

Titolo

Matematica per la vita. Anche dove non te l'aspetti

Autore

M. Degiovanni, R. Lucchetti, A. Marzocchi, M. Paolini

Volume edito a cura della

FONDAZIONE ACHILLE E GIULIA BOROLI

Progetto grafico

Studio CREE – Milano

Realizzazione editoriale

REDINT Studio s.r.l.

Nessuna parte di questo libro può essere riprodotta o trasmessa in qualsiasi forma o con qualsiasi mezzo elettronico, meccanico o altro senza l'autorizzazione scritta dei proprietari dei diritti e dell'editore

info@fondazioneaegboroli.com

www.fondazioneaegboroli.it

On line i libri della collana Homo Sapiens

© 2009 Fondazione Achille e Giulia Boroli

Finito di stampare nel mese di ottobre 2009
a cura di DEAPRINTING – Novara

Edizione fuori commercio

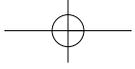
HOMO SAPIENS

M. DEGIOVANNI, R. LUCCHETTI,
A. MARZOCCHI, M. PAOLINI

MATEMATICA PER LA VITA

ANCHE DOVE NON TE L'ASPETTI

FONDAZIONE ACHILLE E GIULIA BOROLI

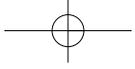


LA FONDAZIONE ACHILLE E GIULIA BOROLI

Nel 1998 Achille Boroli, oggi presidente onorario di De Agostini Editore Spa, ha fondato l'ente che porta il suo nome e quello della moglie Giulia e lo ha dotato di un importante fondo con capitali personali; in questa iniziativa si manifesta la precisa volontà del fondatore di continuare a essere concretamente presente all'interno della società civile con attività di supporto a enti pubblici e privati, laici e religiosi, già operanti nel campo della ricerca scientifica, della charity e della cultura nel senso più ampio del termine. In questo ambito, e più precisamente in conformità con uno degli obiettivi statutari, è nata questa iniziativa editoriale che esprime la volontà di supportare la conoscenza e l'approfondimento dei grandi temi dell'attualità da parte delle più giovani classi di età, al fine di favorire la comprensione del mondo sempre più complesso e problematico in cui viviamo.

Questa iniziativa si affianca a un'altra attività ormai tradizionale della Fondazione, che assegna borse di studio in favore degli studenti meritevoli per favorirne l'iscrizione all'Università.

Editore di successo, animato da una fede intatta nei valori della cultura e della lettura come strumento insuperato di comunicazione, Achille Boroli ha fortemente voluto che la Fondazione realizzasse la collana di libri che oggi presentiamo ai giovani, fiduciosi che l'informazione, la libera riflessione e il pensiero contribuiranno alla formazione dei cittadini del futuro.



SOMMARIO

Prefazione

Alcune strutture

- 11 1. Diamo i numeri
- 19 2. Dimensione
- 21 3. Continuità
- 22 3.1 *Pitagora e i pitagorici*
- 24 4. La teoria della relatività generale
- 26 4.1 *Spazio, tempo e superfici*
- 31 4.2 *Le geometrie non euclidee*

Modelli e previsioni

- 36 1. L'esattezza nel modello matematico
e gli errori sperimentali
- 40 2. Un modello deterministico: la meccanica newtoniana
- 47 3. La fluidodinamica
- 50 4. Previsione dei risultati
- 52 5. Dalla fluidodinamica ai sistemi complessi
- 55 6. Frattali e complessità
- 59 7. Stabilità

La gestione razionale del caso

- 64 1. Un po' di storia
- 67 2. La probabilità classica
- 70 3. Bernoulli e la probabilità frequentista
- 74 4. Eventi e loro operazioni
- 77 5. Dalla probabilità classica alla statistica: le distribuzioni
- 83 6. La probabilità soggettiva e l'impostazione assiomatica
- 87 7. La probabilità condizionata e le applicazioni

Computer e soluzioni approssimate

- 93 1. Il computer e le quattro operazioni
- 93 1.1 *La notazione scientifica*
- 94 1.2 *Gli errori di arrotondamento*
- 96 1.3 *Non solo conti*
- 97 2. Alla ricerca del minimo
- 98 2.1 *Minimi e punti stazionari*

99	2.2	<i>Bolle di sapone</i>
103	2.3	<i>Capire il problema prima di risolverlo</i>
104	2.4	<i>L'importanza di semplificare e generalizzare</i>
107	2.5	<i>Fiocchi di neve</i>
108	2.6	<i>Minimi e convessità</i>
110	3.	Il calcolo numerico
114	3.1	<i>Dal discreto al continuo e ritorno</i>
115	4.	Elaborazione di immagini
117	5.	Sintesi di immagini e realtà virtuale

La matematica nell'arte

123	1.	La matematica nell'architettura e nelle arti figurative
136	2.	Gli elementi matematici nella musica

Giochi e applicazioni

145	1.	Alcune considerazioni sul gioco
147	2.	Alcuni esempi
152	3.	Giochi in forma estesa
153	3.1	<i>L'albero del gioco</i>
157	4.	La teoria non cooperativa
164	5.	La teoria cooperativa
172	6.	Conclusioni

La matematica nelle scienze umane e della vita

175	1.	La crescita delle popolazioni
180	2.	Popolazioni e ambiente
184	3.	Cinetica dei farmaci
187	4.	Predatori e prede

191 **Bibliografia**

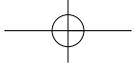
192 **Elenco dei simboli**

193 **Indice analitico**

197 **Indice dei nomi**

PREFAZIONE

Il momento in cui, ne *Il Saggiatore*, Galileo asserisce che il libro della natura è scritto “in lingua matematica” segna, come è noto, un punto di svolta nella concezione della scienza. Al di là dell’affermazione di principio, per circa un secolo il processo di matematizzazione coinvolge, nell’ambito delle scienze della natura, essenzialmente la fisica e, a partire dal XVIII secolo, anche la chimica. I successi conseguiti nel corso di XVIII e XIX secolo inducono, fra fine ’800 e inizio ’900, un tentativo a largo spettro di trattare con linguaggio matematico i più svariati ambiti di studio, anche quelli che, coinvolgendo il comportamento umano, sembravano appannaggio di filosofia e religione. Arriviamo così ai giorni nostri in cui nei campi più svariati sono proposti criteri di valutazione che spesso si traducono in sistemi di numeri, con tutto il bene e il male che si possono immaginare. Lo scopo di questo volume è quello di illustrare come la matematica interagisca con alcuni aspetti della realtà che ci circonda. Il capitolo iniziale è dovuto congiuntamente a Marco Degiovanni e Maurizio Paolini, mentre i capitoli “Modelli e previsioni” e “La gestione razionale del caso” sono di Alfredo Marzocchi. Maurizio Paolini è autore del capitolo “Computer e soluzioni approssimate”, mentre a Roberto Lucchetti si deve il capitolo “Giochi e applicazioni” e a Marco Degiovanni quello “La matematica nelle scienze umane e della vita”. Nel capitolo dedicato alla “Matematica nell’arte”, Alfredo Marzocchi si è occupato di architettura e arti figurative e Marco Degiovanni di musica. Infine Marco Degiovanni ha curato una revisione generale dell’intera opera.



ALCUNE STRUTTURE

1. Diamo i numeri

Se soffriamo di insonnia e la notte non riusciamo a dormire, possiamo contare le pecore... Contare è forse il primo processo matematico che un essere umano compie, magari senza esserne consapevole, ed è anche la prima fonte di infinito – certo, anche solo in senso potenziale –, poiché non si pongono limiti a quanto a lungo si può contare: è possibile andare avanti quanto si vuole. Peraltro, disporre di una quantità illimitata di numeri è fondamentale per poter operare in modo un minimo confortevole.

Vediamo di spiegarci con un esempio. Supponiamo di avere una calcolatrice che può considerare solo numeri con due cifre significative, ossia numeri costituiti da due cifre a piacere seguite unicamente da zeri. I veri calcolatori adottano degli standard che contemplano una quindicina di cifre significative, ma la sostanza del problema non cambia. Tornando alla nostra rudimentale calcolatrice, abbiamo quindi a disposizione tutti i numeri fino a 99. I successivi sono, nell'ordine, 100, 110, 120 fino a 990. Si potrebbe vedere che cosa capita da 1000 in poi, ma per il nostro esempio non è necessario. Inevitabilmente, la nostra calcolatrice opera degli arrotondamenti. Per esempio per lei risulta

$$120 + 3 = 120, \quad 120 + 4 = 120, \quad 120 + 7 = 130$$

perché, non disponendo dei numeri 123, 124 e 127, deve scegliere fra 120 e 130 e opta minimizzando l'errore.

Ora, sappiamo che la somma gode della fondamentale proprietà associativa secondo cui, dati tre numeri x , y e z , si ha sempre

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

Questo fatto autorizza a scrivere semplicemente

$$x + y + z$$

senza ulteriori specificazioni, visto che i due modi di determinare il risultato producono il medesimo effetto. Si tratta di una proprietà senza la quale le espressioni sarebbero infarcite di parentesi e a cui non si rinuncia nemmeno in contesti algebrici molto generali (per esempio, in teoria dei gruppi). Che cosa capita con la nostra calcolatrice, se scegliamo $x = 120$, $y = 3$ e $z = 4$? Da un lato risulta

$$(120 + 3) + 4 = 120 + 4 = 120$$

dall'altro

$$120 + (3+4) = 120 + 7 = 130$$

Questo significa che, per la nostra calcolatrice, non vale la proprietà associativa! In effetti si tratta di un problema che affligge tutti i calcolatori, costretti come sono a procedere ad arrotondamenti, di qualunque potenza essi siano. È sufficiente costruire esempi con numeri abbastanza grandi rispetto al numero di cifre significative che il computer stesso può considerare. Si tratta di un problema insito nella "matematica del finito".

Questa considerazione dovrebbe mettere in guardia da una tentazione oggi ricorrente: dal momento che nelle applicazioni concrete della matematica l'universo numerico è necessariamente finito, per caratteristica intrinseca dei calcolatori, non si può fare a meno dell'infinito? In fondo, in qualunque problema concreto, sono immessi dati gestibili da un computer e si ottengono risultati appartenenti al medesimo universo. È la tipica tentazione del corto circuito, che affligge molti aspetti della vita umana. Una prima risposta, pragmatica, è che risulta più efficiente prendere come riferimento un modello infinito con buone proprietà e poi pensare che il computer commette errori la cui ampiezza andrà stimata, piuttosto che ragionare direttamente in un universo finito. Una seconda risposta è che i numeri bruti dicono poco, se non interagiscono con le idee. E queste emergono se si dispone di un quadro interpretativo stimolante.

Se vogliamo, è anche un primo caso della contrapposizione fra elementare e maneggevole, destinata a ricomparire più volte. La struttura finita è certo elementare, ma è meno maneggevole, mentre quella infinita comporta uno sforzo concettuale maggiore, ma al suo interno si opera con più comodità.

I numeri di cui abbiamo parlato finora, quelli che si incontrano durante il processo del contare: sono i *numeri naturali*. Lo 0 era sconosciuto nell'antichità classica, ciononostante risulta estremamente conveniente includerlo tra i numeri naturali; un insieme di pecore può anche essere vuoto, nel qual caso il processo del contare le pecore si banalizza in zero operazioni!

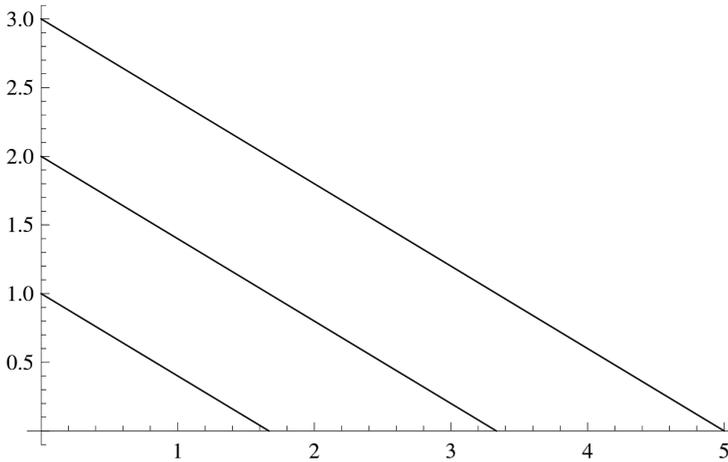


Figura 1. La divisione di un segmento lungo 5 unità in 3 parti uguali. Si riporta, per esempio, sulla perpendicolare per un estremo, l'unità per tre volte e si congiunge l'ultimo punto ottenuto con l'altro estremo. Le parallele a tale congiungente per i primi due punti ottenuti dividono il segmento originario in tre parti uguali.

In conclusione abbiamo un primo insieme di numeri, i numeri naturali, che indicheremo con \mathbb{N} , che comincia con lo 0 e che associamo al generico processo del contare. All'interno dei numeri naturali disponiamo di due operazioni, la somma e il prodotto, il che automaticamente pone la questione delle operazioni inverse, sottrazione e divisione. Entrambe non sono sempre possibili, ma le due impossibilità furono vissute fin dall'antichità in modo diverso. Quanto alla sottrazione, parve del tutto naturale che non si potesse togliere più di quello che era già presente. D'altra parte, il numero fu associato non solo al processo del contare, ma anche a quello del misurare. E nell'ambito della geometria, che si stava allora sviluppando, emergevano invece suggestioni diverse. Se si possiede un segmento lungo 5 unità, esiste una procedura geometrica che con-

sente, per esempio, di dividerlo esattamente in 3 parti uguali. Si tratta di applicare opportunamente il teorema di Talete¹, come indicato in Figura 1. Come fare affinché gli enti numerici non si trovino in posizione di inferiorità rispetto a quelli geometrici?

Una risposta geniale, che introduce idee destinate ad accompagnare lo sviluppo della matematica fino ai giorni nostri, fu fornita dalla scuola di Pitagora². Se l'operazione "5 diviso 3" è impossibile, si conferisce all'operazione impossibile la dignità di oggetto matematico all'interno di un nuovo ambito e si vede se è possibile operare in modo coerente nel nuovo contesto. È la nascita dell'insieme dei numeri razionali (positivi, perché quelli negativi appaiono solo in epoca moderna) e delle frazioni, che non sono altro che divisioni, spesso impossibili, solo indicate, ma su cui si può operare in modo coerente. Per esempio, $10/7$ significa "10 diviso 7". Nel nuovo contesto, un numero naturale come 8 viene identificato con $8/1$, ossia "8 diviso 1". Quindi, nell'ambito dei numeri concepiti dalla scuola pitagorica, si può fare 5 diviso 3 e il risultato è (banalmente) $5/3$.

C'è ancora una sottile differenza tra frazione e numero razionale. È chiaro che divisioni diverse possono produrre il medesimo risultato. Per esempio, "15 diviso 3" dà il medesimo risultato di "30 diviso 6"; è allora necessario introdurre il concetto di frazioni equivalenti e un singolo numero razionale è una collezione di frazioni, tutte equivalenti fra loro e atte a rappresentarlo. È il primo esempio di *relazione di equivalenza*, un concetto pure destinato ad attraversare lo sviluppo della matematica fino ai giorni nostri.

In sintesi, la scuola pitagorica mette in atto un ulteriore passaggio dall'elementare al maneggevole. I numeri naturali sono più elementari dei numeri razionali, ma questi consentono di fare più operazioni.

Il problema delle operazioni impossibili ricompare più volte nello sviluppo della matematica. Per esempio, non esiste nessun numero razionale il cui quadrato sia uguale a 2. Siccome in ambito geometrico è possibile considerare il quadrato di lato unitario e la lunghezza della sua diagonale dovrebbe proprio essere un numero il cui quadrato è 2, il fatto fu vissuto dai pitagorici come un *vulnus* alla perfezione dell'edificio dei numeri razionali (sembra che fosse fatto

¹ Talete (Mileto, 640/624 a.C. – ca. 547 a.C.), filosofo, astronomo e matematico greco. È comunemente considerato il primo filosofo della storia occidentale.

² Pitagora (Samo, ca. 575 a.C. – Metaponto, ca. 495 a.C.), matematico, legislatore e filosofo greco. Fu forse il primo ad elaborare una visione matematizzata della realtà, destinata a influenzare profondamente molti studi successivi.

pure divieto agli aderenti di diffondere la notizia). Anche qui l'unica soluzione è conferire dignità di numero all'operazione impossibile. Nascono così le radici, come $\sqrt{2}$, $\sqrt[3]{4}$, che non sono altro che operazioni inverse dell'elevamento a potenza lasciate indicate, e i numeri irrazionali. Nel nuovo contesto la lunghezza della diagonale del quadrato di lato unitario è (banalmente) $\sqrt{2}$. A differenza della divisione, non fu trovato subito un quadro soddisfacente in cui operare con le nuove entità, e per lungo tempo agli enti geometrici fu conferita dignità concettuale maggiore rispetto a quelli numerici. Solo in epoca rinascimentale, sotto la spinta delle esigenze applicative, si ebbe una fioritura di nuovi studi sugli enti numerici che fu perfezionata nel corso del XIX secolo.

Peraltro la geometria forniva ulteriori sollecitazioni. Quanto è lunga la circonferenza di diametro unitario? Anche qui non fu trovata (infatti non esiste) una risposta che riconducesse il tutto alle operazioni già considerate. Il meglio che si può fare è conferire dignità di numero all'inesistente risposta, controllando se è poi possibile continuare a operare in modo coerente. Siccome in questo caso la convinzione è che vi sia una e una sola entità, si può impiegare un simbolo specifico per denotarla. Fu scelto π . Si tratta di un passaggio che sottointende anche una sorta di presa di possesso, un po' come nel racconto biblico della Genesi, in cui l'atto di dare il nome agli animali suggella la supremazia dell'uomo: "Così l'uomo impose nomi a tutto il bestiame, / a tutti gli uccelli del cielo / e a tutte le bestie selvatiche" (Gen 2, 20).

Tornando alla matematica, a questo punto è opportuna una riflessione sulla frase *controllare che sia possibile continuare a operare in modo coerente*, anche per non dare l'impressione che i matematici, in un empito di gioiosa follia, conferiscano dignità di ente matematico a qualunque operazione impossibile in un dato contesto. Il problema, volendo aumentare la maneggevolezza dello strumento, è quello di non perdere le proprietà formali già acquisite.

Nel caso di frazioni, radici e π tutto funziona, ma talvolta non c'è sbocco. Per esempio, il numero 0 non ammette reciproco, ossia non esiste nessun numero che moltiplicato per 0 dia per risultato 1.

Possiamo dare dignità di numero, in un nuovo contesto allargato, a "reciproco di 0"? Se denotiamo con u questa nuova entità e se nel nuovo ambito continua a valere la proprietà distributiva del prodotto rispetto alla somma, ossia

$$(x + y) \cdot z = x \cdot z + y \cdot z$$

allora risulta

$$1 = 0 \cdot u = (0+0) \cdot u = 0 \cdot u + 0 \cdot u = 1+1= 2$$

che è assurdo, perché $1 \neq 2$. Di conseguenza non esiste nessun allargamento che consenta di dare dignità di numero a “reciproco di 0” e che sia compatibile con la proprietà distributiva. Siccome nessuno ha mai voluto rinunciare alla proprietà distributiva, il reciproco di 0 non si può fare, nemmeno nei contesti più allargati.

Ritornando a radici e π , l'elaborazione di un quadro coerente in cui inserire tali numeri, e anche altri, richiese più di due millenni e si perfezionò nel biennio 1872-73 con i fondamentali lavori di R. Dedekind³, G. Cantor⁴ e K. Weierstrass⁵, che portarono a definire esattamente il concetto di *numero reale*. In questo modo si ricostituiva quella perfetta corrispondenza fra enti numerici ed enti geometrici, vagheggiata dai pitagorici e vanamente inseguita per tanto tempo. Il fatto poi che si potesse pervenire al concetto di numero reale per costruzioni successive, partendo dai numeri naturali e passando per quelli razionali, rivelò una sorprendente possibilità di ricondurre il continuo al discreto, constatazione inaspettata, visto che per millenni il discreto e il continuo erano stati concepiti come realtà inconciliabili. Quello che c'è dietro è il formidabile linguaggio della teoria degli insiemi, elaborata da Cantor e perfezionata nella prima metà del '900; essa fornisce gli strumenti per gestire correttamente il concetto di infinito attuale, ossia la situazione in cui infiniti oggetti sono considerati contemporaneamente. Così come il numero razionale $5/3$ è univocamente determinato dai due numeri naturali 5 e 3, ogni numero reale è univocamente determinato, per esempio, dall'insieme dei numeri razionali di lui maggiori. La costruzione dei numeri reali si basa su questo fatto. Il punto è che, per determinare anche un solo numero reale, occorre considerare contemporaneamente tutti i numeri razionali di lui maggiori, mentre per determinare un numero razionale ba-

³ Julius Wilhelm Richard Dedekind (Braunschweig, 1831 – Braunschweig, 1916), matematico tedesco. Ha fornito contributi importanti in teoria dei numeri e ha proposto una delle prime presentazioni coerenti del sistema dei numeri reali.

⁴ Georg Ferdinand Ludwig Philipp Cantor (San Pietroburgo, 1845 – Halle, 1918), matematico tedesco. Ha fornito importanti contributi in analisi matematica, nell'ambito dei quali ha elaborato quella che oggi si chiama *teoria degli insiemi*, che, tra l'altro, è il primo tentativo riuscito di gestione coerente dell'infinito attuale dopo oltre due millenni di timori e prevenzioni.

⁵ Karl Theodor Wilhelm Weierstrass (Ostenfelde, 1815 – Berlino, 1897), matematico tedesco. Con lui si perfeziona la prima sistemazione rigorosa del calcolo infinitesimale.

stavano due numeri naturali. Solo un linguaggio capace di gestire l'infinito attuale, come la teoria degli insiemi, poteva consentire un simile passaggio. La costruzione che porta a ricondurre i numeri reali ai naturali fu chiamata *arimetizzazione del continuo*, o *arimetizzazione dell'analisi*, in riferimento al calcolo infinitesimale, e spinse il matematico tedesco L. Kronecker⁶ a proferire la celebre frase: "Dio fece i numeri naturali; tutto il resto è opera dell'uomo". A quel punto l'attenzione si rifocalizzò sui numeri naturali e di lì a poco il matematico italiano G. Peano⁷ formulò i suoi famosi cinque assiomi⁸ che tuttora costituiscono il punto di riferimento per le proprietà basilari dei numeri naturali. Tra le conseguenze di questi cinque assiomi si ritrova in particolare l'infinità dei numeri, ovvero il fatto che il processo del contare (che nell'ambito degli assiomi di Peano consiste nel continuare a trovare il successore) non ha mai termine: ogni numero ha un successore (assioma 2), quindi posso continuare a contare e trovo sempre numeri nuovi (assioma 3).

Ritornando alla costituzione dell'insieme \mathbb{R} dei numeri reali, ancora una volta veniva messo in atto un passaggio dall'elementare al maneggevole, perché i numeri razionali sono più elementari dei numeri reali, che però costituiscono un ambiente all'interno del quale sono possibili più operazioni. Anzi, l'ambiente dei numeri reali è così confortevole, che anche i modelli sull'evoluzione delle popolazioni, di cui parleremo nel capitolo "La matematica nelle scienze umane e della vita", descrivono la numerosità delle popolazioni stesse utilizzando numeri reali. Questo significa che, al momento di interpretare i risultati, ci dovrà essere un processo di approssimazione per cui una risposta come "una popolazione di 1000,4 persone" verrà convertita (plausibilmente) in "una popolazione di 1000 persone".

A questo punto i tempi sono maturi per un ritorno all'universo finito del computer, con cui dobbiamo comunque fare i conti (in più di un senso). La mediazione è costituita dal concetto di numero decimale, su cui adesso ci soffermeremo. Al mondo del computer e al

⁶ Leopold Kronecker (Liegnitz, 1823 – Berlino, 1891), matematico e logico tedesco. Importanti i suoi contributi all'algebra e alla teoria dei numeri.

⁷ Giuseppe Peano (Spinetta di Cuneo, 1858 – Torino, 1932), matematico e logico italiano. Ha dato contributi importanti all'analisi matematica, ma è soprattutto noto per la lucidità con cui ha trattato le questioni che, a cavallo fra XIX e XX secolo, coinvolgevano i fondamenti della matematica

⁸ 1. Esiste un numero naturale, 0 (zero); 2. Ogni numero naturale ha un numero naturale successore; 3. Numeri diversi hanno successori diversi; 4. Lo zero non è il successore di alcun numero naturale; 5. Ogni insieme di numeri naturali che contenga lo zero e il successore di ogni proprio elemento coincide con l'intero insieme dei numeri naturali.

modo in cui interagisce con i modelli matematici è dedicato il capitolo “Computer e soluzioni approssimate”.

Quando diciamo che pi greco (π) è uguale a 3,14 avvengono molte cose simultaneamente, che cercheremo di approfondire un po'. Tanto per cominciare si tratta di un'affermazione falsa: sappiamo infatti che in verità π è solo approssimativamente uguale a 3,14. Un altro numero che si comporta in modo simile per certi aspetti è $2/7$, ovvero

0,285714285714...

Ci siamo premurati stavolta di mettere i puntini di sospensione a indicare che in realtà la scrittura 0,285714285714 è solo una approssimazione di $2/7$. Di fatto non possiamo rappresentare $2/7$ in forma decimale, senza dover sprecare infinita carta. Peraltro, se il sistema di scrittura fosse in base 7 o 14 o 21, non ci porremmo nemmeno il problema, visto che in tali basi si può esprimere $2/7$ con un numero finito⁹ di cifre dopo la virgola.

In verità, la scrittura decimale di $2/7$, rapporto tra due numeri naturali, ha un andamento regolare e ripetitivo, che in effetti permette di capire come vanno le cose senza bisogno di scrivere per forza tutte le cifre decimali, e lo stesso discorso vale per tutte le frazioni. In altre parole, possediamo un algoritmo che ci permette di ottenere facilmente tutte le cifre che desideriamo della scrittura decimale di una qualunque frazione.

Per quanto riguarda π invece, curiosando in Internet¹⁰ si trova la seguente pagina¹¹:

Computing Pi to 206 billion digits
by François Labelle

Introduction

Although 10 digits are sufficient for all practical purposes, the computation of Pi to as many digits as pos-

⁹ Per esempio, in base 7 il numero si scrive come 0,2.

¹⁰ L'indirizzo URL completo è molto lungo, ma è sufficiente fare una ricerca con “Google” utilizzando come parole chiave “computing pi to 206 billion digits”.

¹¹ Nel mondo anglosassone si utilizzano in modo rovesciato il “punto” e la “virgola” per la rappresentazione dei numeri, quindi la virgola in 206,158,430,000 serve unicamente a separare gruppi di tre cifre, come in Italia si fa con il punto. In altre parole si tratta di 206 miliardi 158 milioni 430 mila cifre.

sible provides an endless challenge. This page describes the September 1999 computation of Pi to 206,158,430,000 decimal digits by Y. Kanada and D. Takahashi of the University of Tokyo. Storing all these digits in ASCII format would require 303 CD-ROMs!

Si conoscono dunque un numero enorme di cifre decimali di π , largamente superiore a qualunque necessità pratica, tuttavia questo non ci autorizza ad affermare di conoscere tutte le cifre della rappresentazione decimale di π . La situazione è per certi versi simile alla scrittura di $2/7$: abbiamo un algoritmo che ci permette di scrivere tutte le cifre che vogliamo di π , sia pure in tempi sempre maggiori (ed è su tale algoritmo che si basano i programmi per computer come quello citato sopra), solo che non è così semplice e immediato come quello per $2/7$. Nel momento in cui si posseggono algoritmi con tali proprietà, la palla rimbalza nel campo delle applicazioni. In qualunque contesto pratico c'è sempre un margine di errore tollerabile; ci deve essere, anche perché le informazioni note saranno esse stesse affette da errori di misura. A quel punto si tratta di vedere, di volta in volta, quali siano le approssimazioni decimali compatibili con le tolleranze del problema.

Per pagare un terreno circolare di raggio 1 metro che costa 1 euro al metro quadro si dovrà giocoforza ricorrere al vecchio $\pi = 3,14$, non essendo stata ancora coniata la moneta da π euro.

2. Dimensione

Nei libri di testo di biologia per le scuole non viene quasi citato, ma oltre ai cinque sensi canonici ne esiste uno localizzato nell'orecchio interno di cui ci rendiamo scarsamente conto e che è tuttavia di estrema importanza. Gli organi di senso corrispondenti sono i canali semicircolari responsabili della percezione del corpo nello spazio e del suo equilibrio. Alcuni degli esperimenti che Galileo ha immaginato nei suoi scritti avvengono in modo sostanzialmente equivalente all'interno di questi canali, in cui viene rilevato il movimento del fluido chiamato endolinfa, che a sua volta riflette i movimenti, o meglio le accelerazioni, della testa. Il malfunzionamento di questi organi comporta capogiri o addirittura l'impossibilità di rimanere in piedi. Questo avviene, per esempio, dopo una serie di piroette: il risultato è la fastidiosa sensazione che il mondo ci giri at-

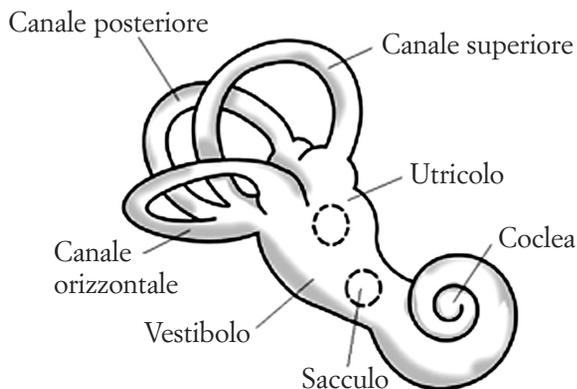


Figura 2. I canali semicircolari nell'orecchio interno.

torno. In ciascun orecchio ci sono tre canali semicircolari, disposti grosso modo secondo tre piani mutuamente ortogonali (Figura 2). È come se avessimo un piccolo sistema di assi cartesiani dentro ciascun orecchio!

Il numero tre non è ovviamente casuale, e corrisponde alle dimensioni dello spazio in cui ci muoviamo. Esso esprime essenzialmente il concetto, più o meno intuitivo, che si può individuare in modo univoco un punto dello spazio specificando le sue tre coordinate, una volta che si sia scelto un sistema di riferimento.

Il sistema di riferimento cartesiano, ovvero basato su tre assi mutuamente ortogonali che si incontrano in un punto (chiamato origine) non è l'unica scelta possibile; per esempio, si può utilizzare latitudine, longitudine e altitudine per individuare un punto nei pressi della superficie della Terra. Si tratta di un sistema di coordinate non cartesiano in cui le linee coordinate, ottenute facendo variare solo una delle tre coordinate, non sono rettilinee. Questa scelta, che privilegia la forma sferica della superficie terrestre, non funziona bene ovunque: nei pressi dei poli, o meglio lungo tutto l'asse di rotazione della Terra, non c'è una corrispondenza biunivoca tra terne di numeri e punti dello spazio, tanto che la longitudine perde di significato. Inoltre risulta necessario scegliere un particolare meridiano (la linea di cambiamento di data) dove la longitudine salta dal valore $+180^\circ$ a -180° . Tuttavia, da un punto di vista locale, il sistema rimane perfettamente valido.

In definitiva, la consapevolezza di vivere in uno spazio tridimensionale può essere considerato un fatto innato, senza con questo nulla voler togliere alla complessità del concetto di dimensione spaziale. Non si può invece dire lo stesso riguardo alla struttura dello spazio, questione che è stata molto dibattuta. Lo spazio euclideo è qualcosa di più di uno spazio tridimensionale. Si assume in effetti di poter fare misure di lunghezza e operazioni come gli spostamenti rigidi. Euclide¹² ha in particolare studiato il caso di uno spazio a due dimensioni¹³, il piano euclideo. Il risultato del suo studio rappresenta il primo esempio storico di sistema assiomatico, vale a dire di un universo di lavoro governato da alcune regole prestabilite (assiomi e postulati) di cui non si pretende a tutti i costi un'aderenza perfetta con la realtà fisica: in effetti gli oggetti della geometria euclidea sono enti astratti che non hanno equivalenti nella realtà se non in modo vago e approssimato. Il famoso quinto postulato¹⁴ era stato correttamente individuato da Euclide come tale: un postulato, quindi non deducibile dagli altri assunti. La cosa era talmente poco ovvia da provocare innumerevoli tentativi di dimostrazione che continuano ancora oggi.

3. Continuità

C'è tuttavia un altro aspetto, molto più sottile della dimensione dello spazio, legato alla struttura fine anche di spazi monodimensionali, cioè di ambienti dove una sola coordinata è sufficiente per individuare una posizione. La continuità di uno spazio euclideo (dimensione tre), di un piano o di una retta (dimensione uno) è un concetto che ha richiesto letteralmente millenni per essere compreso in modo abbastanza completo. Ci interesseremo qui unicamente degli aspetti matematici della questione; la realtà fisica microscopica dello spazio reale è tutt'altra questione, che riguarda fisici e

¹² Euclide, matematico greco, visse molto probabilmente durante il regno di Tolomeo I (367 a.C. ca.-283 a.C.). È sicuramente il più importante matematico della storia antica, e uno dei più importanti e riconosciuti di ogni tempo e luogo. È noto soprattutto come autore degli *Elementi*, la più importante opera di matematica dell'antichità.

¹³ Da un punto di vista puramente concettuale c'è un importante salto logico tra la dimensione uno e la dimensione due, mentre il passaggio dalla dimensione due alla dimensione tre non è altrettanto critico.

¹⁴ Il cui enunciato è: "Se una retta taglia altre due rette determinando dallo stesso lato angoli interni la cui somma è minore di quella di due angoli retti, prolungando le due rette, esse si incontreranno dalla parte dove la somma dei due angoli è minore di due retti".

filosofi. Quello che intendiamo dire è che la matematica ha costruito un modello dello spazio, ma la sua aderenza alla realtà fisica è questione che non riguarda in modo proprio la matematica. Si tratta di un modello rigoroso, su cui si può operare senza pericolo di ambiguità o di incappare in situazioni paradossali.

3.1 Pitagora e i pitagorici

Come abbiamo già avuto modo di osservare nel primo paragrafo, la scoperta dell'esistenza di lunghezze tra loro incommensurabili, ovvero di numeri non razionali, ha causato una grave crisi all'interno della scuola pitagorica¹⁵. In effetti veniva messa in evidenza l'impossibilità di poter misurare le grandezze (per esempio, lunghezze geometriche) utilizzando una medesima piccola unità di lunghezza indivisibile. In altre parole non si riusciva a ricondurre il processo del "misurare" al processo del "contare". D'altra parte, se immaginiamo di disegnare su una retta tutti i numeri razionali, otteniamo un insieme che si addensa ovunque: in apparenza esso riempie tutta la retta.

Ciò rende impossibile visualizzare concretamente l'incompletezza dell'insieme dei numeri razionali. Anche i celebri paradossi di Zenone¹⁶ contro il movimento nascono dalla difficoltà concettuale di afferrare la nozione di continuità di insiemi di grandezze e di insiemi numerici. La costruzione dei numeri reali, che ha infine permesso di superare la dicotomia tra numeri (razionali) e grandezze, è in effetti puramente concettuale e ottenuta con il preciso scopo di trovare una soluzione rigorosa all'apparente contraddizione legata alla scoperta dei numeri irrazionali.

Non si tratta semplicemente di aggiungere un po' di quantità numeriche al solo scopo di dare esistenza alle quantità geometriche incommensurabili (irrazionali) che potevano essere costruite nell'ambito della geometria euclidea, ma di risolvere la questione in modo più radicale. L'insieme che ne deriva, i numeri reali, è di fatto talmente importante da poter essere considerato come pietra an-

¹⁵ La scuola pitagorica, fondata a Crotone intorno al 530 a.C., assunse caratteristiche di setta misterica. Per i pitagorici il numero (inteso come numero naturale secondo la terminologia moderna) è fondamento di ogni cosa, compresa la geometria. È curioso il fatto che proprio il teorema di Pitagora permetta di ottenere il più noto esempio di lunghezze incommensurabili: il lato e la diagonale di un quadrato. Questa scoperta, secondo la leggenda, fu tenuta rigorosamente segreta.

¹⁶ Zenone di Elea (495 a.C.-430 a.C.), filosofo greco presocratico, famoso per aver proposto una serie di paradossi sul concetto di infinito attuale.

golare di quella branca fondamentale della matematica nota con il nome di analisi matematica.

La proprietà essenziale di cui gode l'insieme dei numeri reali è quella che si chiama anche continuità della retta reale: se dividiamo i numeri reali in due sottoinsiemi non vuoti A e B aventi la proprietà che ciascun elemento di A è minore di ciascun elemento di B (cioè si considera una *sezione* dei numeri reali), esiste sempre un numero reale x che separa A e B , cioè tale che tutti gli elementi di A sono minori o uguali a x e tutti gli elementi di B sono maggiori o uguali a x . Questa proprietà non è verificata dai numeri razionali: se, per esempio, costruiamo la sezione (A, B) con B ottenuto prendendo i numeri razionali positivi il cui quadrato è maggiore di 2 e A formato da tutti i numeri razionali rimanenti, ci rendiamo conto che non esiste nessun numero razionale che separi A e B ; tale numero, se esistesse, avrebbe il quadrato uguale a 2, il che è assurdo. Nemmeno l'insieme dei numeri algebrici, formato da tutte le soluzioni di equazioni algebriche di qualunque grado con coefficienti razionali, ben più ampio dell'insieme dei numeri razionali, soddisfa la proprietà di continuità¹⁷.

I numeri reali sono, in effetti, un insieme molto più ricco. La *cardinalità* (concetto introdotto da Cantor per quantificare la numerosità di insiemi infiniti) dell'insieme dei numeri reali, chiamata *cardinalità del continuo*, è strettamente maggiore della cardinalità dell'insieme dei numeri razionali, fatto dimostrato dallo stesso Cantor in modo molto elegante con il cosiddetto procedimento diagonale. La cardinalità dell'insieme dei razionali è invece la stessa di quella dei numeri naturali (cardinalità del numerabile)¹⁸.

In altre parole, mentre è possibile stabilire una corrispondenza biunivoca tra i numeri naturali e i numeri razionali, lo stesso non si può fare con i numeri reali. Cantor tentò a più riprese di dimostrare che non esistono insiemi con cardinalità intermedia tra quella dei numeri naturali e quella dei numeri reali (ipotesi del continuo), fatto che in seguito si scoprì essere sia indimostrabile, sia irrefutabile.

¹⁷ È, per esempio, algebrico il numero $\sqrt{2}$, in quanto soluzione dell'equazione $x^2 - 2 = 0$. Viceversa non è algebrico (si dice allora che è trascendente) il numero π .

