$).

L'esempio precedente mostra la differenza fondamentale tra $\mathbb{Q}$ e $\mathbb{R}$: pur essendo denso in $\mathbb{R}$, $\mathbb{Q}$ è "bucato", nel senso che non esiste un numero razionale che separa l'insieme E dall'insieme dei suoi maggioranti, mentre in $\mathbb{R}$ tale numero esiste, ed è x (il numero il cui quadrato è 2); per rendersene conto, basti osservare che abbiamo appena dimostrato che I, l'estremo superiore di E, che deve verificare $I^2 \geq 2$, non può essere tale che $I^2 > 2$. In altre parole, $\mathbb{R}$ è "continuo" (come del resto è "continua" la retta reale).

Consideriamo ora l'insieme

$$E = \{x \in \mathbb{R} : x \geq 0\}.$$

È facile vedere che E è limitato inferiormente e che il suo estremo inferiore è 0 (che è anche minimo); è altrettanto facile rendersi conto che $M(E) = \emptyset$: se supponiamo che x sia un maggiorante di E (fatto questo che implica $x \geq 0$, dato che 0 appartiene ad E), possiamo subito costruire $x + 1$ che appartiene ancora ad E, ma è più grande di x. In altre parole, E non è limitato superiormente.

Definizione 6.18. Sia E un sottoinsieme di $\mathbb{R}$. E si dice **illimitato superiormente** se $M(E)$ è vuoto. E si dice **illimitato inferiormente** se $m(E)$ è vuoto. E si dice **illimitato** se è sia illimitato superiormente che inferiormente.

Equivalentemente:

i) E è illimitato superiormente se e solo se per ogni $M > 0$ esiste x in E (eventualmente dipendente da M) tale che $x > M$;
ii) E è illimitato inferiormente se e solo se per ogni $M < 0$ esiste x in E (eventualmente dipendente da M) tale che $x < M$.

Infatti, la i) afferma che nessun numero positivo M appartiene ad $M(E)$ (il che implica automaticamente che nessun numero negativo vi può appartenere), mentre la ii) afferma che nessun numero negativo M appartiene ad $m(E)$ (cosicché nessun numero positivo vi può appartenere).

Come trattare (dal punto di vista di estremo superiore ed inferiore) gli insiemi illimitati? Il risultato chiave è il Teorema 6.13. In altre parole, vogliamo "assegnare" un estremo superiore ad un insieme illimitato superiormente, facendoci guidare dalla monotonia: se F è un qualsiasi sottoinsieme di E limitato superiormente, mentre E non lo è, l'estremo superiore di E (da definire) deve essere maggiore dell'estremo superiore di F. Per capire come fare, riprendiamo l'insieme $E = \{x \in \mathbb{R} : x \geq 0\}$,

e, fissato y in $\mathbb{R}$ con $y \geq 0$, consideriamo $F_y = \{x \in \mathbb{R} : 0 \leq x \leq y\} = [0, y]$. Evidentemente, F_y è contenuto in E; d'altra parte, l'estremo superiore di F_y è proprio y (che è anche massimo), per cui se vogliamo che il Teorema 6.13 valga, deve essere

$$y = \sup F_y \leq \sup E\,,$$

e quindi il definendo estremo superiore di E deve essere maggiore di ogni numero reale positivo; volendo, deve essere "il più grande" numero reale positivo. Ma un tale numero non esiste! Siccome non esiste, lo "inventiamo", o meglio inventiamo un simbolo che abbia per noi il significato dell'inesistente "più grande numero reale positivo". Tale simbolo è $+\infty$, che si legge "più infinito", ed è quello che scegliamo per definire l'estremo superiore di E e, in generale, l'estremo superiore di un insieme illimitato superiormente. Ripetiamo — *ad nauseam* — che $+\infty$, ed il suo speculare $-\infty$, che si legge "meno infinito" e viene usato per definire l'estremo inferiore di un insieme illimitato inferiormente, non sono numeri: sono due **simboli**, che usiamo per rappresentare qualcosa di più grande (o più piccolo) di tutti i numeri reali. Pertanto, non ha senso dire che un insieme illimitato superiormente ammette massimo: il suo estremo superiore è $+\infty$, che non è un numero reale.

Esercizio 6.19. Se l'estremo superiore di un insieme illimitato superiormente è $+\infty$, quanto vale il minimo dell'insieme vuoto? E quanto il massimo dell'insieme vuoto? Una volta risposto, la risposta è coerente con il Teorema 6.13?

A questo punto possiamo affermare (usando l'Assioma 6.8 e la definizione precedente) che ogni sottoinsieme E di $\mathbb{R}$ ammette sia estremo superiore che estremo inferiore. Inoltre, se F è un sottoinsieme di E si ha

$$\inf E \leq \inf F \leq \sup F \leq \sup E\,.$$

Quest'ultimo fatto è molto utile nel caso in cui si voglia dimostrare che un insieme E è illimitato superiormente (o inferiormente); invece di dimostrarlo direttamente, è sufficiente far vedere che esiste un suo sottoinsieme illimitato superiormente (o inferiormente).

Esercizio 6.20. Sia

$$E = \left\{ x \in \mathbb{R} : x \geq 0 \text{ e } \binom{2[x]}{[x]} \geq 2^{[x]} \right\}\,.$$

Calcolare l'estremo superiore e l'estremo inferiore di E.

Risposta 6.20: E contiene $\mathbb{N}$ (si veda l'Esercizio 2.15). Non c'è quindi bisogno di dimostrare che E è l'insieme dei numeri positivi per affermare che $\inf E = 0 = \min E$ e $\sup E = +\infty$.