¹⁸ La cardinalità di \mathbb{N} , il più piccolo numero cardinale transfinito, è stata indicata da Cantor con il simbolo \aleph_0 (si legge "alef-con-zero"; alef è la prima lettera dell'alfabeto ebraico), notazione tutt'oggi utilizzata. I cardinali subito maggiori sono \aleph_1 , \aleph_2 ecc. A causa di difficoltà legate alla cosiddetta ipotesi del continuo non è possibile identificare la cardinalità dei numeri reali, generalmente indicata con la lettera c , con uno di questi.

4. La teoria della relatività generale

Nel 1915 Albert Einstein¹⁹ pubblicò la sua teoria della relatività generale, in cui a partire dai fondamenti gettati dalla teoria della relatività speciale del 1905, cerca di ricondurre le leggi della gravitazione a effetti dovuti alla curvatura dello spazio-tempo provocata dalla massa dei corpi celesti.

Tale approccio ha costituito un capovolgimento totale rispetto alle teorie precedenti: le orbite dei pianeti non si incurvano, come conseguenza della forza di gravitazione che perturba un moto altrimenti rettilineo uniforme in base alla legge $F = ma$, sono invece traiettorie rettilinee²⁰ in uno spazio curvo, come schematicamente illustrato nella Figura 3. Pensare allo spazio tridimensionale in cui viviamo come uno spazio curvo è tutt'altro che intuitivo; in effetti la curvatura dello spazio dovuta alla gravitazione del pianeta Terra è talmente piccola da non essere minimamente percepibile e gli effetti diventano importanti solo in presenza di grandi masse come le stelle o su enormi distanze (milioni di anni luce).

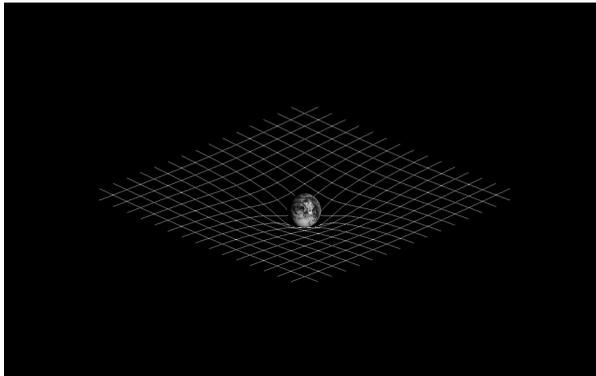


Figura 3. Immagine schematica di come la massa della Terra incurvi lo spazio-tempo circostante.

¹⁹ Albert Einstein (Ulma, 1879 – Princeton, 1955), fisico tedesco. Ha rivoluzionato la concezione del mondo fisico con le sue teorie della relatività ristretta (1905) e della relatività generale (1915). Ricevette il premio Nobel per la fisica nel 1921 per la spiegazione dell'effetto fotoelettrico.

²⁰ Nel contesto delle varietà riemanniane, ovvero spazi dotati di una metrica, si parla più propriamente di geodetiche per riferirsi a percorsi di minima lunghezza, che corrispondono alle rette della geometria euclidea.

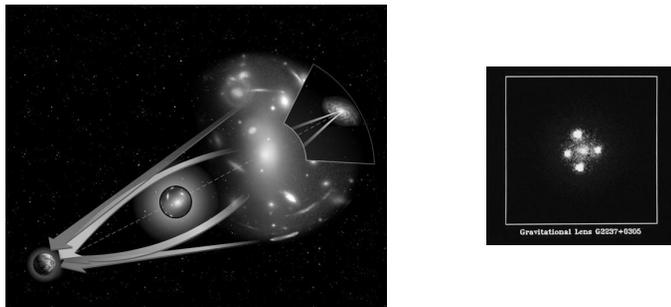


Figura 4. L'effetto lente gravitazionale è schematicamente mostrato nel disegno di sinistra. La fotografia a destra mostra la croce di Einstein: le quattro "stelle" ai vertici del rombo sono in realtà immagini diverse di uno stesso quasar che si trova dietro a una galassia massiccia. Quest'ultima è la causa dell'effetto lente.

Si tratta di una situazione analoga alla difficoltà di percepire la curvatura della superficie terrestre in base all'esperienza diretta: a causa della grande differenza di scala tra la dimensione di un uomo e la dimensione della Terra la prima sensazione che si ha è di una superficie piatta. Solo misure indirette, come il famoso esperimento di Eratostene²¹ o una foto panoramica da un punto lontano dalla superficie terrestre (fatta da un astronauta in orbita) possono convincere della sfericità della Terra. Il fatto che oggi nessuno metta più in dubbio la sfericità della Terra non deve farci concludere che l'uomo contemporaneo abbia una maggiore sensibilità nella percezione concreta della realtà tangibile. Si tratta più semplicemente del fatto che fin da piccoli siamo informati su questo fatto a scuola.

La curvatura dello spazio della relatività generale non discrimina tra pianeti e raggi luminosi. Anche questi ultimi vengono deviati (molto meno vistosamente a causa della elevata velocità della luce), cosa che ha permesso una notevole verifica sperimentale della teoria. In ef-

²¹ Eratostene di Cirene (Cirene, 276 a.C. – Alessandria d'Egitto, 194 a.C.), matematico, astronomo, geografo e poeta greco. Misurò per primo con grande precisione il raggio della Terra. Elaborò il crivello di Eratostene, un metodo per l'individuazione dei numeri primi. Per il suo esperimento considerò le città di Alessandria e Siene (l'odierna Assuan) che stanno circa sullo stesso meridiano. Siene si trova sul tropico del Cancro, cosicché a mezzogiorno del solstizio d'estate il sole si trova allo zenit e i suoi raggi raggiungono il fondo di un pozzo. Misurando alla stessa ora dello stesso giorno la lunghezza dell'ombra proiettata da un palo nella città di Alessandria, Eratostene ha potuto calcolare la differenza di longitudine tra le due città e quindi, in base alla loro distanza (252.000 stadi), la lunghezza di un meridiano terrestre. L'errore nella sua valutazione è inferiore al 3%.

fetti, quando transita vicino a un corpo molto massiccio un raggio luminoso proveniente da una stella lontana subisce una deviazione simile a quella dovuta alla rifrazione in una lente. L'effetto lente che ne deriva provoca in particolari situazioni lo sdoppiamento apparente di alcuni corpi celesti. Uno di questi casi è illustrato a destra nella Figura 4; a sinistra nella stessa figura viene illustrato schematicamente l'effetto: il nostro occhio vede provenire da due direzioni sensibilmente diverse raggi luminosi che hanno la stessa origine.

4.1 Spazio, tempo e superfici

Sono ora necessarie alcune considerazioni. Anzitutto nella teoria della relatività speciale, e quindi anche nella relatività generale, bisogna riferirsi allo spazio-tempo come a un tutt'uno, non come a due entità separate e indipendenti come invece avveniva prima di Einstein, e come anche il senso comune suggerirebbe. Il comportamento sensibilmente diverso dello spazio rispetto al tempo è conseguenza di una particolare scelta della metrica²² nel *continuum* spazio-temporale. Trascureremo però questo aspetto importante e ci limiteremo a parlare di uno spazio curvo dotato di una più usuale metrica riemanniana. Probabilmente Einstein si sarebbe trovato in seria difficoltà nello sviluppare la sua teoria della relatività generale se non avesse potuto utilizzare le teorie matematiche da poco introdotte da Bernhard Riemann²³. Si è trattato di un caso esemplare della simbiosi che intercorre, spesso con modalità sorprendenti, tra teorie matematiche e applicazioni alla fisica o più in generale al mondo reale.

Per esemplificare i concetti principali che stanno alla base della teoria delle varietà riemanniane ci avvarremo di un'analogia con una ben nota forma d'arte giapponese. Nella Figura 5 è rappresentato il modello più conosciuto di origami, l'arte giapponese di piegare la carta: si tratta di piegare opportunamente un foglio di carta quadrato, in genere bianco da una parte e colorato dall'altra. È vietato ef-

²² La *metrica di Minkowski* è una particolare nozione di distanza in cui sono ammesse anche distanze negative e nulle. Un raggio di luce, per esempio, collega due eventi (cioè due punti dello spazio-tempo) a distanza nulla. Due eventi associati a tempi differenti nella stessa posizione spaziale hanno distanza positiva. Due eventi temporalmente simultanei in posizioni spaziali diverse si trovano invece a una distanza negativa.

²³ Bernhard Riemann (Breselenz, 1826 – Selasca, 1866), matematico e fisico tedesco. Contribuì in modo determinante a molti ambiti della matematica, in particolare in geometria e analisi complessa. Da lui prende il nome la famosa *ipotesi di Riemann*, tuttora indimostrata.

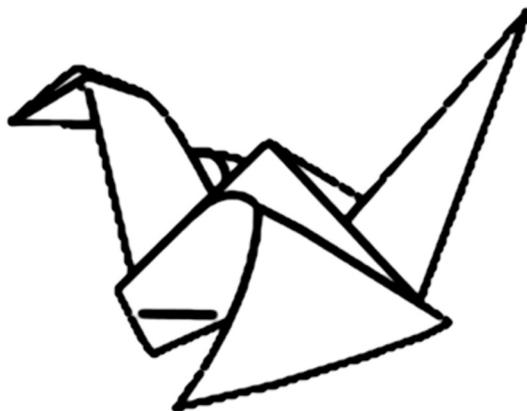


Figura 5. La gru è la più classica costruzione dell'arte dell'origami.

fettuare tagli e non si può usare la colla. Cos'hanno in comune il foglio quadrato iniziale e il modello finale di gru della Figura 5? Nonostante l'aspetto completamente diverso, per un ipotetico abitante bidimensionale del foglio di carta non ci sarebbe alcuna differenza. Immaginiamo che l'universo di tale abitante sia il foglio di carta e che quindi egli non abbia alcuna percezione di una terza dimensione. Non avrebbe modo di effettuare misure al di fuori del suo universo e di conseguenza non percepirebbe le pieghe presenti nel modello di origami.

Abbiamo un primo esempio di una varietà riemanniana: una superficie alla cui struttura metrica intrinseca siamo interessati, ovvero a tutte le proprietà che dipendono solo da misure di lunghezza che si possono effettuare rimanendo sempre confinati sulla superficie. Il foglio di carta, piegato o non piegato, è una superficie piatta dal punto di vista della sua struttura metrica, ovvero è indistinguibile da un piano. La superficie laterale di un cilindro o la falda di un cono sono anch'esse piatte per un loro abitante, tant'è vero che possiamo ottenerle arrotolando un foglio di carta. Esse sono indistinguibili da un piano, perlomeno da un punto di vista locale²⁴.

²⁴ Dobbiamo assumere che gli ipotetici abitanti della superficie di un cilindro (o di un cono) non facciano osservazioni globali, cioè osservazioni che mettano in luce il fatto che un giro completo attorno al cilindro permetterebbe di ritornare al punto iniziale.

La superficie della Terra (che immaginiamo perfettamente sferica) è invece una superficie intrinsecamente curva, anche se ne consideriamo solo una porzione. Se scegliamo tre punti sulla superficie terrestre e li colleghiamo con geodetiche, vale a dire con percorsi di lunghezza minima, otteniamo la versione sferica di un triangolo. Se misuriamo i tre angoli al vertice di questo triangolo e li sommiamo abbiamo però una sorpresa: il risultato è maggiore di 180° ! La discrepanza rispetto al valore 180 diventa molto evidente se i tre punti sono molto distanti tra loro: se, per esempio, scegliamo il Polo Nord e due punti sull'Equatore rispettivamente sul meridiano 0 (di Greenwich) e 90, otteniamo un triangolo con tre angoli retti, la cui somma è 270° , che supera 180° di 90° . Al contrario si avrà una discordanza estremamente piccola, ma mai nulla, se i tre punti sono vicini: per esempio, per tre città della Lombardia. Se ci limitiamo a una ristretta area geografica la geometria della superficie terrestre è quasi (ma non esattamente) euclidea.

La costruzione del triangolo che abbiamo appena effettuato fa uso solamente delle proprietà metriche intrinseche, e questo mostra che effettivamente abbiamo una superficie intrinsecamente diversa da un piano. È questo fatto a rendere difficoltosa la rappresentazione cartografica della superficie terrestre. Tutto funziona piuttosto bene finché ci si limita a rappresentare su un foglio di carta piccole porzioni della superficie; i problemi nascono quando si vogliono rappresentare interi continenti, se non tutto il globo terrestre. Si possono fare svariate scelte nella tecnica di rappresentazione.

Per esempio, nella Figura 6 è rappresentata la proiezione cilindrica di Mercatore²⁵, che provoca vistose amplificazioni delle regioni vicine ai poli, mentre nella Figura 7 si vede la proiezione di Buckminster Fuller²⁶ in cui la superficie terrestre viene proiettata su un icosaedro successivamente sviluppato sul piano. Qualunque scelta si faccia sarà tuttavia impossibile evitare deformazioni più o meno evidenti dei dettagli geografici.

È proprio l'analogia con la cartografia che ha suggerito ai matematici l'introduzione dei termini di carta locale e di atlante: una carta

²⁵ Mercatore, nome latinizzato di Gerardus Mercator (Rupelmonde, 1512 – Duisburg, 1594), matematico, astronomo e cartografo fiammingo. Noto per i suoi studi sulla cartografia.

²⁶ Richard Buckminster Fuller (Milton, 1895 – Los Angeles, 1983), inventore, architetto e designer statunitense. Inventò, tra l'altro, le cupole geodetiche e la rappresentazione cartografica detta Dymaxion Map, o proiezione di Fuller.

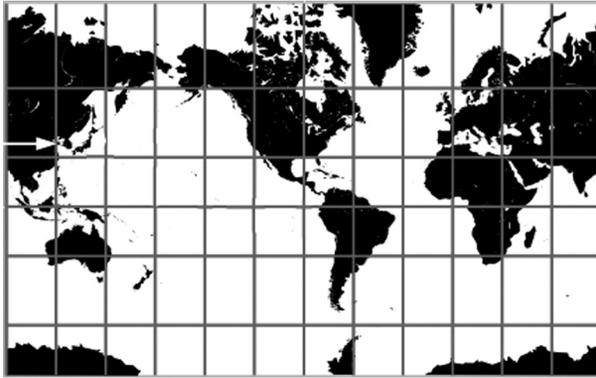


Figura 6. Proiezione cilindrica di Mercatore, le zone vicine ai poli subiscono una forte espansione e appaiono molto più grandi di quanto non sono in realtà.

locale consiste in una corrispondenza biunivoca tra una porzione della superficie e una porzione del piano cartesiano, con l'effetto di fornire un sistema di coordinate locali. Un atlante è una collezione di carte locali che ricoprono tutta la superficie e aventi sovrapposizioni parziali tra di esse. Nelle zone della superficie appartenenti a più di una carta locale è necessario fornire una descrizione di come passare dal sistema di coordinate riferite a una carta a quello riferito a un'altra carta. Quello che si ottiene è una varietà bidimensionale. La generalizzazione a tre o più dimensioni non presenta difficoltà concettuali, si tratta semplicemente di aumentare opportunamente il numero di coordinate.

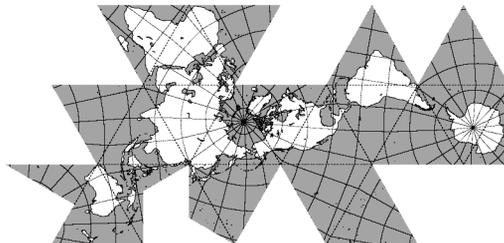


Figura 7. Planisfero Dymaxion, ottenuto tramite la proiezione di Fuller.

La struttura metrica riemanniana può essere aggiunta a una varietà bidimensionale per mezzo del tensore metrico g_{ij} definito su ciascuna carta locale. Si tratta di una matrice simmetrica 2×2 , le cui componenti variano in funzione delle coordinate della carta, che permette di calcolare la lunghezza di un piccolo spostamento.

Se indichiamo con il vettore a due componenti $\xi = (\xi_1, \xi_2)$ un piccolo spostamento in termini di coordinate locali, la lunghezza s dello spostamento si otterrà con la formula $s = g_{11}\xi_1\xi_1 + g_{12}\xi_1\xi_2 + g_{21}\xi_2\xi_1 + g_{22}\xi_2\xi_2$ che viene più concisamente scritta, utilizzando la convenzione di Einstein di somma rispetto agli indici ripetuti, come $s = g_{ij}\xi_i\xi_j$. Utilizzando il tensore metrico è possibile calcolare, usando le tecniche dell'analisi, la lunghezza di una curva qualsiasi e questo permette di individuare le geodetiche come curve di lunghezza minima. La curvatura intrinseca di una superficie dotata di una metrica riemanniana è rappresentata da un oggetto denotato con R^i_{jkl} : si tratta del tensore di Riemann, un tensore a quattro indici ottenuto a partire dal tensore metrico tramite opportune operazioni di derivazione. Ricordando che per una superficie ogni indice può assumere i due valori 1 e 2, si hanno un totale di 2^4 componenti per tale tensore, ciascuna dipendente dal punto sulla superficie. Le componenti diventano n^4 per varietà di dimensione n generica. Grazie però a proprietà di simmetria e ad altre identità note, il numero di componenti indipendenti di tale tensore si riduce a

$$\frac{1}{12}n^2(n^2 - 1) \quad (1)$$

In dimensione 1 tale numero si riduce a 0, questo significa che il tensore di Riemann è nullo, ovvero che una linea curva non ha curvatura intrinseca. Ciò esprime il fatto che un abitante di una linea non ha nessun modo per riconoscere eventuali incurvamenti del suo universo monodimensionale. In dimensione 2 la formula (1) fornisce come risultato 1. Ciò significa che la curvatura di una superficie (ora in qualche modo misurabile dai suoi abitanti) si esprime con un unico valore che in generale può variare da punto a punto. Tale valore è essenzialmente espresso dalla "curvatura gaussiana", elaborata dal tedesco Carl Gauss²⁷. Per superfici contenute

²⁷ Carl Friedrich Gauss (Braunschweig, 1777 – Gottinga, 1855), matematico, astronomo e fisico tedesco. Ha fornito contributi determinanti a molti campi della matematica.

nello spazio euclideo tridimensionale tale curvatura può essere calcolata facendo il prodotto delle due curvature principali²⁸: la curvatura gaussiana di una sfera di raggio unitario è 1. Un piano ha entrambe le curvature principali nulle, quindi (come ci si aspetterebbe) la sua curvatura gaussiana è nulla. La superficie di un cilindro (o di un cono) ha una curvatura principale nulla, quindi la sua curvatura gaussiana è essa stessa nulla, lo stesso valore ottenuto per il piano. Tale risultato è consistente con il fatto che gli abitanti della superficie non sono in grado di distinguere tra il cilindro (il cono) e un piano. Abbiamo un universo piatto. Il numero di componenti indipendenti nel tensore di Riemann aumenta molto quando la dimensione dello spazio aumenta: per $n = 3$ si hanno 6 componenti indipendenti che diventano 20 in dimensione $n = 4$. Osserviamo che lo spazio-tempo di Einstein ha dimensione 4, quindi le proprietà di incurvamento dello spazio-tempo della relatività generale non sono di fatto molto maneggevoli.

4.2 Le geometrie non euclidee

Ritornando al caso delle superfici (dimensione $n = 2$), ricordiamo che le informazioni sulla curvatura possono essere riassunte da un singolo numero, la curvatura di Gauss, definita punto per punto. Particolare interesse assumono le superfici per cui tale numero K è di fatto costante ovunque. La costanza della curvatura K permette di effettuare gli spostamenti rigidi, e questo in definitiva permette di costruire una geometria analoga alla geometria euclidea in cui le geodetiche prendono il posto delle rette. A conti fatti²⁹ si scopre che rimangono validi tutti gli assiomi della geometria euclidea con l'eccezione del quinto postulato.

Nel caso $K = 0$ anche il quinto postulato rimane valido, e si riottiene la geometria euclidea. Nel caso $K > 0$ si ottiene la geometria sferica (Figura 8). Un buon modello di questa geometria si ottiene considerando la superficie di una sfera con la metrica usuale. Le geodetiche sono i cerchi massimi e non esistono rette parallele.

Nel caso $K < 0$ si ottiene la geometria iperbolica (Figura 9). Data una retta (ovvero una geodetica) e un punto fuori di essa, esistono

²⁸ Si parlerà ancora di curvature principali nel paragrafo 2 del capitolo "Computer e soluzioni approssimate".

²⁹ Ci possono essere altre discrepanze dovute alla struttura in grande della superficie, per esempio sulla superficie di un cilindro ci saranno alcune geodetiche chiuse, corrispondenti a circonferenze ottenute intersecando il cilindro con piani ortogonali all'asse.

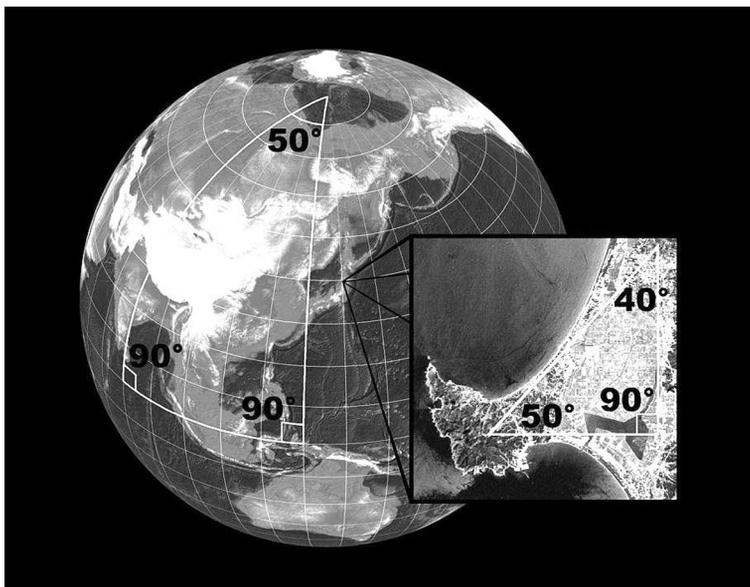


Figura 8. La superficie terrestre come modello di geometria sferica. La somma degli angoli interni di un triangolo è maggiore di 180° .

infinite rette passanti per quel punto che non intersecano la prima retta. Modelli di geometria iperbolica, cioè realizzazioni concrete di superfici con le proprietà descritte, sono stati costruiti in particolare da Jules-Henri Poincaré³⁰. Si tratta però di costruzioni in cui la struttura metrica non è quella abituale.

Nel modello del disco di Poincaré (vedi Figura 9 a destra) la superficie coincide con la parte interna di un cerchio unitario e le geodetiche sono gli archi di cerchio che raggiungono ortogonalmente il bordo del disco. Alcuni disegni di Maurits Cornelis Escher si riferiscono a questo modello di geometria iperbolica; uno di questi è rappresentato nella Figura 10 del capitolo “La matematica nell’arte”. È possibile costruire anche un modello, simile alla sfera della geometria sferica (la cosiddetta pseudosfera rappresentata nella Figura

³⁰ Jules Henri Poincaré (Nancy, 1854 – Parigi, 1912), matematico, fisico teorico e filosofo naturale francese. Poincaré è considerato un enciclopedico e in matematica l’ultimo universalista, dal momento che eccelle in tutti i campi della disciplina noti ai suoi giorni.

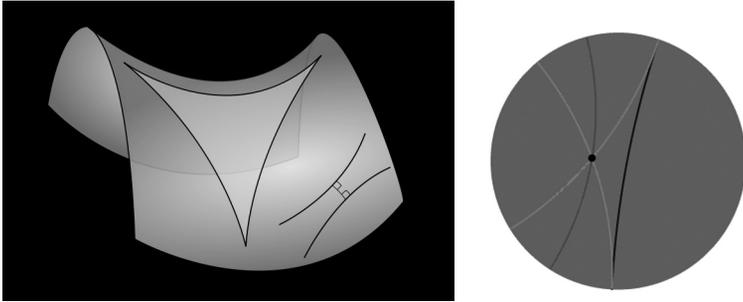
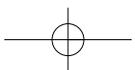


Figura 9. Nella geometria iperbolica la somma degli angoli di un triangolo è inferiore a 180° . A destra è rappresentato il modello del disco di Poincaré; data una retta e un punto fuori di essa ci sono infinite rette parallele passanti per tale punto.



Figura 10. La pseudosfera.

10), in cui si costruisce una superficie nello spazio e si utilizza la metrica usuale, tuttavia tale superficie contiene dei punti singolari che rendono questo modello meno interessante rispetto ai modelli proposti da Poincaré.



MODELLI E PREVISIONI

La matematica appare a molti una scienza esatta: tutti i tifosi di calcio conoscono bene l'espressione "certezza matematica" dello scudetto o della retrocessione. Può apparire quindi strano parlare di errore in ambito matematico: secondo quanto si percepisce della matematica, senz'ombra di dubbio un ragionamento, una dimostrazione, una formula contenenti un errore sono da buttare. In logica si può mostrare che, se si ammette che anche solo uno degli enunciati della matematica possa essere sempre falso, allora sarebbe corretto dedurre qualunque cosa: l'implicazione sarebbe sempre vera (almeno secondo le regole della logica), e quindi, sempre secondo le regole della logica, si potrebbe correttamente dedurre tutto ciò che si vuole. D'altro canto, sappiamo che a volte le previsioni della scienza non sono rispettate, e quindi da qualche parte si annidano degli errori.

Ma non è di errori interni alla matematica che si parla, bensì del modo in cui essa cerca di trattare gli errori esterni; si parla, cioè, delle situazioni concrete alle quali la matematica si applica. Le applicazioni della matematica sono cresciute in maniera notevole durante il '900, anche se è percezione diffusa il fatto che ciò risalga a molto prima. Probabilmente la celebre frase che Galileo³¹ scrive ne *Il Saggiatore*:

“La filosofia è scritta in questo grandissimo libro che continuamente ci sta aperto innanzi a gli occhi (io dico l'universo), ma non si può intendere se prima non s'impara a intender la lingua, e conoscer i caratteri, ne' quali è scritto. Egli è scritto in lingua matematica, e i caratteri son triangoli, cerchi e altre figure geometriche, senza i quali mezzi è impossibile a intenderne umanamente parola; senza questi è un aggirarsi vanamente per un oscuro laberinto.”

³¹ Galileo Galilei (Pisa, 1564 – Arcetri, 1642), fisico, astronomo, matematico e filosofo italiano. È considerato il padre del metodo sperimentale e della scienza moderna.

ha contribuito a diffondere un'impressione di ubiquità della matematica che non si è rivelata tale fino agli inizi del '900. Se è vero, da un lato, che la fisica dipende in maniera essenziale dalla matematica per la formalizzazione dei suoi concetti, è altrettanto vero che la tecnica ha fatto a meno per secoli della matematica, crescendo molto di più per via empirica – per tentativi, invenzioni (ed errori) – che per via deduttiva. Tuttavia, anche restringendosi solamente alla fisica, la prima grande discrepanza con la matematica riguarda l'essenza stessa dell'agire del fisico: la misurazione.

1. L'esattezza nel modello matematico e gli errori sperimentali

Ogni misurazione è affetta da errori, per molti versi ineliminabili e per ragioni che vanno fino al cuore della fisica e nelle quali non vogliamo addentrarci. Eppure, nella pratica millenaria a ogni misurazione (di una quantità scalare) è associato un numero. Vale la pena di notare che questa non è una richiesta matematica, né che sia sempre chiaro quale tipo di numero sia opportuno usare. Esistono misurazioni non numeriche (per esempio, misure di scarpe 40-41), ma in genere, quando le cose devono essere fatte in modo preciso, si tende a pensare a un numero “con molte cifre decimali”. La scelta di servirsi del numero per rappresentare il risultato di una misura è quindi una convenzione, e si potrebbe forse affermare che il numero è il più semplice “modello matematico”. Per modello si intende un qualunque ente, concreto o astratto, creato per simulare, rappresentare o predire i comportamenti di un certo fenomeno o comunque di una situazione da descrivere o studiare. A seconda del tipo di grandezza fisica da rappresentare, si usano numeri diversi: se le grandezze sono discrete, i numeri interi sono i più indicati (ma non i soli: per esempio, in fisica si usano spesso numeri “seminteri”, come $1/2$, $3/2$, $5/2$ ecc.); se invece le grandezze sono supposte continue la parte del leone è svolta dal numero reale.

La modellizzazione del continuo è forse la più profonda e importante concettualizzazione matematica, e le sue implicazioni sono spesso ignorate a fronte dei vantaggi che i numeri reali offrono, come già osservato nel capitolo iniziale.

Tuttavia, anche assumendo convenzionalmente un tipo di numeri – per esempio, quelli reali –, il passo successivo nell'indagine dei fenomeni rivela immediatamente la necessità di ulteriori scelte.

A parte, forse, alcuni sistemi veramente elementari come le particelle subatomiche (che sono entità comunque non misurabili in senso convenzionale), ogni sistema fisico appare all'indagine composto di parti, spesso a loro volta composte da altre parti, e così via. Appare evidente allo sperimentatore il fatto che per descrivere il comportamento del sistema serve conoscere quello delle parti.

Quindi, volendo essere "rigorosi", la descrizione completa di un sistema sarebbe ricondotta a quella dei suoi componenti elementari. È questo l'atteggiamento cosiddetto riduzionista, che nella pratica spesso si rivela impossibile da perseguire e anche inutile ai fini che ci si propone. Un semplice esempio chiarirà cosa vogliamo dire. L'espressione "il treno arriva in stazione" può essere sensata o completamente priva di senso a seconda del tipo di modello che si usa per la stazione e per il treno: se essi sono rappresentati con dei punti geometrici, l'espressione ha un senso univoco, ma se si usa per descriverli un segmento o un arco di curva, allora l'interpretazione potrebbe essere più difficile: se la motrice del treno sta fra l'inizio e la fine della stazione — intesa come segmento —, questa descrizione potrebbe non essere sufficiente per parlare di salita e discesa dei passeggeri, perché alcuni vagoni non sono raggiungibili, e così via. Apprendiamo allora che il tipo di predizione del fenomeno suggerisce la sua rappresentazione. Se, sempre nell'esempio del treno, diamo per scontato che sovrapporre i punti equivalga a poter far salire tutti i passeggeri è sufficiente descrivere treno e stazione con dei punti, ma se si è interessati a un flusso di salita e discesa dei passeggeri, allora no.

Altra questione, ben più complessa e filosofica, è se la conoscenza del fenomeno sia completa solo con la descrizione ultima delle parti o meno, o addirittura se la realtà del fenomeno necessiti quella delle sue parti. Questo esula dallo scopo di questa nostra indagine, e il lettore troverà le più diverse proposte nei testi filosofici. Noi ci accontentiamo di ricordare che la correttezza delle previsioni di un modello dipende non solo dalla bontà delle deduzioni, ma anche dall'interpretazione di dette previsioni in relazione alla struttura del modello.

Anche qui un semplice esempio chiarirà quanto vogliamo dire. Se siamo interessati al moto della Luna attorno alla Terra, possiamo decidere di rappresentare Terra e Luna con dei punti materiali, che sono punti della geometria con l'aggiunta della massa. Se poi siamo interessati a una descrizione non esageratamente precisa possiamo scegliere la meccanica newtoniana per la descrizione del fenomeno (ri-

nunciando cioè alla relatività generale, che è più precisa), la quale fornisce la curva (l'orbita della Luna attorno alla Terra), e, volendo, tutte le altre quantità orbitali desiderate. Tutto questo ci permette di conoscere la posizione della Luna sulla volta celeste in un momento preciso, ma non ci darà grandi informazioni sulle maree. Se avessimo voluto aver conto anche di queste, avremmo dovuto includere l'azione del Sole, e, probabilmente, anche rinunciare a rappresentare la Terra come un punto.

Quali sono i vincoli che vanno comunque imposti a un modello matematico? Senza dubbio la *non contraddittorietà*: se un modello matematico produce risultati contraddittori significa che vi è una contraddizione nelle premesse o un errore di deduzione. Però la non contraddittorietà è temperata dall'interpretazione, nel senso che se un modello prevede risultati contraddittori fuori dal suo campo di applicabilità, allora questa contraddizione non toglie la validità del modello in altre situazioni. Per esempio, in molti modelli biologici si preferisce usare variabili reali, anziché intere, per misurare la numerosità delle specie, come risulterà anche nel capitolo dedicato a "Scienze umane e della vita". Può quindi accadere che un modello preveda equilibrio predatori-prede se vi sono $100\sqrt{3} = 173,205 \dots$ predatori (per esempio tigri), che è contraddittorio in quanto 0,205... tigri è meno di una tigre, quindi è morta e non può essere un predatore, e allora le tigri sono 173 e non è una situazione di equilibrio. Anche se questo esempio fa un po' ridere, l'espressione "equilibrio" è in realtà da interpretarsi in modo diverso, e cioè come "oscillazione numerica molto piccola" rispetto all'entità delle specie presenti, con un "piccola" che a rigore andrebbe specificato.

Quali sono gli ambiti nei quali la modellizzazione matematica risulta più efficace? Nella fisica si è soliti parlare di teorie: dire che esse sono dei modelli non è stato ben visto in quanto riveste tutta la disciplina di un vago alone di agnosticismo. Tuttavia, nello studio dei fenomeni complessi, quali quelli atmosferici o climatici, è apparso sempre più evidente che una teoria paragonabile alla meccanica newtoniana o alla relatività è, per ora, molto al di là dalla nostra portata, e quindi il concetto di modello come "teoria cosciente della sua fallacia" si è fatto sempre più strada. Esempi tipici sono i modelli legati alla turbolenza, in cui si sostituisce alle variabili fisiche "esatte" la loro media in un dato intervallo di tempo, e si studia l'evoluzione di questo valore medio nel tempo. Anche le variabili esatte verificano un sistema ben preciso (anch'esso frutto di modellizzazione), le

celebri equazioni di Navier-Stokes³², ma le nuove variabili mediate non soddisfano un sistema univocamente determinato, ma necessitano di un “termine di chiusura”, cioè una relazione aggiuntiva che il modello esatto non può prevedere a causa dell’incertezza sul termine di media. Questa equazione viene aggiunta sulla base di considerazioni specifiche del problema in esame e anche avendo in vista la risolubilità del problema, in quanto il sistema di Navier-Stokes è spesso intrattabile sia analiticamente, sia numericamente.

Fare una classificazione esaustiva dei modelli matematici più spesso applicati è difficile e senz’altro va oltre gli scopi di questo volume. Tuttavia esistono delle considerazioni che si possono ancora fare a livello generale. Una di queste è il concetto di validazione del modello. Si parla di validazione quando si confrontano le previsioni del modello usato con le misurazioni sperimentali sul fenomeno studiato. Se le previsioni trovano corrispondenza con il fenomeno osservato, si può dire che il modello è accettato e si possono considerare le sue previsioni attendibili, sempre all’interno del campo di variabilità nel quale il modello ha senso.

Per esempio, in taluni casi la meccanica newtoniana produce previsioni in ottimo accordo con i fenomeni che cerca di descrivere, e quindi la si accetta, nei limiti intrinseci del modello, che sono quelli di velocità piccole rispetto a quella della luce e masse grandi rispetto a quelle atomiche.

Naturalmente le previsioni restano tali e possono essere smentite dall’osservazione; una caratteristica dei modelli della scienza è proprio quella di essere smentibili (spesso si usa il termine “falsificabili”): a differenza di una previsione astrologica sul carattere di una persona, una previsione scientifica può sempre rivelarsi sballata, mentre per una previsione astrologica smentita si riesce sempre a mettere in campo una spiegazione *ad hoc*.

In generale il modello è tanto più buono quanto più cerca di andare alle radici del fenomeno, e tanto più rischia di essere errato quanto meno indaga il fenomeno in esame. Un esempio è il celebre indovinello che chiede di spiegare la legge con la quale sono generati numeri in risposta a delle password numeriche. Se a 12 il sistema risponde 6, a 10 risponde 5, a 8 risponde 4 e a 6 risponde 3, si po-

³² Claude-Louis Navier (Digione, 1785 – Parigi, 1836), ingegnere e scienziato francese, noto soprattutto per i suoi contributi alla fluidodinamica. George Gabriel Stokes (Skreen, 1819 – Cambridge, 1903), matematico e fisico irlandese, noto per i suoi contributi alla fluidodinamica.

trebbe arguire (modello) che la risposta sia $n/2$, se la password è n . Eppure il sistema risponde 7 se si dà 4 in ingresso, perché la formula “vera” era: «conta le lettere, in italiano, della password». In questo caso, sapendo che il sistema dipende dall'impostazione della lingua, avremmo forse evitato il semplice modello $n/2$ ³³.

Infine, non possiamo non tacere alcune considerazioni piuttosto ovvie; siccome la matematica applicata offre la possibilità di effettuare una scelta, non è sempre evidente a quali criteri ci si atterrà: un buon modello, ma impraticabile nelle previsioni, potrebbe essere scartato a favore di uno più rozzo, ma più semplice e trattabile. Oppure, a parità di risultati, il modello più semplice è in genere preferito, sia se è più semplice nell'elaborazione delle previsioni, sia se è più diretto e facile da estendere.

I modelli più utilizzati nelle applicazioni si possono dividere in due categorie: *modelli deterministici* e *modelli statistico-probabilistici*. Non è una classificazione esaustiva, ma copre buona parte del panorama esistente. In un modello deterministico lo sperimentatore ha a sua disposizione dati iniziali forniti dall'osservazione, a partire dai quali il modello prevede univocamente il comportamento del fenomeno negli istanti di tempo successivi. In un sistema statistico-probabilistico, le quantità non sono note esattamente, né è possibile prevederle con esattezza, ma si riescono ad avere informazioni sulla probabilità che queste variabili giacciono in dati intervalli. Vediamo allora alcuni semplici esempi di questi modelli, per capire più da vicino come funzionano.

2. Un modello deterministico: la meccanica newtoniana

Il modello deterministico per eccellenza è la meccanica newtoniana. In questo modello, nella sua formulazione più elementare, cioè quella dei corpi puntiformi o rigidi, si suppone che le quantità misurate siano note con errore nullo. Questo è un fatto importante sul quale torneremo in seguito. Il modello prevede che, note (sempre con assoluta esattezza) le masse delle parti e le forze agenti sul sistema meccanico oggetto dell'esame, il modello sia in grado di prevedere con errore nullo le quantità meccaniche di interesse in ogni istante successivo. Per “quantità meccanica di inte-

³³ Purtroppo numerosi sedicenti “test di intelligenza” sono basati sull'errata convinzione che la legge soggiacente si possa individuare in modo unico, il che è assolutamente falso.

resse” in genere si intende la cosiddetta “legge oraria del movimento”, che è una funzione $t \mapsto \mathbf{u}(t)$ nella quale il vettore $\mathbf{u}(t)$ determina, per esempio, le posizioni di tutti i punti del sistema. Ai nostri fini basta considerare il caso di un singolo punto di massa costante: note allora la massa del punto e la forza \mathbf{F} impressa su di esso, la legge di Newton³⁴ afferma che, in un riferimento inerziale, vale la legge

$$m\mathbf{a}(t) = \mathbf{F}(t) \quad (1)$$

dove $\mathbf{a}(t)$ indica l’accelerazione del punto all’istante t .