Definizione 6.21. L'insieme dei numeri reali x maggiori (maggiori o uguali) o minori (minori o uguali) di un fissato numero reale a vengono anch'essi detti intervalli, ma illimitati (superiormente o inferioremente a seconda dei casi). Come per gli intervalli limitati, se a è un fissato numero reale abbiamo le seguenti notazioni:

$$\begin{aligned}
[a, +\infty) &= \{x \in \mathbb{R} : x \geq a\}\,, \\
(a, +\infty) &= \{x \in \mathbb{R} : x > a\}\,, \\
(-\infty, a] &= \{x \in \mathbb{R} : x \leq a\}\,, \\
(-\infty, a) &= \{x \in \mathbb{R} : x < a\}\,.
\end{aligned}$$

Si noti che il "lato" infinito è sempre aperto, coerentemente con il fatto che $+\infty$ e $-\infty$ non sono numeri reali. In particolare,

$$\mathbb{R} = (-\infty, +\infty)\,.$$

7. Radici, potenze. Funzioni elementari.

Nel paragrafo 4 abbiamo "costruito" un numero reale non razionale: la lunghezza della diagonale del quadrato unitario, e abbiamo visto che tale numero, sia esso x, gode della proprietà di avere il quadrato uguale a 2. Come è noto, per semplificare le notazioni tale numero si definisce la radice quadrata di 2, e si indica $x = \sqrt{2}$. Più che sulla notazione (standard), mettiamo l'accento sul fatto che x è "la" radice quadrata di 2: ovvero, l'unico numero positivo il cui quadrato è due. È evidente che di numeri reali il cui quadrato vale due ne esiste un altro (ed uno solo): $-x$, ovvero $-\sqrt{2}$. E questo a motivo del fatto che il quadrato di x ed il quadrato di $-x$ coincidono. Per evitare fraintendimenti, scriviamo $\sqrt{2}$, sottintendendo il segno $+$, intendendo così che tra i due numeri reali il cui quadrato vale due scegliamo il positivo. Continuando con le nostre costruzioni geometriche, e sempre usando il Teorema di Pitagora, possiamo costruire $\sqrt{n}$ per ogni n naturale[29], ovvero l'unico numero reale positivo il cui quadrato vale n. Ci chiediamo ora se, fissato un qualsiasi numero reale y, esista un unico numero reale positivo x tale che $x^2 = y$. È evidente che se $y < 0$ tale numero x non esiste, dato che $x^2 \geq 0$ per ogni x reale. Limitiamoci pertanto agli $y \geq 0$. Quello che si dimostra è il seguente fatto: dato un qualsiasi $y \geq 0$ esiste un unico numero reale $x \geq 0$ tale che $x^2 = y$.

Esercizio 7.1. Sia $y \geq 0$ e sia $E = \{x \in [0, +\infty) : x^2 \leq y\}$. Dimostrare che E non è vuoto, che ammette estremo superiore S e che $S^2 = y$.

Definiamo $\sqrt{y}$ l'unico numero reale positivo il cui quadrato è y. Che cosa abbiamo ottenuto? Abbiamo ottenuto una **funzione**: vale a dire una legge che, ad ogni numero

[29] Come?

reale $y \geq 0$ associa l'unico numero reale $x \geq 0$ tale che $x^2 = y$. Scrivendo in maniera compatta:

$$\begin{array}{rcccl} \sqrt{\cdot} & : & [0,+\infty) & \to & [0,+\infty) \\ & & y & \mapsto & \sqrt{y}\,. \end{array}$$

Prima di procedere oltre, definiamo in maniera corretta cosa sia una funzione.

Definizione 7.2. Sia A un sottoinsieme di $\mathbb{R}$. Una funzione f da A in $\mathbb{R}$ è una legge che, ad ogni elemento x di A, associa uno ed un solo elemento y di $\mathbb{R}$. Tale elemento y si dice "immagine di x tramite f", e si indica con $y = f(x)$. In simboli:

$$\begin{array}{rcccl} f & : & A & \to & \mathbb{R} \\ & & x & \mapsto & y = f(x)\,. \end{array}$$

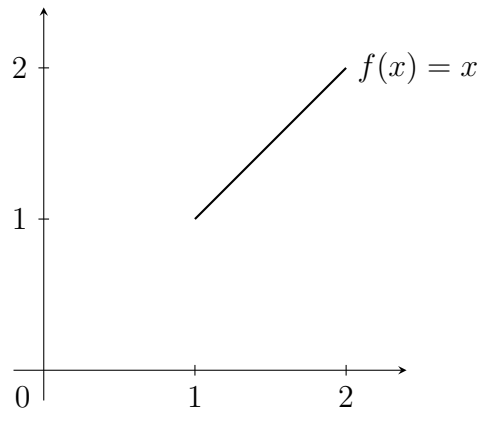
L'insieme A si dice **dominio**, o **insieme di definizione** della funzione f, mentre x si dice **argomento** della funzione f. Associato ad una funzione esiste un secondo sottoinsieme di $\mathbb{R}$, detto **immagine** di A tramite f, che è l'insieme dei valori assunti da $f(x)$ al variare di x in A:

$$\mathrm{Im}(f) = f(A) = \{y \in \mathbb{R} : \text{esiste } x \text{ in } A \text{ tale che } f(x) = y\}\,.$$

La maniera migliore di "rappresentare" una funzione da A in $\mathbb{R}$ è quella di ricorrere al suo **grafico**, che non è altro che il sottoinsieme del piano cartesiano[30] formato dalle coppie $(x, f(x))$ al variare di x in A:

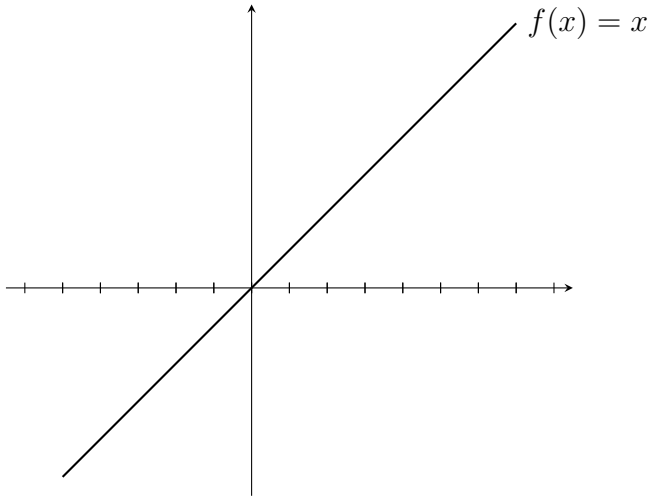
$$G(f) = \{(x,y) \in \mathbb{R} \times \mathbb{R} : x \in A\,,\ y = f(x)\}\,.$$

Ad esempio, se $f : [1,2] \to \mathbb{R}$ è la funzione definita da $f(x) = x$ per ogni x in $[1,2]$, il grafico di f è il seguente:

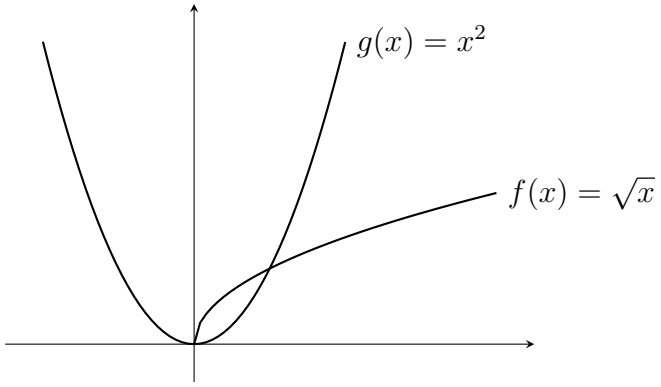


[30] Qualsiasi cosa esso sia...

e, più in generale, il grafico (o meglio parte di esso) della funzione $f : \mathbb{R} \to \mathbb{R}$ definita da $f(x) = x$ per ogni x in $\mathbb{R}$ (una tale funzione si dice "identità" dato che lascia ogni x invariato) è il seguente:



Nel caso della funzione $f(x) = \sqrt{x}$, il dominio è $A = [0, +\infty)$, mentre l'immagine è $f(A) = [0, +\infty)$; il grafico è il seguente, assieme al grafico della funzione $g(x) = x^2$, definita su tutto l'asse reale, e a valori in $[0, +\infty)$.



Esercizio 7.3. Sia x un numero reale qualsiasi: quanto vale $\sqrt{x^2}$?

Svolto l'esercizio precedente? Quale è la risposta? Chi ha risposto $\pm x$? Avendo appena detto che una funzione (come lo è la radice quadrata) associa uno ed un solo numero reale al proprio argomento, è chiaro che $\pm x$ è la risposta sbagliata. Chi ha risposto x? Dal momento che la funzione radice quadrata ha come immagine l'intervallo $[0, +\infty)$, è chiaro che la risposta è sbagliata nel caso in cui, ad esempio, $x = -1$. In questo caso, infatti, x^2 vale 1, e quindi $\sqrt{x^2} = \sqrt{1} = 1$, dal momento

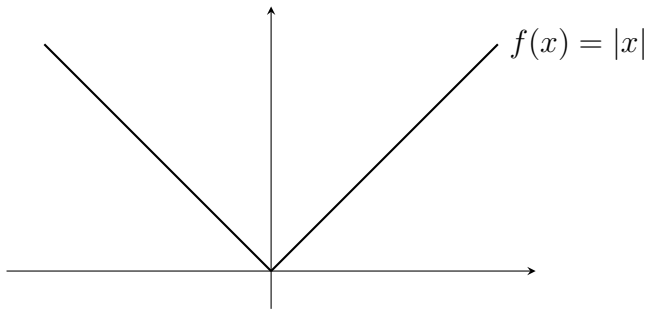
che 1 è l'unico numero reale positivo il cui quadrato vale 1. Se facciamo altri esempi, arriviamo al seguente risultato[31]:

$$\sqrt{x^2} = \begin{cases} x & \text{se } x \geq 0, \\ -x & \text{se } x < 0. \end{cases}$$

Per aiutare[32] i matematici, è stata inventata una notazione apposita per indicare la funzione $\sqrt{x^2}$. Il **valore assoluto**, o **modulo**, indicato con $|\cdot|$, è la funzione da $\mathbb{R}$ in $\mathbb{R}$ definita da

$$|x| = \begin{cases} x & \text{se } x \geq 0, \\ -x & \text{se } x < 0. \end{cases} \tag{7.1}$$

Come si vede abbastanza facilmente, il dominio di $|\cdot|$ è tutto $\mathbb{R}$, mentre l'immagine è $[0, +\infty)$. **Avvertenza:** il lettore è fortemente invitato a non pensare al modulo di x come "x privato del segno". Il valore assoluto è definito in (7.1), ed è la funzione che lascia invariato x se x è non negativo, mentre gli associa il valore $-x$ se x è negativo. Per rendersi conto del perché la frase "x privato del segno" sia sbagliata, osserviamo che se non diciamo quanto vale x non ha senso parlare del suo segno: quale è il segno di $-x$ se x vale -2? In altre parole, se il valore assoluto di x fosse veramente "x privato del segno", allora $|-x|$ sarebbe x, il che è falso se ad esempio $x = -1$. Il grafico della funzione modulo è il seguente:



Notiamo, *en passant*, che $|x - y|$ calcola la "distanza" tra il punto x e il punto y sulla retta reale, indipendentemente dal fatto che x si trovi prima o dopo y; vale a dire, $|x - y|$ è la lunghezza del segmento che congiunge x a y. La proprietà più importante della funzione modulo è la cosiddetta "disuguaglianza triangolare": per ogni x e y in $\mathbb{R}$ si ha

$$|x + y| \leq |x| + |y| \,. \tag{7.2}$$

[31] Altrettanto corretto è lo spostamento dell'uguale: x se $x > 0$ e $-x$ se $x \leq 0$.