Tutti conosciamo questa legge, ma come si adopera? La prima idea che viene è questa: nota la forza, si ricava l’accelerazione dalla formula, e una volta nota l’accelerazione si dovrebbe trovare il moto, cioè la legge oraria (che, lo ripetiamo, è una funzione del tempo): quando, per esempio, la forza è costante (moto rettilineo uniformemente accelerato).

Questo modello, che per più di due secoli è stato considerato assolutamente esatto (e, quindi, neanche un modello, ma la “realtà”), è stato esteso dalla teoria della relatività, che opera correzioni significative quando sono coinvolte velocità prossime a quella della luce. C’è però un problema, che conferisce al modello un notevole interesse matematico.

Non sempre la forza \mathbf{F} è nota come esplicita funzione del tempo t . Il più delle volte la forza impressa sul punto proviene da un “campo di forze”, ossia da una funzione che associa a ogni punto dello spazio (non solo a ogni istante di tempo) la forza impressa sul punto transigente in quella posizione e in quell’istante. Se indichiamo sempre con \mathbf{F} (cosa errata, ma comoda) questa funzione, l’equazione (1) si riscrive

$$m\mathbf{a}(t) = \mathbf{F}(t, \mathbf{x}(t)) \quad (2)$$

dove \mathbf{x} indica la posizione del punto. La scrittura $\mathbf{F}(t, \mathbf{x}(t))$ sta a significare che per conoscere l’entità della forza serve conoscere, oltre al tempo t , anche il punto $\mathbf{x}(t)$ dove la si vuole calcolare. Que-

³⁴ Isaac Newton (Woolsthorpe-by-Colsterworth, 1642 – Londra, 1727), matematico, fisico e filosofo inglese. Ha stabilito le leggi fondamentali della meccanica classica e la legge di gravitazione universale.

sto fatto è del tutto normale: la legge di gravitazione universale, ma anche l'elettromagnetismo, sono forze di questo tipo.

A questo punto il problema si può delineare così: se conosco la posizione del punto in un istante, posso ricavare la forza e quindi l'accelerazione come prima: a causa di ciò, però, il punto si sposterà, e quindi la forza cambierà anch'essa, perché in punti diversi in generale la forza sarà diversa. Come possiamo fare, quindi, a ricavare la legge oraria? Sembra che ci troviamo in un circolo vizioso.

Fu Newton a dar risposta a questo problema, e in ciò consistette il suo più grande contributo alla matematica. Newton scoprì che, nota la legge oraria della posizione, era facile ricavare quella della velocità e successivamente l'accelerazione, ossia il percorso inverso di quello che ci serve. Gli apparve chiaro che c'erano delle regole semplici per operare su una funzione, così come sono semplici certe operazioni sui numeri. Del resto, già abbiamo detto che l'incognita fondamentale del problema è (o era per Newton) la legge oraria, che è una funzione.

Il passaggio da posizione a velocità, e da velocità ad accelerazione, è detto *derivazione* ed è oggetto degli studi di quinta superiore in molte scuole. Per esempio, se l'accelerazione segue la legge oraria del moto uniformemente accelerato (che si ha quando è costante, e quindi nel caso semplice di una forza costante), allora, come sappiamo già dai tempi di Galileo, la legge oraria della posizione è

$$\mathbf{x}(t) = \mathbf{x}_0 + \mathbf{v}_0 t + \frac{1}{2} \mathbf{a} t^2$$

dove \mathbf{x}_0 e \mathbf{v}_0 sono rispettivamente la posizione e la velocità all'istante $t = 0$, e le operazioni di derivazione producono la legge oraria delle velocità

$$\mathbf{v}(t) = \mathbf{v}_0 + \mathbf{a} t$$

Ma non vogliamo entrare in dettagli troppo tecnici, per cui seguiremo Newton passo passo. Egli indicò questa operazione con un punto posto sopra la funzione da operare, così $\dot{\mathbf{x}}$ indica la derivata della posizione, dunque la velocità.

Naturalmente questo non cambia di una virgola il problema; anzi, se la forza cambia da posto a posto, a maggior ragione cambierà la velocità man mano che il punto si sposterà. Newton continua però nel suo ragionamento e deduce che la stessa operazione permette di pas-

sare da velocità ad accelerazione. E in effetti, se ci si pensa un attimo, l'accelerazione non è altro che la "velocità di variazione della velocità", così come la velocità è la "velocità di variazione della posizione". Per questo motivo indica con due punti (") il passaggio da posizione ad accelerazione: $\mathbf{a} = \ddot{\mathbf{x}}$. Ma a questo punto appare un fatto nuovo: la legge di Newton si può riscrivere

$$m\ddot{\mathbf{x}}(t) = \mathbf{F}(t, \mathbf{x}(t))$$

e assume l'aspetto di un'equazione nella quale l'incognita non è un numero, come nelle normali equazioni algebriche o trigonometriche, ma una *funzione* e prende il nome di equazione *differenziale*³⁵.

Se si tiene presente che ai tempi di Newton le equazioni algebriche erano, da duecento anni almeno, un piatto forte della matematica, si capisce come questo problema abbia suscitato subito un grande interesse nel mondo della matematica, al punto forse da distogliere in po' l'attenzione dal problema iniziale. Ma la legge di Newton non avrebbe avuto, forse, la fortuna che ha avuto se non avesse previsto con straordinaria precisione risultati in uno specifico campo, di grande interesse per gli studiosi dell'epoca: la meccanica dei pianeti. Nel '600 e nel '700, infatti, scienza e tecnica erano molto più legate di quanto non siano oggi, e la spiegazione delle orbite dei pianeti ebbe grande risonanza praticamente presso tutti gli scienziati. Se, per esempio, una teoria fisica, come quella della "materia oscura" o dell'"energia oscura", riuscirà nei prossimi anni a spiegare la coesione delle galassie, essa sarà sicuramente interessante, ma avrà probabilmente meno impatto, per tutti gli scienziati odierni, di quello avuto dalla teoria di Newton.

La matematica dopo Newton ha studiato per decenni le equazioni differenziali nell'intento di determinare tutte le *funzioni* soluzioni di una data equazione, fatto certamente importante ma non essenziale, se si hanno in mente le applicazioni. Quello che inoltre è accaduto nello sviluppo storico è che il concetto di funzione al tempo di Newton e Leibniz³⁶ era un po' diverso da quello odierno, e si restringeva

³⁵ Il nome "differenziale" deriva ovviamente da "differenza", ed è legato al concetto di velocità, che si calcola prendendo le differenze di posizioni successive nel tempo, calcolando la velocità media e cercando di prenderne un "valore limite" quando l'intervallo di tempo considerato diventa arbitrariamente piccolo.

³⁶ Gottfried Wilhelm von Leibniz (Lipsia, 1646 – Hannover, 1716), filosofo e matematico tedesco. Uno dei fondatori, assieme a Newton, del calcolo infinitesimale.

alle cosiddette funzioni elementari (polinomi, radicali, funzioni trigonometriche ed esponenziali, le loro inverse, tutte le loro possibili composizioni, insomma quelle che si incontrano a scuola).

Mentre per le equazioni algebriche si cominciava a disporre di una teoria soddisfacente (si sapeva, per esempio, che era inutile cercare formule risolutive contenenti solo radicali per equazioni dal quinto grado in su), la teoria delle equazioni differenziali presentava aspetti misteriosi. Per esempio, era facile riconoscere che le traiettorie dei pianeti attorno al Sole (trascurando le reciproche influenze) erano ellissi con il Sole in uno dei fuochi (la prima legge di Keplero³⁷), però era molto più difficile trovare la legge oraria.

L'idea successiva fu quindi quella dell'approssimazione. Già ai tempi di Newton si sapeva che le funzioni si possono descrivere per mezzo di funzioni più semplici, a patto di commettere qualche errore. Per esempio, un arco della funzione coseno può essere approssimato, nel suo punto di massimo, con un arco di parabola (e ve n'è una sola che la descrive con il minimo scarto), ma anche con una funzione di quarto grado, e così via: aumentando il grado della funzione approssimante si riduce l'errore, almeno nei punti vicini al punto in cui si approssima.

Il matematico inglese Brook Taylor³⁸ dimostrò che numerose funzioni si potevano descrivere esattamente, ma considerando una serie infinita di funzioni polinomiali approssimanti. Fu chiaro anche che queste ultime, le cosiddette funzioni analitiche, comprendevano anche nuove funzioni: infatti, come si vede quando si studia il calcolo differenziale in quinta superiore, l'operazione di derivazione trasforma una funzione elementare in una funzione elementare, ma l'operazione inversa (che prende il nome di integrazione) invece no; questa poteva quindi essere una causa del fatto che di certe equazioni differenziali non si riuscivano a trovare le soluzioni.

Tra l'altro, lo sviluppo in serie di Taylor (questo è il nome che ha preso la scoperta di cui abbiamo riferito) offre un vantaggio anche alla fisica: se si è in possesso della posizione e della velocità iniziale e se si conosce la soluzione sotto forma di serie infinita, si può calcolare la posizione in un istante abbastanza vicino a quello iniziale con precisione voluta, anche se non "esattamente". Infatti, se

³⁷ Keplero, nome italianizzato di Johannes Kepler (Weil, 1571 – Ratisbona, 1630) astronomo tedesco. Formulò le leggi del moto dei pianeti intorno al Sole e compì studi di ottica.

³⁸ Brook Taylor (Edmonton, 1685 – Londra, 1731), matematico inglese. Ha fornito contributi fondamentali all'approssimazione delle funzioni.

$t < 1$, le potenze via via più elevate di t come t^2, t^3, \dots produrranno dei contributi che saranno via via più piccoli, e da un certo punto in poi potranno essere trascurati.

Anche in questa fase la meccanica dei pianeti si rivelò fonte di ispirazione: J.-L. Lagrange³⁹ calcolò uno sviluppo in serie che descriveva la legge oraria dei pianeti, e successivamente il fisico e matematico J.-B. Fourier⁴⁰ generalizzò l'idea di Taylor considerando funzioni approssimanti non polinomiali ma trigonometriche, che si prestavano molto meglio a descrivere fenomeni periodici come le orbite dei corpi celesti, e infine nel XIX secolo l'astronomo F.W. Bessel⁴¹ applicò queste idee ai pianeti scoprendo le funzioni che portano il suo nome e che sono usate pressoché ovunque in fisica. Si sa che i matematici, risolvendo un problema, ne creano altri dieci: uno dei problemi che si posero con le equazioni differenziali era l'esistenza e l'unicità della soluzione.

Il problema dell'esistenza di una soluzione di un'equazione differenziale non è sempre un capriccio matematico: nella stessa meccanica dei pianeti, il campo di forze dell'attrazione gravitazionale che esprime la forza fra due punti materiali di masse m_1 e m_2

$$\mathbf{F}(\mathbf{r}) = -Gm_1m_2 \frac{\mathbf{r}}{|\mathbf{r}|^3}$$

(dove \mathbf{r} è il vettore congiungente i due punti e $|\mathbf{r}|$ il suo modulo o lunghezza) non è definita se $\mathbf{r} = \mathbf{0}$, ossia se i due punti materiali coincidono.

È chiaro che in quelle condizioni il modello del punto materiale viene meno: ogni corpo celeste ha delle dimensioni, che non possono essere trascurate in caso di collisione. Tuttavia, la dimostrazione matematica che la collisione in certi casi non avviene, cosa che succede se la soluzione esiste per tutti i tempi futuri, è rassicurante. In maniera analoga si può vedere la questione dell'unicità: potrebbe

³⁹ Joseph-Louis Lagrange (Torino, 1736 – Parigi, 1813), matematico italiano. Ha fornito contributi fondamentali in meccanica analitica e calcolo delle variazioni.

⁴⁰ Jean-Baptiste-Joseph Fourier (Auxerre, 1768 – Parigi, 1830), matematico e fisico francese. Fondamentali i suoi studi sulla propagazione del calore e sull'approssimazione di funzioni.

⁴¹ Friedrich Wilhelm Bessel (Minden, 1784 – Königsberg, 1846), matematico e astronomo tedesco. Ha fornito contributi fondamentali in astronomia, servendosi di funzioni speciali che portano il suo nome.

infatti darsi che, in corrispondenza a certi dati iniziali, non esista una sola soluzione fornita dal modello. Se si crede nella bontà infinita di un modello, ciò è inaccettabile, ma se si pensa che un modello matematico possa mancare di ulteriori specificazioni, allora si può accettare questa eventualità. Evidentemente in questo caso non sono fornite ulteriori informazioni che permettano di discriminare tra l'una e l'altra soluzione.

I progressi ottenuti attraverso gli sviluppi in serie aprono un percorso che porta naturalmente all'analisi numerica moderna. È chiaro che cercare le soluzioni sotto forma di serie può essere insoddisfacente, visto che è tecnicamente impossibile calcolare infiniti termini. E se, limitandosi a una somma finita, si deve commettere un errore non è detto che sia questo l'unico modo o quello più efficiente. Torneremo fra poco su questo fatto.

Quanto a Newton e ai primi successi delle equazioni differenziali, il fatto che, una volta individuate le soluzioni dell'equazione e trovata quella che rispetta le date condizioni iniziali (nel caso della meccanica, la posizione iniziale e la velocità iniziale), essa sia definita per ogni istante successivo, ha promosso la celebre frase di Pierre-Simon de Laplace⁴²:

“Un'Intelligenza che, per un dato istante, conoscesse tutte le forze da cui è animata la natura e la situazione rispettiva degli esseri che la compongono, se per di più fosse abbastanza profonda per sottomettere questi dati all'analisi, abbraccerebbe nella stessa formula i movimenti dei più grandi corpi dell'universo e dell'atomo più leggero: nulla sarebbe incerto per essa e l'avvenire, come il passato, sarebbe presente ai suoi occhi.”

Con la teoria di Newton fu possibile prevedere l'esistenza del pianeta Nettuno nel 1846 sulla base delle sole perturbazioni dell'orbita di Urano, scoperto solo circa 50 anni prima. Le eclissi di Sole furono previste con precisione sempre maggiore e infine i calcoli meccanici hanno permesso, unitamente agli altri progressi delle conoscenze, il volo spaziale. Divenne naturale, anche senza bisogno di Nettuno, cercare di estendere lo spirito del principio “forza = massa \times accelerazione” anche ad altri ambiti della meccanica, come la fluidodinamica, iniziando da Eulero.

⁴² Pierre-Simon de Laplace (Beaumonten-Auge, 1749 – Parigi, 1827), matematico, fisico e astronomo francese. Diede contributi fondamentali nel campo della meccanica celeste.

3. La fluidodinamica

Se si vuole descrivere un corpo continuo, quale un fluido, il modello del punto materiale non è necessariamente il più conveniente. Un'altra descrizione utile si ha mediante una nuova funzione che descrive le proprietà fisiche (come, nei fluidi, velocità, temperatura o pressione) in un determinato punto del corpo e in un determinato istante. Anche questo è un campo vettoriale, come lo era il campo di forze di cui abbiamo parlato, ma con l'aggravante di essere incognito: anzi, esso è l'incognita del problema. Se si cerca di riformulare un'equivalente della legge di Newton in questo ambito ci si rende conto che le mutue interazioni all'interno del corpo non possono non essere tenute in considerazione, e da ciò emerge il concetto di pressione. La pressione è l'esempio forse più semplice di come un corpo continuo deformato risponde alla sollecitazione: per esempio, una compressione. Anche questa è una modellizzazione, molto sottile se vogliamo, ma estremamente efficace. Senza entrare troppo in dettagli, è noto che le differenze di pressione da un punto a un altro del corpo sono sorgenti di movimento, e quindi la "forza dovuta alla pressione" si traduce nel concetto matematico di gradiente di pressione, e va aggiunta alle forze esterne, come la gravità, che agiscono sul fluido. La massa si sostituisce con la densità ρ e la legge di Newton assume in questo ambito la forma⁴³

$$\rho \dot{v} = \mathbf{f} - \text{grad } p \quad (3)$$

nella quale v è il campo incognito di velocità e \mathbf{f} è quello delle forze note agenti sul fluido, mentre $\text{grad } p$ rappresenta il gradiente di pressione. È anche questa una forma di derivazione, ma non è questo il punto che ci interessa qui. Infine, come per Newton il punto indica la derivazione rispetto al tempo. Questa equazione si chiama equazione di Eulero⁴⁴.

Oltre a questa condizione, poi, si deve esprimere il fatto che la massa totale del fluido si conserva, che porta a un'ulteriore condi-

⁴³ È bene sottolineare che non si tratta della stessa equazione, ma di un'equazione analoga, introdotta come modello ispirato alla generalizzazione della legge (1). Dimostrare che dalla meccanica dei punti, aumentando il loro numero e diminuendo le masse, segue l'equazione data è cosa molto difficile e non ancora ben compresa.

⁴⁴ Eulero, nome italianizzato di Leonhard Euler (Basilea, 1707 – San Pietroburgo, 1783), matematico e fisico svizzero. Uno degli scienziati più importanti del XVIII secolo, ha fornito contributi decisivi in molti campi della matematica.

zione sul campo di velocità e sulla densità. Ma torniamo alle differenze con la meccanica newtoniana. Qui ora la difficoltà matematica in gioco è diversa: essa non sta tanto nel fatto che la forza esterna \mathbf{f} possa divenire un campo più o meno complicato, ma risiede nella pressione p . L'analogia con la meccanica newtoniana, infatti, porta sì a concludere che deve esistere una forza indotta dalla pressione, ma non specifica in alcun modo come questa dipenda dalle variabili in gioco. Pertanto scopriamo che possiamo modellizzare fluidi diversi specificando come questa forza dipenda dalle variabili. In questa fase non è più così evidente quale sia il principio ispiratore che dobbiamo seguire, e quindi si apre una fase di riflessione sul modello più opportuno da scegliere.

La scelta più semplice è quella di supporre che il fluido sia incomprimibile, ossia che abbia sempre la stessa densità in ogni punto del corpo. Allora la pressione si scopre essere determinata univocamente dal problema (ma resta un'incognita il suo valore in ogni punto), e l'analisi del modello può proseguire. Si incontra a questo punto una difficoltà nuova: l'equazione è molto più difficile, in quanto i campi incogniti v e p dipendono sia dallo spazio che dal tempo, e quindi le derivate che compaiono nella legge (3) sono derivate diverse (si chiamano derivate parziali; incidentalmente, anche la derivazione rispetto al tempo assume un significato più complicato in questo ambito). La teoria matematica delle equazioni alle derivate parziali è estremamente più complessa di quella delle equazioni "ordinarie" del tipo (2). Esistenza e unicità sono spesso problemi ancora irrisolti e, di conseguenza, la bontà delle approssimazioni non è più così sicura come nel caso ordinario. A questo si aggiunge un ulteriore problema che affronteremo tra poco.

Il fatto che le sole forze in gioco nel fluido siano dovute al gradiente di pressione ha comunque permesso ai matematici di dedurre interessanti conseguenze sulle possibili soluzioni, tra le quali il fatto che, secondo questo modello, se un fluido in quiete non presenta vortici, ossia zone di rotazione locale del fluido su se stesso, allora non è possibile che si formino negli istanti successivi. Questa previsione, che è, ripeto, una deduzione matematica, è in conflitto con l'esperienza in maniera stridente: basta vuotare una bottiglia piena d'acqua rovesciandola perché essa sia smentita. Dunque il modello è incompleto e va corretto (anche se l'approssimazione del fluido perfetto – così si chiama quello "euleriano" – continua a essere usata quando la presenza di vortici non è attesa o non è rilevante). Un mo-

do di superare questo inconveniente è quello di introdurre il concetto di viscosità: si tratta di un fenomeno complesso, a rigore diverso da materiale a materiale, se lo si considera dal punto di vista microscopico. In vista però del fatto che si cerca una modellizzazione macroscopica, che non pretende di descrivere ciò che avviene su scala molecolare, possiamo decidere di semplificare il modello introducendo, per così dire, il “più semplice” termine matematico che renda ragione del fenomeno viscoso.

Per renderci conto in maniera semplice di come ciò avvenga, torniamo per un attimo al semplice punto materiale. Immaginiamo che esso si muova in un mezzo viscoso a velocità costante. Perché esso continui nel suo moto, è necessario imprimere una forza che controbilanci la resistenza viscosa del mezzo. Sperimentalmente, e anche intuitivamente, questa forza dipenderà dalla velocità alla quale si vuol muovere il punto, e il più semplice modello che possiamo concepire è quello di una forza proporzionale alla velocità. Quindi la forza di resistenza viscosa per un punto materiale avrà l'espressione

$$\mathbf{F} = -\mu\mathbf{v} \quad (4)$$

dove μ è un coefficiente che rende conto della viscosità del materiale. Se si vuole passare però a un mezzo continuo, la situazione si complica lievemente, in quanto i vari punti che lo compongono possono essi stessi essere in movimento a causa del fatto che il corpo si espande o si contrae. Se, per esempio, una goccia d'acqua si contrae in una direzione, i suoi punti avranno velocità diverse e non è ben chiaro a quale velocità si debba attribuire la forza viscosa. Risulta infatti che la velocità “giusta” da considerare è la differenza di velocità tra i vari punti, cioè la velocità relativa, e siccome nel modello continuo i punti sono infinitamente vicini (forza e debolezza di questo tipo di modello), ne segue che le forze di attrito viscoso dipendono dalle derivate spaziali del campo di velocità.

Senza entrare nei dettagli dei calcoli, che non sono tuttavia complessi, il risultato finale che otteniamo nel modello viscoso conduce ad aggiungere un termine contenente le derivate seconde del campo di velocità (analogamente a quanto è capitato per la pressione, che ha avuto bisogno del concetto di gradiente) e che prende il nome di “laplaciano del campo di velocità”, indicato con $\Delta\mathbf{v}$, ma che per noi è solo un simbolo.

Pertanto l'equazione di Eulero (3) si modifica come segue:

$$\rho \dot{\mathbf{v}} = \mathbf{f} - \text{grad } p + \mu \Delta \mathbf{v} \quad (5)$$

L'equazione testé scritta è molto famosa e prende il nome di equazione di Navier-Stokes. Essa costituisce il più semplice e tuttavia potente modello di previsione dei fenomeni fluidodinamici.

Questa equazione è comunque molto diversa dall'equazione di Newton della meccanica del punto, almeno sotto il profilo dell'analisi. Vediamo perché. Il primo rilievo immediato è che si tratta di un'equazione alle derivate parziali. Questo implica che la soluzione incognita è un campo dipendente non solo dal tempo, ma anche dalle coordinate spaziali. In questo primo ordine di idee non esiste un risultato di esistenza e unicità come si può ottenere per le equazioni differenziali ordinarie della meccanica del punto materiale. Ma questo è solo un primo aspetto. Il più importante è in realtà un altro: siccome la soluzione è un campo di velocità, si tratta di una funzione definita nello spazio su un certo dominio, che è poi lo spazio occupato dal fluido. Questo dominio è spesso un'incognita ulteriore del problema (si pensi a come si può muovere la superficie di un fluido), e in ogni caso va considerato noto con precisione assoluta. A ciò si aggiunge il fatto che anche la condizione iniziale è una funzione definita su un dominio, e la sua conoscenza assoluta è solo un'approssimazione, in realtà tutt'altro che scontata, come vedremo. Infine, non si deve trascurare un ultimo aspetto: anche quando il fluido è contenuto in un recipiente, e quindi la sua forma è conosciuta, non è affatto chiaro come esso si debba comportare rispetto al recipiente; esso potrebbe aderire o scivolare, e questa è in definitiva una modellizzazione ulteriore che lo sperimentatore si ritrova.

4. Previsione dei risultati

Abbiamo detto che una grande fortuna della meccanica celeste consistette nella precisione delle sue previsioni. Senza entrare nei dettagli, ciò fu dovuto essenzialmente al fatto che si riuscì a trovare una rappresentazione opportuna, sotto forma di sviluppo in serie convergente, della soluzione del problema, nella quale il trascurare gli ultimi (anche se infiniti) termini produceva un errore trascurabile. Una simile opportunità non è affatto presente in fluidodinamica, ma soffermiamoci ancora un momento sulla meccanica ce-

leste. Quando i corpi in gioco sono molti, come nel sistema solare, anche gli sviluppi in serie diventano impraticabili, e si fa strada un modo di pensare diverso. L'idea, che è descritta nel capitolo "Computer e soluzioni approssimate", consiste nel sostituire il problema ottenuto dal modello con un nuovo problema, detto approssimato, più trattabile del primo e però da esso distinto. Spesso questo nuovo problema è discreto, ossia non è descritto da una funzione ma da un numero finito (anche molto alto) di numeri, il che lo rende implementabile su un computer. Nasce così l'*analisi numerica*.

Per esempio, per il problema della meccanica newtoniana, un'idea (non l'unica), consiste nel sostituire la velocità istantanea con la velocità media presa su un intervallo di tempo h , finito (e in genere piccolo rispetto all'intervallo di tempo studiato), cioè l'equazione (2) viene sostituita da

$$m \frac{\mathbf{v}(t+h) - \mathbf{v}(t)}{h} = \mathbf{F}(\mathbf{x}(t)) \quad (6)$$

A questo punto viene messo in moto uno schema ricorsivo: note le condizioni iniziali e fissato $h > 0$, $\mathbf{x}(t_0)$ e $\mathbf{v}(t_0)$, si suppone che il moto sia uniformemente accelerato con accelerazione $\mathbf{F}(\mathbf{x}(t_0))/m$ negli istanti fra t_0 e $t_0 + h$, cosa che non è vera esattamente, producendo la posizione $\mathbf{x}_h(t_0 + h)$ dalle note formule

$$\mathbf{x}_h(t_0 + h) = \mathbf{x}_h(t_0) + \mathbf{v}(t_0)h + \frac{1}{2} \frac{\mathbf{F}(\mathbf{x}(t_0))}{m} h^2$$

e si usa poi la (6) per calcolare la velocità all'istante $t_0 + h$:

$$\mathbf{v}_h(t_0 + h) = \mathbf{v}(t_0) + \frac{h}{m} \mathbf{F}(\mathbf{x}(t_0))$$

A questo punto sono disponibili due nuovi valori di posizione e velocità all'istante $t_0 + h$ e si può ripetere lo schema per calcolare i valori all'istante $t_0 + 2h$, e così via.

Appaiono evidenti a questo punto due osservazioni: la prima è che il problema *approssimato* è un *nuovo* problema, distinto dal precedente (e che quindi potrebbe necessitare di studi matematici non in-

feriori di quelli dedicati al problema di partenza); la seconda è che si ha l'impressione che il problema approssimato tenda a quello di partenza quando h è piccolo.

Questo secondo aspetto è molto delicato: infatti, se può essere (relativamente) facile chiarire che cosa significhi che una *funzione reale* tende a un numero mediante il concetto di limite in maniera rigorosa o comunque lasciandolo intuire, risulta molto più difficile farlo con un intero problema, e su questo vertono molti degli studi analitici sull'analisi numerica. In questo senso si introduce il concetto di errore, inteso (per esempio) come modulo della differenza tra i valori della soluzione dell'equazione differenziale e del valore dello schema iterato $|x(t) - x_h(t)|$ a un dato istante t . Il termine "errore" è qui pericoloso, perché non va confuso con gli errori sperimentali o di modello, ai quali si può però aggiungere: si tratta di una discrepanza tra le soluzioni di due problemi distinti.

Abbiamo compiuto questa breve digressione sull'analisi numerica per il semplice fatto che essa si rivela l'unico ausilio per trattare problemi legati all'equazione della fluidodinamica di Navier-Stokes, ma rimangono numerose ulteriori complicazioni: in questo ambito, infatti, non solo il tempo va "discretizzato", ma anche il dominio spaziale, e le condizioni iniziali⁴⁵. Tutto ciò rende il problema approssimato molto più lontano dal problema di partenza, o comunque a esso più "fragilmente legato", e in definitiva fa sì che anche in mancanza di fatti nuovi la previsione può risentire di errori di approssimazione o sperimentali in maniera significativa. È questo un motivo per cui le previsioni numeriche meteorologiche non sono affidabili oltre cinque giorni, mentre conosciamo al decimo di secondo la data della prossima eclisse totale di Sole. Si potrebbe in definitiva pensare che sia comunque una questione di forza: parafrasando Laplace, se disponessimo di un computer abbastanza potente, saremmo in grado di prevedere il tempo anche fra vent'anni. Potrebbe essere vero, a patto di capire bene il significato di "abbastanza potente". E qui nascono le difficoltà.

5. Dalla fluidodinamica ai sistemi complessi

Poco prima dell'avvento dell'analisi numerica, i matematici non avevano ancora perso la speranza di ridurre i problemi fluido-

⁴⁵ E naturalmente, se si usa una implementazione su computer, anche la rappresentazione dei singoli numeri reali va sostituita da una approssimazione razionale.

dinamiche a sviluppi in serie, sperando di aumentare la precisione delle previsioni. Del resto, ancora nella prima metà del '900 nessuno si sarebbe sognato di chiedere a un matematico di risolvere un problema tecnico di idraulica: le leggi matematiche dedotte dalle equazioni di Eulero e di Navier-Stokes costituivano, per così dire, delle linee-guida, che restavano però ben lontane dal sostituire la pratica ingegneristica. Accadde così che lo studioso americano Edward Lorenz⁴⁶ ebbe l'idea di semplificare il problema di Navier-Stokes, in una situazione molto particolare, per riottenere un sistema di equazioni differenziali ordinarie, alle quali l'analisi numerica poteva offrire una maggiore varietà di strumenti. Senza entrare nei dettagli, che ci porterebbero fuori dal discorso, basti ricordare che le incognite del problema (di fatto un problema termofluidodinamico, cioè nel quale anche la temperatura entra in gioco) sono le ampiezze dei primi "modi normali" della soluzione nello spazio. L'idea è quella sopra accennata di Fourier, cioè lo sviluppo in serie di funzioni trigonometriche; la soluzione appare come una sovrapposizione (somma) di questo tipo di funzioni nella variabile spaziale, in maniera analoga a quanto avviene per un suono composto in musica (dove invece queste funzioni variano nel tempo: le loro intensità, cioè i coefficienti della somma, determinano il suono completo e, in generale, le frequenze più alte sono associate ad ampiezze via via meno rilevanti dal punto di vista energetico, per cui si possono trascurare in prima approssimazione). Lorenz pensò che troncando lo sviluppo in serie di Fourier della soluzione, senza di fatto stimare l'errore compiuto, e inserendo la soluzione monca nelle equazioni di Navier-Stokes della termofluidodinamica si sarebbe ottenuto un sistema di equazioni differenziali dal quale dedurre alcune previsioni. Lorenz scelse tre ampiezze incognite, visto che potevano rappresentarsi con un sistema non troppo difficile. Se indichiamo con $x(t)$, $y(t)$, $z(t)$ i valori delle tre funzioni, il sistema assume la forma

$$\begin{cases} \dot{x} = -\sigma x + \sigma y \\ \dot{y} = rx - y - xz \\ \dot{z} = xy - bz \end{cases}$$

dove σ , r , b sono parametri provenienti dal sistema fisico in esame.

⁴⁶ Edward Norton Lorenz (West Hartford, 1917 – Cambridge (Massachusetts), 2008), matematico statunitense. È il pioniere della teoria del caos.

Dal punto di vista newtoniano, la situazione sembra peggiore: abbiamo non una ma tre funzioni incognite, però perlomeno non ci sono i due punti (incidentalmente, le equazioni di Newton sono più difficili perché l'incognita è vettoriale). Il sistema di Lorenz, in ogni caso, non fornisce soluzioni esatte, né sviluppi in serie convergenti in maniera utile per delle applicazioni. Però l'analisi numerica è applicabile e produce un risultato per certi versi sorprendente. Per alcuni valori dei parametri, infatti, condizioni iniziali molto vicine danno origine a soluzioni qualitativamente molto diverse, e soprattutto oscillanti in maniera assolutamente irregolare.

Quando si diffusero nella comunità scientifica, i risultati di Lorenz furono presi come una conferma del fatto che la turbolenza era insita nelle equazioni della fluidodinamica. Il problema della turbolenza era ben noto ai matematici da più di un secolo: l'osservazione dei fenomeni fluidodinamici portava a concludere che in certi casi da una situazione semplice e regolare si sviluppava improvvisamente un moto molto complicato e privo di struttura. Per esempio, se si pompa dolcemente dell'acqua in un tubo cilindrico, essa fluisce in maniera regolare e le traiettorie delle particelle assomigliano alla forma del tubo; ma se si aumenta la pressione sopra un valore critico l'acqua comincia a formare vortici, le particelle oscillano in maniera non periodica e il flusso totale si riduce.

Sono state sviluppate molte teorie per dare una spiegazione matematica della turbolenza e, soprattutto, dei criteri pratici per tenerla sotto controllo, dato che comporta notevoli sprechi. Con la modesta spiegazione di Lorenz (che matematicamente, però, non aggiunge molto, visto che non considera l'intera soluzione ma solo un suo troncamento) si aggiunse un tassello alla comprensione di questo aspetto del concetto emergente di complessità.

Il risultato di Lorenz ha rilevanza ai fini del nostro discorso soprattutto per un altro punto, cioè la dipendenza dalle condizioni iniziali, più che per la previsione dell'irregolarità delle oscillazioni. Gli studi seguenti a Lorenz hanno infatti mostrato che in quasi tutti i sistemi non lineari si presenta il fenomeno della divergenza delle orbite: pur partendo da dati iniziali pressoché identici, dopo breve tempo le corrispondenti soluzioni si allontaneranno molto, in maniera tale che richiedere una precisione significativa sul risultato previsto richiede una precisione inimmaginabile sulle condizioni iniziali. Senza entrare nei dettagli, la precisione richiesta per avere una discrepanza d di una quantità dopo t secondi tende a 0 come

$d e^{-\sigma t}$, cioè esponenzialmente, e, il che è peggio, con un coefficiente σ in generale piuttosto elevato. Dobbiamo infine precisare che questo fenomeno non riguarda la soluzione approssimata dell'analisi numerica, ma la soluzione esatta dell'equazione differenziale di Lorenz, e la sua discretizzazione (la sostituzione del modello matematico iniziale con una sua approssimazione) aggiunge in realtà altra incertezza alla previsione.

Sorpresa nella sorpresa, nel sistema di Lorenz e in molti altri sistemi accade anche questo: se il valore dei parametri (non delle condizioni iniziali) viene cambiato di pochissimo, il fenomeno della divergenza delle orbite potrebbe sparire di colpo. Vi è quindi non solo una sensibilità rispetto alle condizioni iniziali, ma anche rispetto ai valori dei parametri del sistema. Questo da un certo punto di vista è ancora più radicale: infatti, per il fenomeno della divergenza delle orbite esiste, in generale, un teorema che dà la stima della precisione richiesta ai dati iniziali per avere una precisione voluta, come abbiamo detto poco fa, ma per la sensibilità rispetto al valore dei parametri non c'è in generale alcuna garanzia che le intere orbite corrispondenti allo stesso dato iniziale con parametri diversi siano simili per tempi lunghi (per tempi brevi invece sì).

Ecco quindi che la frase di Laplace, pur vera se vogliamo, prende un sapore amarognolo: conoscere la "situazione degli esseri" (cioè le condizioni iniziali) per poter prevedere il futuro con precisione assoluta è impossibile nella pratica, e l'approssimazione non serve se non per pochi casi fortunati. Inoltre, anche il "conoscere tutte le forze" (ossia i valori dei parametri) è cosa impossibile e produce risultati devastanti anche commettendo il minimo errore. Ma crediamo che i matematici si sarebbero arresi? Ovviamente no.

6. Frattali e complessità

Un punto che si muove sotto l'azione di forze, come nel caso della meccanica newtoniana, percorre in genere, a parte eventuali urti con altri punti, una curva regolare, senza spigoli o rimbalzi. È una conseguenza del fatto che l'accelerazione varia con continuità, e quindi del fatto che la forza varia con continuità. Tra l'altro, anche se la forza salta improvvisamente, come quando si sente di colpo la forza centrifuga entrando in una curva, la traiettoria non presenta spigoli. Queste però non sono le uniche curve continue che si possono pensare, e ancora una volta le idee più innovative si devono ai matema-

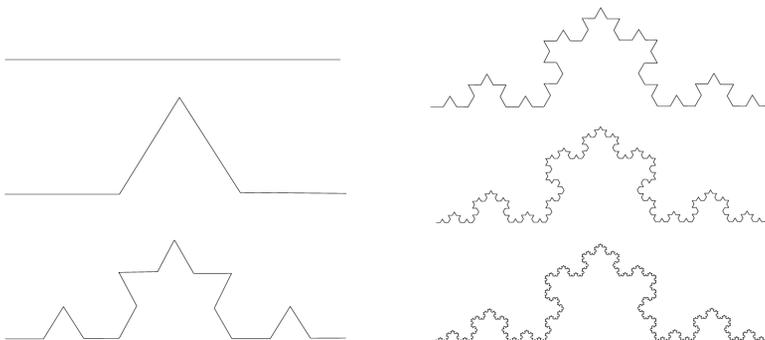


Figura 1. La costruzione della curva di von Koch.

tici . Se una curva ha uno spigolo, allora la sua retta tangente salta improvvisamente: per esempio, se la curva è una legge oraria della posizione, ciò significa che la velocità salta improvvisamente (come in un rimbalzo). Dunque *non* è possibile ottenere la velocità dalla posizione con l'operazione di derivazione: in questo caso si dice che (limitatamente al punto di rimbalzo) la curva non è derivabile.

Naturalmente si può pensare che i rimbalzi siano molti, anche moltissimi: ma è possibile che una curva abbia rimbalzi in ogni punto? Questo pare più arduo: ci vuole molta fantasia per immaginare una curva del genere (oltretutto una curva continua, senza cioè che il corpo scompaia in un punto e riappaia in un altro). Non a caso passarono quasi due secoli prima che si scoprisse che era possibile! Il matematico tedesco Karl Weierstrass trovò un esempio di funzione continua non derivabile in nessun punto: un vero mostro matematico, una curva che non si può nemmeno disegnare. Per anni si pensò che queste curve fossero semplicemente delle curiosità, finché venne scoperto il cosiddetto *moto browniano*, uno strano moto a zigzag di punti immersi in un fluido, dovuto ai continui urti da parte delle sue molecole. La frequenza di questi urti era così alta che si poteva modellizzare il fenomeno ammettendo urti in ogni istante, e anche il "mostro" di Weierstrass trovò un'applicazione, perlomeno concettuale.