[32] Si fa per dire...

Per quanto riguarda la differenza, il modulo si comporta in maniera un po' più complicata: per ogni x e y in $\mathbb{R}$ si ha

$$|x-y| \geq \big||x|-|y|\big| \,. \tag{7.3}$$

Infine, osserviamo che se a è non negativo, dire $|x| \leq a$ è equivalente a dire $-a \leq x \leq a$, mentre dire $|x| \geq a$ è equivalente a dire che $x \geq a$ oppure $x \leq -a$.

Esercizio 7.4. Risolvere le seguenti disequazioni:

$$\begin{gathered} |x| \leq x+2\,, \\ |x| \leq |x-2|\,, \\ |x-4| < |x-1|\,, \\ |2x-1| \leq |x-1|+|x-2|\,. \end{gathered}$$

Risposta 7.4: 1) La disequazione è equivalente ai due sistemi

$$\begin{cases} x \leq x+2 \\ x \geq 0 \end{cases} \qquad \begin{cases} -x \leq x+2 \\ x < 0 \end{cases}$$

Il primo ha come soluzione $x \geq 0$, dato che la prima disequazione è sempre risolta; il secondo ha come soluzione $-1 \leq x < 0$; facendo l'unione, la disequazione è verificata per $x \geq -1$. 2) Questo esercizio si risolve in maniera semplice rispondendo alla domanda "Quali sono i punti della retta reale che distano più da 2 che da 0?". 3) $x > \frac{5}{2}$. 4) $x \leq -1$.

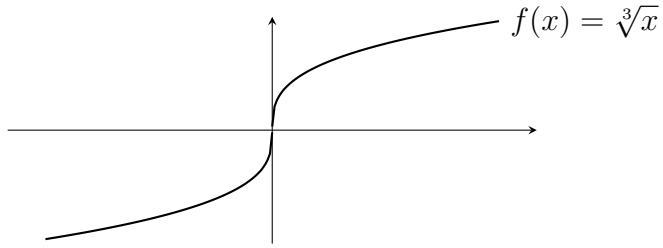
Riprendiamo ora il discorso sulle radici quadrate. Abbiamo visto (o meglio, abbiamo detto) che esiste ed è unica la radice quadrata (positiva) di un numero reale non negativo; in particolare, esiste $\sqrt{2}$, che tra l'altro sappiamo anche "costruire". L'esigenza di lavorare con le radici quadrate proviene dalla necessità di risolvere equazioni della forma $x^2 = a$, con $a \geq 0$. Ovviamente, il quadrato non è l'unica potenza che sappiamo fare. Ad esempio, possiamo considerare x^3, x^4, ed in generale x^n con n naturale. Così come all'elevazione al quadrato era legata la necessità di risolvere l'equazione $x^2 = 2$, è legittimo chiedersi se si possano risolvere equazioni della forma $x^n = a$, se tali equazioni siano risolubili, e sotto quali condizioni su n e su a. Ad esempio, chiediamoci se sia possibile determinare, se esiste, un numero reale x tale che $x^3 = 2$. In questo caso l'intuizione geometrica non ci aiuta più[33], ma fortunatamente abbiamo strumenti analitici sufficientemente potenti (o, equivalentemente, conosciamo sufficientemente bene la retta reale) per poter affermare che tale numero reale esiste, ed è anche unico. A guidarci nella "dimostrazione" dell'esistenza ed unicità di tale x è l'Esercizio 7.1: preso l'insieme

$$E = \{x \in \mathbb{R} : x \geq 0 \text{ e } x^3 \leq 2\}\,,$$

[33]Il problema di costruire x, detto anche "problema della duplicazione del cubo", è uno dei più antichi della geometria classica; il fatto che non abbia soluzione, ovvero che non esista un modo di costruire x con "riga e compasso", è stato dimostrato solo agli inizi del secolo diciannovesimo.

si dimostra facilmente[34] che E non è vuoto, che è limitato superiormente, e che il suo estremo superiore S soddisfa l'equazione $S^3 = 2$.

A questo punto, cosa ci impedisce di sostituire 2 con qualsiasi numero reale y? Niente: ed infatti siamo in grado di dimostrare che per ogni numero reale y esiste un unico numero reale x tale che $x^3 = y$. Osserviamo che siccome la terza potenza di un numero negativo è negativa, non abbiamo più la restrizione sulla positività di y. Chiamando x "radice cubica di y", o "radice terza di y", indicata in simboli con $x = \sqrt[3]{y}$, abbiamo così definito un'altra funzione: $\sqrt[3]{\cdot} : \mathbb{R} \to \mathbb{R}$, che a x associa la sua radice cubica. Il grafico di $\sqrt[3]{x}$ è il seguente:



A questo punto possiamo generalizzare: dato un numero naturale $n \geq 1$, e dato un numero reale y (positivo nel caso in cui n sia pari), esiste un unico numero reale x (dello stesso segno di y), tale che $x^n = y$. Definiamo x la "radice n-sima di y", e scriviamo $x = \sqrt[n]{y}$. Una notazione alternativa per la radice n-sima, anch'essa molto usata, è quella di elevare x all'esponente frazionario $\frac{1}{n}$. In questa maniera resta valida la legge, soddisfatta dalle potenze intere, secondo la quale $(x^a)^b = x^{ab}$. Infatti, per definizione di radice n-sima,

$$x = (x^{\frac{1}{n}})^n = x^{\frac{n}{n}} = x^1 \,.$$

Esercizio 7.5. Per quali x in $\mathbb{R}$ si ha

$$\sqrt[n]{x^n} = x \,?$$

Un'osservazione: il fatto che per n dispari la radice n-sima sia definita su tutto $\mathbb{R}$, e non su $[0, +\infty)$ come nel caso di n pari, è un caso fortunato: vedremo tra breve come in generale si potrà parlare di elevamento a potenza solo nel caso in cui la base sia non negativa.

Avendo definito la radice n-sima di x come x elevato alla $\frac{1}{n}$, possiamo ora definire la potenza con esponente razionale di un numero reale: dato $\frac{p}{q}$, numero razionale, definiamo

$$x^{\frac{p}{q}} = \sqrt[q]{x^p} \,. \tag{7.4}$$

[34] Provateci!

Dobbiamo fare delle restrizioni su p e q, o su x? Sì, dobbiamo obbligatoriamente chiedere che x sia non negativo, perché non possiamo a priori decidere della parità di q. Inoltre, se $\frac{p}{q}$ è negativo, dobbiamo chiedere che x sia diverso da zero. In definitiva, dato un qualsiasi numero razionale $\frac{p}{q}$ ed un qualsiasi numero reale non negativo x (con $x \neq 0$ se $\frac{p}{q}$ è negativo), è ben definito il numero reale $x^{\frac{p}{q}}$. Alternativamente, se $x \geq 0$, e y è un razionale, è ben definito il numero reale x^y (a patto che x sia diverso da zero se y è negativo). Come numero reale x^y soddisfa tutte le proprietà delle potenze; ad esempio $x^{y+z} = x^y \cdot x^z$, e $(x \cdot y)^z = x^z \cdot y^z$ (nel primo caso $x \geq 0$ e y e z sono razionali, mentre nel secondo x e y sono non negativi e z è razionale). Una delle proprietà più importanti dell'elevamento a potenza razionale è la monotonia: se $x \geq 1$, e se $y > z > 0$ sono due numeri razionali, allora $x^y > x^z$. In altre parole, fissato $x \geq 1$, la funzione che a y razionale positivo associa x^y cresce al crescere di y.

Esercizio 7.6. Verificare che se $0 \leq x \leq 1$, e se $y > z > 0$, allora $x^y < x^z$. Cosa si può dire di x^y e x^z se $y < z < 0$?

Siamo ora pronti a fare il passo successivo: dato x numero reale non negativo (per il momento maggiore di 1), e dato y reale (per il momento positivo), in che modo definire x^y? Ad esempio, quanto vale (o, meglio, cosa è) $2^{\sqrt{2}}$? Non essendo $\sqrt{2}$ un numero razionale, non possiamo più passare per le radici n-sime, come abbiamo fatto per la potenza ad esponente razionale, ma possiamo usare il fatto che i razionali sono densi in $\mathbb{R}$. Siccome per ogni $\varepsilon > 0$ esistono due numeri razionali α_ε e β_ε tali che $\sqrt{2} - \varepsilon < \alpha_\varepsilon < \sqrt{2} < \beta_\varepsilon < \sqrt{2} + \varepsilon$, se vogliamo che la monotonia si conservi anche per la potenza ad esponente reale, dovrà essere

$$2^{\alpha_\varepsilon} < 2^{\sqrt{2}} < 2^{\beta_\varepsilon} .$$

Siccome α_ε e β_ε sono arbitrariamente vicini, lo saranno anche 2^{α_ε} e 2^{β_ε}[35], che, in qualche modo, "stringono" tra di loro un unico numero reale, che è quello che definiamo come $2^{\sqrt{2}}$.

Questo procedimento, che può essere reso rigoroso, si può estendere a qualsiasi numero reale $x \geq 0$ ed a qualsiasi numero reale y: l'idea è proprio quella di usare da un lato la monotonia della potenza ad esponente razionale (che garantisce il fatto che, avvicinandosi ad y reale con dei razionali i risultati non "oscillino" troppo), e dall'altro la densità dei razionali nei reali (che permette di approssimare bene quanto si vuole y con dei razionali).

[35] Perché?

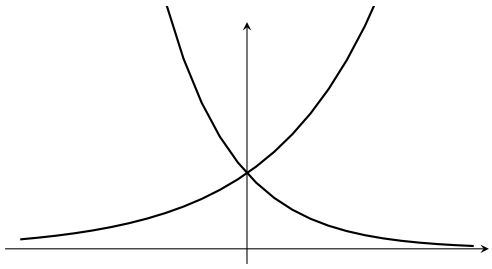
Una maniera alternativa, ma equivalente, di definire $2^{\sqrt{2}}$ è quella di usare in maniera diversa la monotonia e la densità dei razionali: detto

$$E = \{x \in \mathbb{R} : x = 2^q \text{ con } q \text{ razionale e } q \leq \sqrt{2}\},$$

è facile vedere che E è limitato superiormente (ad esempio ogni x di E è minore di 4). Definiamo allora

$$2^{\sqrt{2}} = \sup E\,.$$

Con lo stesso procedimento, fissato $x \geq 0$ e y reale è possibile definire x^y (sempre con l'avvertenza che se $x = 0$, allora y deve essere non negativo). La prossima figura mostra il grafico di 2^x (la curva crescente) e di $\left(\frac{1}{3}\right)^x$ (la curva decrescente). Le due curve si incrociano nel punto $x = 0$ e $y = 1$ (dal momento che $z^0 = 1$ per ogni z).