Ma l'interesse dei matematici era rivolto a un altro aspetto, quello geometrico. Sembrava incredibile che esistessero curve che non si possono nemmeno disegnare: eppure gli esempi si moltiplicavano. Il più celebre per la sua semplicità è probabilmente la *curva di von Koch*, costruita come segue. Prendiamo un segmento di lunghezza

data e togliamone il terzo centrale, sostituendolo con due segmenti uguali in modo da formare un triangolo equilatero (vedi Figura 1). Osserviamo che la spezzata che abbiamo costruito ha lunghezza $4/3$ di quella iniziale. Poi ripetiamo questa operazione con ognuno dei quattro segmenti che abbiamo ottenuto: abbiamo una nuova spezzata, che avrà lunghezza $4/3$ della precedente, cioè $16/9$ di quella iniziale, e continuiamo l'operazione. Troveremo spezzate costituite da pezzi sempre più corti e numerosi, con lunghezza totale sempre maggiore (dopo n iterazioni ci saranno 4^n segmenti ciascuno di lunghezza $1/3^n$ di quello iniziale per una lunghezza totale pari a $(4/3)^n$ volte quella iniziale. A dispetto del fatto che la lunghezza diventa sempre maggiore, la curva non si allontana all'infinito ma diventa sempre più spigolosa, e si può, in maniera ragionevole, definire la curva limite ideale corrispondente a un numero infinito di iterazioni. Anche la curva di von Koch è indisegnabile: ci si rende conto di ciò pensando che nelle varie iterazioni il segmentino iniziale diviene sempre più corto, e quindi nella curva limite non si può nemmeno capire in quale direzione ci si debba muovere (disegnare è un'operazione meccanica, quindi newtoniana, a meno di voler considerare il tremolio della mano...). Oltretutto, ha "lunghezza infinita".

Queste erano e restarono a lungo delle curiosità. Un aspetto però interessante è quello della dimensione: tutti sanno che una linea ha dimensione 1, una superficie 2, un volume 3, ma se si chiede cosa sia la dimensione si rischia di ottenere curiose risposte. Il matematico Felix Hausdorff⁴⁷ ne scoprì una definizione interessante e, almeno apparentemente, facile. Prendiamo un quadrato e dividiamone il lato in n parti. Il quadrato si divide allora in n^2 quadrati, ciascuno copia esatta del quadrato di partenza. Questo è per Hausdorff indice del fatto che il quadrato ha dimensione 2, l'esponente di n^2 . Se ripetiamo l'idea con un cubo, dividendo lo spigolo in n parti, troviamo n^3 cubetti, e quindi la dimensione del cubo è 3. Se invece del cubo si ha una sfera, la situazione si complica un po', ma non gravemente, e non ci interessa approfondire la questione per non perdere di vista il ragionamento.

Cerchiamo ora di ripetere la stessa idea con la curva di von Koch. Qui è più complicato, perché essa ha lunghezza infinita, e quindi il metodo di Hausdorff non funziona a prima vista. Ma esso si può

⁴⁷ Felix Hausdorff (Breslavia, 1868 – Bonn, 1942), matematico tedesco. Ha fornito contributi importanti sulla teoria degli insiemi e sull'analisi funzionale.

riformulare così: nel quadrato iniziale si vedono delle copie simili al quadrato stesso con rapporto di similitudine delle aree uguale a $(1/n)^2$ e lato $1/n$. Dalla figura della curva di von Koch appare che vi sono delle copie della stessa curva evidentemente riscalate di $1/3$, nelle quali ogni riscalamento si ottiene dividendo in 4 la lunghezza iniziale. Generalizzando l'idea di Hausdorff, dovremmo trovare l'esponente x tale che

$$\left(\frac{1}{3}\right)^x = \frac{1}{4}$$

ossia $3^x = 4$, da cui $x = \log 4 / \log 3$. Appare perciò che è possibile una generalizzazione del concetto di "dimensione" che si applica anche alla curva di von Koch, e che questa dimensione può non essere intera!

Nessuna meraviglia, quindi, se i fisici si tennero alla larga da queste stranezze, ma esse riemersero in maniera inaspettata proprio nelle applicazioni. Queste curve vennero rapidamente generalizzate a oggetti più complessi e simili a superfici o polveri, e presero il nome di "frattali". La loro dimensione (sulla quale non c'è ancora una definizione univoca) resta un indicatore importante della loro natura.

Torniamo all'esempio di Lorenz. Negli anni '70 si capì che numerosi sistemi come quello di Lorenz, che presentavano il comportamento irregolare al variare dei parametri in essi contenuti, potevano essere descritti usando il concetto di frattale. Possiamo qui solo accennare all'idea: abbiamo visto che in sistemi di questo tipo le orbite corrispondenti a punti vicini tendono a divergere con una velocità esponenziale, con un fattore σ all'esponente. Questo non è sempre vero, ma supponiamo per semplicità che sia così. Se il sistema dipende da alcuni parametri, il numero σ dipenderà anch'esso da questi parametri, e ci si può chiedere come. La risposta è che in generale questa relazione sarà un frattale. Altrettanto accade per altri indicatori importanti della natura del sistema.

Ecco quindi che si delinea una situazione problematica per lo sperimentatore: anche ammettendo che il modello sia ragionevolmente giusto, non sarà possibile in generale avere una misura esatta dei valori dei parametri (né delle condizioni iniziali), per cui ogni successiva miglioria delle misurazioni potrebbe avere un valore devastante sulle previsioni. Anche le previsioni numeriche del modello, che per loro natura, abbiamo visto, sostituiscono al modello uno schema di-

verso, rischiano di portare a deviazioni enormi dal sistema originario, al punto di non poter più parlare di previsione. Questo e altri fenomeni hanno ricevuto il nome significativo di “caos deterministico”, a sottolineare che anche sistemi nei quali in linea di principio è possibile la conoscenza “esatta” dello stato del fenomeno possono presentare comportamenti così complessi da rendere vani i vantaggi del determinismo. In ogni caso, però, un progresso di conoscenze c'è: per esempio, avere informazioni sulla dimensione frattale degli insiemi che governano il sistema può rivelare analogie di comportamento inaspettate.

7. Stabilità

Anche lasciando da parte i sistemi complessi, vi è un ultimo aspetto da considerare se si vuol parlare di previsione di risultati da modelli matematici.

Abbiamo visto che i matematici del XIX secolo erano a caccia di soluzioni esatte delle loro equazioni differenziali. Alcune le trovarono anche, ma ciò non basta ancora per le applicazioni. Infatti, come appare chiaro dal discorso che abbiamo fatto, queste soluzioni dipenderanno in generale dalle condizioni iniziali a esse associate. Adesso che ci siamo fatti furbi, ci possiamo chiedere: ma come variano queste soluzioni variando le condizioni iniziali? Infatti, nelle situazioni reali non conosceremo con esattezza i dati iniziali, e quindi la soluzione calcolata potrebbe differire da quella vera. Ora, se le due soluzioni (quella vera e quella approssimata) saranno molto diverse, lo sperimentatore si troverà di fronte a un dilemma: sono le misurazioni dei dati iniziali troppo grossolane, oppure la soluzione prevista dal modello non è di alcuna utilità? Se poi accade che, pur affinando le misure, la divergenza persiste, appare chiaro che la soluzione “esatta” prevista dal modello è un puro gioco matematico.

Anche qui un esempio può chiarire meglio il discorso. Tra le più semplici soluzioni di un problema meccanico ci sono quelle che non dipendono dal tempo, che si chiamano *posizioni di equilibrio*. Il problema dei corpi celesti non ha posizioni di equilibrio, ma molti altri sì, specie in ambito macroscopico: tutta la statica studia le soluzioni di equilibrio di sistemi meccanici. Pensiamo allora a posizionare una sfera sulla cima di una collinetta. Se essa è soggetta solo al proprio peso, il modello matematico prevede che la sfera sulla cima della collinetta, lasciata ferma, resti in equilibrio. La stessa analisi matemati-

ca, però, prevede anche in questo caso che, se la posizione del punto di contatto della sfera non sarà esattamente sulla cima della collinetta, allora la sfera non starà in equilibrio e rotolerà giù. In matematica si usa dire che quella posizione di equilibrio è instabile. Si tratta di una pura astrazione, dal punto di vista sperimentale, perché chiaramente non sarà mai possibile centrare la posizione esatta (o dare velocità iniziale nulla), per cui lo sperimentatore si può addirittura chiedere se questa soluzione esista.

A parte queste considerazioni filosofiche, appare chiaro che le soluzioni di utilità pratica devono essere in qualche modo stabili, vale a dire “resistenti a piccole variazioni dei dati iniziali”, nel senso che, se le condizioni iniziali variano di poco, allora anche le corrispondenti soluzioni dovranno variare di poco (se poi, nell'esempio della sfera, alla forza peso si aggiungono altre forze, il risultato può cambiare radicalmente, ma in questo caso cambia anche il modello; se, per esempio, si aggiunge dell'attrito, una posizione instabile potrebbe diventare stabile).

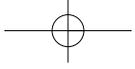
Chiarire che cosa significa questo “di poco” è un interessante concetto matematico. Siccome infatti le previsioni di questo tipo di modelli sono funzioni del tempo, si pone il problema di chiarire cosa voglia dire che due funzioni sono “vicine”, o “differiscono di poco”. Se pensiamo ai grafici di due funzioni, questo non appare affatto evidente: i singoli valori delle due funzioni in un valore x possono variare molto con x , per cui si richiede una certa “uniformizzazione”. L'interessante scoperta, del XX secolo, in questo senso è che bisogna rassegnarsi: esistono modi diversi di misurare la distanza fra funzioni, ma tutte queste distanze, o perlomeno quelle utili, condividono una serie di proprietà che accomunano le funzioni ai punti. Così come si può definire in maniera semplice la distanza fra punti, e dedurre alcune relazioni, altrettanto si può fare (ma non in modo univoco) per le funzioni, e dedurre delle proprietà molto simili a quelle della distanza fra punti.

Ma c'è di più: le funzioni, o perlomeno quelle “che servono”, assomigliano, per certi versi, anche ai vettori. Sappiamo che i vettori si possono decomporre in direzioni ortogonali. Pensiamo, per esempio, a una forza che agisce su un'automobile. Se la decomponiamo in una componente parallela al terreno e una perpendicolare, l'intuizione fisica ci suggerisce che la componente parallela influenzerà il moto dell'auto, mentre quella perpendicolare la sua aderenza al terreno. In questo senso le due componenti si divido-

no i compiti e sono responsabili di effetti diversi. Se riprendiamo ora quanto abbiamo detto a proposito dello sviluppo in serie trigonometrica di Fourier, troviamo che certe funzioni periodiche si possono decomporre in somme di funzioni trigonometriche, ma tali che ognuna di esse si occupa di un aspetto diverso in un certo problema fisico.

Un esempio è quello del suono, che sarà ripreso nel capitolo dedicato all'arte. Nella generazione di un suono complesso, le varie funzioni trigonometriche di Fourier s'interpretano come le armoniche del suono stesso, che in molti casi si possono distinguere e riconoscere: esse giocano il ruolo delle componenti dei vettori nel caso meccanico. Il fatto interessante a questo proposito è che, mentre le forze, come gli altri vettori della fisica, si possono decomporre al più lungo tre direzioni perpendicolari, le componenti trigonometriche di una funzione sono, in generale, infinite. Siccome 3 è il numero di dimensioni dello spazio, i matematici parlano di "vettori di dimensione infinita" per le funzioni. Qui il termine "dimensione" ha un'accezione diversa da quella che abbiamo illustrato nei frattali: è un esempio della cattiva abitudine dei matematici — e non solo loro — di usare parole uguali per concetti diversi in contesti diversi. La stabilità è un tipico problema "in dimensione infinita": si può studiarla in relazione alle condizioni iniziali, come abbiamo visto, ma anche in relazione ai parametri di un sistema o addirittura alla forma matematica del sistema stesso (spesso gli sperimentatori non sono certi, per esempio, che le forze in gioco in un sistema meccanico siano solo quelle che hanno inserito; però, se il sistema accetta piccole perturbazioni di forze non specificate, essi saranno sicuri che gli errori — modellistici, se vogliamo — che hanno introdotto non influiranno qualitativamente sulle previsioni del sistema). Oltre a questa importanza applicativa, essa è fonte di interessanti concetti e studi matematici.

In conclusione, abbiamo visto di quali difficoltà sia cosperso il cammino di coloro che vogliono prevedere il futuro usando la matematica. Bisogna per questo stare attenti da un lato a non eccedere con la fiducia e analizzare criticamente ogni passo che viene compiuto senza sottacere i possibili errori, e dall'altro a non commettere l'errore, ben più grave, di sancire l'impossibilità di effettuare delle previsioni in ambito complesso. In cambio di ogni difficoltà, la matematica ricompensa, a modo suo: ponendo interessanti sfide e nuovi problemi.



LA GESTIONE RAZIONALE DEL CASO

Tutti abbiamo lanciato un dado, l'abbiamo osservato mentre rotolava e abbiamo desiderato che apparisse, alla fine del suo movimento, il risultato utile per noi. Forse pochi di noi si sono chiesti se fosse possibile dosare la velocità del lancio in modo da ottenere il risultato voluto: i più avranno soffiato sulla mano, o adottato qualche altra scaramanzia.

Certo il dado obbedisce alle leggi della meccanica, al pari del moto dei pianeti, che ammette previsioni, come nel caso delle eclissi, di sorprendente esattezza, eppure il dado è imprevedibile, lo sentiamo, perché forse abbiamo coscienza del fatto che non possiamo dosare la velocità del lancio (per inciso, i bambini più piccoli fanno questo abbastanza inconsciamente e spesso "barano" cercando di lasciar cadere il dado dalla mano, in generale con scarso esito perché in quel caso l'altezza della mano incide ancora molto), e forse anche perché osserviamo che quando rotolano non riusciamo nemmeno più a distinguere le loro facce. Non parliamo poi di un'urna che gira con 90 palline: credo proprio che a nessuno venga in mente la prevedibilità dell'estrazione⁴⁸.

Sappiamo che il dado non è prevedibile perché è un sistema complesso e dà origine a risultati che dipendono così sensibilmente dalle condizioni iniziali al punto che il semplice battito del cuore nella mano che lancia il dado, a parità di tutte le altre condizioni, può portare a due risultati completamente diversi.

Ci si chiede allora che cosa possa dire la matematica su una situazione così aggrovigliata: e in effetti riconoscere che la matematica possa giocare un ruolo fondamentale anche in questioni che attengono al caso puro e semplice ha richiesto molti secoli. La matematizzazione del concetto di *probabilità*, l'idea centrale del nostro di-

⁴⁸ In ogni caso sono state inventate truffe sia con i dadi – i dadi truccati – sia con le urne, in questo caso barando sulla superficie delle palline.

scorso, è avvenuta in maniera esplicita solo nel XVII secolo ed è ancor oggi oggetto di numerose discussioni.

1. Un po' di storia

Anche se gli antichi non possedevano un fondamento matematico della probabilità, sicuramente sapevano valutare ciò che è più plausibile da ciò che non lo è. Siccome l'esito di una caccia non è uniforme in tutte le direzioni (è certamente più probabile trovare prede in corrispondenza di corsi d'acqua, o particolari piante, o in certi mesi dell'anno), l'uomo sin dagli albori della civiltà ha associato un concetto di probabilità alle scelte che doveva compiere, in generale però senza farne una teoria.

I primi che cercarono di fare un po' d'ordine in questo concetto pare siano stati i filosofi greci. Secondo Carneade⁴⁹, ben noto ai conoscitori dei *Promessi sposi*, si possono distinguere vari casi del "probabile" (usiamo le virgolette apposta per cercare di distinguere questo nome da quello in uso comune oggi):

1. ciò che è solo probabile;
2. la rappresentazione che appare vera;
3. la rappresentazione persuasiva e non contraddetta;
4. la rappresentazione persuasiva e non contraddetta ed esaminata da ogni parte.

Come si vede, si tratta piuttosto di una classificazione di ciò che appare "vero" che di ciò che *dichiaratamente* non è prevedibile. Forse questa analisi nasceva dalla constatazione di alcuni avvenimenti che si credevano sicuri, ma che in realtà avevano dato esiti imprevisti, come crolli di ponti. Abbiamo parlato dei dadi: non è un caso se il primo interesse per la "probabilità" sia nato in relazione ai giochi d'azzardo ("azzardo" deriva dall'arabo *az-zabr*, e in arabo *zabr* è il dado). Secondo lo storico romano Svetonio l'imperatore Claudio (10 a.C.-54 d.C.) scrisse un trattato sui giochi d'azzardo, andato perduto. Anche il celebre poeta Ovidio (43 a.C.-18 d.C.) riferisce dell'esistenza di scritti sulle proprietà dei pezzi dei giochi d'azzardo, aggiungendo però che "valevano tanto quanto i pezzi stessi", e cioè poco o niente. Si trattava probabilmente di scritti simili alle moderne cabale del lotto.

⁴⁹ Carneade di Cirene (214 a.C. – 120 a.C.), filosofo greco. Fu esponente dello scetticismo.

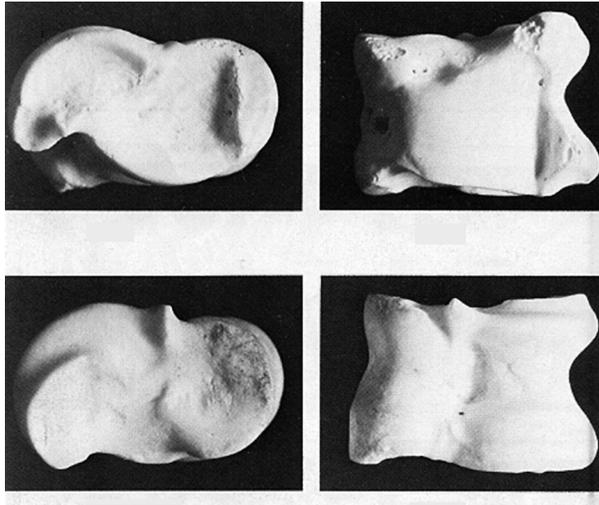


Figura 1. Le facce di un astragalo.

Ma quali erano i giochi più comuni all'epoca? Due sono noti ancora oggi: i dadi e "testa o croce", mentre il terzo non è più diffuso: l'astragalo. L'astragalo è un osso della zampa della pecora, più precisamente del tarso, che ha la proprietà di avere quattro facce molto diverse tra loro, e quindi ben riconoscibili; lanciando un astragalo come un dado, compariva sempre una delle quattro facce (Figura 1), però non tutte si presentano in modo equiprobabile: la statistica mostra che due delle facce si presentano con probabilità circa doppia delle altre due. Pare che greci e romani non arrivarono mai a questa constatazione, a meno che siano andati perduti gli scritti che ne trattano. Si sa invece che il gioco più diffuso consisteva nel lanciare quattro astragali e osservare le combinazioni uscite, e pare che si scommettesse su quelle che avevano quattro facce tutte diverse.

Nel Medioevo, a quanto si sa, i giochi d'azzardo più diffusi erano quelli con i dadi. Uno dei più semplici e diffusi era la "zara" (di chiara derivazione araba come dice il nome), che consisteva nel lanciare tre dadi e scommettere, all'atto del lancio, sulla somma dei risultati⁵⁰. Interessante il fatto che alcuni risultati erano esclusi dal gioco, come

⁵⁰ Anche Dante conosceva la zara, che cita nel VI canto del Purgatorio: "Quando si parte il gioco de la zara, / colui che perde si riman dolente, / ripetendo le volte e tristo impara, / con l'altro se ne va tutta la gente...".

il 3 e il 18 con tre dadi, perché ritenuti troppo poco “probabili” e fuorvianti. Non si poteva così puntare sul 3, e se usciva il 3, che era la “zara”, il lancio era considerato nullo; stessa sorte accadeva talora alla somma 4 e alla somma 17, pure ritenute troppo poco frequenti. Curiosamente i commentatori dell’epoca motivavano questa esclusione dicendo che le somme rispettivamente 4 e 17 “si possono ottenere in un sol modo”, il che è falso perché i modi possibili sono tre: uno per ogni dado che non dà 1 o, rispettivamente, 6. Sempre a proposito della zara, ci è rimasto un solo trattato del ’200 nel quale si intravede quello che potremmo oggi chiamare la statistica del gioco: esso contiene una tabella nella quale sono indicate tutte le “forze” e le “debolezze” delle varie somme, intese come il numero di combinazioni che danno la somma voluta (le “forze”) e, per differenza quelle che non la danno (le “debolezze”). Pare però che i giocatori non si curassero molto del fatto che alcune combinazioni, come il 10 e l’11, fossero più probabili di altre, né assolutamente sembra che le somme meno frequenti fossero ricompensate meglio: evidentemente il lancio dei tre dadi era ritenuto così complesso da non poter dare origine a differenze nella somma. I nostri giocatori d’azzardo, poi, avevano escogitato interessanti variazioni della zara: siccome un lancio dura molto poco, era frequente scommettere di vincere dopo un certo numero di lanci, per rendere il gioco più attraente.

Capitava però (e possiamo anche immaginare perché, visto che in certe circostanze il gioco d’azzardo era vietato per motivi di ordine pubblico) di dover interrompere il gioco prima del termine pattuito. E qui si poneva il problema di come ripartire la posta se, per esempio, su un certo numero di mani svolte r erano andate al primo giocatore e s al secondo (ovviamente il più delle volte erano andate a vuoto). L’esistenza di questa domanda fa capire che la soluzione proporzionale (cioè dividere la posta in parti proporzionali a r e a s) non fosse soddisfacente, il che significa, perlomeno per coloro che si sono posti questo problema, che le varie uscite non erano percepite come equivalenti (cioè, vincere una mano con un 5 non era come vincerla con un 10, perché il 5 era più raro).

Ma c’è di più: anche famosi matematici dell’epoca, come Luca Pacioli⁵¹, rispondevano semplicemente di dividere la posta in parti proporzionali, e ciò era sbagliato per un altro motivo. Supponiamo pu-

⁵¹ Luca Pacioli (Borgo San Sepolcro, 1445 – Roma, 1514), matematico italiano. Autore della *Summa* (1494), compendio del sapere matematico del suo tempo e del trattato di geometria *De divina proportione* (1509).

re che le vincite alla zara siano equiprobabili (ossia che si scommetta sempre sul 10, cosa che avrebbe creato non pochi turbamenti), che si pattuiscono 6 mani, con una posta di 80 scudi, e che, arrivati al 5 a 3 per un giocatore, si debba sospendere il gioco. Pacioli avrebbe detto di dare 50 scudi al primo giocatore e 30 al secondo. Invece la teoria corretta dimostra, come vedremo più avanti, che la probabilità che due o più eventi indipendenti si verifichino è uguale al prodotto delle probabilità degli eventi. Siccome le mani sono supposte equiprobabili, il secondo giocatore ha probabilità $1/2$ di vincere ciascuna mano, ed essendo indipendenti ha probabilità $(1/2) \cdot (1/2) \cdot (1/2) = 1/8$ di vincerle tutte e tre, che è quello che deve fare per vincere, potendo solo vincere per 6 a 5. Dunque la probabilità che il primo vinca è $7/8$, ossia 7 volte quella del secondo. Secondo le probabilità, dunque, il primo giocatore avrebbe dovuto ricevere il settoplo del secondo, cioè 70 scudi contro 10, che è ben diverso dal risultato di Pacioli⁵²! La nostra storia si conclude con Girolamo Cardano⁵³, che scrisse un libro intitolato *De ludo aleae* (Sul gioco dei dadi) e parla abbastanza esplicitamente di “casi favorevoli” e “casi possibili”. Siamo alle porte delle idee di Blaise Pascal⁵⁴, il primo matematico a dare una veste allo stesso tempo teorica e operativa al concetto di probabilità, che verrà denominata “probabilità classica”.

2. La probabilità classica

Torniamo al dado. Se il lancio (e il dado, soprattutto) non è “disonesto”, non si vede perché alcuni risultati debbano essere preferibili ad altri (per inciso, il dado non è simmetrico, perché i disegni incisi sulle facce sono diversi, e quindi il baricentro non è esattamente nel centro del cubo, ma questo pare non importare ai giocatori e ai probabilisti). Ammettiamo allora che questi risultati siano “ugualmente possibili” o, come diremo in seguito, equiprobabili. La nascita della probabilità classica si deve, oltre a Blaise Pascal, a Pier-

⁵² In ogni caso, di errori celebri sulla probabilità la matematica abbonda: Leibniz affermò che è ugualmente probabile fare 12 e 11 con due dadi, mentre d'Alembert sostenne che gettando due volte una moneta, la probabilità di avere almeno una croce è $2/3$ invece di $3/4$, che è il risultato corretto.

⁵³ Girolamo Cardano (Pavia, 1501 – Roma, 1576), matematico, medico e astrologo italiano. Famosa la sua disputa con Tartaglia, riguardante la risoluzione delle equazioni di terzo grado.

⁵⁴ Blaise Pascal (Clermont-Ferrand, 1623 – Parigi, 1662), matematico, fisico, filosofo e teologo francese. In ambito scientifico, ha fornito contributi importanti alla teoria dei fluidi ed è uno dei fondatori, assieme a Fermat, della probabilità classica.

re di Fermat⁵⁵. In un celebre scambio di lettere i due concordano nel dare questa definizione di probabilità di un certo risultato (sempre in relazione ai giochi d'azzardo):

“La probabilità [classica] è il rapporto fra il numero di casi favorevoli al risultato diviso per il numero dei casi totali, purché questi siano ugualmente possibili.”

Iniziamo così a vedere da un lato il germe di una definizione operativa (il modo di calcolare la probabilità) e dall'altro lato il tarlo delle definizioni di probabilità, ossia la circolarità (la probabilità che viene definita in termini di se stessa, visto che si deve far ricorso alla locuzione “ugualmente possibili”). Certo, la simmetria di un dato problema, come quello del dado, può suggerire che i vari risultati siano equiprobabili, ma non dobbiamo scordare che si tratta di una situazione ideale: con dei dadi reali la situazione potrebbe essere diversa, e la definizione classica rivelarsi non applicabile.

Nonostante queste debolezze, la definizione classica di probabilità si prestò a numerose elaborazioni non banali, tra cui quella celebre del cavaliere de Méré. Questo personaggio, amico, pare, di Pascal, si lamentava di perdere in un gioco d'azzardo, mentre era convinto che avrebbe dovuto vincere. Il problema era semplice: se si lancia un dado un certo numero di volte, la probabilità che appaia almeno una volta un 6 tende a salire, dunque dopo un certo numero di lanci supera il 50%. De Méré sapeva quanti erano questi lanci per un 6 con un dado, ma non sapeva quanti erano per il 12 con due dadi⁵⁶.

Vediamo di risolvere dapprima il problema, più semplice, con il 6 e un dado. Quanti sono i casi possibili su n lanci? Siccome a ogni lancio può uscire una qualunque delle 6 facce del dado, si avranno 6 possibilità dopo il primo lancio, altre 6 per ognuna delle 6 precedenti, ossia $6 \cdot 6 = 36$, poi $6 \cdot 6 \cdot 6 = 6^3 = 216$ per 3 lanci, e quindi 6^n dopo n lanci. Queste sequenze sono tutte ugualmente possibili, visto che il dado è supposto simmetrico (e non truccato). Vediamo ora quanti sono i casi favorevoli: non sono facili da calcolare, ma è agevole calcolare i casi sfavorevoli, dati da quelle sequenze di n uscite nelle quali non esce mai il 6, e quindi sono 5^n , perché ap-

⁵⁵ Pierre de Fermat (Beaumont-de-Lomagne, 1601 – Castres, 1665), matematico francese. Ha anticipato alcune idee del successivo calcolo infinitesimale ed è uno dei fondatori della geometria analitica, assieme a Cartesio, e della probabilità classica, assieme a Pascal.

⁵⁶ De Méré scommetteva di realizzare un 12 con due dadi entro un certo numero di lanci.

punto il 6 è “bandito” dai casi sfavorevoli. Ma allora i casi favorevoli sono la differenza $6^n - 5^n$, e la probabilità di avere almeno un 6 dopo n lanci è

$$p_n = \frac{6^n - 5^n}{6^n} = 1 - \left(\frac{5}{6}\right)^n$$

Come si vede, questo numero aumenta all'aumentare di n , e un semplice conto mostra che per $n = 3$ viene 0,42 e per $n = 4$ viene 0,52. Quindi con 4 lanci è più probabile che esca almeno un 6 invece che non appaia. Questo, l'abbiamo detto, de Méré lo sapeva, ma, siccome far 12 con due dadi è 6 volte più difficile che far 6 con un dado, pensava che per aver più probabilmente un 12 con due dadi che il contrario fossero necessari $6 \cdot 4 = 24$ lanci. Così scommetteva e alla lunga perdeva. Il calcolo che abbiamo svolto, invece, si estende facilmente al caso di de Méré, e spiega il motivo delle sue perdite. Che siano 6 o 36 le combinazioni diverse del dado (infatti se i dadi sono due le combinazioni diverse in ogni lancio sono 36), quello che conta e resta è l'equiprobabilità: quindi stavolta ci saranno 36^n sequenze totali e $36^n - 35^n$ favorevoli, e il calcolo di Pascal e Fermat fornisce

$$P_n = \frac{36^n - 35^n}{36^n} = 1 - \left(\frac{35}{36}\right)^n$$

Un semplice calcolo aritmetico mostra adesso che $P_{24} = 0,491$ e che $P_{25} = 0,505$. Quindi la risposta alla domanda di de Méré non è 24 ma 25, e si capisce perché alla lunga egli perdesse.

Come vediamo, la definizione classica di probabilità permette di risolvere problemi non banalissimi.

Siamo in ogni caso ben lontani da una vera e propria matematizzazione della probabilità: per esempio, veniva tacitamente accettato il fatto che se per due risultati diversi si hanno probabilità diverse, quello con probabilità maggiore è considerato più “plausibile” di quello con probabilità minore, il che è vero, ma andrebbe dimostrato; in quanto è evidente se il numero di casi totali è uguale nei due risultati, ma meno se i casi totali sono diversi. Problemi simili, come quelli che abbiamo citato sopra a proposito di Luca Pacioli, furono

trattati da Christiaan Huygens⁵⁷ nel suo trattato *De ratiociniis in ludo aleae* (Sulla ragione nel gioco del dado), sempre in relazione alla definizione classica di probabilità e, ovviamente, al gioco d'azzardo.

3. Bernoulli e la probabilità frequentista

È unanimemente riconosciuto che il fondatore della teoria matematica della probabilità, e quindi del primo tentativo di “matematizzare il caso”, fu Jakob Bernoulli⁵⁸ nel suo trattato *Ars coniectandi* (L'arte della congettura). In quest'opera si trova una differente e (apparentemente) più flessibile definizione di probabilità, la definizione frequentista:

“La probabilità di un risultato è il limite del rapporto fra il numero di occorrenze del risultato e il numero delle ripetizioni dell'esperimento, purché queste siano infinitamente ripetibili e indipendenti.”

Cosa vuol dire, innanzitutto, il “limite”? Senza entrare in dettagli troppo tecnici, l'idea è quella di considerare la cosiddetta frequenza relativa di un risultato, che non è nient'altro che il rapporto $f_{R,N}$ tra il numero N di esperimenti (lanci del dado) effettuati e il numero di volte N_R nel quale il risultato è comparso, nell'ipotesi di indipendenza che discuteremo dopo. In simboli,

$$f_{R,N} = \frac{N_R}{N}$$

Abbiamo indicato la frequenza con $f_{R,N}$ per sottolineare che essa dipende dal risultato osservato ma anche dal numero di ripetizioni dell'esperimento. Se lanciamo un dado 10, 100 e 1000 volte, osserviamo che la frequenza relativa non è in generale la stessa, e, sempre in generale, il suo valore tende a stabilizzarsi verso un dato numero: che secondo Bernoulli è il “valore limite”, cioè la probabilità, il numero che si raggiungerebbe se si potessero fare infinite ripetizioni (indipendenti) dell'esperimento. Questa definizione è più flessibile di quella classica (e la implica, come vedremo), perché permette di

⁵⁷ Christiaan Huygens (L'Aia, 1629 – L'Aia, 1695), matematico, astronomo e fisico olandese. Importanti i suoi studi sulla natura ondulatoria della luce.

⁵⁸ Jakob Bernoulli (Basilea, 1654 – Basilea, 1705), matematico e scienziato svizzero. È il fondatore del moderno calcolo delle probabilità e del calcolo delle variazioni.

calcolare la probabilità di risultati in situazioni non così simmetriche come quella del dado.

Certo, a un purista fa accapponare la pelle per vari motivi. Primo, si deve postulare l'esistenza del limite, ossia che in nessuna situazione reale (se si può dar senso a questa parola) la frequenza relativa oscilla in modo imprevedibile senza attestarsi su alcun valore, tuttavia, dato che si hanno a disposizione comunque un numero finito di ripetizioni, questo fatto non fa poi così male (si chiama *legge empirica dei grandi numeri*).

Fa più male il fatto che non si possa avere un calcolo esatto della probabilità di risultati complessi, proprio a causa del fatto che le ripetizioni sono in numero finito, né la frequenza potrebbe dar suggerimento di quale sia il limite. Per esempio, se si calcola la frequenza del risultato 6 con un dado, la frequenza relativa dopo 1000 lanci potrebbe dare 0,1665 e "suggerire" $1/6$ che è 1,6666... Ma se la probabilità fosse $\sqrt{7}/8$, è possibile scoprirlo con altrettanta sicurezza? E se fosse un numero irrazionale molto vicino a $\sqrt{7}/8$? Ma quello che fa più male di tutto è la richiesta di indipendenza. È semplice da richiedere ma pesa molto: chi ci garantisce che dopo 100.000 lanci di un dado, esso non si sia completamente smussato ai vertici e risponda in modo diverso che all'inizio? Potremmo sostituire al dado vecchio uno nuovo, ma dovremmo ammettere che è uguale al precedente, oppure lanciare 100 volte 100 dadi, ma l'esperimento è diverso, a meno che non postuliamo anche che i dadi siano uguali.

Oltretutto l'indipendenza esclude, in linea di principio, situazioni evidentemente non ripetibili. Qual è la probabilità che un ponte crolli? Qual è la probabilità che una ditta fallisca? Qual è la probabilità che un paziente sopravviva a un'operazione? Sono domande non banali, alle quali la definizione frequentista di probabilità offre il fianco.

Ma allora perché ha avuto così successo? A mio avviso, perché con il suo aiuto si sono potute dimostrare certe relazioni tra le probabilità di risultati, e quindi a cominciare a far matematica, quella vera.

Vediamo, per esempio, una delle proprietà più importanti della probabilità: l'additività. Essa consiste semplicemente in questo: se due risultati R_1 e R_2 si escludono (ossia se non risultano mai contemporaneamente, come le usuali somme nei dadi), allora la probabilità che venga l'uno oppure l'altro risultato è la somma delle probabilità, che potremmo scrivere in formula così:

$$p(R_1 \text{ oppure } R_2) = p(R_1) + p(R_2) \quad (1)$$

La dimostrazione di questo fatto è semplice: siccome i due risultati si escludono, il numero totale di volte in cui viene uno dei due risultati, che indichiamo con N_{R_1} oppure N_{R_2} , è dato dalla somma delle volte nelle quali è uscito R_1 con le volte nelle quali è uscito R_2 , cioè

$$N_{R_1 \text{ oppure } R_2} = N_{R_1} + N_{R_2}$$

Se adesso dividiamo per il numero totale di ripetizioni N troviamo

$$f_{R_1 \text{ oppure } R_2, N} = \frac{N_{R_1 \text{ oppure } R_2}}{N} = \frac{N_{R_1}}{N} + \frac{N_{R_2}}{N} = f_{R_1, N} + f_{R_2, N}$$

A questo punto, se al crescere di N , i tre numeri tendono alle rispettive probabilità, come si suppone per la legge empirica dei grandi numeri, non ci vuol molto a convincersi che questi valori verificheranno proprio la (1)⁶⁰. Anche la formula del “complementare” si dimostra analogamente. Se N_R è il numero di volte nelle quali è uscito il risultato R , allora $N - N_R$ è il numero di volte nelle quali non è uscito R . Quindi la sua frequenza relativa è

$$f_{\text{non è uscito } R, N} = \frac{N - N_R}{N} = 1 - \frac{N_R}{N}$$

e quindi, quando N cresce indefinitamente troviamo

$$p(\text{Non esce } R) = 1 - p(R)$$

Quindi, se sappiamo che la probabilità di fare 2 con due dadi è $1/36$ e quella di fare 3 con due dadi è $2/36$, allora la probabilità di fare

⁵⁹ Questa formula non è sempre vera se i risultati non si escludono. Pensate ai fatti “oggi nevica” e “oggi piove” e che su 10 giorni, 3 abbia piovuto, 2 abbia nevicato e in uno abbia prima piovuto e poi nevicato (nei restanti non ha né piovuto né nevicato). Se si contano i giorni nei quali “ha nevicato oppure ha piovuto” viene 6, ma i giorni in cui ha piovuto sono 4 e quelli in cui ha nevicato 3, per cui la formula non vale.

⁶⁰ In realtà, se si dà una definizione precisa di “tendere a”, che in matematica si chiama limite, allora il fatto che il limite della somma sia uguale alla somma dei limiti diviene un teorema e si può quindi dimostrare. Noi non entriamo nei dettagli per non appesantire il discorso.

“maggiore o uguale a 4” è $33/36$. Perché? Perché non fare “maggiore o uguale a 4” significa “meno di 4”, e “meno di 4” significa 2 oppure 3. Ma il 2 e il 3 chiaramente si escludono, per cui la probabilità di fare 2 oppure 3 è $3/36$, e quindi quella di non fare 2 oppure 3 è il complementare, cioè $1 - 3/36 = 33/36$.

Con la teoria di Bernoulli è quindi possibile per la prima volta parlare compiutamente di calcolo delle probabilità e la sua teoria, in aggiunta, ricomprendeva la vecchia definizione di probabilità classica. Vediamo perché. Supponiamo che i vari risultati possibili siano 6^n (non N) e siano tutti equiprobabili. Siccome questi risultati si escludono vicendevolmente, la probabilità che esca almeno uno di essi sarà la somma delle probabilità dei singoli risultati:

$$p(\text{Almeno uno degli } R_k) = p(R_1) + p(R_2) + \dots + p(R_n)$$

Ma la probabilità a primo membro della formula appena scritta è uno, perché un risultato esce sempre: se poi tutte le probabilità a destra sono uguali, ciascuna di esse deve valere $1/n$. Dunque

$$p(R_k) = \frac{1}{n}$$

Adesso prendiamo una combinazione che ci interessa (sarebbe un “risultato composto”, ma su questo torneremo fra un attimo) e supponiamo che si verifichi quando si verificano m dei possibili n risultati. Per esempio, questa combinazione potrebbe essere “fare 6 con due dadi”: essa avviene nelle forme $1 + 5$, $2 + 4$, $3 + 3$, $4 + 2$, $5 + 1$ distinguendo primo e secondo dado, dunque in questo caso $n = 36$ (le 36 uscite dei due dadi) e $m = 5$ (queste cinque possibilità).

Chiaramente le forme $1+5$ ecc. si escludono, per cui la probabilità di fare 6 è la *somma* delle 5 probabilità che la “generano”. Siccome esse sono tutte uguali (e quindi pari a $1/36$), la probabilità cercata sarà $5/36$, cioè proprio il rapporto fra i 5 casi favorevoli elencati e i 36 casi possibili. Nel caso generale, evidentemente, lo stesso ragionamento porta a m/n , che è la formula di Pascal e Fermat.

Con la sua teoria Bernoulli fu in grado di calcolare diverse probabilità, alcune sorprendenti. Per esempio, riuscì a stabilire che con più di 23 persone è più probabile che due di esse festeggino il loro compleanno nello stesso giorno che il contrario: in genere si pensa che

ne servano molte di più⁶¹. Ma il merito dell'impostazione bernoulliana è anche un altro: quello di puntare i riflettori su un nuovo soggetto della probabilità, che è l'evento.

4. Eventi e loro operazioni

I matematici si resero ben presto conto che occorreva a questo punto fare qualche distinzione, un po' pignola se vogliamo, ma utile per capire il prosieguo del discorso. Supponiamo di lanciare ancora una volta il nostro dado e che esca 6. Il 6 è un "risultato", abbiamo detto finora; ma se ci chiediamo quanto valga la probabilità che esca "il 6 oppure il 5", ci accorgiamo che questo non suona come un risultato, perché una volta lanciato il dado esce 6, oppure 5, oppure qualcos'altro: "esce il 6 oppure esce il 5" è quello che si definisce un evento.

Non è una scelta felice: nel linguaggio comune, "evento" potrebbe essere anche qualcosa che accade e basta e si potrebbe confondere con il risultato, ma tant'è, neanche in matematica tutte le ciambelle escono con il buco, e questo termine è ormai entrato nell'uso comune. Con riferimento al dado, chiamiamo allora 1, 2, 3, 4, 5, 6 i risultati, mentre locuzioni come "esce meno di 3", "esce un numero dispari" ecc. denotano degli eventi. C'è ovviamente un legame fra i due concetti, ed è questo: per fare un evento occorrono dei risultati. Per esempio, "esce un numero dispari" è costituito da 1, 3, 5, e così via. Ci rendiamo subito conto che (quasi) tutto ciò che possiamo formulare a parole è costituito da risultati, e a volte eventi diversi a parole sono costituiti dagli stessi risultati.

Per fare un esempio astruso, dei sei simboli del dado, 1, 3 e 5 hanno il punto nel centro della faccia, mentre gli altri no. Quindi "esce un numero dispari" ed "esce un simbolo con il punto al centro della

⁶¹ Il trucco sta nel fatto che non si sa quale sia quel giorno. Infatti la probabilità che su n persone ognuno festeggi per sé è facile da calcolare in modo classico pensando di "scegliere" il giorno di compleanno uno alla volta in modo da sovrapporsi: il primo ha 365 giorni liberi, il secondo solo 364, il terzo 363, e così via. Le sequenze totali di compleanni sono invece 365^n . Dunque la probabilità di avere compleanni tutti diversi è

$$\frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdot \frac{362}{365} \cdot \dots$$

e il suo complementare a uno, cioè

$$1 - \frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdot \frac{362}{365} \cdot \dots$$

scende sotto 1/2 per $n = 23$.

faccia” sono costituiti dagli stessi risultati. Ci rendiamo allora conto che conviene di più individuare gli eventi semplicemente con i risultati che li compongono: così abbiamo l’evento $\{1, 2, 3\}$, l’evento $\{4, 5\}$ e così via: usiamo le parentesi grafie per elencare i risultati che compongono l’evento.

Già, e che dire di $\{6\}$? È un evento perché ha le graffe, e la sua interpretazione è semplice “esce 6”. Questo tipo di eventi, che si chiamano elementari, sono insidiosi perché si confondono con i risultati. Benché dicano la stessa cosa 6 e $\{6\}$ sono diversi: il primo è un risultato, il secondo un evento.

C’è ancora un’insidia: che dire di questo evento, che solo un matematico può concepire: “esce un numero composto dispari”, con un dado, beninteso? Se ci pensiamo un attimo, è impossibile: il più piccolo numero composto dispari è 9, e non c’è sul dado. Quindi l’evento non è costituito da nessun risultato. Ci chiediamo se abbia senso una cosa del genere, potremmo decidere di no, ma ci resta un dubbio: se definiamo gli eventi con frasi del linguaggio comune, chi ci garantisce che siano possibili? Lasciamo per un momento la questione in sospeso.

Il motivo fondamentale per l’introduzione degli eventi è stato però un altro. Se essi sono individuati da frasi del linguaggio (che poi individuano risultati), è comune pensare di connetterli con “e”, “oppure”, o negarli: insomma, comporli come si fa con delle operazioni. Per esempio, nella roulette si punta spesso su “rosso dispari”, che significa “esce un numero rosso” ed “esce un numero dispari”; nei dadi, come abbiamo già visto, “esce il 6 oppure esce il 5”. Appare evidente che le operazioni con gli eventi sono associate a operazioni logiche, di cui le frasi che individuano gli eventi sono proposizioni.

A questo punto il quadro è completo e fornisce una ben precisa risposta: ma allora gli eventi sono insiemi! Con questa semplice idea tutto quanto abbiamo detto finora si incasella: le operazioni con gli eventi si trasferiscono sulle operazioni con gli insiemi di risultati che gli eventi stessi individuano. Per esempio, l’evento “ A oppure B ” diventerà l’unione $A \cup B$, perché i risultati che danno l’evento A oppure l’evento B saranno quelli che danno A unitamente a quelli che danno B ; l’evento “ A e B ” diventerà l’intersezione $A \cap B$, perché solo i risultati che producono entrambi gli eventi saranno quelli di interesse.

E infine, siccome tra gli insiemi c’è pure, a torto o a ragione, l’insieme vuoto, possiamo convenire che l’insieme vuoto corrisponde a

quegli eventi assurdi del tipo “esce un numero dispari composto” dal lancio di un solo dado. Tra l’altro, l’insieme vuoto ha anche un altro vantaggio: può infatti capitare che l’intersezione di due insiemi sia vuota, e quindi se l’insieme vuoto non fosse un evento, certe intersezioni di insiemi non si potrebbero fare, il che complicherebbe la situazione.

In questo modo, invece, le operazioni di unione e intersezione di eventi danno sempre per risultato un evento. Oltretutto, ciò non guasta neanche la definizione frequentista di probabilità, perché essa chiede solo di contare le volte in cui avviene un dato evento. Se esso non è costituito da risultati, è chiaro che non avverrà mai, e quindi il suo N_R sarà sempre 0⁶².

Quando la probabilità si rivolge al gioco d’azzardo, il discorso fin qui non fa una piega. Però, una volta chiarite le basi del concetto di probabilità, non ci volle molto a capire che essa poteva essere di grande aiuto anche al di fuori del gioco. Un fatto di cui ci si rese conto presto è che anche il concetto di misurazione poteva esser fatto rientrare nell’ambito della probabilità. Se, per esempio, facciamo misurare la lunghezza di un pezzo a differenti persone o con differenti strumenti di misura, può capitare che si ottengano risultati diversi; non radicalmente diversi, ma diversi. Le cause possono essere le più disparate: tarature dello strumento di misura, distrazione dello sperimentatore, oppure ancora piccole variazioni del pezzo stesso.

Lo stesso discorso vale, per diversi pezzi prodotti da una stessa macchina: in teoria essi dovrebbero essere tutti uguali, ma in pratica non è così. Perciò, se si deve produrre un pezzo con determinate caratteristiche, bisogna valutare la probabilità che esso non sia conforme a quanto richiesto e aggiustare le cose in modo che questa sia ragionevolmente bassa⁶³.

A prima vista, calcolare la probabilità di fare 7 con due dadi non pare molto diverso dal calcolare la probabilità che la misura della lunghezza di un pezzo meccanico sia superiore, per esempio, a 7 mm. Eppure una differenza sostanziale c’è: non ci aspettiamo certo di fare 7,5 con due dadi, ma è ben possibile che la misura del pezzo sia

⁶² Ben diverso è il problema del viceversa: può ben succedere che un evento, secondo la definizione frequentista, abbia probabilità zero e il suo N_R non sia 0! Questo è un fatto delicato sul quale non ci soffermiamo.

⁶³ Siccome spesso ridurre la probabilità che un pezzo non sia “a norma” ha dei costi, bisogna anche che non sia inutilmente bassa, per non sprecare risorse.

7,5 mm, o 7,13 mm. Siamo in altre parole abituati a associare al processo di misurazione un numero reale⁶⁴.

Questo fatto ha delle conseguenze, prima fra tutte il fatto che esistono infiniti eventi possibili. Infatti, se indichiamo con X la misura del pezzo, “ X misura più di 7 mm” è un evento, ma anche “ X misura più di 7,1 mm” è un evento, e così pure “ X misura più di 7,05 mm”, e così via. Ma allora sorge un dubbio: siccome, nel caso del gioco d’azzardo, tutte le collezioni di risultati sono eventi, la stessa cosa sarà vera anche nelle misurazioni?

I matematici sanno bene che quando c’è di mezzo l’infinito, bisogna stare attenti. E, in effetti, i problemi sorgono subito. Se si ammette che troppi insiemi di risultati sono eventi, si rischia di ottenere risultati non coerenti con quanto ci aspettiamo dalla probabilità (per esempio, l’additività che abbiamo visto poco fa), e dunque certi insiemi di risultati non possono essere considerati eventi.

Detta così, sembra una sconfitta, almeno dal punto di vista della semplicità, ma di fatto le cose stanno in maniera diversa: se potessimo effettuare un’analisi approfondita, ci potremmo rendere conto che i “non-eventi” che abbiamo escluso sono situazioni assolutamente non verificabili nella pratica, ma puri e semplici “mostri” matematici legati alla definizione di numero reale. In questo contesto le operazioni fra eventi che abbiamo visto sopra acquisiscono maggiore importanza, perché si decide di chiamare eventi degli insiemi di risultati per i quali le operazioni di unione, intersezione e passaggio al complementare siano sempre definite. In altre parole, facendo l’intersezione o l’unione di due eventi, o il complementare di un evento, non risulterà mai un non-evento”⁶⁵.

5. Dalla probabilità alla statistica: le distribuzioni

Spostandoci sulla teoria della misurazione, un problema con grande ricaduta pratica riguarda non solo le misure dei pezzi meccanici, ma le misure legate alle persone. Esso si pose per la pri-

⁶⁴ Nella pratica, una misura dà per risultato sempre un numero razionale. però, come è noto dalla geometria, certe costruzioni geometriche sono ammissibili (per esempio, intersecare una retta e una circonferenza) solo se si affiancano ai numeri razionali anche quelli irrazionali. È un discorso lungo sul quale non possiamo essere esaurienti, ma il risultato è che i numeri reali, anche se più complicati e difficili da capire, sono più efficienti e comodi da usare per rappresentare il risultato di un processo di misura.

⁶⁵ Per lo meno un numero finito di unioni o intersezioni: se si vogliono unire o intersecare infiniti eventi, la cosa si complica, ma ha una via d’uscita, che però ci porterebbe troppo lontano dal nostro discorso.

ma volta nella storia della scienza in relazione al servizio militare, ma in un periodo storico, la rivoluzione industriale, nel quale questo tipo di discorsi sarebbe venuto all'ordine del giorno.

Oggi il servizio di leva non c'è più, ma molte persone che l'hanno effettuato, soprattutto se di statura alta, ricordano che non avranno trovato la divisa della loro taglia. Del resto è chiaro: non esistono molte persone più alte di 2 m, quindi perché ogni caserma deve avere divise di quella taglia? D'altro canto è chiaro che se nessuna caserma ha divise così grandi, quei pochi individui alti non troveranno mai la loro uniforme. Il problema è quindi: preso un soldato, qual è la probabilità che esso sia più alto di 2 m? È, chiaramente, la stessa domanda si pone sulla statura fra 190 e 200 cm, fra 180 e 190, e così via.

La definizione frequentista di probabilità offre subito una risposta a questo problema: misurare la statura di un numero abbastanza alto di persone e calcolare la frazione di quelle che sono più alte di 2 metri, ossia la frequenza relativa dell'evento "essere più alto di 2 metri". Siccome questa frequenza tende alla probabilità per un numero grande di ripetizioni, ossia in questo caso di misure delle persone, essa può essere presa come un valore attendibile della probabilità. Il processo che abbiamo definito si chiama statistica, e dal '700 a oggi si è sviluppato fino a diventare una disciplina scientifica con numerosissime applicazioni. Cerchiamo di analizzare un po' più a fondo la situazione.

La definizione frequentista di probabilità contiene una grande limitazione, perché parla di ripetizioni indipendenti dell'evento. Siccome per ordinare le uniformi in caserma ci interessa una stima della probabilità e non il suo valore esatto, non è grave il fatto di limitarsi a un numero finito di ripetizioni dell'evento (in questo caso, di misurazioni), ma il fatto che le stature delle persone siano indipendenti non è scontato. Per esempio, in Italia i friulani sono mediamente più alti dei calabresi. Se facciamo una statistica con troppi friulani, il valore della frequenza di persone più alte di 2 m sarà superiore a quello reale dell'intera popolazione italiana, e se lo adottiamo per una caserma calabrese rischiamo di avere troppe uniformi grandi. D'altro canto, misurare la statura di tutte le persone è dispendioso e a volte non praticabile (pensiamo a un sondaggio elettorale, nel quale bisogna essere veloci nel dare i risultati, quindi serve un campione limitato ma rappresentativo), ed ecco spiegata una ragione di certi flop della statistica (specie in campo elettorale, appunto). Ma non ci vuole molto per commettere errori di questo genere: se operiamo

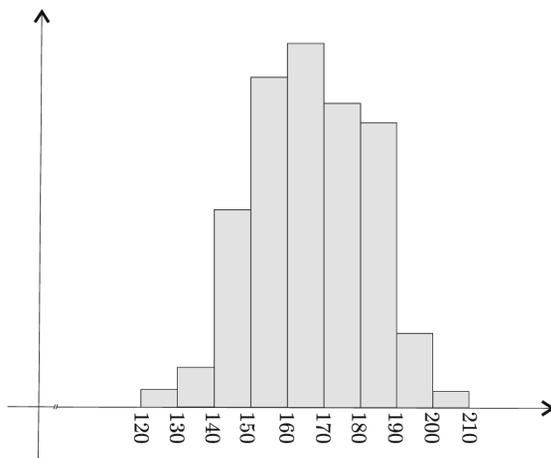


Figura 2. Un possibile istogramma della statura.

la rilevazione dei consumi di un certo prodotto all'uscita di un supermercato, otterremo risultati che trascurano le fasce più giovani (di solito gli studenti non vanno al supermercato), e quindi non c'è una vera indipendenza.

Torniamo alla statura. L'idea più immediata per avere un quadro della situazione è la seguente: dividere le stature possibili in intervalli e misurare la frequenza relativa dei vari intervalli nel campione preso in esame, ossia la percentuale di persone con statura compresa nei vari intervalli. Si ottiene così un istogramma, relativo al campione utilizzato. Se poi il campione aumenta, queste percentuali tendono sempre più al valore della probabilità di avere una statura nell'intervallo esaminato, almeno in ipotesi di indipendenza. In generale il diagramma che si ottiene è del tipo indicato in Figura 2.

Naturalmente la scelta di dividere le stature di 10 cm in 10 cm è arbitraria e dettata da motivi pratici: se abbiamo in mente le uniformi, è chiaro che 175 e 179 cm non costituiranno una grave differenza. Ma se dobbiamo o vogliamo essere più precisi, allora dovremmo dividere l'intervallo delle misure possibili⁶⁶ in intervallini più piccoli,

⁶⁶ Anche l'intervallo delle misure possibili è arbitrario: in teoria sarebbe un numero reale qualunque. Ma siccome non ci aspettiamo soldati alti meno di 20 cm (o addirittura con statura negativa!) o alti più di 3 m, possiamo scegliere entro questi limiti.



Figura 3. Un possibile istogramma della statura più raffinato del precedente.

per esempio di ampiezza 1 cm. Si capisce subito che, se non vogliamo rischiare di avere frequenze nulle dobbiamo aumentare la numerosità del campione per avere una stima significativa. Per esempio, con gli intervalli di 10 cm abbiamo misurato 10.000 persone sapendo che vi sono 30 intervallini, ma con le stesse persone e gli intervalli di 1 cm gli intervallini sono circa 300, quelli meno “popolati” rischiano di dare 0, anche se in realtà vi sono casi fra la popolazione reale⁶⁷. Se però siamo in grado di aumentare la numerosità del campione in maniera sufficiente, l’istogramma che otterremo assomiglierà a una curva più di quello precedente, e quindi al limite (pensando cioè di avere a disposizione infiniti campioni) si constata che l’istogramma “tende” a una curva, come in Figura 3.

Ma torniamo al primo istogramma e chiediamoci: qual è la probabilità che una statura sia compresa tra 170 e 190 cm? Se assumiamo che l’istogramma sia sufficientemente preciso da dare una stima della probabilità, è molto facile: dobbiamo sommare la frequenza relativa all’intervallo 170-180 cm e quella relativa all’intervallo 180-190 cm. Correttamente, perché gli eventi “statura compresa fra 170 e 180 cm”

⁶⁷ Quello della numerosità sufficiente per rappresentare i casi estremi (le cosiddette “code”) in maniera adeguata è un altro problema classico della statistica, che è spesso in conflitto con i limiti di costo e di tempo e anche etiche che le varie situazioni impongono. Se per stabilire che un certo farmaco è letale per lo 0,001% della popolazione servono vent’anni di test, è giusto non commercializzarlo per vent’anni, anche se sarebbe utile “quasi” per tutti? Su una popolazione come quella italiana questo significa 600 persone...

e “statura compresa fra 180 e 190 cm” si escludono, e quindi la probabilità dell’unione è la somma delle rispettive probabilità. In termini di grafico, dobbiamo calcolare le aree delle due barre relative (nell’ipotesi di avere indicato la frequenza in ordinata) e sommarle.

Se ci poniamo lo stesso problema con il secondo istogramma, quello “al centimetro”, le “barre” diventano 20, ma la sostanza non cambia: bisogna calcolare la loro area e sommare tutto. E se, al limite, avessimo una curva continua? Ci rendiamo conto subito che la risposta sarebbe data dall’area della regione compresa fra la curva, l’asse delle ascisse e le due rette verticali ($x = 170$ e $x = 190$ nel nostro esempio). A questo proposito la matematica ha avuto qui una simpatica sorpresa: già da secoli i matematici amavano cercare di studiare aree e volumi di figure curve, e in particolare Newton e Leibniz, nel ’600, avevano fornito gli strumenti per calcolare le aree anche in condizioni assai generali, attraverso un processo detto “integrazione”. In questo modo gli statistici ebbero la pappa pronta, per così dire.

Questo, assieme alla passione che hanno i matematici per “tirare al limite” le conseguenze, fece pendere la bilancia più verso le curve che verso gli istogrammi, almeno per quanto riguardava i calcoli teorici, e nacque il concetto di *distribuzione*. Se due quantità avevano la stessa curva di distribuzione, chiaramente avrebbero fornito sempre la stessa probabilità in relazione agli stessi intervalli. Se, per esempio, misuriamo il diametro maggiore del cranio, ne facciamo un grafico “al millimetro” e constatiamo che, raffinando sempre più gli intervalli, la curva che ne esce è la stessa, allora la probabilità di sapere che l’altezza stia fra 170 e 190 cm sarà uguale a quella che il diametro maggiore del cranio sia compreso fra 270 e 300 mm (ci può essere quindi un riscaldamento). Quindi studiare la distribuzione permetteva di studiare proprietà di numerose situazioni contemporaneamente.

Agli statistici del ’700 non sfuggì un fatto sorprendente: se si studia la distribuzione della statura, dei diametri del cranio, della lunghezza degli arti, e se se ne prende un istogramma abbastanza “fine”, in modo da intuire a quale curva si tenda al crescere della numerosità del campione, la curva che esce è sempre la stessa! A essere onesti, non proprio la stessa: due curve potevano differire per una traslazione e un riscaldamento delle ascisse (dal quale derivava un riscaldamento delle ordinate, perché l’area complessiva sotto la curva, che è la frequenza relativa complessiva, deve essere pari a 1, perché corri-

sponde al 100% dei casi esaminati). Fu il grande matematico Carl Gauss che intuì che la funzione “comune” a tutte queste misure sulle persone era la funzione data da $f(x) = \exp(-x^2)$. Per la precisione, la funzione “capostipite” era

$$N(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

che ha area totale sotto la curva uguale a 1. Questa distribuzione, di importanza fondamentale per tutta la statistica e la probabilità, si chiama *distribuzione normale* o *gaussiana*. Tra l'altro, Gauss ammise implicitamente che la probabilità che la statura fosse negativa era diversa da 0, perché questa curva non si annulla mai.

Chiaramente, si tratta di una modellizzazione: mai nessuna frequenza relativa darà risultati diversi da 0 per stature negative, ma ciononostante la gaussiana si usa perché è comoda, e gli eventuali errori che si commettono da questa assunzione sono compensati dalla complicazione di scegliersi una distribuzione più realistica ma meno semplice. I matematici seppero dare una risposta a questo apparente miracolo. Essi mostrarono che quando una certa quantità è il risultato di una sovrapposizione di numerosi effetti casuali indipendenti, la distribuzione del risultato finale è la gaussiana, anche se gli effetti generatori non sono gaussiani.

Non vogliamo entrare nel tecnico, ma ci interessa il significato. La statura è il risultato macroscopico di una serie di processi microscopici, ciascuno distribuito a suo modo, ma se essi sono indipendenti, allora la loro media, che fornirà la statura, sarà distribuito come una gaussiana. È questo il “teorema del limite centrale”, dimostrato da Aleksandr Ljapunov⁶⁸ nel XX secolo, ma intuito nel 1733 da Abraham de Moivre⁶⁹ e riscoperto più tardi da Laplace.

In ogni caso, non tutti gli esperimenti casuali hanno una distribuzione di tipo gaussiano. Per esempio, consideriamo l'esperimento di prendere a caso un elettrodomestico (anche un cellulare o un com-

⁶⁸ Aleksandr Michajlovič Ljapunov (Jaroslavl', 1857 – Odessa, 1918), matematico e fisico russo, noto soprattutto per i suoi risultati sulla stabilità dei sistemi meccanici e sul calcolo delle probabilità.

⁶⁹ Abraham de Moivre (Vitry-le-François, 1667 – Londra, 1754), matematico francese. È noto per i suoi contributi sui numeri complessi e sulla probabilità.

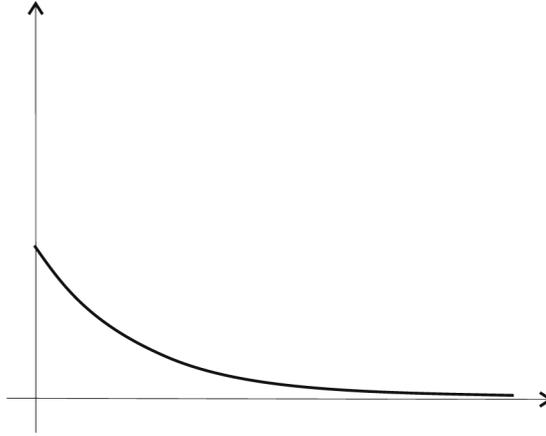


Figura 4. Il grafico della distribuzione esponenziale.

puter) da un campione e di misurare il momento in cui esso ha il primo guasto. Ci sarà una probabilità che si guasti dopo un giorno, dopo 10 o dopo vari anni, e osserviamo che in questo caso essa deve essere dedotta dall'esperienza in quanto non c'è nessuna ragione teorica che ci possa aiutare nelle deduzioni. Quello che si osserva è una distribuzione di tipo "esponenziale", per cui la probabilità che il primo guasto avvenga a un tempo T decresce al crescere di T (e questo è giusto, perché è purtroppo via via più difficile trovare oggetti che durino molto senza rompersi). Il grafico di questa distribuzione, rappresentato in Figura 4, segue la legge

$$E(x) = a e^{-x/a}$$

ed è radicalmente diverso da quello della distribuzione normale.

6. La probabilità soggettiva e l'impostazione assiomatica

Si dovette aspettare il XX secolo per avere una definizione di probabilità che rispondesse, almeno parzialmente, ai dubbi che alcuni esprimevano in relazione all'uso spregiudicato del termine probabilità in ambiti particolari ma importanti, come gli eventi rari (terremoti, inondazioni) e tutte quelle situazioni nelle quali l'esperienza influisce in maniera essenziale sull'oggetto in esame (interventi chirurgici, prove distruttive ecc.).

La probabilità soggettiva di un evento, dovuta allo studioso italiano Bruno De Finetti⁷⁰, si definisce così:

“La probabilità di un evento è il premio (compreso tra 0 e 1) che una persona ritiene equo scommettere sapendo di vincere 1 se l’evento si verifica e 0 se non si verifica. Le probabilità degli eventi devono però essere tali che non si possa ottenere con una sequenza di scommesse una vincita certa o una perdita certa.”

A tutta prima sembra una definizione complessa, però in effetti non è così. Sicuramente è “soggettiva”, in quanto legata a una definizione di scommessa, ma contiene un cavillo. Supponiamo che l’evento sia “fare 6 con il dado”. Sappiamo che (o crediamo che sia così) la probabilità frequentista di questo evento, assumendo il dado equo, è $1/6$. Questo implica che su un numero grande N di ripetizioni, circa $N/6$ avranno il 6, e il “circa” diviene sempre meno circa al crescere di N . Supponiamo che un ipotetico scommettitore sia convinto (perché magari sospetta che il dado sia truccato) che la probabilità di fare 6 con il dado sia $1/2$; allora lui scommette (nel senso che paga) $0,5$ a ogni lancio del dado, e vince 0 quando il dado non dà 6 e 1 quando dà 6. Ma noi sappiamo che il dado è equo, per cui su N lanci solo approssimativamente $N/6$ daranno 6, e quindi vincerà $N/6$, dato che vince 1 a ogni uscita del 6. Però lui ha speso $N/2$, perché ha giocato $0,5$ a ognuno degli N lanci. Ma allora avrà perso $N/2 - N/6 = N/3$ in tutto, e al crescere di N questa perdita cresce sempre più. Allo stesso modo, se il nostro scommettitore scommettesse, diciamo, $1/12$ a ogni lancio, alla lunga prenderebbe $N/6$ e avrebbe pagato $N/12$, con un guadagno netto di $N/12$ ⁷¹. Dunque la probabilità di fare 6 con un dado equo deve essere $1/6$.

A tutta prima sembra un’inutile complicazione, ma non lo è dal punto di vista del principio, perché questa definizione non richiede né indipendenza, né ripetibilità dell’esperimento, anche se il cavillo continua a parlare di “sequenza di scommesse”. Naturalmente il cavillo resta, e quindi potrebbe ben darsi che la probabilità soggettiva sia diversa dalla probabilità frequentista, o addirittura

⁷⁰ Bruno De Finetti (Innsbruck, 1906 – Roma, 1985), matematico e statistico italiano. A lui si deve la formulazione della concezione soggettiva operativa della probabilità.

⁷¹ Per chi ha qualche dimestichezza con i giochi d’azzardo, come la roulette, per esempio, saprà che in questo caso viene pagato non una somma fissa (1 nel nostro caso) ma un determinato multiplo della somma puntata, e quindi il truccetto sopra descritto non funziona.

non definibile. Spieghiamo meglio con un esempio. Se un paziente deve valutare l'affidabilità di un intervento chirurgico su di sé, quanto scommetterà? Chiaramente, se l'esito dell'intervento è sfavorevole potrebbe avere un danno enorme e addirittura morire, nella migliore delle ipotesi il nostro paziente non si sentirà autorizzato a scommettere, e in ogni caso pazienti diversi potrebbero dare indicazioni diverse. Invece il medico, per il quale i pazienti sono tutti uguali, farà una statistica con la frequenza delle operazioni riuscite, e pubblicherà questo dato, ben sapendo che ogni paziente fa storia a sé.

La probabilità soggettiva ha avuto il pregio di svincolarsi da quella frequentista, ma il difetto di sollevare un vespaio di discussioni, prima fra tutte quella sulla confrontabilità del valore attribuito alla probabilità (la "scommessa"), e oggi non molti l'accettano. È un po' come il dibattito sul libero arbitrio: se la probabilità è soggettiva, allora ciascuno la interpreta come vuole e non serve a molto, mentre se si accetta il "cavillo", allora non è più soggettiva, e anzi il "cavillo" richiede di sapere già (come abbiamo visto nel nostro semplice esempio) quale sia la probabilità frequentista.

Tuttavia su molte questioni i frequentisti e i soggettivisti si trovano d'accordo. Per esempio, sul fatto che la probabilità debba essere un numero positivo (o nullo), che la probabilità di un evento certo sia 1, e che la probabilità dell'unione disgiunta sia uguale alla somma delle probabilità.

Anche l'impostazione soggettiva permette di dedurre dal "cavillo" queste tre proprietà. Vediamo come. Che per ogni evento E si abbia $p(E)$, la sua probabilità, maggiore o uguale a zero è ovvio perché la "scommessa" varia tra 0 e 1. Sia adesso U un evento che si verifica sempre (ossia che contiene tutti i risultati possibili). Allora, a ogni ripetizione, il nostro scommettitore riceverà sempre 1, e quindi se scommettesse meno o più di 1 vincerebbe o perderebbe sempre, e dunque il "cavillo" impone di dire che $p(U) = 1$. Prendiamo poi due eventi complementari, e chiamiamoli E ed \bar{E} . Chiaramente uno dei due si verifica, per cui $E \cup \bar{E}$ si verifica sempre, e quindi quale che sia la scommessa sull'evento E , per esempio $1/6$, la scommessa sull'evento complementare \bar{E} deve essere il complementare a 1 della scommessa su E , cioè $5/6$, perché se così non fosse, uno scommettitore che puntasse sia su E , sia su \bar{E} avrebbe una vincita certa o una perdita certa perché, comunque vada, a ogni turno vince 1 (da E o dal suo opposto).

Dunque

$$p(E) + p(\overline{E}) = 1$$

Lo stesso ragionamento vale ovviamente se gli eventi sono 3 o più, basta che siano necessari (nel senso che almeno uno si verifica) e disgiunti (con il che se ne verifica sempre esattamente uno). Allora siano A e B due eventi disgiunti. Non è difficile constatare, con qualche diagramma dei corrispondenti insiemi, che A , B e $A \cup B$ sono necessari e disgiunti, per cui

$$p(A) + p(B) + p(\overline{A \cup B}) = 1$$

però anche $A \cup B$ e $\overline{A \cup B}$ sono ovviamente necessari e disgiunti, per cui

$$p(A \cup B) + p(\overline{A \cup B}) = 1$$

Sottraendo membro a membro troviamo allora

$$p(A) + p(B) - p(A \cup B) = 0$$

che è proprio l'additività.

A partire dalla teoria frequentista, erano stati poi scoperti e dimostrati vari teoremi riguardanti la probabilità. Fu il matematico russo Andrej Kolmogorov⁷² che propose di dare un'impostazione assiomatica alla teoria della probabilità, come in geometria, richiedendo un numero (possibilmente basso) di assiomi dai quali far discendere le proposizioni note come semplici teoremi, e quindi spostando il discorso sulla bontà o meno dei soli assiomi. Il bello fu che gli assiomi richiesti si rivelarono solo tre: il fatto che $p(E) \geq 0$; il fatto che la probabilità di un evento certo sia 1; l'additività, però in una forma più forte di quella che abbiamo visto, nella quale possano comparire infiniti eventi disgiunti⁷³. In formule

1. $p(E) \geq 0$ per ogni evento E ;
2. $p(U) = 1$, dove U è l'evento "universale" che contiene tutti i risultati;

⁷² Andrej Nikolaevič Kolmogorov (Tambov, 1903 – Mosca, 1987), matematico russo. Ha fornito contributi fondamentali al calcolo delle probabilità, la meccanica classica e la complessità computazionale.

⁷³ Per la precisione, una quantità numerabile di eventi disgiunti.

3. Se $E_1, E_2, \dots, E_n, \dots$ è una successione di eventi a due a due disgiunti ($E_i \cap E_j = \emptyset$), allora

$$p(E_1 \cup E_2 \cup \dots \cup E_n \cup \dots) = p(E_1) + p(E_2) + \dots + p(E_n) + \dots \quad 74$$

A partire da questi tre assiomi, si possono ricavare tutti i teoremi sui quali i probabilisti moderni si trovano d'accordo, anche se, ovviamente, viene meno, o potrebbe venir meno, qualche interpretazione. Si tratta appunto di teoremi: non si possono negare se non si negano i tre assiomi fondamentali.

7. La probabilità condizionata e le applicazioni

Grazie a quello che siamo riusciti a definire fin qui è ora possibile parlare brevemente di qualche applicazione della probabilità anche fuori dall'ambito dei giochi d'azzardo. Un concetto molto utile e importante è quello di probabilità condizionata, che si definisce così: dato un evento, diciamo A , a probabilità non nulla, definiamo probabilità di un secondo evento B condizionatamente a A il numero

$$p(B | A) = \frac{p(B \cap A)}{p(A)}$$

(la scrittura " $B | A$ " si legge " B dato A "). Cerchiamo di capire cosa significa: da un punto di vista lessicale, essa va letta così: è la probabilità che si verifichi B sotto la condizione che si verifichi A , o condizionatamente al verificarsi di A (che non è certo). Per esempio, è una probabilità condizionata quella di morire se un aereo precipita: in questo caso, non si tratta della probabilità di morire durante il viaggio, ma quella nel caso in cui l'aereo precipiti. Da un punto di vista frequentista è chiaro perché sia così: se supponiamo che il numero N di ripetizioni dell'esperimento sia alto, cosicché la probabilità si possa sostituire con la frequenza relativa, abbiamo

$$p(B | A) = \frac{N_{B \cap A} / N}{N_A / N} = \frac{N_{B \cap A}}{N_A}$$

⁷⁴ Anche supponendo, o immaginando, cosa sia un'unione infinita, si tratterebbe di specificare che cosa si intenda con la somma infinita a destra di questa formula. Ovviamente, si può (si tratta del concetto di serie numerica), ma non vogliamo entrare troppo in dettagli matematici e lasciamo all'immaginazione anche questo concetto. Così come il fatto, peraltro evidente, che se da E_i in poi tutti gli eventi sono vuoti, allora la loro probabilità è zero (si può dimostrare, ovviamente) e gli zeri non aggiungono nulla alla somma infinita, per cui risulta la solita legge "finita"

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) \quad \text{se } E_1 \cap E_2 = \emptyset$$

Ecco perché si interpreta $p(B | A)$ come probabilità condizionata: stiamo contando i casi in cui è accaduto B tra quelli nei quali è accaduto A , quindi condizionatamente all'essersi verificato A .

Nella pratica spesso si conoscono più le probabilità condizionate di quelle assolute, come nel caso dell'aereo di cui sopra. Dalla definizione di probabilità condizionata discende allora subito che

$$p(B \cap A) = p(B | A)p(A) \quad (2)$$

e quindi potremmo usare questa formula per calcolare qualcosa che non conosciamo, come $p(B \cap A)$, che sappiamo interpretare come “accade sia l'evento B sia l'evento A ”.

Per continuare con il nostro esempio, se supponiamo che la probabilità di morire se l'aereo precipita sia del 99,99%, quindi, almeno soggettivamente, molto alta, sappiamo (da dati certi) che la probabilità di un incidente aereo, cioè $p(A)$, è dello 0,043 per 100.000 decolli, ossia lo 0,0000043 su un singolo decollo. Se stiamo prendendo l'aereo e ci chiediamo: “Morirò o no per un incidente?”, devono succedere due cose: l'aereo deve precipitare (A) e devo morire (B), cioè $A \cap B$. La probabilità di morire è quindi

$$p(A \cap B) = 0,9999 \cdot 0,0000043 = 0,000004299 \approx \frac{1}{2325813}$$

ossia, in parole povere, di 1 su 2.000.000⁷⁵ nel nostro esempio semplificato. Questo esempio spiega perché la gente ha paura di volare (confonde la probabilità di morire, molto bassa, con quella condizionata, molto alta). Con l'interpretazione di $p(B | A)$ come probabilità che B si verifichi “a causa di A ”, possiamo definire A e B indipendenti se $p(B | A) = p(B)$, ossia se l'evento B ha la stessa probabilità di verificarsi sia che si verifichi A o meno. La formula appena scritta ci dice allora che per due eventi indipendenti si ha

$$p(B \cap A) = p(B) \cdot p(A)$$

⁷⁵ Secondo i calcoli delle compagnie aeree, un passeggero che si imbarcasse da oggi in poi ogni giorno a caso su un aereo dovrebbe “attendere” 21.000 anni per avere una significativa probabilità di morire per incidente aereo in quel lasso di tempo. Naturalmente, non sa quando...

formula che abbiamo usato quando abbiamo parlato più sopra dell'errore di Pacioli.

L'equazione (2) si presta anche a un'altra interessante interpretazione. Siccome $A \cap B = B \cap A$, possiamo scrivere

$$p(A | B)p(B) = P(A \cap B) = p(B \cap A) = p(B | A)p(A)$$

e quindi possiamo ricavare

$$p(A | B) = \frac{p(B | A)p(A)}{p(B)}$$

È la formula di Bayes⁷⁶ che, alla luce della nostra interpretazione della probabilità condizionata, comporta alcune conseguenze notevoli. Se la scrittura $p(B | A)$ significa “la probabilità dell'evento B condizionatamente al verificarsi di A ” ($B | A$ non è un evento ma solo una scrittura, è bene dirlo), allora la scrittura $p(A | B)$ significherà “la probabilità dell'evento A condizionatamente al verificarsi di B ”, e quindi può rispondere alla domanda “ammesso che si sia verificato A , qual è la probabilità che sia stato B a causarlo?” Nel nostro esempio essa risponderebbe alla domanda: “Se io morirò, qual è la probabilità che sia stato per l'incidente aereo?”. Nel nostro caso, per calcolarla serve la probabilità che io muoia (s'intende, come s'intendeva prima, nell'arco di tempo del viaggio), che non è facile da calcolare. Da dati ricavati in Internet (tanto stiamo facendo un esempio) si ricava che per uno statunitense medio la probabilità di morire a 17 anni nell'arco di un anno è circa 0,0009. Se pensiamo che il viaggio duri un giorno, possiamo tranquillamente dividere per 365 questa probabilità, supponendo tutti equiprobabili i giorni della nostra morte, ottenendo $p(B) = 0,0000246$ (notate che, giustamente, è più alta di quella di prima perché qui stiamo supponendo possibili anche altre cause di morte). Quindi, applicando la formula di Bayes, troviamo

$$p(A | B) = 0,9999 \cdot 0,0000043 / 0,0000246 \approx 0,098 = 9.8\%$$

⁷⁶ Thomas Bayes (Londra, 1702 – Tunbridge Wells, 1761), matematico inglese. Con la teoria che da lui prese il nome formalizzò il calcolo delle probabilità che tiene conto della modificazione dell'informazione originaria derivante dal risultato sperimentale.

Quindi, se moriremo nella giornata odierna, possiamo concludere che al 10% circa ciò potrebbe essere dovuto all'incidente aereo (e quindi al 90% circa ad altre cause, come un incidente nel tragitto all'aeroporto, una causa naturale, un omicidio nell'aeroporto o nell'aereo ecc).