Come è evidente dal grafico[36], qualsiasi sia $A > 0$ con $A \neq 1$, la curva A^x è definita su tutto $\mathbb{R}$ e ha come immagine l'intervallo illimitato $(0, +\infty)$.

Esercizio 7.7. Dimostrare che se $f(x) = 2^x$, detto $E = \mathrm{Im}(f) = f(\mathbb{R})$, allora $\inf E = 0$ e $\sup E = +\infty$.

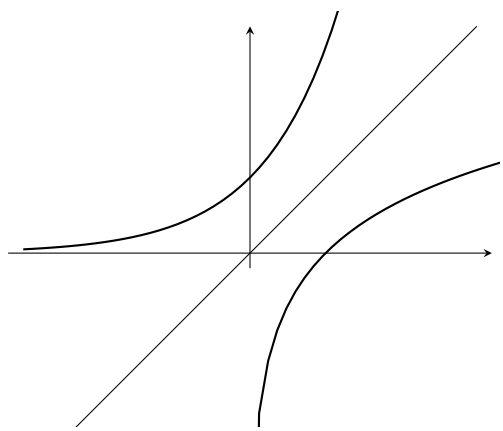
Risposta 7.7: L'insieme E contiene l'insieme dei numeri naturali della forma 2^n, e tale insieme è illimitato superiormente essendo $2^n > n$ per ogni n in $\mathbb{N}$ (dimostrarlo per induzione!). Per mostrare che 0 è l'estremo inferiore, si osservi che E contiene solo numeri positivi (cosicché l'insieme dei minoranti di E contiene $(-\infty, 0]$) e numeri della forma 2^{-n}, che diventano arbitrariamente piccoli al crescere di n.

Sia ora $f : A \to \mathbb{R}$ una funzione. Per definizione, f associa ad ogni elemento x di A un unico numero reale $y = f(x)$, che ovviamente appartiene ad $f(A)$. Una domanda che ci si può porre è se, fissato y in $f(A)$ (per il quale esiste sicuramente **almeno** un x di A tale che $y = f(x)$), esista un **unico** x di A tale che $y = f(x)$. La risposta è, in generale, negativa. Ad esempio, la funzione f definita su $\mathbb{R}$ da $f(x) = 1$ (ovvero la funzione che assume costantemente il valore 1) non verifica questa proprietà, dato che a $y = 1$ corrisponde non un unico x, bensì infiniti. Un altro esempio è la funzione

[36] Il che vuol dire che potrebbe anche non essere vero...

$f(x) = x^2$, definita per x reale. Infatti, se y è un qualsiasi numero reale positivo, esistono esattamente due numeri reali, $\sqrt{y}$ e $-\sqrt{y}$, che hanno come immagine y. Analogamente per $f(x) = |x|$. Le funzioni $f(x) = x$, $f(x) = x^3$ e $f(x) = 2^x$, invece, godono di questa proprietà: ogni numero reale nella loro immagine proviene da uno ed un solo numero del dominio.

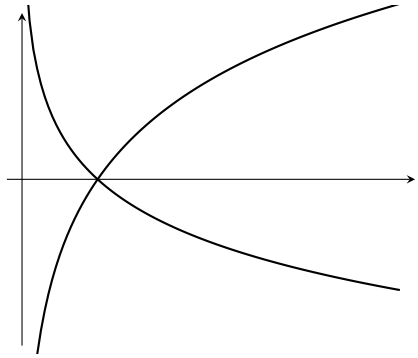
Definizione 7.8. Sia $f : A \to \mathbb{R}$ una funzione. Se per ogni y di $f(A)$ esiste un unico x in A tale che $y = f(x)$, la funzione f si dice **invertibile**. Data una funzione invertibile resta determinata una funzione $g : f(A) \to A$, detta funzione **inversa**, che è la funzione che a y in $f(A)$ associa l'unico x di A tale che $y = f(x)$. La funzione g è dunque caratterizzata dall'essere tale che, per ogni x di A, $g(f(x)) = x$, mentre per ogni y di $f(A)$ si ha $f(g(y)) = y$. In genere, con una notazione che può trarre in inganno, la funzione inversa viene denotata con f^{-1}. Dal punto di vista del grafico, dato il grafico di una funzione invertibile $y = f(x)$, il grafico della funzione inversa (sugli stessi assi con i nomi cambiati), si ottiene "ribaltando" il grafico rispetto alla retta $y = x$; si veda la figura.



Definizione 7.9. Sia $A > 0$, con $A \neq 1$. La funzione A^x, definita su $\mathbb{R}$ a valori in $(0, +\infty)$, è invertibile. Definiamo **logaritmo in base** A la funzione inversa di A^x, e indichiamo tale funzione con $\log_A$. Come funzione, dunque, $\log_A : (0, +\infty) \to \mathbb{R}$ è caratterizzata dall'essere

$$\log_A(A^x) = x\,, \quad \forall x \in \mathbb{R}\,, \qquad A^{\log_A(y)} = y\,, \quad \forall y \in (0, +\infty)\,.$$

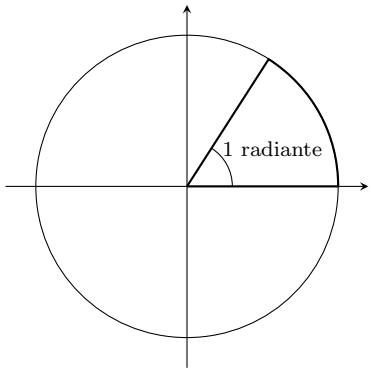
Il grafico che segue rappresenta le curve $\log_2(x)$ (la curva crescente) e $\log_{\frac{1}{3}}(x)$ (la curva decrescente).