La formula di Bayes permette anche di valutare la più probabile fra diverse possibili cause di un dato effetto, che è poi la versione più importante della formula. Per giungervi, però, serve un ultimo passo. Supponiamo che un dato evento, diciamo A , possa verificarsi in seguito a due (o più, noi per semplicità ne considereremo solo due) cause B_1 e B_2 , necessarie e disgiunte, nel senso detto sopra, ossia che una e una sola delle due si verifica. Questo si scrive in formule

$$B_1 \cup B_2 = U, \quad B_1 \cap B_2 = \emptyset$$

dove U è l'evento "universale", che raccoglie tutti i risultati possibili (l'"universo del discorso" in termini di insiemi). Dall'esame della Figura 5 si vede ora subito che $A \cap B_1$ e $A \cap B_2$ sono disgiunti (e, naturalmente, si può dimostrare usando le proprietà degli insiemi), così come si vede che la loro unione è tutto l'evento A . Dunque, per la proprietà di additività,

$$p(A) = p(A \cap B_1) + p(A \cap B_2)$$

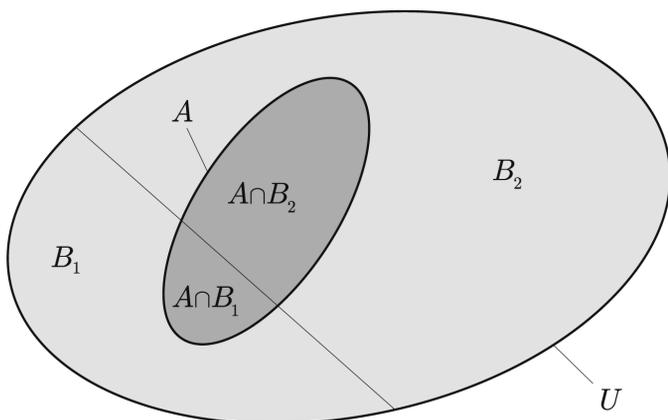


Figura 5. Teorema della probabilità totale.

Dalla definizione di probabilità condizionata segue ora⁷⁷

$$p(A \cap B_1) = p(A | B_1)p(B_1) \quad p(A \cap B_2) = p(A | B_2)p(B_2)$$

e quindi troviamo

$$p(A) = p(A | B_1)p(B_1) + p(A | B_2)p(B_2)$$

Questo risultato è interessante perché dice come calcolare la probabilità totale quando un evento proviene da due cause diverse, e trova numerose applicazioni. Supponiamo, per esempio, di avere due macchinari che producono pezzi tutti uguali. Il primo produce una frazione del 2% di pezzi difettosi, il secondo solo una frazione dell'1%. Supponiamo poi che il primo macchinario produca il 60% dei pezzi prodotti e l'altro il 40%. Possiamo calcolare allora la probabilità totale che un pezzo sia difettoso. Infatti B_1 e B_2 sono gli eventi "il pezzo è prodotto dal primo macchinario" e "il pezzo è prodotto dal secondo macchinario" e sono rispettivamente 0,6 e 0,4 (il 60% e il 40% che si diceva). Poi A è l'evento "il pezzo è difettoso". Ora $p(A | B_1)$ è la probabilità di avere un pezzo difettoso condizionatamente al fatto che sia stato prodotto dal primo macchinario, cioè appunto il 2%, ossia 0,02, e analogamente $p(A | B_2) = 0,01$. Dalla formula trovata abbiamo allora

$$p(A) = 0,6 \cdot 0,02 + 0,4 \cdot 0,01 = 0,016$$

che corrisponde all'1,6% dei pezzi.

Si capisce poi che se gli eventi B_1, B_2, \dots, B_n sono n anziché 2 (ma sempre necessari e disgiunti), la formula vale anche per n .

Riscriviamo ora la formula di Bayes scambiando A e B , giusto per comodità,

$$p(B | A) = \frac{p(A | B)p(B)}{p(A)}$$

⁷⁷ Le formule appena scritte valgono anche se $p(B_1)$ o $p(B_2)$ sono nulli, il che non era possibile con la definizione che abbiamo dato noi di probabilità condizionata. Siccome in quel caso comunque il prodotto è 0, e sia $A \cap B_1$ che $A \cap B_2$ sono più piccoli rispettivamente di B_1 e B_2 , ne segue che $p(A \cap B_1) \leq p(B_1) = 0$ e $p(A \cap B_2) \leq p(B_2) = 0$.

essa continuerà a valere per ogni evento B . Sostituiamo a B uno dei B_1, B_2 di sopra e teniamo conto dell'espressione di $p(A)$. Ne segue

$$p(B_1 | A) = \frac{p(A | B_1)p(B_1)}{p(A | B_1)p(B_1) + p(A | B_2)p(B_2)}$$

e analogamente per B_2 . Questa formula è utile in quanto serve per trovare il colpevole. Supponiamo di avere estratto dal cesto degli interi pezzi prodotti un pezzo difettoso: qual è la probabilità che sia stato il primo macchinario a produrlo? Si tratta di trovare appunto $p(B_1 | A)$. A questo punto è facile: $p(A)$ lo conosciamo già, e quindi

$$p(B_1 | A) = \frac{0,02 \cdot 0,6}{0,016} = 0,75$$

per cui la probabilità cercata (il "colpevole" è il primo macchinario) è del 75%.

La formula di Bayes è un utile strumento anche per valutare la probabilità che una certa malattia derivi da una certa causa, e la statistica aiuta in questo senso a produrre stime sempre più accurate delle probabilità condizionate. Analoghe considerazioni valgono nel campo economico e delle ricerche di mercato così come, nel nostro esempio, nell'ambito della valutazione dell'efficienza aziendale.

In definitiva, le applicazioni del calcolo delle probabilità sono oggi quasi ovunque distribuite, e costituiscono un interessante punto di convergenza fra teoria e pratica: da un lato la matematica produce i fondamenti per l'elaborazione, e dall'altro la statistica i dati provenienti dalla popolazione e dall'esperienza.

In ogni caso non va mai dimenticato che la probabilità non dà certezza, ma una ragionevole valutazione dell'attendibilità di un evento, proprio come intendevano, in ultima analisi, gli antichi.

COMPUTER E SOLUZIONI APPROSSIMATE

1. Il computer e le quattro operazioni

Una volta si diceva che occorre saper “leggere, scrivere e far di conto”. Anche oggi l’utilità di saper “leggere e scrivere” è indubbia, ma la situazione è diversa riguardo al saper “far di conto”. Si tratta delle quattro operazioni: addizione, sottrazione, moltiplicazione e divisione. E indubbio che oggi queste le sa fare molto meglio il computer; anzi, il computer è stato inventato proprio allo scopo di alleviare l’uomo dal tedio di noiosi riporti, prestiti e quant’altro. Le macchine calcolatrici meccaniche degli anni ’60 del secolo scorso facevano addizioni e sottrazioni (essenzialmente mimando la tecnica che si imparava a scuola) e, con fatica, le moltiplicazioni e le divisioni. Da allora di acqua ne è passata molta sotto i ponti almeno dal punto di vista dell’informatica. Analizziamo il legame esistente tra i computer odierni e le *operazioni elementari*⁷⁸.

1.1 La notazione scientifica

A differenza delle prime macchine calcolatrici, i computer utilizzano la cosiddetta notazione scientifica per rappresentare i numeri. In un testo di fisica il numero di Avogadro viene scritto (in modo approssimato) come $1,022 \times 10^{23}$, che significa 1022 seguito da 20 zeri, un po’ noioso da scrivere per intero! La parte 1,022 è chiamata mantissa e contiene quelle che chiameremo le prime cifre significative del numero, mentre il 23 è l’esponente: rappresenta essenzialmente l’ordine di grandezza del numero. Nel computer avviene sostanzialmente la stessa cosa; una differenza, che ha una certa rilevanza, è la base utilizzata: 2 invece che 10. Un vantaggio importante nell’utilizzo della notazione scientifica consiste nella possibilità di operare con numeri molto grandi o molto piccoli. Il numero di cifre utilizzate per la mantissa è strettamente correlato all’errore che si

⁷⁸ Con questo termine saranno indicate le quattro operazioni più qualche altra.

commette nella rappresentazione delle grandezze in gioco. L'errore s'intende misurato in termini relativi, in altre parole quello che conta è la percentuale di errore relativamente alla misura esatta.

1.2 Gli errori di arrotondamento

Il computer esegue le quattro operazioni essenzialmente con le stesse regolette che si insegnavano a scuola: allineando le cifre, facendo i riporti ecc. Viene però deciso una volta per tutte il numero di cifre significative che si intende utilizzare – cioè la lunghezza della mantissa – per impedire che nel corso di numerose operazioni si ottengano numeri sempre più lunghi da rappresentare. Questo vincolo comporta inevitabilmente l'introduzione di un piccolo errore, l'*errore di arrotondamento*.

La maggior parte dei computer moderni utilizza normalmente 8 byte, ovvero 64 bit, per memorizzare (diremo, con un termine più corretto, rappresentare) un numero reale; di questi bit, 52 vengono usati per la mantissa, 11 per l'esponente e 1 per il segno⁷⁹. L'errore di arrotondamento che in questo modo viene generato a ogni operazione elementare rende inaffidabili le cifre della scrittura decimale a partire dalla diciassettesima (approssimativamente). In altre parole, il risultato di un'operazione elementare fatta dal computer è affidabile solo nelle prime 16 cifre circa della rappresentazione in base 10. Notiamo che internamente il computer opera in base 2 e solo i risultati finali dei programmi sono convertiti in base 10. Per quanto sia piccolo, in alcune circostanze l'errore di arrotondamento può avere effetti disastrosi, come può essere evidenziato dal seguente semplice esempio: vogliamo risolvere l'equazione di secondo grado

$$x^2 - 2px + q = 0 \quad (1)$$

con i valori

$$p = 500000000,16667 \text{ e } q = 333333333,33333$$

più precisamente siamo interessati alla soluzione più piccola, in effetti con i valori indicati la nostra equazione ha due soluzioni reali positive. Utilizzando la formula risolvente, e facendo i calcoli con un computer otteniamo:

⁷⁹ Questi valori sono prescritti dallo standard internazionale IEEE 754, a cui aderisce la quasi totalità dei processori esistenti oggi.

$$x_1 = p - \sqrt{p^2 - q} = 0,33333331346512 \quad (2)$$

Il valore effettivo ottenuto può dipendere dal computer utilizzato! Per capire quanto buono sia il risultato possiamo sostituire il valore trovato nell'equazione di partenza; invece di zero otteniamo però $x^2 - 2px + q = 19,86765939$. Se facciamo la stessa prova, ma usando il valore $x_1 = 1/3$ otteniamo $-5,96046447754 \cdot 10^{-08}$, uno scarto ben più piccolo, che ci convince del fatto che la soluzione corretta è molto più vicina a $1/3$ (ovvero “zero virgola tre periodico”) che al numero ottenuto con la formula risolvente.

Il computer utilizzato per fare questi conti opera internamente con circa 16 cifre, mentre la soluzione ottenuta con la formula risolvente è sbagliata già nell'ottava cifra significativa e ciò dopo solamente quattro operazioni elementari!

È successo che le prime tre operazioni forniscono un risultato intermedio che, a causa dell'errore di arrotondamento, si discosta leggermente dal valore esatto, e fin qui nulla di male, se non che l'ultima sottrazione amplifica questo piccolo errore di un fattore superiore a 10^9 , cioè di 1 miliardo di volte. Per fortuna questo fenomeno (dovuto all'instabilità dell'algoritmo risolutivo scelto) non è così frequente e soprattutto può essere previsto e quindi evitato, permettendoci di utilizzare con fiducia la soluzione di problemi che richiedono l'esecuzione di svariati miliardi di operazioni successive.

Tornando al problema proposto come esempio, possiamo manipolare la formula risolvente (2) come segue:

$$x_1 = (p - \sqrt{p^2 - q}) \frac{p + \sqrt{p^2 - q}}{p + \sqrt{p^2 - q}} = \frac{q}{p + \sqrt{p^2 - q}} \quad (3)$$

La nuova formula è, dal punto di vista matematico, del tutto equivalente alla prima; non lo è però se si vuole tenere conto degli errori di arrotondamento, in altre parole, le operazioni elementari svolte dal computer non verificano le usuali proprietà delle quattro operazioni, o, se preferite, le verificano solo in modo approssimato.

In particolare la formula (3), se utilizzata nel nostro particolare contesto, non è affetta da instabilità numerica e il suo utilizzo per risolvere il problema (1) porta al risultato (essenzialmente corretto) $x_1 = 0,33333333333333$.

1.3 Non solo conti

A onor del vero, oggi il computer non si limita affatto a eseguire miriadi di operazioni elementari, anzi il suo utilizzo per fare i conti sembra essere passato in secondo piano. Quando si compone una lettera, si invia una e-mail o si naviga in Internet utilizzando il computer, le quattro operazioni tra numeri reali sono coinvolte solo in minima parte, o non lo sono affatto. Per questi compiti si sfrutta infatti la capacità dei computer moderni di essere programmati per effettuare procedure ripetitive, magari estremamente complesse, in cui la capacità di far di conto è relativamente poco importante, o non lo è affatto.

Anche nell'ambito della matematica (escludendo utilizzi del computer di tipo marginale, come la scrittura degli articoli di ricerca, la comunicazione con i colleghi, lo svolgimento di pratiche amministrative ecc.) si possono individuare due modi di utilizzare il computer nettamente distinguibili: da una parte rimane la necessità di utilizzare la potenza di calcolo nella forma primitiva di saper svolgere con efficienza le quattro operazioni⁸⁰, dall'altra è possibile un utilizzo del computer che aiuti il matematico a effettuare i passaggi formali che sono alla base del suo lavoro. Un esempio di questo tipo di attività è fornito dagli ambienti di *calcolo simbolico*, in cui non si pretende di sostituire a ogni numero (reale) una sua (inevitabilmente non esatta) rappresentazione tramite la notazione scientifica, ma si procede piuttosto con manipolazioni formali di espressioni in modo analogo a come si effettuano i passaggi per risolvere una equazione, o per calcolare una derivata. In particolare, in tali ambienti le frazioni sono rappresentate come rapporto tra numeri interi, senza effettuare la divisione con conseguente troncamento del risultato dopo un certo numero di cifre. L'ambiente *Derive*⁸¹ è un *computer algebra system* (CAS) molto utilizzato per la didattica nelle scuole secondarie. Molto conosciuto è anche *Mathematica* di Stephen Wolfram e non mancano ambienti di calcolo di questo tipo sviluppati secondo la filosofia open source, primo fra tutti *Maxima*⁸².

C'è da dire che per motivi di comodità praticamente tutti i vari ambienti di lavoro per la matematica, pur nascendo come ambienti di

⁸⁰ Si vedano i paragrafi precedenti.

⁸¹ Della Texas Instruments, sviluppato in *MuLisp* a partire dall'ambiente *MuMath* degli anni '80.

⁸² Successore del glorioso *Macsyma*, sviluppato a partire dal 1982 presso il Massachusetts Institute of Technology.

calcolo numerico (come *Matlab*, *Scilab*, *Octave*) o come ambienti di calcolo simbolico (come i già citati *Derive*, *Mathematica*, *Maxima*) contengono contaminazioni di tipo simbolico i primi e di tipo numerico i secondi, rendendo la distinzione nei due ambiti un po' meno netta. Tuttavia è estremamente importante avere chiara la differenza tra un procedimento effettuato per via numerica rispetto allo stesso procedimento effettuato tramite manipolazioni simboliche.

2. Alla ricerca del minimo

Molti dei problemi risolubili con l'ausilio del computer si collocano nell'area dell'*ottimizzazione*, termine con il quale si indica la ricerca della configurazione ottimale rispetto a una qualche funzione costo al variare di opportuni parametri. Per *configurazione* qui intendiamo uno stato del sistema, che a seconda dei casi può essere una particolare scelta economica, oppure la forma assunta da un corpo elastico, o ancora una configurazione di equilibrio di un sistema complesso, e così via.

Nell'ambito della meccanica dei corpi continui, per esempio, il lettore ha già avuto modo di apprendere che il problema da risolvere assume generalmente la forma di un'equazione differenziale alle derivate parziali; in molti casi questo tipo di problemi ammette una riformulazione in termini di minimizzazione di energia potenziale. La nozione di *ottimalità* sottintende la ricerca di quella particolare configurazione che risulta essere migliore delle altre configurazioni ammissibili in base a opportuni criteri di valutazione.

Assumeremo che la bontà di una particolare configurazione si possa riassumere con un singolo numero reale. Questo numero può rappresentare, a seconda dei contesti, un costo economico, un'energia potenziale, elastica o termica, la quantità di difetti in un materiale ecc. Per fare un esempio molto semplice mettiamo una pallina in una scodella. Una volta lasciata libera di muoversi e dopo un periodo transitorio⁸³, la pallina si fermerà nel punto più basso (il fondo) della scodella: ha di fatto scelto la posizione che rende minima la sua energia potenziale gravitazionale. In questo caso la configurazione del sistema è individuata dalla posizione della pallina, che si può

⁸³ In questo tipo di problemi si tende a trascurare le situazioni dinamiche transitorie, focalizzando l'attenzione sullo stato assunto dal sistema quando si è stabilizzato.

esprimere con due numeri indicanti le coordinate spaziali; questi due numeri rappresentano lo stato; l'energia potenziale complessiva si rappresenta in questo caso in funzione dello stato. Si tratta di una funzione in due variabili che assume valori nell'insieme dei numeri reali.

Vale qui la pena osservare che la scelta di considerare problemi di minimo anziché di massimo è puramente convenzionale: è infatti immediato trasformare un problema di massimo in un problema di minimo semplicemente cambiando il segno della quantità da massimizzare.

2.1 Minimi e punti stazionari

Quando si studiano i massimi e minimi di una funzione di una variabile $f : \mathbb{R} \rightarrow \mathbb{R}$ si scopre che i punti cercati verificano una proprietà di stazionarietà: l'annullamento della derivata prima. In formula $f'(x) = 0$.⁸⁴

Una proprietà analoga ($\nabla f(x) = 0$, annullamento di tutte le derivate parziali) vale per funzioni di più variabili $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Il problema di minimo si trasforma così in un'equazione o un sistema di equazioni. Sul legame tra punti di minimo e punti stazionari ci sono due importanti considerazioni da fare. Anzitutto la proprietà di stazionarietà $f'(x) = 0$ o $\nabla f(x) = 0$ non distingue tra minimi globali e minimi locali; questi ultimi sono punti ottimali solo rispetto a piccole perturbazioni dello stato del sistema. Il fatto che tali soluzioni siano accettabili o no dipende dal problema specifico, se si vuole minimizzare un costo economico si è interessati essenzialmente al minimo globale, mentre una configurazione fisica di equilibrio può corrispondere a minimi locali.

In secondo luogo un punto stazionario non è necessariamente un punto di minimo o di massimo (locale). Se si opera in più variabili, può essere un cosiddetto *punto di sella*, cioè un punto che risulta di minimo se ci si restringe solo a spostamenti in certe direzioni, mentre sarà di massimo se ci si restringe in altre direzioni. Un valico di

⁸⁴ La derivata prima $f'(x)$ rappresenta la pendenza del grafico della funzione nel punto di coordinate $(x, f(x))$, si noti che il valore di tale pendenza non può essere ricavato dal solo valore della funzione in x , ma è necessaria la conoscenza della funzione almeno in un piccolo intorno di x ; tale intorno può tuttavia essere arbitrariamente piccolo. Un intorno è un insieme contenente tutti i punti aventi distanza da x minore di una quantità positiva arbitraria. La definizione rigorosa dice che $I \subseteq \mathbb{R}$ è un intorno di x se esiste $\delta > 0$ tale che per ogni $y \in \mathbb{R}$, $|y - x| < \delta \Rightarrow y \in I$. Un discorso analogo, che omettiamo, vale per il gradiente $\nabla f(x)$ di una funzione di due o più variabili, dove x rappresenta un vettore di numeri e non un singolo numero reale.

montagna è un punto di minimo tra due creste montuose ed è un punto di massimo tra le due vallate sui versanti opposti del passo, non è dunque né minimo né massimo e tuttavia è un punto stazionario, essendo il terreno del passo pressoché pianeggiante. Si tratta in effetti di un punto di sella per la funzione in due variabili che misura la quota in funzione della latitudine e della longitudine.

Presenteremo ora in modo un po' più diffuso un paio di esempi particolarmente affascinanti, ricordando altresì che una fetta veramente considerevole di problemi concreti si esprime in ultima analisi tramite un problema di ottimizzazione.

2.2 Bolle di sapone

Le bolle di sapone, il gioco da bambini per antonomasia, rappresentano una situazione per certi versi particolarmente semplice: si tratta di una pellicola molto sottile di acqua e sapone che assume configurazioni particolari.

In questo caso, la configurazione cercata è la forma assunta da una superficie; descrivere una superficie è solo in apparenza semplice: generazioni di matematici vi si sono dedicati e tra questi spiccano i nostri Renato Caccioppoli⁸⁵ ed Ennio de Giorgi⁸⁶. In ogni caso è chiaro che non ci bastano alcuni numeri per individuare una superficie, ne servono una quantità infinita, stiamo lavorando in uno spazio di dimensione infinita.

Questo spazio di dimensione infinita non va confuso con lo spazio tridimensionale in cui “galleggiano” le bolle, si tratta bensì di uno spazio astratto di configurazioni possibili. In tali contesti, l'intuito geometrico rischia di non aiutare molto e la formalizzazione matematica diventa uno strumento potente.

La pellicola di una bolla di sapone è caratterizzata microscopicamente dalla presenza di forze tra le molecole costituenti che complessivamente danno origine alla cosiddetta tensione superficiale, la stessa che permette a certi insetti di camminare sull'acqua e che conferisce la forma sferica alle gocce di pioggia.

⁸⁵ Renato Caccioppoli (Napoli, 1904 – Napoli, 1959), matematico italiano. Ha fornito importanti contributi all'analisi funzionale e alla teoria delle equazioni differenziali. Ha anche introdotto un concetto di ipersuperficie che ha aperto la strada allo studio delle ipersuperfici minime in ogni dimensione. Gli ultimi giorni della sua vita sono rievocati nel film *Morte di un matematico napoletano*.

⁸⁶ Ennio De Giorgi (Lecce, 1928 – Pisa, 1996), matematico italiano. Ha fornito contributi fondamentali nello studio delle ipersuperfici minime e nella teoria delle equazioni differenziali alle derivate parziali, nel cui ambito ha dimostrato uno storico risultato di regolarità, ottenuto indipendentemente anche da John Nash.

Piuttosto che trattare il problema dal punto di vista di un equilibrio di forze (tensione superficiale e forza di pressione) risulta conveniente introdurre un'energia di tipo elastico e osservare che l'equilibrio delle forze corrisponde a una configurazione in cui tale energia assume valore minimo. Nel caso delle bolle di sapone l'energia dovuta alla tensione superficiale risulta proporzionale alla superficie della bolla, la qual cosa esprime chiaramente l'idea intuitiva che una pellicola di sapone cerca di assumere la configurazione in cui l'estensione superficiale risulti minima.

Nel caso delle usuali bolle c'è poi da tenere conto del fatto che l'aria racchiusa all'interno della bolla non può fuoriuscire, e fornisce quindi un vincolo di volume. Siccome la pressione esercitata dalla superficie della bolla sull'aria contenuta all'interno è estremamente piccola, si può con buona approssimazione immaginare che il volume interno sia una quantità prestabilita. In altre parole il problema diventa: trovare la forma della superficie chiusa di estensione minima che racchiuda un volume prescritto. L'intuito (e l'esperienza fatta da bambini) fa subito sospettare che la soluzione sia una sfera, e in effetti è così, anche se una dimostrazione rigorosa di questo fatto è tutt'altro che semplice. D'altra parte ci si può domandare quale configurazione assuma una pellicola di sapone racchiusa da una curva prestabilita, come nell'esempio di destra della Figura 1.

In questo caso non c'è più un volume d'aria imprigionato, e il problema è semplicemente quello di trovare la superficie di estensione minima che abbia un bordo prestabilito. Il problema della ricerca di superfici minime è noto con il nome di problema di Plateau⁸⁷. Non sempre l'intuizione fornisce la soluzione corretta a questo problema: per esempio, la forma assunta da una pellicola con bordo su due circonferenze uguali affacciate (Figura 2, sinistra) non è cilindrica, ma si incurva verso l'interno e la superficie che si ottiene è una catenoidale, cioè una curva catenaria⁸⁸ ruotata attorno all'asse delle due circonferenze.

Matematicamente questa superficie, come peraltro tutte le superfici minime, ha curvatura media nulla in ogni suo punto, ovvero le sezioni della superficie con piani ortogonali in un dato punto presentano vicino al punto una curvatura che varia al ruotare del piano ortogo-

⁸⁷ Joseph-Antoine-Ferdinand Plateau (Bruxelles, 1801 – Gand, 1883), fisico belga. Dai suoi studi su capillarità e tensione superficiale nasce il problema omonimo.

⁸⁸ La catenaria, avente equazione $y = \cosh x = \frac{e^x + e^{-x}}{2}$, è per inciso la soluzione di un altro problema di minimo: la forma assunta da una corda sospesa agli estremi.

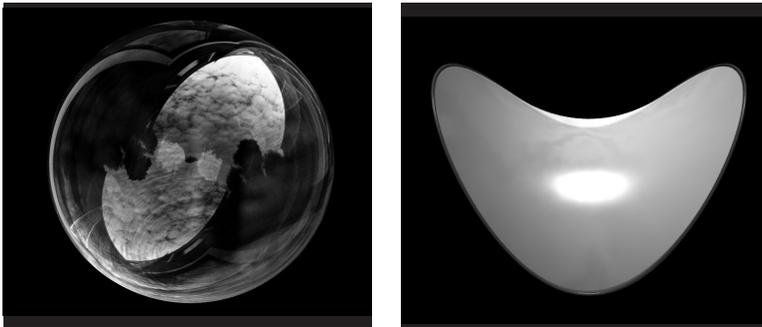


Figura 1. Una bolla di sapone (a sinistra) e una pellicola di sapone con bordo asse dalle curve in cui la superficie interseca i due piani normali.

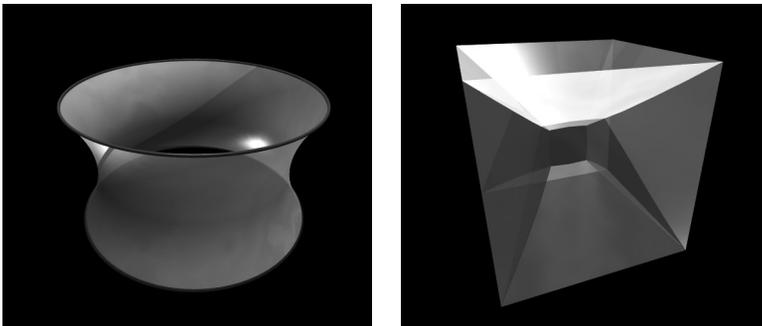


Figura 2. Catenoide (sinistra) e superficie con diramazioni con bordo sugli spigoli di un cubo (destra).

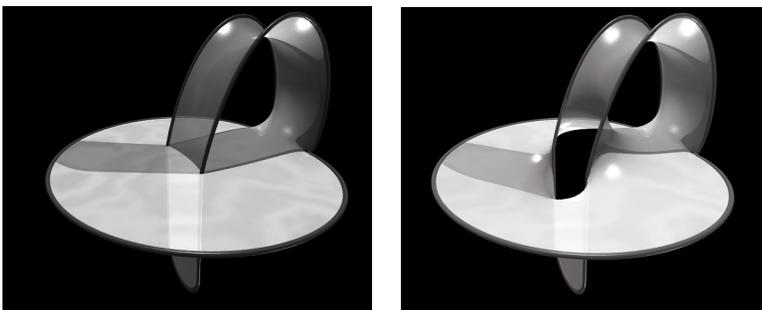


Figura 3. A sinistra una superficie (impossibile) con autointersezione. A destra una soluzione senza autointersezioni. Le immagini di questa pagina sono tratte da <http://web.math.unifi.it/users/paolini/>.

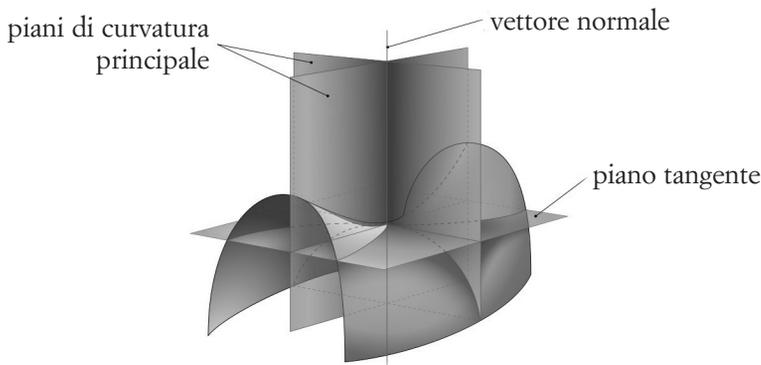


Figura 4. Le curvature principali della superficie a forma di sella sono individuate dalle curve in cui la superficie interseca i due piani normali.

nale. Tale variazione oscilla tra un valore minimo (negativo) e un valore massimo (positivo) e questi valori estremi hanno somma nulla, cioè sono uguali in valore assoluto e di segno opposto. Nella Figura 4 sono rappresentate le due curve che corrispondono alle curvature principali, intersezione della superficie con due particolari piani ortogonali. Nel paragrafo 4 del primo capitolo abbiamo incontrato la nozione di curvatura gaussiana, ottenuta come prodotto delle due curvature principali. Nonostante l'apparente parentela, la curvatura gaussiana e la curvatura media rappresentano caratteristiche sostanzialmente diverse della superficie: mentre la prima è una nozione intrinseca legata alla metrica della superficie, nella seconda entra in gioco anche il modo in cui la superficie è disposta nello spazio.

In genere una superficie minima è regolare, cioè ha un aspetto estremamente liscio e se ne facciamo uno zoom in un suo punto somiglierà sempre più a un piano. Quando questo non avviene, si parla di punti singolari. Per esempio, una superficie con bordo sugli spigoli di un cubo presenta linee triple (incontro fra 3 superfici con angolo di 120°) e punti quadrupli, dove si incontrano ben 4 superfici con orientazioni differenti (Figura 2 a destra). Tra i motivi per cui questo particolare problema è stato ed è tutt'ora una vera palestra per la matematica c'è lo strano comportamento dello spazio euclideo di dimensione 8 (lo spazio di cui abbiamo esperienza diretta

ha dimensione 3) in cui è possibile trovare una ipersuperficie minima (di dimensione 7) che sia bordo di un insieme e con un punto singolare⁸⁹. Negli spazi di dimensione più bassa questo non è possibile.

Un altro motivo di interesse è la grande varietà di approcci possibili al problema, ciascuno valido in determinati contesti, come accenneremo nel prossimo capitolo.

Infine risulta stimolante il legame particolarmente stretto con tutti i problemi fisici in cui la tensione superficiale compare come ingrediente non trascurabile. In questo senso l'esempio delle bolle di sapone risulta essere solo il più emblematico, ma situazioni del tutto simili si hanno per esempio per la forma della superficie libera dell'acqua in un capillare, di una goccia che cade, di tensostrutture in architettura, di membrane elastiche ecc.

2.3 Capire il problema prima di risolverlo

Il primo ostacolo da superare per risolvere un problema è capirlo. Il problema di Plateau di trovare la superficie di area minima che abbia un contorno assegnato ci offre un buon esempio di come ci possano essere risposte anche molto diverse a seconda di come il problema viene impostato.

L'approccio più classico consiste nel descrivere la superficie incognita come grafico di una funzione f definita su un certo sottoinsieme Ω del piano \mathbb{R}^2 a valori in \mathbb{R} con il vincolo che il suo valore sia assegnato per punti che stanno sul bordo di Ω ⁹⁰. Questa impostazione, la prima storicamente affrontata, è estremamente importante poiché si colloca in modo assolutamente naturale nell'ambito del calcolo delle variazioni e ne diventa di fatto la più tipica delle applicazioni. Si tratta del modo matematicamente più economico di scrivere il problema, che consiste nel minimizzare l'integrale

$$S(f) = \int_{\Omega} \sqrt{1 + |\nabla f|^2} \, dx \quad (4)$$

La "funzione" S , che rappresenta l'area della superficie, dipende da un oggetto, denotato dalla lettera f , che indica una intera funzione e

⁸⁹ Il cono di equazione $x_1^2 + x_2^2 + x_3^2 + x_4^2 = x_5^2 + x_6^2 + x_7^2 + x_8^2$, esempio fornito nel 1969 da E. Bombieri, E. De Giorgi, E. Giusti.

⁹⁰ Sono necessarie anche altre ipotesi, in particolare legate alla regolarità di f su cui però non insistiamo.

non un punto di \mathbb{R} o di \mathbb{R}^n : stiamo lavorando in uno spazio di dimensione infinita. L'equazione di Eulero-Lagrange⁹¹ fornisce una condizione di stazionarietà per la superficie cercata, che si dimostra essere equivalente al già citato annullamento della curvatura media. Questo approccio presenta delle limitazioni, la prima delle quali dovuta al fatto che solo alcune superfici possono essere descritte come grafico di una funzione. In particolare, la catenoidale della Figura 2 (p. 101) non è il grafico di una funzione, se considerata globalmente; può però essere suddivisa in varie porzioni, ciascuna delle quali, dopo un'opportuna scelta del sistema di assi coordinati, diventa un grafico e quindi soluzione del problema di minimo summenzionato. Altri approcci, più moderni, permettono di ampliare l'ambito di applicabilità e anche di distinguere tra diverse situazioni fisiche: la superficie può autointersecarsi, come la superficie di sinistra nella Figura 3 a p. 101 (configurazione non ammissibile per una pellicola di sapone, che invece potrebbe assumere la configurazione della superficie di destra nella stessa figura), oppure può presentare ramificazioni, come succede nel caso illustrato dalla superficie di destra nella Figura 2.

2.4 L'importanza di semplificare e generalizzare

Nel problema delle bolle di sapone si cerca una superficie, ovvero un oggetto di dimensione 2, contenuta nello spazio tridimensionale, che ha dimensione 3. Possiamo cercare di generalizzare il problema facendo variare questi due numeri.

Abbiamo in effetti citato il caso di superfici di dimensione 7 nello spazio euclideo di dimensione 8. L'intuizione geometrica non aiuta molto, ma le formule matematiche si prestano piuttosto bene a questo tipo di generalizzazione. Possiamo però anche fare l'operazione opposta di diminuire i due parametri, per esempio considerando oggetti di dimensione 1 (linee) contenute nel piano euclideo (dimensione 2). Il problema che si ottiene è ben più semplice di quello originario; d'altra parte è sicuramente il caso di capire con precisione quello che succede nel caso semplice se si vuole affrontare con qualche speranza il problema più complesso.

⁹¹ L'equazione di Eulero-Lagrange per una funzione $S(x, f, p)$ dove $p = \nabla f$ indica il gradiente di f , si scrive come $\text{div } S_p = S_f$ dove S_p rappresenta il vettore delle derivate parziali di S rispetto a ciascuna componente di $p = \nabla f$ e l'operatore div è la divergenza rispetto alla variabile spaziale x . Nel nostro caso la S non dipende direttamente dal valore f della funzione, quindi il termine S_f non è presente. Si otterrà $\sum_{i=1}^n \frac{\partial}{\partial x_i} \frac{\partial S}{\partial f_{x_i}} = 0$.

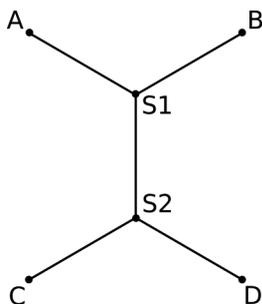


Figura 5. Quattro città collegate in modo ottimale.

Qual è il modo più economico di costruire delle strade che colleghino tra di loro quattro città collocate ai vertici di un quadrato? Supponiamo per semplicità che il costo sia proporzionale alla lunghezza della strada, per cui si tratta di minimizzare la lunghezza complessiva della rete stradale. Ci convinciamo facilmente che non ci potranno essere curve, posto che si può sempre drizzare una curva guadagnando sicuramente sulla lunghezza complessiva. In più ci possono essere degli incroci. L'insieme dei quattro lati del quadrato non è una soluzione, perché si può ridurre il costo eliminando completamente un lato pur continuando a mantenere collegate le quattro città. Otteniamo un sistema di strade avente costo 3 (immaginando che il lato del quadrato abbia lunghezza 1 e che il costo unitario sia anch'esso 1). Se tracciamo le due diagonali del quadrato abbiamo una lunghezza complessiva di $2\sqrt{2}$, che, calcolatrice alla mano, fornisce una soluzione ancora migliore con quattro strade e un incrocio. Tuttavia la soluzione esatta del problema, illustrata nella Figura 5, è un'altra ancora. Essa è formata da un totale di 5 strade e due incroci a Y, S_1 e S_2 , in cui si incontrano tre strade. Il costo complessivo di $1 + \sqrt{3}$, che risulta ancora più basso.

I due incroci a tre vie della soluzione sono chiamati punti di Steiner⁹² e le tre vie formano tre angoli uguali di 120° . Come si può no-

⁹² Jakob Steiner (Berna, 1796 – Berna 1863), matematico svizzero. Ha posto le basi della geometria sintetica moderna. Utilizzando il processo chiamato “simmetrizzazione di Steiner” ha risolto per primo il problema isoperimetrico dimostrando rigorosamente che il cerchio è la curva chiusa di lunghezza fissata che racchiude la massima area (problema di Didone).

tare la soluzione ottimale del problema non è proprio così intuitiva, d'altra parte vorremmo essere sicuri che non si possa fare ancora meglio della soluzione di Figura 5. In altre parole vorremmo *dimostrare* che non si può fare meglio.

Il primo, fondamentale, passo per produrre una dimostrazione rigorosa consiste nel dimostrare che esiste una soluzione. Ciò può apparire scontato, tuttavia si tratta quasi sempre del passaggio matematico più delicato, in particolare perché richiede una corretta impostazione del problema. Non a caso uno dei sette problemi del millennio⁹³ richiede tra l'altro di dimostrare l'esistenza di una soluzione dell'equazione di Navier-Stokes per la fluidodinamica, già incontrata nel paragrafo 4 del capitolo "Modelli e previsioni".

Un esempio emblematico per capire la delicatezza delle dimostrazioni di esistenza ci viene dalla famosa questione dell'incommensurabilità della diagonale del quadrato. Per i pitagorici era inconcepibile che la diagonale di un quadrato di lato unitario non fosse esprimibile sotto forma di frazione. Da un punto di vista più moderno esprimiamo questo fatto dicendo che la diagonale del quadrato unitario non esiste, se si pretende di lavorare nell'insieme dei numeri razionali. Quest'affermazione fa letteralmente a pugni con la nostra intuizione geometrica, e la scappatoia è stata l'introduzione dell'insieme dei numeri reali, formalizzato in modo soddisfacente solo in tempi recenti (XVIII e XIX secolo).

Una volta stabilita l'esistenza di una soluzione per il problema delle quattro città, questione appunto delicata su cui non ci soffermiamo, il più è fatto: si tratta semplicemente di stabilire una serie di proprietà che tale soluzione deve avere grazie al fatto di essere ottimale. In particolare si dimostra che:

- tutti i tratti di strada devono essere rettilinei;
- gli unici incroci possibili sono quelli del tipo della Figura 5 con angoli di 120° ; quest'ultimo fatto si dimostra facendo vedere che un altro tipo di incrocio, per esempio un incrocio a X, può essere leggermente modificato in modo da diminuire la lunghezza complessiva delle strade, contraddicendo quindi l'ipotesi di ottimalità della soluzione proposta.

⁹³ All'inizio del nuovo millennio il Clay Mathematics Institute ha elencato sette problemi matematici irrisolti e giudicati di grande rilevanza. Tra questi, oltre alla già citata equazione di Navier-Stokes, spicca l'ipotesi di Riemann, unico dei sette problemi a essere anche presente tra i famosi 23 problemi proposti da Hilbert all'inizio del '900.

Stabilito questo, si tratta di costruire tutte le possibili configurazioni di strade che soddisfano tali proprietà e selezionare quella di costo minimo. Per inciso il nostro problema ha due soluzioni distinte, la seconda delle quali si ottiene ruotando di 90° il disegno della Figura 5.

2.5 *Fiocchi di neve*

Nel 1890 circa Jožef Stefan⁹⁴ avviò lo studio di una classe di problemi che da lui hanno poi preso il nome. Il problema di Stefan studia l'evoluzione nel tempo della temperatura di un materiale in cui avviene una transizione di fase: un esempio tipico è il congelamento dell'acqua. Lo studio della transizione di fase acqua-ghiaccio è paradigmatico di come un modello fisico possa risultare sufficientemente corretto in un'ampia casistica di situazioni concrete e tuttavia fallire in modo evidente quando è applicato in particolari situazioni. Nel modello proposto da Stefan la superficie di separazione tra le due fasi viene modellizzata prescrivendo che la sua temperatura corrisponda alla temperatura di transizione (0°C nel caso della transizione acqua-ghiaccio) e che venga tenuto conto del rilascio/assorbimento del calore latente a causa della transizione di fase dovuta all'avanzamento/recessione della superficie di separazione. In alcune situazioni è necessario introdurre ulteriori ingredienti nel modello, motivate da considerazioni fisiche fatte su una scala intermedia tra la scala molecolare e quella macroscopica. In particolare, risulta utile introdurre un'energia localizzata proprio sulla superficie di separazione tra le fasi e che risulta in definitiva proporzionale all'estensione della superficie stessa; in questo modo la situazione diventa alquanto simile al problema di Plateau delle bolle di sapone. In condizioni macroscopiche normali l'energia così introdotta è talmente piccola che la posizione della superficie di separazione non ne risulta di fatto influenzata. Diversa è la situazione quando si considerano problemi su scala molto piccola, oppure circostanze in cui il modello di Stefan porterebbe a una superficie molto estesa (per esempio, negli esperimenti di sottoraffreddamento in cui si riesce ad avere acqua a temperature inferiori a 0°C).

Il fenomeno della formazione dei fiocchi di neve, non ancora pienamente compreso, rientra in queste situazioni critiche, con l'aggiunta

⁹⁴ Jožef Stefan (Sveti Peter, 1835 – Vienna, 1893), fisico e matematico sloveno. Fondamentali i suoi studi sul corpo nero e le transizioni di fase.

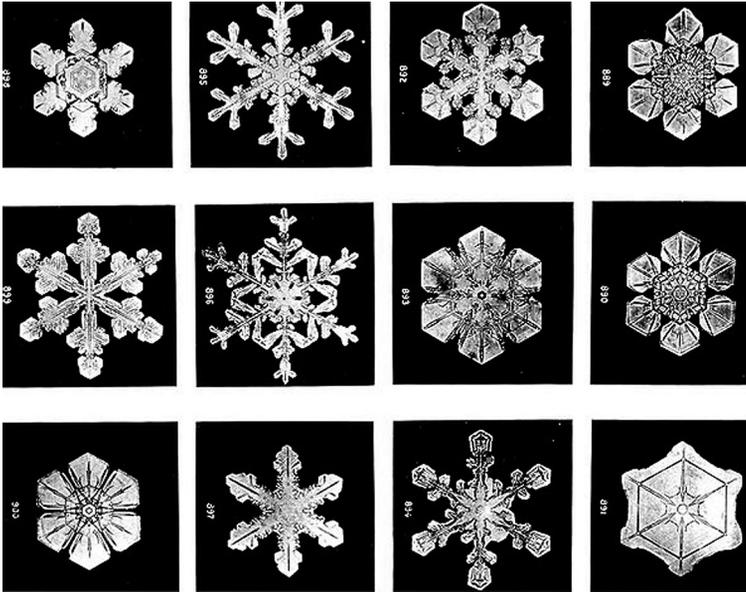


Figura 6. La simmetria su base esagonale dei fiocchi di neve.

di un importante ingrediente: la struttura cristallina del ghiaccio, che corrisponde a un prisma a base esagonale, fa sì che l'energia associata alla superficie di transizione dipenda in realtà dall'orientazione della superficie stessa. In questo modo risultano più convenienti dal punto di vista energetico le orientazioni parallele alle facce del prisma esagonale che rappresenta la struttura cristallina del ghiaccio. È questa forte dipendenza dalla orientazione la responsabile delle strutture "esagonate" dei fiocchi di neve: una situazione fortemente instabile, con la conseguente grande varietà di forme che si osservano in natura (Figura 6).

2.6 Minimi e convessità

A ben guardare la formazione dei fiocchi di neve non si presenta a prima vista come un problema di ottimizzazione, poiché la fase cinetica di transizione è importante per la determinazione della forma finale. D'altra parte la tensione superficiale assume un ruolo di fondamentale importanza, tanto da giustificare lo studio di versioni semplificate del problema in cui si vuole studiare in particola-

re gli effetti dovuti alla forte dipendenza dall'orientazione dell'energia di transizione acqua-ghiaccio.

Curiosamente c'è una forte parentela con una situazione che apparentemente non ha nulla a che spartire. Siamo al comando di un'imbarcazione a vela e ci troviamo in mare aperto in presenza di vento proveniente da nord. Idealizzando al massimo supporremo la vela perfettamente piatta e che la barca possa muoversi solamente lungo la direzione in cui è orientata. Dirigiamo la prua verso una direzione generica indicata dal vettore unitario ξ e orientiamo la vela in modo ottimale per ottenere la massima spinta possibile nella direzione prescelta. Se ξ punta esattamente a nord la vela non sarà di alcun aiuto e converrà ammainarla (oppure orientarla in modo perfettamente allineato con il vento così da non ricevere alcuna spinta). In tale situazione (ξ orientato verso nord) la velocità che si riesce a ottenere sarà nulla. Ora disegniamo su una mappa nautica l'insieme E dei punti che possono essere raggiunti in un tempo prefissato, per esempio in un minuto, puntando la prua nelle varie direzioni. L'insieme che si ottiene in questo modo avrà grosso modo la forma illustrata nella Figura 7: si tratta di un insieme non convesso, il che ha una conseguenza molto interessante: l'insieme di punti raggiungibili in un minuto ammettendo di poter scegliere la traiettoria da seguire è più grande di E ; precisamente si tratta del suo inviluppo convesso⁹⁵.

Un punto y dell'inviluppo convesso di E è contenuto in un segmento avente entrambi gli estremi x_1 e x_2 in E : ebbene, è possibile raggiungere y seguendo una traiettoria che mescola le due direzioni x_1 e x_2 con la giusta proporzione. In questo modo è possibile muoversi controvento utilizzando la ben nota andatura a zig-zag, di bolina. Osserviamo che l'insieme che si ottiene costruendo l'inviluppo convesso di E presenta dei tratti perfettamente rettilinei, precisamente si tratta di quelle parti dell'inviluppo convesso esterne a E . Il fatto quindi che l'insieme di punti raggiungibile con la barca a vela in un dato tempo presenti delle regioni rettilinee non è dovuto a una forma particolarissima delle leggi aerodinamiche, ma semplicemente al fatto che si deve convessificare una regione inizialmente non convessa. Questa è la stessa ragione per cui l'anisotropia (dipendenza dall'orientazione) di molti materiali, in particolare materiali con

⁹⁵ Un insieme si dice convesso se contiene sempre il segmento congiungente due suoi punti qualunque. L'inviluppo convesso di un insieme E è il più piccolo insieme convesso che contiene E .

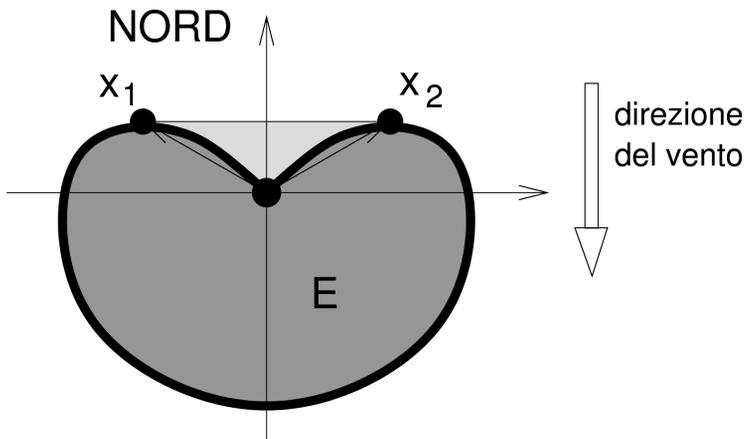


Figura 7. La zona raggiungibile da una barca a vela in presenza di vento da nord.

struttura cristallina, presenta ampi tratti rettilinei. In molti casi si ha in effetti un'anisotropia descritta da un insieme poligonale (poliedrale in dimensione 3).

3. Il calcolo numerico

La maggior parte dei problemi che nascono da situazioni concrete, compresi quelli descritti nelle pagine precedenti, pur rappresentabili in qualche modo con modelli matematici, non sono risolvibili in modo esatto tramite manipolazioni formali, cioè non ne conosciamo una formula risolvente. Per trovare comunque una soluzione, seppur approssimata, ci viene in aiuto la grande capacità di calcolo del computer. In effetti in situazioni concrete non è essenziale saper scrivere in modo rigorosamente esatto la soluzione, ma basterà un'approssimazione sufficientemente accurata.

Dal punto di vista matematico il modello comprende equazioni differenziali che legano tra loro una o più quantità, variabili da punto a punto nel dominio del problema, che rappresentano lo stato del sistema. Le variabili in gioco sono le coordinate spaziali e, per problemi di tipo evolutivo, il tempo. Queste si suppongono continue, ovvero rappresentabili con numeri reali. Il dominio in cui è definito il problema corrisponde all'oggetto che si vuole studiare; la forma del dominio gioca un ruolo importante ed è uno dei dati del problema.

Il processo risolutivo consiste nel sostituire il modello matematico iniziale (chiamato *modello continuo*) con una sua opportuna approssimazione (*modello discreto*). Quest'operazione introduce inevitabilmente un errore che viene chiamato *errore di discretizzazione*. Quantificare, o meglio fornire una stima dell'errore di discretizzazione così commesso è tanto importante quanto la risoluzione approssimata stessa, il che giustifica appieno l'esigenza di uno studio matematico accurato del processo di discretizzazione.

Riprendendo l'esempio delle bolle di sapone potremmo pensare alla superficie di una pellicola di sapone come composta da tanti piccoli triangolini piatti incollati tra di loro. Cercheremo poi la configurazione di minima energia (minima estensione nel caso delle bolle di sapone) nell'ambito di tali superfici sfaccettate. Il problema assume subito un aspetto completamente diverso: al posto di un numero infinito di gradi di libertà, cioè di informazioni che è necessario indicare per descrivere in modo completo una superficie, ora abbiamo un numero finito, anche se molto grande, di triangolini di cui dobbiamo individuare la posizione dei vertici. In altre parole ora la

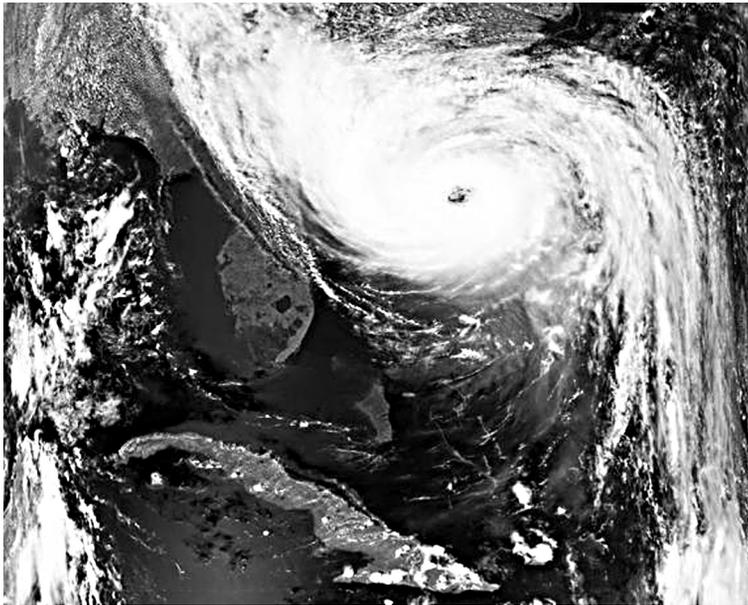


Figura 8. Immagine meteorologica dell'uragano Hugo.

soluzione viene descritta da un numero finito di quantità numeriche che dovranno soddisfare un sistema di equazioni di grande dimensione, e questo è pane per i denti di un computer. Quanto maggiore è la bontà dell'approssimazione desiderata, tanto più piccoli dovranno essere i triangolini con cui dobbiamo suddividere la superficie e di conseguenza aumenterà la dimensione del problema da risolvere.

La tecnica illustrata per l'esempio delle bolle di sapone può essere adattata a tutti i problemi che portano a modelli matematici in qualche modo simili, nei quali la somiglianza va intesa rispetto alle equazioni matematiche coinvolte, non rispetto al problema concreto da risolvere. Uno dei punti di forza della matematica consiste infatti nella capacità di mettere in luce somiglianze nascoste in realtà apparentemente del tutto diverse. L'approccio appena illustrato si applica in innumerevoli ambiti a svariati problemi concreti, tra cui per esempio:

- previsioni meteorologiche (Figura 8);
- resistenza di strutture in campo edilizio;
- ricostruzione di immagini mediche a partire da dati sperimentali di vario tipo (TAC, NMR, Figura 9);
- studio delle caratteristiche di nuovi materiali in ambito industriale, medico o sportivo (Figura 10);

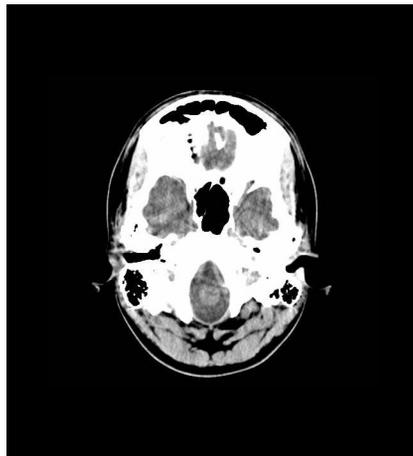


Figura 9. Immagine di una tomografia computerizzata del cervello.

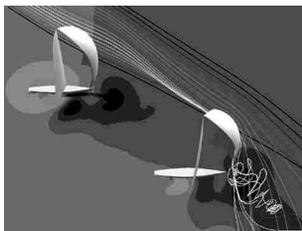


Figura 10. La vittoria del team svizzero di Alinghi nella Coppa America 2003 è anche merito delle simulazioni numeriche del professor Alfio Quarteroni e del suo gruppo. Si veda anche il sito <http://mox.polimi.it>.

- simulazione di realtà complesse come alternativa a costosi o indesiderabili esperimenti (galleria del vento, operazioni chirurgiche, esperimenti su cavie ecc.).

Gli esempi citati hanno una caratteristica in comune, che influenza poi il processo di modellizzazione: le quantità che si vogliono calcolare corrispondono a caratteristiche che variano da punto a punto e in alcuni casi anche nel tempo, come, per esempio, lo spostamento dalla posizione di equilibrio di un materiale elastico, la velocità e la pressione di un fluido, la temperatura di un corpo. In tutti i casi si suppone che l'oggetto di studio sia un continuum spaziale (e temporale), vale a dire che non ci sia una qualche *granularità di fondo*. Tale assunto viene posto pragmaticamente nel processo di modellizzazione allo scopo di ottenere un modello matematico che sia al contempo abbastanza vicino alla realtà da studiare e sufficientemente semplice da permettere di essere studiato e risolto.

Non si pretende di ottenere una descrizione esatta della realtà, anzi in molti casi è palese il contrario: per esempio, i modelli per la fluidodinamica non tengono conto del fatto che il fluido, liquido o gas, è composto da molecole; lo stesso avviene per la teoria dell'elasticità; in ambito biologico in molti casi non si considera il comportamento delle singole cellule, ma solo di quantità mediate su una scala più grande. Le quantità medie che si ottengono (temperatura, pressione ecc.) sono tipicamente legate tra loro nel modello attraverso operatori differenziali, che ne indicano le variazioni; le equazioni che ne derivano prendono il nome di equazioni alle derivate parziali.

Come già accennato, un modello matematico di questo tipo, benché si presti bene a una prima analisi di tipo matematico, non permette in molti casi di ottenere una soluzione esplicita. L'approssimazione numerica consiste allora nel sostituire gli operatori differenziali con una loro versione discreta, allo stesso modo con cui si può approssimare una velocità istantanea con una velocità media, permettendo il calcolo esplicito di una soluzione approssimata al prezzo di un'enorme quantità di calcoli da fare. Nel gergo matematico si dice che il modello matematico è ambientato in spazi (i cui punti rappresentano intere funzioni) di dimensione infinita, mentre la sua versione approssimata si colloca in spazi di dimensione finita e risulta quindi numericamente trattabile.

3.1 Dal discreto al continuo e ritorno

Schematicamente possiamo decomporre il processo che porta alla soluzione (approssimata) di un problema concreto in una prima fase in cui la realtà viene matematizzata con la definizione di un opportuno modello matematico seguita dal processo di discretizzazione numerica appena descritto.

Nella prima fase si opera un processo di livellamento che permette di nascondere i dettagli microscopici e ottenere una descrizione macroscopica tramite quantità che descrivono lo stato del sistema. In molti casi la realtà microscopica è di tipo discreto (atomi, molecole, cellule ecc.). Il modello matematico continuo consiste in una specie di omogeneizzazione: si perde un po' nella accuratezza del modello a vantaggio di una descrizione molto più sintetica. D'altra parte l'effettiva soluzione del problema matematico richiede come abbiamo visto un processo di discretizzazione, che comporta apparentemente un ritorno alla situazione iniziale. Ci si domanda allora se non sia il caso di abbandonare del tutto la fase intermedia del modello matematico continuo cercando di descrivere direttamente la realtà discreta. Recentemente questo punto di vista è diventato piuttosto di moda: gli automi cellulari sono un esempio. Ci sono situazioni in cui questo approccio integralmente discreto può essere vantaggioso, tuttavia c'è il rischio di perdere di vista aspetti importanti della realtà a favore di una sua visione eccessivamente dettagliata. In effetti è possibile collegare la realtà discreta a una sua approssimazione anch'essa discreta, ma su una scala molto dissimile, o basata su una struttura spazio-temporale differente, solo osservando che si ha a che fare con due approssimazioni diverse dello stesso modello continuo.

4. Elaborazione di immagini

Le immagini possono avere varie origini e l'elaborazione richiesta può essere molto diversificata⁹⁶. L'utilizzo di computer e tecniche di elaborazione anche molto sofisticate è essenziale per questo tipo di compiti. Ecco alcuni esempi.

- Immagini prese da telecamere (autovelox, zone a traffico limitato). In questo caso ci si può aspettare un ambiente noto: per esempio, uno sfondo sempre uguale che può rappresentare una strada o un incrocio, con il soggetto (per esempio l'auto) in una posizione grosso modo nota a priori e con una orientazione anch'essa sostanzialmente prestabilita. Abbiamo perciò un certo bagaglio di conoscenze a priori, che rivestono una notevole importanza nell'agevolare il lavoro di analisi. Nel caso in esame, le informazioni che si vogliono estrarre sono molto specifiche: per esempio, è richiesta la rilevazione della targa dell'autoveicolo.
- Fotografie di un dipinto, di una scultura o in generale di un'opera d'arte in condizioni di degrado. In questo caso una richiesta tipica consiste nel restaurare l'immagine, cercando di ricostruire le zone mancanti (colore rovinato, intonaco scrostato...) e di recuperare l'aspetto originale.
- Immagini generiche da classificare. Può essere il caso di un motore di ricerca in Internet che deve indicizzare le immagini trovate in base al loro contenuto. In questo caso viene richiesto un lavoro più vago di riconoscimento del contenuto dell'immagine.

Identificheremo l'immagine da analizzare, che per semplicità supporremo in bianco e nero, con una funzione $g : \Omega \rightarrow [0, 1]$ rappresentante il valore dell'intensità luminosa (livello di grigio) e definita su un dominio $\Omega \subseteq \mathbb{R}^2$ che solitamente è un rettangolo. L'azione richiesta è estremamente variegata in base alle specifiche esigenze. In ogni caso comunque possiamo mettere in evidenza alcuni aspetti ricorrenti.

- L'azione richiesta non è mai formalizzabile in modo completo e non ambiguo. Buona parte delle scelte viene lasciata a tentativi basati su parametri che possono essere scelti in modo alquanto arbitrario. Entra in gioco in modo rilevante anche la cosiddetta intelligenza artificiale, con l'intento di cercare di imitare in qualche modo i meccanismi con cui il nostro cervello elabora le informazioni ricevute dagli occhi.

⁹⁶ Con il termine "elaborazione" intendiamo qui vari aspetti legati all'analisi di immagini.

- Una parte di analisi quasi sempre richiesta è l'estrazione dei contorni delle figure e la segmentazione dell'immagine. Anche azioni apparentemente banali come il ritaglio della silhouette di una persona, un oggetto, un volto, azioni eseguite dal nostro cervello senza che ce ne rendiamo conto, risultano estremamente complesse da automatizzare. La semplice estrazione di linee in cui l'intensità luminosa varia repentinamente non è sufficiente a questo scopo, poiché entrano in gioco effetti visivi, fisici e fisiologici che complicano molto la questione; per esempio, ombre, occlusione di parti di oggetti da parte di altri oggetti, presenza di spigoli vivi (come gli angoli delle case), cambiamenti repentini nella colorazione. Non mancano poi situazioni effettivamente ambigue, dove non è possibile un'interpretazione univoca della scena⁹⁷.

L'estrazione di contorni può essere affrontata con tecniche di tipo locale, analizzando per esempio, la variazione di intensità luminosa nelle diverse zone dell'immagine, oppure con un approccio globale in cui si cerca di riconoscere o ritagliare i vari oggetti della scena, ricostruendola tramite una o più funzioni e classificando con un indice numerico la bontà della identificazione effettuata. Nella costruzione di tale indice entrano in gioco vari fattori, tra cui la corrispondenza dei contorni degli oggetti ricostruiti con le discontinuità di g (intensità luminosa) e una buona corrispondenza tra g e l'intensità luminosa u dell'immagine ricostruita. Nell'immagine ricostruita risultano penalizzate regioni dove ci sono forti variazioni di u che non corrispondono a bordi di oggetti. È possibile costruire vari modelli, compresi alcuni in cui si cerca di ricostruire in qualche modo i contorni di oggetti occlusi da altri. Alla fine si ottiene un problema di ottimizzazione simile a quelli già discussi in precedenza.

Quest'approccio variazionale comporta in genere un impegno computazionale estremamente elevato; si tratta comunque di modelli tutt'ora in corso di studio e di miglioramento. Un approccio pragmatico al problema tende a utilizzare una combinazione di entrambi i metodi (locale e globale), unitamente a una buona dose di buon senso e di successive approssimazioni. Ci sono sicuramente ambiti specializzati in cui si è riusciti a raggiungere un buon livello di riconoscimento, basti pensare ai lettori ottici di codici a barre, alla lettura

⁹⁷ Le ben note illusioni ottiche nascono da una esasperazione voluta di queste situazioni di ambiguità nell'interpretazione della scena.

ra ottica di moduli, all'utilizzo in robot industriali impiegati nelle catene di montaggio.

5. Sintesi di immagini e realtà virtuale

Per certi versi la sintesi di un'immagine è l'operazione contraria dell'analisi descritta nella sezione precedente. Si vuole in questo caso ottenere un'immagine quanto più realistica possibile a partire da una scena tridimensionale descritta in modo quantitativo completo.

Un ambito di applicazione tipico si trova nella progettazione ingegneristica (CAD/CAM: *computer aided design/computer aided modelling*). In questo caso si desidera avere (possibilmente in modo interattivo) una visione tridimensionale dell'oggetto che si sta progettando sufficientemente accurata da permettere un immediato impatto visivo. L'industria cinematografica è un altro ambito tipico di applicazione, che porta all'estremo le tecniche di realtà virtuale. Molti film recenti contengono parti ottenute in modo sintetico: gli effetti speciali, che in passato venivano ottenuti utilizzando modellini in scala, trucchi cinematografici, scenografie finte, sono oggi quasi invariabilmente realizzati tramite elaborazione a computer delle scene interessate. È anzi sempre più frequente la produzione di film interamente costruiti in modo sintetico. Citiamo infine l'ambito dei videogiochi. Un moderno gioco su computer richiede la rappresentazione accurata sullo schermo di una scena in tre dimensioni di ciò che il giocatore vede. Si desidera una rappresentazione realistica della scena mantenendo però una velocità di elaborazione che permetta l'interattività del gioco. Nel caso del CAD/CAM è sufficiente avere una visione stilizzata della scena ricostruita; ciò permette di avere dei tempi di elaborazione sufficientemente rapidi da consentire la manipolazione interattiva dell'oggetto rappresentato. Per esempio, in una finestra del computer compare la scena in 3D sintetizzata con possibilità di cambiare interattivamente il punto e la direzione di vista. In altre finestre ci potranno essere rappresentazioni schematiche (piante, prospetti e quant'altro) o anche ulteriori viste in 3D.

Nel caso di scene virtuali si utilizzano tecniche di *shading*: si calcola per ogni oggetto la sua proiezione prospettica e si colorano i pixel con un'intensità luminosa che dipende da varie quantità, tra cui l'inclinazione della superficie rispetto al punto di vista e alle sorgenti luminose, oltre alle caratteristiche della superficie stessa.

Per l'utilizzo in ambito cinematografico o artistico è invece necessario ottenere un realismo ancora maggiore: per esempio, tenendo conto delle ombre, di caratteristiche particolari delle superfici: riflessione, rifrazione, diffusione e iridescenza. La superficie può essere liscia o ruvida e quindi diffondere la luce in modo diverso in funzione dell'angolo di incidenza. Per converso si è disposti a spendere tempi di elaborazione anche estremamente lunghi per ottenere il risultato desiderato.

Analizzeremo nel seguito un po' più in dettaglio la tecnica di *raytracing* (tracciamento dei raggi) utilizzata in genere per ottenere immagini molto realistiche. Un software libero⁹⁸ che si basa su questa tecnica è PovRay⁹⁹.

Nella Figura 11 troviamo un'immagine ottenuta dagli autori con PovRay. L'idea di fondo è quella di seguire in senso inverso i raggi di luce che arrivano all'occhio o alla pellicola fotografica. In altre parole si considera ogni punto della (immaginaria) pellicola fotografica (ciascun pixel dell'immagine sintetizzata) e si cerca di risalire lungo il raggio luminoso corrispondente, seguendo una particolare direzione a partire dalla posizione della macchina fotografica fino a intersecare il primo oggetto che si incontra nella scena.

A questo punto si tratta di capire quale intensità luminosa (e colore) emette tale oggetto in quel punto e in quella particolare direzione; questo dipenderà da vari fattori, i più importanti dei quali sono la posizione delle sorgenti luminose e la posizione di eventuali altri oggetti che si interpongono. Entrano in gioco anche le caratteristiche della superficie: se la superficie è parzialmente riflettente, si dovrà studiare in particolare il raggio di luce che incide la superficie con l'angolo giusto. La scena viene descritta utiliz-

⁹⁸ Il movimento del software libero (free software), fondato da Richard Stallman nel 1985 comprende programmi per computer che garantiscono all'utilizzatore le quattro libertà:

1. di eseguire il programma, per qualsiasi scopo (libertà 0);
2. di studiare come funziona il programma e adattarlo alle proprie necessità (libertà 1)
3. di ridistribuire copie in modo da aiutare il prossimo (libertà 2);
4. di migliorare il programma e distribuirne pubblicamente i miglioramenti (e le versioni modificate in genere), in modo tale che tutta la comunità ne tragga beneficio (libertà 3).

La maggior parte del software libero è disponibile gratuitamente in Internet, ma non va confuso con il freeware (software gratuito). Conseguenza importante delle libertà 1 e 3 è la possibilità dell'utilizzatore di accedere, se lo desidera, al codice sorgente, cioè al *progetto* del software, agevolando così un'estesa attività collaborativa. L'obbligo di permettere l'accesso al codice sorgente ha motivato l'uso del più pragmatico termine *open source* (sorgente aperta) come alternativa a free software. Benché per certi versi siano sovrapponibili, le due accezioni differiscono nelle premesse etiche e filosofiche. Per maggiori dettagli facciamo riferimento a www.fsf.org.

⁹⁹ Disponibile in Internet all'indirizzo www.povray.org.



Figura 11. Esempio di immagine sintetizzata utilizzando PovRay.

zando un insieme di primitive, ossia di costrutti semplici, per la definizione degli oggetti e immaginando lo spazio tridimensionale dotato di tre coordinate spaziali x , y , z , che nel caso di PovRay sono disposte convenzionalmente in modo da misurare rispettivamente posizioni in orizzontale (coordinata x), in verticale (coordinata y) e in profondità (coordinata z). È anche necessario indicare posizione e orientamento della macchina fotografica oltre alla posizione e caratteristiche delle varie sorgenti di luce. Per fare un esempio, consideriamo un pixel della Figura 11 corrispondente al basamento cilindrico nero; tale oggetto viene definito nella descrizione della scena con la primitiva:

```
#declare updmf=2;
[...]
cylinder {
  <0,0,0>
  <0,updmf,0>,2.0
  texture {pigment {Black}}
  finish {
    ambient 0.1
    diffuse 0.6
    phong 1
    phong_size 100
    reflection 0.25
  }
}
```

in cui è descritto l'oggetto come un cilindro avente i centri dei due cerchi di base rispettivamente nell'origine e in una posizione a quota "updmf" ($= 2$) e aventi raggio 2^{100} . La parte seguente descrive il tipo di materiale di cui è composta la superficie del cilindro, in particolare vi troviamo l'informazione sul colore (uniformemente nero, in questo caso) seguita da caratteristiche fisiche della superficie, come la riflettività, la quantità di luce diffusa e la presenza di effetto Phong¹⁰¹, che provoca le caratteristiche macchie luminose nelle zone dell'oggetto dove si riceve direttamente la luce riflessa da una delle sorgenti luminose. Il "pavimento" della scena è composto da due piastrelle di forma e materiale diversi; esse sono disposte secondo una famosa piastrellatura non periodica ideata da Roger Penrose¹⁰² (si veda anche la figura 9 del capitolo "La matematica nell'arte"). Le forme ricordano quelle di un aquilone (piastrelle di marmo chiaro) e di una freccia (piastrelle scure). I due tipi di piastrelle sono descritti una sola volta e individuati dai nomi *kitem* e *dartm*, dopodiché ogni istanza viene generata con una sequenza del tipo:

```
object { kitem
  rotate rot*y
  translate tras
}
```

dove l'oggetto di riferimento *kitem* viene opportunamente ruotato attorno all'asse *y* (verticale) e traslato del vettore *tras*. Il lettore interessato può trovare la descrizione completa della scena all'indirizzo <http://dmf.unicatt.it/~paolini/penrose/>.

Siccome un'animazione non è altro che una sequenza di fotogrammi, la tecnica di *raytracing* può essere utilizzata anche per la costruzione di animazioni; in questo caso nella descrizione della scena verrà utilizzata la variabile particolare "clock" indicante l'istante di tempo a cui deve corrispondere la scena man mano generata. Il processo di generazione dei fotogrammi è troppo lento per permetterne l'utilizzo in tempo reale, ma risulta conveniente nell'ambito della

¹⁰⁰ L'uso di una variabile, updmf, permette di riutilizzare lo stesso valore in vari punti nella descrizione della scena. Nel nostro caso sarà utilizzata anche come quota iniziale per il tronco di cono collocato sopra il cilindro.

¹⁰¹ Dal nome dell'ideatore Bui Tuong Phong dell'Università dell'Utah.

¹⁰² Roger Penrose (Colchester, 1931), fisico, matematico e filosofo britannico. Ha dato importanti contributi alla cosmologia, si occupa di giochi matematici ed è autore di vari libri divulgativi.

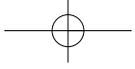
produzione cinematografica, dove i tempi di produzione di un film hanno una durata di parecchi mesi.

In ambito cinematografico sono utilizzate dapprima tecniche tipiche del CAD/CAM per ottenere in tempi rapidi e in modo interattivo una buona descrizione della scena di cui si vuole il *rendering*, dopodiché il tutto viene dato in pasto a una batteria di supercomputer che si spartiscono il compito di generare i singoli fotogrammi in alta qualità. Va notato che il lavoro di rendering dei fotogrammi è perfettamente parallelizzabile (ci può essere addirittura un computer per ciascun singolo fotogramma). Per dare un'idea della mole di calcolo richiesto per la produzione di un film di animazione citiamo da *Wikipedia* un paragrafo riguardante *Shrek 2* (prodotto dalla DreamWorks Pictures, e campione di incassi tra i film di animazione, seguito da *Alla ricerca di Nemo* della Pixar):

“Per realizzare le animazioni sono state impiegate oltre 330 workstation HP con sistema operativo Linux¹⁰³, mentre per la composizione finale delle scene digitali sono stati utilizzati 780 server sempre HP con Linux.”

Anche PovRay può essere utilizzato per la costruzione di animazioni. Citiamo un interessante esempio matematico ottenuto da Jos Leys, Etienne Ghys e Aurélien Alvarez (École Normale Supérieure di Lyon, Francia): il video di 2 ore *Dimensions* da essi prodotto può essere scaricato liberamente dal sito www.dimensions-math.org e illustra in modo perfetto le possibilità offerte dalla tecnica di *raytracing*.

¹⁰³ Il sistema operativo GNU/Linux è il più noto esempio di software libero. Il progetto GNU (acronimo ricorsivo che sta per “GNU is Not Unix”) di Stallman prevedeva inizialmente lo sviluppo di un ambiente di lavoro completo basato sul software libero, tuttavia il kernel (la parte centrale del sistema operativo) non era ancora pronto quando nel 1991 Linus Torvalds dell’Università di Helsinki pubblicò la prima versione del kernel Linux. Pur non essendo parte del progetto GNU, Linux è un software libero; dal connubio del nuovo kernel di Torvalds e la grande quantità di applicativi già sviluppati nell’ambito del progetto GNU è nato l’ambiente GNU/Linux.



LA MATEMATICA NELL'ARTE

1. La matematica nell'architettura e nelle arti figurative

Se si pensa che l'arte sia una pura manifestazione di creatività, indipendente da regole e deduzioni logiche, si stenterà a vedere un rapporto fra la matematica e le arti figurative o la musica. In realtà le cose stanno in modo ben diverso: in particolare nell'antichità e nel Rinascimento, le arti figurative seguivano dettami e canoni ben precisi e, anche se un artista non dimostrava teoremi, non avrebbe avuto dubbi sul profondo legame tra le forme geometriche trattate e studiate dalla matematica e le forme fondamentali del suo operare. Cerchi, rette, poligoni, piani, sfere, ma senza dubbio anche ovali e coniche erano note agli artisti, ma anche il concetto di simmetria e l'inevitabile legame con il "bello" che questo concetto porta con sé. Fregi e decorazioni, sebbene non da tutti ritenuti opere d'arte per la loro ripetitività, sono intrinsecamente anche oggetti matematici, e la matematica che li regola è tutt'altro che banale.

Vogliamo trattare qui solo alcuni punti di contatto tra matematica e arte: essere esaustivi sarebbe veramente difficile e ci porterebbe troppo lontano. Per esempio, non tratteremo dell'arte rinascimentale nel senso che riguarda la composizione di figure sacre, che spesso sono collocate con precise regole nelle quali certe figure geometriche (basta pensare al triangolo) ricoprono ruoli fondamentali. Vogliamo partire da un semplice rettangolo. Esso avrà un lato più lungo e un lato più corto, per cui sarà sempre possibile ricavare un quadrato dal lato più corto all'interno del rettangolo, toglierlo e osservare il rettangolo rimanente. Se questo rettangolo è simile al rettangolo iniziale, chiameremo il rettangolo iniziale aureo.

Di rettangoli aurei ve ne sono infiniti, perché se ora ricaviamo nuovamente dal rettangolo più piccolo un quadrato, resterà un rettangolo ancora più piccolo ma simile al primo, e così via. però il rapporto fra i lati di un rettangolo aureo non può essere qualunque, e ha un valore ben noto ai matematici. Vediamo di calcolarlo.

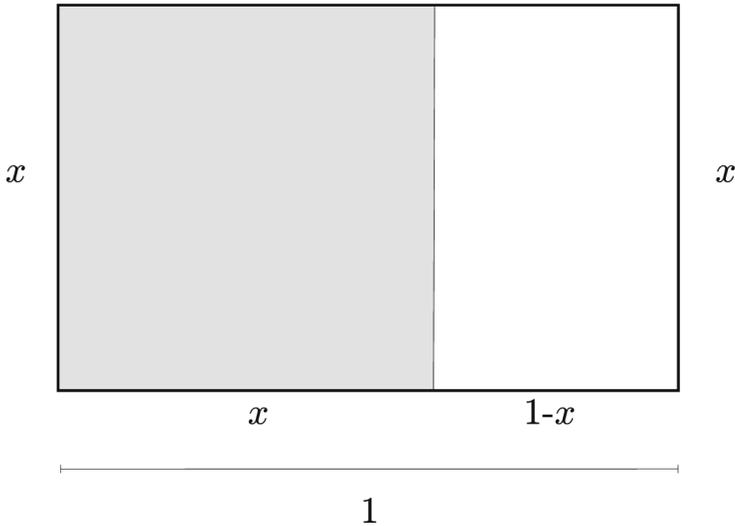


Figura 1. Costruzione di un rettangolo aureo.