Le due curve si intersecano nel punto $x = 1$, $y = 0$. Una delle proprietà "storicamente" più importanti dei logaritmi, conseguenza del fatto che $A^{x+y} = A^x \cdot A^y$, è il trasformare i prodotti in somme[37]:

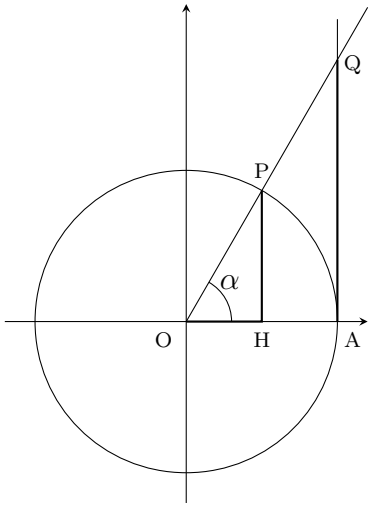
$$\log_A(x \cdot y) = \log_A(x) + \log_A(y)\,, \quad \forall x, y \in (0, +\infty)\,.$$

Nell'introdurre radici n-sime, potenze ad esponente razionale e reale, funzioni esponenziali e logaritmi, siamo partiti dal numero "costruibile" $\sqrt{2}$, la diagonale del quadrato di lato 1. Esistono altri modi di "costruire" esplicitamente funzioni. Uno di questi consiste nell'utilizzare la circonferenza di centro l'origine e raggio 1, e gli angoli misurati in radianti. Ricordiamo che un angolo al centro di una circonferenza di raggio 1 misura un **radiante** se sottende un arco di circonferenza di lunghezza 1. Con questa convenzione, l'angolo retto misura $\frac{\pi}{2}$ radianti, l'angolo piatto π radianti, mentre l'angolo giro misura 2π radianti. Inoltre, gli angoli sono misurati muovendosi lungo la circonferenza in verso antiorario a partire dal punto $x = 1$, $y = 0$. Nella figura che segue, è mostrato l'angolo che misura un radiante.



[37]E non le somme in prodotti, come talvolta si legge nei compiti scritti. Per esercizio, dimostrare che se il logaritmo della somma fosse il prodotto dei logaritmi, la funzione logaritmo sarebbe identicamente nulla...

Sia ora α un angolo (misurato in radianti a partire dal punto A di coordinate $x = 1$, $y = 0$) compreso tra 0 e 2π, e consideriamo il punto P tale che l'arco di circonferenza AP abbia lunghezza α. A partire da P tracciamo la perpendicolare all'asse x, che incontra tale asse nel punto H. Definiamo **seno** dell'angolo α, indicato con $\mathrm{sen}(\alpha)$, la lunghezza con segno del segmento PH, e **coseno** dell'angolo α, indicato con $\cos(\alpha)$, la lunghezza con segno del segmento OH. Infine, tracciata la tangente alla circonferenza in A, sia Q il punto di intersezione tra la semiretta OP e la tangente. Definiamo **tangente** di α, indicata con $\mathrm{tg}(\alpha)$, la lunghezza con segno del segmento AQ.



Dalla similitudine dei triangoli OPH e OQA si ottiene subito la relazione fondamentale tra seno, coseno e tangente:

$$\mathrm{tg}(\alpha) = \frac{\mathrm{sen}(\alpha)}{\cos(\alpha)}\,, \quad \forall \alpha \neq \frac{\pi}{2}\,,\ \frac{3\pi}{2}\,.$$

Abbiamo così definito tre funzioni: $\mathrm{sen} : [0, 2\pi] \to \mathbb{R}$, $\cos : [0, 2\pi] \to \mathbb{R}$ e $\mathrm{tg} : [0, 2\pi] \setminus \{\frac{\pi}{2}, \frac{3\pi}{2}\} \to \mathbb{R}$. È facile vedere che $\mathrm{Im}(\mathrm{sen}) = [-1, 1] = \mathrm{Im}(\cos)$, cosicché seno e coseno soddisfano le maggiorazioni fondamentali

$$|\mathrm{sen}(\alpha)| \leq 1\,, \qquad |\cos(\alpha)| \leq 1\,,$$

mentre $\mathrm{Im}(\mathrm{tg}) = \mathbb{R}$. Infine, grazie al Teorema di Pitagora,

$$\mathrm{sen}^2(\alpha) + \cos^2(\alpha) = 1\,.$$

Che succede se l'angolo α non è compreso tra 0 e 2π? Ad esempio, se $\alpha = 2\pi + 1$? È chiaro che il punto P che ottieniamo è lo stesso che otteniamo per $\alpha = 1$, e quindi

$\operatorname{sen}(2\pi + 1) = \operatorname{sen}(1)$, e lo stesso vale per coseno e tangente. Più in generale, se k è un intero relativo, e α è un numero reale qualsiasi, si ha

$$\operatorname{sen}(\alpha + 2k\pi) = \operatorname{sen}(\alpha)\,, \quad \cos(\alpha + 2k\pi) = \cos(\alpha)\,, \quad \operatorname{tg}(\alpha + 2k\pi) = \operatorname{tg}(\alpha)\,.$$