Siccome non ci interessano figure simili, possiamo sempre pensare che il primo rettangolo abbia lato lungo pari a uno. Chiamiamo x allora il lato corto del primo rettangolo (vedi Figura 1); dalla figura si vede subito che il lato corto del secondo rettangolo è pari a $1 - x$, e che quello lungo misura x . Se essi devono essere simili, deve succedere che i rapporti dei lati coincidono, ossia

$$\frac{1}{x} = \frac{x}{1-x}$$

Da questa relazione, siccome x non è ovviamente né zero né uno, si ricava

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

Siccome x deve essere positivo, esso vale

$$x = \frac{\sqrt{5} - 1}{2} \approx 0.618033$$

Questo numero x , indicato spesso con φ , è ben noto ai matematici: è legato alla celebre successione di Fibonacci

1, 1, 2, 3, 5, 8, 13, 21, ...

ottenuta partendo con la coppia (1, 1) e calcolando ogni successivo valore come somma degli ultimi due trovati ($1 + 2 = 3$, $2 + 3 = 5$ ecc.) nel seguente modo: il rapporto fra due successivi valori della successione si avvicina sempre di più a φ .



Figura 2. Il Partenone e i rettangoli aurei in esso contenuti.

Perché dovrebbe interessare un rettangolo aureo? In effetti non si sa, ma in numerose opere dell'architettura e dell'arte si ritrova riprodotto in modo così evidente da non lasciare adito a dubbi. Consideriamo la Figura 2: non si può dubitare che nella costruzione di questo importante edificio il rettangolo aureo abbia giocato un ruolo fondamentale. Sappiamo poi bene che i greci erano molto più interessati ai rapporti fra le grandezze di quanto non lo siamo noi oggi, che abbiamo un'unità di misura ben definita e stabile; il fatto, quindi, di aver usato il rapporto aureo, oltretutto irrazionale, sta a significare che il rettangolo aureo fosse collegato in qualche modo non banale con il concetto di bello ed elegante.

Facendo un salto nei secoli, un altro interessante esempio di uso del rettangolo aureo è nella Figura 3, che nessuno faticherà a riconoscere. Qui è più difficile esser certi della veridicità dell'affermazione che Leonardo da Vinci volesse dare al viso della Gioconda le pro-

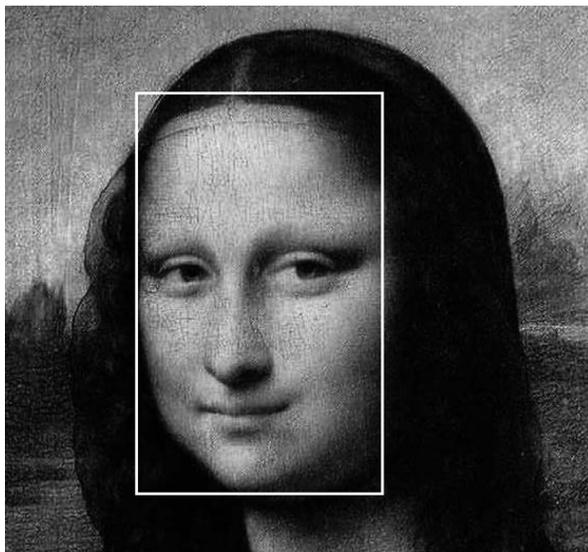


Figura 3. Il viso della Gioconda segue approssimativamente un rettangolo aureo.

porzioni di un rettangolo aureo; in effetti, osservando bene la figura, non è proprio esatto. Certamente, comunque, Leonardo conosceva la sezione aurea, per cui il fatto che la Gioconda abbia un rapporto anche solo “quasi” aureo nel suo viso potrebbe, in definitiva, confermare il legame fra questo numero e la bellezza.

Tra le ricerche, invece, poco plausibili, si devono annoverare tutte quelle che si prefiggono di trovare rapporti aurei in parti del corpo in natura. In questo ambito, infatti, vi è una notevole variabilità nelle misure e nelle proporzioni, che solo un’indagine statistica può descrivere in maniera accurata (conformemente a quanto detto per la statura nel capitolo “La gestione razionale del caso”), ma in aggiunta è difficile osservare il rapporto aureo con sufficiente precisione: se infatti il rapporto, per esempio, tra distanza fra gomito e punta delle dita e fra polso e punta delle dita (come per chi scrive) è 0.74, non è aureo. Trovare un attore famoso che abbia questo rapporto pari a 0.618 è pura tautologia, in quanto si definisce “bello” ciò che in realtà si vuole dimostrare che è bello.

Invece, molto probabilmente aurea è la scelta “asimmetrica” della torre del Palazzo della Signoria a Firenze, come mostra la figura 4, che divide il rettangolo fra terra e l’inizio dei merli esattamente come



Figura 4. La collocazione della torre campanaria del Palazzo della Signoria a Firenze e il rettangolo aureo.

abbiamo fatto nella Figura 1. Consco di questo significato è anche un pittore dell'arte moderna, Piet Mondrian¹⁰⁴, che “cita” nelle sue opere il rettangolo aureo come esempio dell'importanza delle forme nella sua concezione pittorica. Un esempio tipico (riportato in bianco e nero) è la composizione della Figura 5, nella quale la scelta del rettangolo aureo è evidente.

Il calcolo dei rettangoli aurei non è stato l'unico contatto dei pittori classici con la matematica; un altro, forse più legato alla loro pratica e tecnica di disegno, è la prospettiva. Chiunque abbia provato a comporre un disegno tecnico in prospettiva sa che vi sono delle costruzioni geometriche non banali, e che chi le ha inventate è stato sicuramente ben ferrato in matematica. Un buon testimone di rapporti tra arte e matematica è Piero della Francesca¹⁰⁵. Anche se è riconosciu-

¹⁰⁴ Piet Mondrian (Amersfoort, 1872 – New York, 1944), pittore olandese. Esponente dell'astrattismo, impostò le sue opere sugli elementi del piano, della linea e del colore. Teorizzò il neoplasticismo.

¹⁰⁵ Piero della Francesca (Sansepolcro, 1415/1420 – Sansepolcro, 1492), pittore italiano. Figura centrale del '400, fu anche teorico della pittura (*De quinque corporibus regularibus*, 1480 ca., *De prospectiva pingendi*, 1490 ca.).

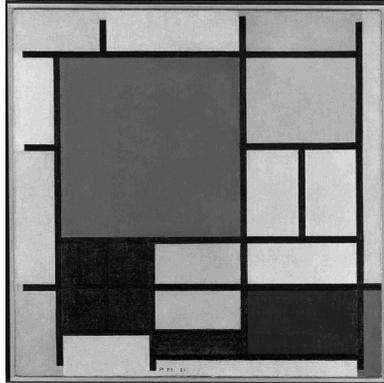


Figura 5. P. Mondrian, *Composition with Red, Yellow, Blue and Black* (1921).

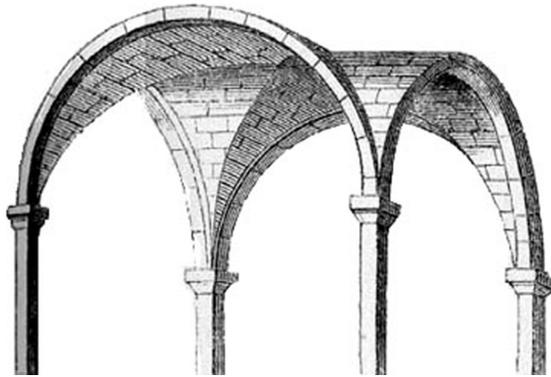


Figura 6. Piero della Francesca calcolò la superficie esatta della volta a crociera.

to, giustamente, molto di più per la sua produzione pittorica che matematica, Piero della Francesca si interessò della matematica al livello dei più importanti studiosi della sua epoca, studiando e risolvendo equazioni di terzo e quarto grado (la cui formula risolutiva generale doveva essere scoperta e pubblicata di lì a un secolo circa), problemi non banali di geometria, come il calcolo del volume e della superficie della volta a crociera (la superficie che si ottiene intersecando due cilindri ortogonali con lo stesso raggio e aventi gli assi di simmetria incidenti, si veda la Figura 6), un problema che richiede il calcolo integrale, solitamente affrontato nei corsi universitari.

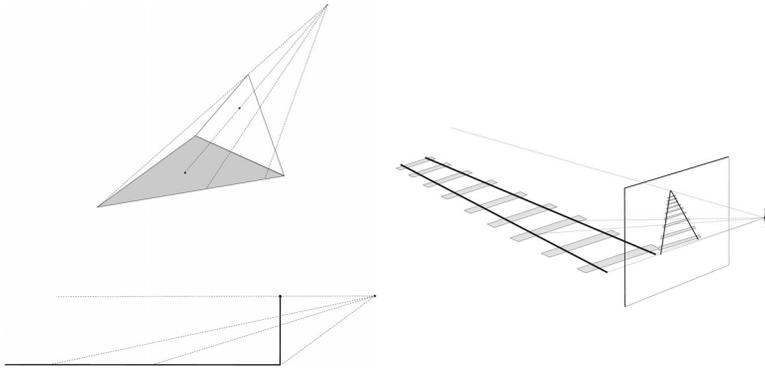


Figura 7. Deformazione dei rapporti di lunghezza in una proiezione (a sinistra). Il “punto all’infinito” appare nella prospettiva di un binario (a destra).

Tra l’altro, la sistematizzazione della prospettiva come scienza, dovuta poi anche all’architetto Leon Battista Alberti¹⁰⁶, a Leonardo da Vinci e a numerosi altri tra matematici e artisti (tra cui dobbiamo citare l’incisore Albrecht Dürer¹⁰⁷), diede impulso anche alla matematica pura.

Nella similitudine, infatti, si conservano i rapporti fra le lunghezze dei segmenti, mentre se proiettiamo un segmento su un altro segmento (Figura 7, a sinistra), questi rapporti non si conservano. Eppure non è possibile trasformare una figura in un’altra figura qualsiasi per proiezione da un punto, per cui, argomentarono i matematici, qualcosa si deve conservare. Ci volle del tempo per scoprire questa quantità, anche a causa di un altro fatto: se per esempio proiettiamo un segmento su un piano perpendicolare a esso da un punto situato alla stessa quota dell’estremo del segmento (Figura 7, a sinistra), osserviamo che il segmento si trasforma in una semiretta. Dunque, per proiezione, le lunghezze possono anche divenire “infinite”. Ma c’è di più: se proiettiamo su un piano, come accade nella visione comune, due rette parallele, quali due binari (Figura 7, a destra), otteniamo due rette incidenti, private del punto di incidenza perché corrisponde al “punto all’infinito”, il punto di fuga, appunto, della prospettiva.

¹⁰⁶ Leon Battista Alberti (Genova, 1404 – Roma, 1472), architetto e umanista italiano, fu anche autore di trattati teorici di architettura, scultura e pittura.

¹⁰⁷ Albrecht Dürer (Norimberga, 1471 – Norimberga, 1528), pittore e incisore tedesco. Affianco alla sua produzione figurativa trattati sulla prospettiva.

Il fatto che un punto “sparisse” e “riapparisse” per proiezione indusse il matematico francese Girard Desargues¹⁰⁸ a ritenere come “incidenti” anche due rette parallele e, equivalentemente, di accogliere fra i “punti” anche i punti “all’infinito”. Con questa modifica della definizione di incidenza, la geometria acquisiva una simmetria che era sfuggita per più di venti secoli: per due punti passa sempre una e una sola retta e due rette si intersecano sempre in uno e un solo punto¹⁰⁹. Grazie a questa modifica i teoremi della geometria diventavano “simmetrici” fra punto e retta e fra “condurre” e “intersecare” (si tratta della cosiddetta *dualità*). La nuova geometria di Desargues venne chiamata, a buon titolo, geometria proiettiva.

Il matematico Michel Chasles¹¹⁰ scoprì poi che la famosa “quantità conservata” per proiezione era il “birapporto” di quattro punti allineati

$$(ABCD) = \frac{\overline{CA} \cdot \overline{DB}}{\overline{CB} \cdot \overline{DA}}$$

(cosa, già nota al matematico greco Pappo di Alessandria¹¹¹, ma senza l’inquadramento moderno in termini di prospettiva). La geometria proiettiva ha da allora approfondito i suoi studi ed è ancora una disciplina matematica in evoluzione.

Un altro ambito naturale di contatto fra matematica e arte è, come dicevamo all’inizio, quello della simmetria. Molte persone sono convinte dell’esistenza di un profondo legame fra le simmetrie e la bellezza, e uno dei modi più comuni di osservarla si trova nei fregi e nei pavimenti.

Se abbiamo un pavimento fatto con mattonelle quadrate (supposto infinito), esso ammette molte simmetrie: per esempio, è possibile traslare tutto il piano di una mattonella e ritrovarlo identico (si intende, perché il pavimento è supposto infinito), ma anche ruotandolo di 90°, oppure ribaltandolo rispetto agli assi o alle diagonali delle mattonelle si ottiene lo stesso risultato.

¹⁰⁸ Girard Desargues (Lione, 1591 – Lione, 1661), matematico francese, considerato uno dei fondatori della geometria proiettiva.

¹⁰⁹ Desargues dovette, ovviamente, anche accogliere come “retta” l’insieme di tutti i punti all’infinito, ossia l’“orizzonte” della prospettiva, che chiamò retta all’infinito o retta propria.

¹¹⁰ Michel Chasles (Épernon, 1793 – Parigi, 1880), matematico francese, condusse importanti studi sulla geometria proiettiva.

¹¹¹ Pappo di Alessandria, uno dei più importanti matematici del periodo tardo ellenistico, a cui si devono alcune anticipazioni della geometria proiettiva. Intorno al 320 d.C. scrisse un commento all’*Almagesto* di Claudio Tolomeo.

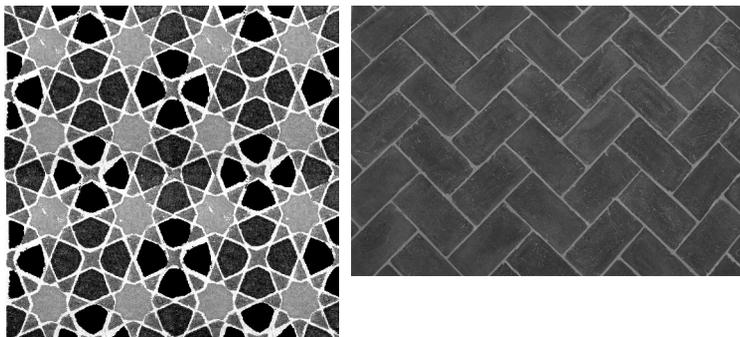


Figura 8. Una tassellatura dei pavimenti dell'Alhambra di Granada (a sinistra). Pavimento a "spina di pesce" (a destra).

La Figura 8, a sinistra, riproduce questo tipo di pavimento, ma in maniera più artistica. Esso si trova in uno dei più straordinari monumenti dell'arte araba, l'Alhambra di Granada. Se invece abbiamo un pavimento a "spina di pesce" (Figura 8, a destra) la situazione è diversa, perché stavolta ruotandolo di 90° non otterremo la stessa figura, per cui le sue simmetrie sono diverse. Analogo ragionamento si può fare per un muro di mattoni, che è diverso ancora.

Ma che c'entra la matematica, a parte le simmetrie? Il fatto è che due simmetrie, come due rotazioni, si possono comporre, ossia effettuare di seguito, e questa operazione di composizione, denotata con \circ , è, per motivi matematici generali, sempre associativa, ossia, indicando con $\sigma_1, \sigma_2, \sigma_3$ tre rotazioni che trasformano in sé il pavimento,

$$\sigma_1 \circ (\sigma_2 \circ \sigma_3) = (\sigma_1 \circ \sigma_2) \circ \sigma_3.$$

Per un matematico questo fatto è ricco d'interesse, in quanto permette di creare un'analogia fra quanto accade nei numeri, ambito originario di questo tipo di proprietà, e quanto accade per oggetti radicalmente diversi (almeno a prima vista, in certi casi: le rotazioni di 90° sono legate ai numeri un po' come le lancette di un orologio che segnano solo le 3, le 6, le 9 e le 12).

Le proprietà di questi "gruppi" di simmetrie affascinarono a tal punto i matematici da creare una disciplina, l'*algebra astratta*, che giocò e gioca ancora oggi un grande ruolo nella matematica e fuori di essa (per esempio, nella fisica teorica). Lo studio di queste simmetrie produsse numerosi risultati di notevole interesse.

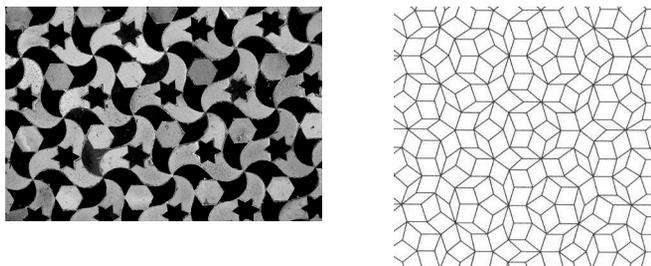


Figura 9. Una tassellatura a simmetria esagonale dei pavimenti dell'Alhambra di Granada (a sinistra). Il "pavimento quasi-periodico" di Penrose (a destra).

Per esempio, il geologo e cristallografo russo Evgraf Fëdorov¹¹² dimostrò che nel piano ci possono essere solo 17 di questi gruppi distinti, e tutti questi gruppi sono rappresentati nell'Alhambra!

La Figura 9, a sinistra, riporta un pavimento con simmetria esagonale. Si può dimostrare matematicamente che non è possibile costruire un pavimento periodico a simmetria pentagonale. Suscitò dunque molto scalpore nel mondo matematico la scoperta, a opera del matematico e fisico Roger Penrose, di un "pavimento" a simmetria pentagonale, ovviamente non periodico, ma "quasi-periodico" in un senso che qui è un po' difficile spiegare, ma che si capisce osservando la Figura 9, a destra.

Dai "pavimenti" (che si chiamano in realtà tassellature) era affascinato anche uno degli artisti più conosciuti e amati dai matematici, Maurits Cornelis Escher¹¹³. Le sue opere, sebbene non siano molto considerate in ambito artistico perché ritenute povere di uno stile originale per il suo tempo, rappresentano una pietra miliare nei rapporti fra matematica e arti figurative. Escher, che visitò l'Alhambra nel 1922, diede vita a numerose tassellature del piano con motivi, come il celebre disegno *Angeli e demoni* rappresentato in Figura 10, a sinistra. La matematica interessava Escher, benché si dichiarasse ignorante in materia, e negli anni più fecondi della sua vita, gli anni '50, egli venne in contatto con il matematico anglo-canadese Harold Coxeter¹¹⁴, il qua-

¹¹² Evgraf Stepanovič Fëdorov (Orenburg, 1853 – 1919), matematico e mineralogista russo.

¹¹³ Maurits Cornelis Escher (Leeuwarden, 1898 – Laren, 1972), incisore e grafico olandese. I suoi artifici ottici conferiscono un'atmosfera fantastica alle sue composizioni di dettagli realistici.

¹¹⁴ Harold Scott MacDonal Coxeter (Londra, 1907 – Toronto, 2003), matematico inglese. Svolse la maggior parte della sua attività in Canada, fornendo contributi di rilievo in algebra e geometria.

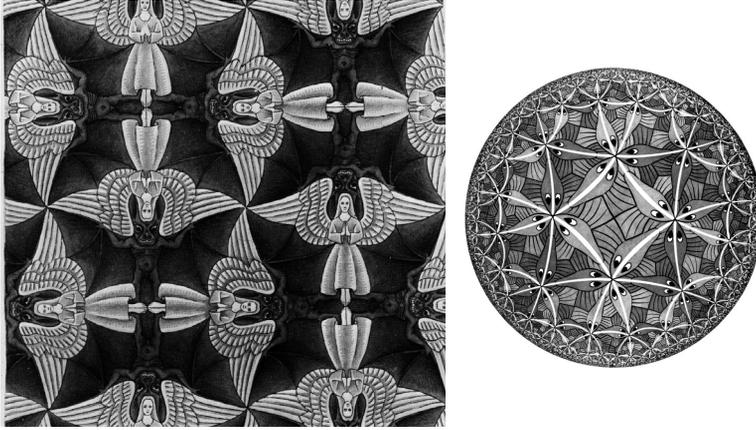


Figura 10. Dal disegno *Angeli e demoni* di M.C. Escher (a sinistra). M.C. Escher, *Limite III del cerchio*, 1959 (a destra).

F

le lo convinse a studiare tassellature che avessero come geometria di riferimento non quella euclidea, nella quale vale il postulato delle parallele, ma una geometria non euclidea, nella quale le “rette” (che stavolta sono curve, a differenza di quanto avviene nella geometria proiettiva) hanno la proprietà che per un punto esterno a una retta esistono infinite rette parallele alla retta data (la cosiddetta “geometria iperbolica”).

È straordinario il fatto che Escher sia riuscito a dare di questo concetto matematico molto difficile da comprendere e visualizzare, una visualizzazione bella come il celebre *Limite III del cerchio* del 1959 (Figura 10, a destra). In questo disegno le rette sono le congiungenti delle “spine dorsali” dei pesci e i punti sono gli ordinari punti del cerchio.

Due rette, quindi due archi di curva, sono parallele se non hanno punti in comune, e si può vedere bene che per un punto esterno a una data retta esistono più rette non incidenti la retta: si tratta del “modello di Poincaré” della geometria iperbolica, dal nome del celebre matematico francese Henri Poincaré.

A seguito della diffusione, anche e soprattutto per i suoi riflessi filosofici, della teoria della relatività, l'intero mondo della cultura si vide investito del concetto della “quarta dimensione”, e il fatto che per Einstein la quarta dimensione fosse di tipo temporale aumentò l'interesse della filosofia e del mondo artistico per questo tipo di



Figura 11. S. Dalí, *Crocifissione, corpo ipercubico*, 1954.

concetti. L'esistenza, per così dire, di una quarta dimensione (spaziale o temporale) deve aver stuzzicato non poco il celebre pittore surrealista Salvador Dalí¹¹⁵. In molte delle sue opere compaiono espliciti riferimenti al “tempo curvo”, spesso rappresentati con orologi piegati o sciolti, e nel quadro *Crocifissione, corpo ipercubico* del 1954 il Cristo è raffigurato crocifisso a uno strano solido, che non è altro che lo sviluppo tridimensionale del cubo quadridimensionale, o ipercubo.

Se si prende un cubo (per esempio, di cartone) e lo si “apre”, nel piano le 6 facce del cubo si dispongono a forma di croce. Dalí associò alla croce l'immagine della crocifissione e salì, per così dire, di una dimensione: geometricamente l'ipercubo, infatti, è costituito da 8 cubi solidi, i quali non si possono saldare nello spazio perché hanno bisogno di un'ulteriore dimensione, ma che danno origine a un ipercubo in maniera molto naturale. Questo è visibile come segue: con-

¹¹⁵ Salvador Dalí (Figueres, 1904–1989), pittore e scrittore spagnolo, protagonista del surrealismo.

sideriamo un piano cartesiano e un quadrato con vertici i punti $(1, 1)$, $(1, -1)$, $(-1, -1)$, $(-1, 1)$; il lato di questo quadrato è 2. Se ora consideriamo, accanto alle coordinate (x, y) , la "quota" z pari a 1 o -1 , in quanto vogliamo costruire un cubo, troviamo gli 8 vertici $(1, 1, 1)$, $(1, 1, -1)$, $(1, -1, 1)$, $(1, -1, -1)$, $(-1, -1, 1)$, $(-1, -1, -1)$, $(-1, 1, 1)$ e $(-1, 1, -1)$. Come vediamo, vi sono tutte le combinazioni di $+1$ e -1 nelle tre coordinate, che costituiscono poi la dimensione del cubo.

Se vogliamo passare alla rappresentazione dell'ipercubo, dobbiamo aggiungere una quarta coordinata, pari a ± 1 , ottenendo quindi 16 vertici (l'ipercubo ha infatti 16 vertici), che non stiamo a scrivere. Notiamo però che le facce del cubo sono individuate dai punti che hanno una coordinata costante: per esempio, $+1$ nella seconda, e sono dunque 4 perché tali sono i modi di inserire ± 1 nelle altre due coordinate: la faccia è $(1, 1, 1)$, $(-1, 1, 1)$, $(1, 1, -1)$, $(-1, 1, -1)$. Nell'ipercubo, quindi, se vogliamo l'equivalente delle facce, dovremo fissare una delle coordinate, e le altre tre saranno libere, dando origine a 8 punti: per esempio, la faccia $y = 1$ sarà data da

$$(1, 1, 1, 1), (-1, 1, 1, 1), (1, 1, -1, 1), (-1, 1, -1, 1), (1, 1, -1, -1), (-1, 1, -1, -1), (1, 1, 1, -1), (-1, 1, 1, -1)$$

e si può mettere in corrispondenza con il cubo tridimensionale di vertici

$$(1, 1, 1), (-1, 1, 1), (1, -1, 1), (-1, -1, 1), (1, -1, -1), (-1, -1, -1), (1, 1, -1), (-1, 1, -1)$$

ottenuto eliminando la seconda coordinata costante $y = 1$. Quindi possiamo dire che le facce dell'ipercubo sono cubi. Quanti sono? Ce n'è uno per ogni modo di fissare una delle coordinate, dunque 8 in tutto.

Gli spigoli sono invece individuati da 2 coordinate costanti nel cubo, come $(1, 1, 1)$ e $(1, 1, -1)$, e quindi ve ne sono 12, in quanto dobbiamo estrarre dalle 3 coordinate delle coppie che teniamo costanti, che in tutto sono 3, e per ciascuna di queste coppie abbiamo 2 possibilità (± 1) per le coordinate, cioè 4, e quindi in tutto 12 spigoli. Per l'ipercubo è simile, ma ora gli spigoli saranno quadrati con 2 coordinate costanti. Siccome dalle 4 coordinate si possono estrarre stavolta 6 coppie che devono restare costanti, ognuna potendo dive-

nire ± 1 , avremo 24 quadrati come spigoli dell'ipercubo. Un calcolo analogo mostra che vi sono 32 "iperspigoli", ossia segmenti di dimensione 1 nei quali tre coordinate sono costanti.

Se proseguiamo con questo tipo di ragionamento osservando, per esempio, quali facce del cubo siano adiacenti, in termini di coordinate, e riportando tutto nella quarta dimensione con la coordinata aggiuntiva, riusciamo a costruire lo sviluppo dell'ipercubo così come l'aveva visualizzato (o sentito descrivere) Dalí.

Quella della "quarta dimensione" non è che uno dei concetti matematici che hanno affascinato gli artisti, e non v'è dubbio che con lo sviluppo delle ricerche matematiche, aiutate in questi anni dalla rapida diffusione delle informazioni in tutto il mondo, il mondo delle arti figurative sappia e voglia trovare sempre più agganci con la matematica.

2. Gli elementi matematici nella musica

Nel corso dei secoli la musica in Occidente ha sviluppato forme espressive diverse, che hanno spinto a modificare leggermente l'intonazione delle note utilizzate. Fin dall'antichità l'uomo ha imparato a riconoscere maggiori o minori affinità fra le altezze dei vari suoni, correlandole ai rapporti fra le lunghezze delle corde o delle canne che venivano adoperate per produrre i suoni stessi.

Il complesso di osservazioni empiriche ha trovato una prima razionalizzazione con Pitagora mediante l'abbinamento intervallo-frazione che ben aderiva allo spirito del suo sistema filosofico, che demandava ai rapporti fra numeri interi l'interpretazione dell'intera realtà. Questa descrizione, a cui non erano estranei risvolti di tipo mistico-religioso, ha trovato conferma in un contesto scientifico moderno nell'acustica ottocentesca. La grandezza fisica che caratterizza l'altezza di un suono è la frequenza dell'onda sonora e la frazione, con cui fin dall'antichità veniva caratterizzato l'intervallo fra due suoni, è esattamente il rapporto fra le frequenze dei suoni che costituiscono l'intervallo.

La questione di stabilire quale debba essere la frequenza delle note utilizzate nella pratica musicale coinvolge quindi anche acustica e matematica. Occorre inoltre precisare che, a causa della risposta di tipo logaritmico del nostro apparato auditivo, l'addizione nel senso comune degli intervalli corrisponde alla moltiplicazione dei rapporti corrispondenti.

I risultati sulle serie di Fourier, già menzionati nel capitolo “Modelli e previsioni”, consentono d’interpretare un dato suono come la sovrapposizione di un primo suono elementare, della frequenza data e detto fondamentale o primo armonico, con un suono elementare di frequenza doppia, detto secondo armonico, uno di frequenza tripla, detto terzo armonico, e così via secondo i multipli interi della frequenza del suono dato. A rigore questo è vero se il suono è prodotto da un oggetto approssimativamente unidimensionale (canne, corde ecc.), ed è comunque questa la situazione più frequente. Perciò due suoni si sovrappongono e si amalgamano bene, se hanno molti armonici in comune.

Nell’ambito dei vari intervalli, un elevatissimo grado di affinità è sempre stato riconosciuto alle coppie di suoni le cui frequenze sono una doppia dell’altra. È naturale, perché in questo caso gli armonici del suono più acuto sono tutti presenti, sia pure con diversa ampiezza, fra gli armonici del suono più grave. Si parla in questo caso di intervallo di ottava e la somiglianza è tale che note poste a distanza di una o più ottave vengono designate con lo stesso nome.

Accettato questo fatto, mai messo in discussione in più di due millenni, l’individuazione di tutte le note ritenute ammissibili è ricondotta alla suddivisione dell’ottava in parti più piccole. L’intero universo delle note sarà poi ottenuto riportando la suddivisione di un’ottava-modello nelle altre ottave. Questo non esclude che, in particolari effetti quali il portamento e il glissando, sia espressamente contemplato di far udire addirittura un continuo di altezze.

La prima suddivisione dell’ottava in Occidente è attribuita alla scuola pitagorica, ma un’analoga suddivisione è riscontrabile anche nell’antica civiltà cinese. Essa si basa sull’intervallo che, dopo l’ottava, produce la più elevata sensazione di affinità, ossia la quinta, corrispondente al rapporto $3/2$: in effetti, si ha la maggiore affinità se il suono più acuto ha frequenza tripla rispetto a quello più grave, nel qual caso gli armonici del suono più acuto sono tutti presenti anche fra gli armonici del suono più grave.

Viceversa, rispetto al caso dell’intervallo di ottava, sono in maggior numero gli armonici del suono più grave che non stanno fra quelli del suono più acuto. Per questo motivo l’affinità che si produce con la frequenza tripla è minore rispetto a quella della frequenza doppia. Siccome poi abbiamo convenuto di identificare suoni a distanza di ottava, il rapporto $3/1$ viene considerato equivalente a $3/2$.

La procedura di suddivisione consiste nel partire dalla nota più bassa dell'ottava-modello e poi generare gli altri suoni salendo e scendendo ripetutamente di un intervallo di quinta. In questo modo i suoni escono ben presto dall'ottava-modello. Tenuto però conto della somiglianza attribuita a suoni a distanza di ottava, si possono abbassare o alzare le note ottenute di una o più ottave, in modo da farle rientrare nell'ambito dell'ottava-modello.

In quante parti suddividere l'ottava è questione in una certa misura convenzionale, e soluzioni diverse sono state adottate, dalla scala pentatonica tipicamente orientale alla scala di sette suoni a cui siamo abituati in Occidente e su cui ci soffermeremo. È invece assolutamente convenzionale stabilire quanti passi fare verso il basso e quanti verso l'alto, e quali nomi attribuire alle note ottenute. Conveniamo di chiamare Do la nota di partenza e di fare un passo verso il basso, ottenendo una nota di nome Fa, e cinque verso l'alto, ottenendo nell'ordine Sol, Re, La, Mi, Si.

Fa	Do	Sol	Re	La	Mi	Si
$\frac{2}{3}$	1	$\frac{3}{2}$	$\frac{9}{4}$	$\frac{27}{8}$	$\frac{81}{16}$	$\frac{243}{32}$
	$\frac{3}{2}$	$\frac{3}{2}$	$\frac{3}{2}$	$\frac{3}{2}$	$\frac{3}{2}$	$\frac{3}{2}$

Nella seconda riga è indicata la frequenza della nota rispetto alla nota iniziale (il Do), mentre la terza riga indica il rapporto fra le frequenze delle due note vicine.

Riportate nell'ottava-modello che va dal Do iniziale a quello di frequenza doppia, esse si dispongono nell'ordine

Do	Re	Mi	Fa	Sol			
1	$\frac{9}{8}$	$\frac{81}{64}$	$\frac{4}{3}$	$\frac{3}{2}$			
	$\frac{9}{8}$	$\frac{9}{8}$	$\frac{256}{243}$	$\frac{9}{8}$			
			Sol	La	Si	Do	
			$\frac{3}{2}$	$\frac{27}{16}$	$\frac{243}{128}$	2	
			$\frac{9}{8}$	$\frac{9}{8}$		$\frac{256}{243}$	

Come in precedenza, nella seconda riga è indicata la frequenza della nota rispetto alla nota iniziale (il Do in basso), mentre la terza riga indica il rapporto fra le frequenze delle due note vicine. A questo

punto il modello va riprodotto partendo dagli altri Do, in modo da creare una sequenza potenzialmente infinita del tipo

[...], Si, Do, Re, Mi, Fa, Sol, La, Si, Do, Re, [...]

Fra note contigue compaiono solo due tipi di intervallo: 1) il semitono (diatonico) pitagorico, di rapporto $256/243$, che si trova fra Mi e Fa e tra si e Do; 2) il tono pitagorico, di rapporto $9/8$, negli altri casi. Questo sistema di suddivisione dell'ottava, assolutamente coerente e razionale, è rimasto in auge sino alla fine del Medioevo ed è particolarmente adatto al canto e a una musica in cui viene prodotta una sola nota alla volta.

Nel passaggio dal Medioevo all'epoca moderna si sono invece affermate espressioni artistiche, come la polifonia, che comportano di produrre più suoni contemporaneamente. Tenuto conto che si amalgamano meglio i suoni con un buon numero di armonici in comune, risulta che due suoni nell'ambito di un ottava si sovrappongono in modo gradevole se il rapporto fra le frequenze è del tipo $3/2$, $5/4$, $6/5$ oppure 2 diviso uno di questi tre numeri. Nel sistema pitagorico gli intervalli di questo tipo sono solo la quinta, che già conosciamo, e la quarta, corrispondente al rapporto $4/3$, che c'è per esempio fra Do e Fa.

Ma le esigenze dell'arte spingevano a sperimentare altre soluzioni. Per esempio, fra le note Do e Mi sussiste un intervallo di terza maggiore, corrispondente al rapporto $81/64$, non particolarmente gradevole. Se però il Mi viene fatto leggermente calare, si può arrivare al rapporto $80/64 = 5/4$, che dà invece un suono molto piacevole. Tra l'altro in questo modo l'intervallo di terza minore fra Mi e Sol, corrispondente al rapporto $32/27 = 96/81$ si allarga, con la discesa del Mi, a $96/80 = 6/5$, pure assai gradevole:

$$\frac{81}{64} \searrow \frac{80}{64} = \frac{5}{4} \qquad \frac{32}{27} = \frac{96}{81} \nearrow \frac{96}{80} = \frac{6}{5}$$

Questo piccolo intervallo, corrispondente al rapporto $81/80$ di cui occorre far calare il Mi per ottimizzare la resa delle terze Do-Mi e Mi-Sol, fu chiamato comma sintonico o di Didimo.

Una grande sfida, con cui si confrontarono i teorici della musica dal Rinascimento in poi, fu quella di contemperare le esigenze, in effetti incompatibili, di avere una quinta di rapporto $3/2$ e una terza mag-

giore di rapporto $5/4$ (quindi una terza minore di rapporto $6/5$). Che le esigenze siano incompatibili lo si può immaginare dal fatto che la quinta di $3/2$ determinava già tutto univocamente. In modo un po' più artistico, si può considerare la sequenza

		La		
Sol				Sol
	Mi			
		Re		
	$\frac{5}{6}$	$\frac{4}{3}$	$\frac{2}{3}$	$\frac{4}{3}$

che potrebbe costituire un semplice basso nella tonalità di Sol maggiore. Sotto sono indicati i rapporti fra una nota e quella precedente, in modo da avere una terza minore di $6/5$ (compare $5/6$, perché è discendente), una quarta di $4/3$ e una quinta di $3/2$ (compare $2/3$, perché è discendente). Purtroppo il prodotto delle quattro frazioni riserva una brutta sorpresa, perché

$$\frac{5}{6} \times \frac{4}{3} \times \frac{2}{3} \times \frac{4}{3} = \frac{80}{81} < 1 !$$

Questo significa che, attenendosi alle ampiezze degli intervalli ottimali per la sovrapposizione di suoni, nel giro di quattro note il basso risulterebbe calante, in effetti di un comma sintonico.

Una prima soluzione, ideata dal grande teorico Gioseffo Zarlino¹¹⁶ nella seconda metà del XVI secolo, si potrebbe definire del “capro espiatorio”. L'idea proposta da Zarlino era quella di ottimizzare la sequenza di terze, alternativamente maggiori e minori,

Fa	La	Do	Mi	Sol	Si	Re
$\frac{5}{4}$	$\frac{6}{5}$	$\frac{5}{4}$	$\frac{6}{5}$	$\frac{5}{4}$	$\frac{6}{5}$	

riportando poi tutto nell'ottava-modello. In questo modo rimangono automaticamente ottimali anche le quinte Fa-Do, La-Mi, Do-Sol, Mi-Si e Sol-Re, mentre tutto il miglioramento è pagato dalla quinta Re-La e dalla terza minore Re-Fa. In particolare, la quinta Re-La corrisponde a un rapporto $40/27 = 80/54$ ed è più stretta, rispetto al rapporto ottimale $81/54 = 3/2$, nella misura di un comma sintonico $81/80$.

¹¹⁶ Gioseffo Zarlino (Chioggia, 1517 – Venezia, 1590), uno dei maggiori teorici musicali di tutti i tempi per i suoi studi che impostarono i principi della tonalità moderna.

Una soluzione migliore, con cui sono tuttora accordati alcuni organi, si rifà invece a un principio “ugualitario”, che riprende la procedura pitagorica, stringendo leggermente tutte le quinte, in modo da arrivare a una terza maggiore ottimale, anziché eccedente di un comma sintonico. L'aspetto positivo è che la terza maggiore, per esempio Do-Mi, è raggiunta in quattro passaggi: Do-Sol-Re-La-Mi. Pertanto, per assorbire il comma sintonico, è sufficiente stringere ogni quinta solo di un quarto di comma sintonico, con un effetto di stonatura sulle quinte molto minore di quello avvertibile sulle terze pitagoriche. Ne traggono beneficio soprattutto la polifonia e l'armonia, con la loro esigenza di sovrapporre più terze. Questa suddivisione dell'ottava prende il nome di *sistema mesotonico* (a un quarto di comma sintonico). Più precisamente, poiché la terza maggiore Do-Mi è ottenuta abbassando di due ottave il Mi prodotto alla fine della sequenza Do-