Queste proprietà si esprimono dicendo che le funzioni seno, coseno e tangente sono **periodiche** di **periodo** 2π. Pertanto, le funzioni seno e coseno si possono considerare definite su tutto $\mathbb{R}$ (sempre con la stessa immagine), mentre la tangente è definita su $\mathbb{R}$ privato dei multipli dispari di $\frac{\pi}{2}$. Una volta estese ai reali, le funzioni seno e coseno soddisfano le relazioni, dette formule di addizione,

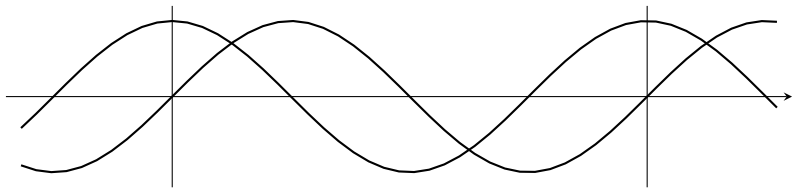
$$\begin{aligned} \operatorname{sen}(\alpha \pm \beta) &= \operatorname{sen}(\alpha)\cos(\beta) \pm \operatorname{sen}(\beta)\cos(\alpha)\,, \\ \cos(\alpha \pm \beta) &= \cos(\alpha)\cos(\beta) \mp \operatorname{sen}(\alpha)\operatorname{sen}(\beta)\,. \end{aligned}$$

Partendo da queste formule, e dalla relazione fondamentale tra seno, coseno e tangente, è facile vedere che

$$\operatorname{tg}(\alpha + k\pi) = \operatorname{tg}(\alpha)\,, \quad \forall k \in \mathbb{Z}\,,$$

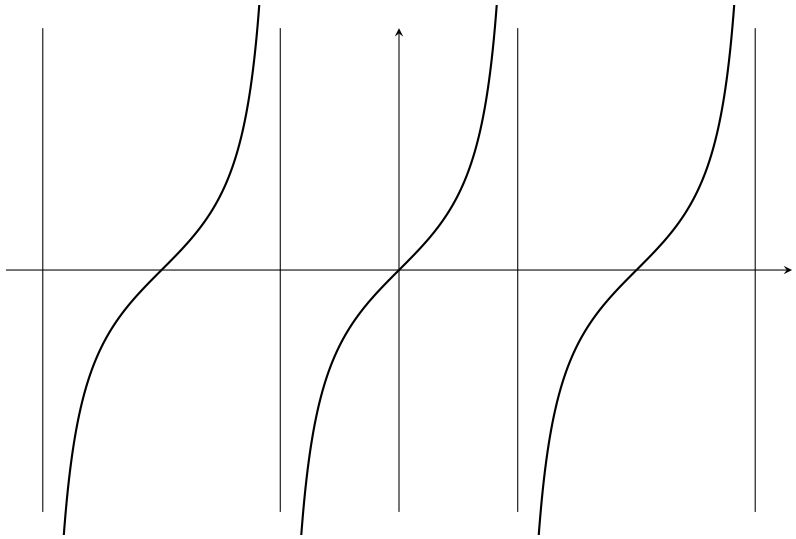
e quindi la tangente è periodica di periodo π.

Graficamente abbiamo la seguente situazione[38]:

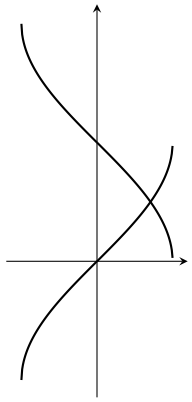
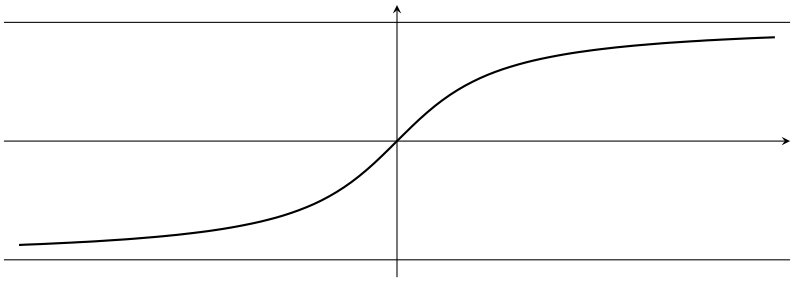


mentre per la tangente abbiamo:

[38] Al lettore decidere quali delle due curve è il grafico di $\operatorname{sen}(x)$, quale di $\cos(x)$, sapendo che la prima verticale è $x = 0$.



Essendo periodiche (e quindi assumendo infinite volte lo stesso valore), le funzioni seno, coseno e tangente non sono ovviamente invertibili. Se, però, restringiamo nella maniera seguente l'insieme di definizione: $\operatorname{sen} : [-\frac{\pi}{2}, \frac{\pi}{2}] \to [-1, 1]$, $\cos : [0, \pi] \to [-1, 1]$ e $\operatorname{tg} : (-\frac{\pi}{2}, \frac{\pi}{2}) \to \mathbb{R}$, le funzioni risultano essere invertibili. Abbiamo allora l'**arcoseno**, indicato con $\operatorname{arc\,sen}$, l'**arcocoseno**, indicato con $\operatorname{arc\,cos}$, e l'**arcotangente**, indicato con $\operatorname{arc\,tg}$. Le prime due funzioni sono definite su $[-1, 1]$, a valori in $[-\frac{\pi}{2}, \frac{\pi}{2}]$ e $[0, \pi]$ rispettivamente, mentre l'arcotangente è definita su tutto $\mathbb{R}$ a valori in $(-\frac{\pi}{2}, \frac{\pi}{2})$. Graficamente, abbiamo:



per arcoseno ed arcocoseno, mentre il grafico dell'arcotangente è: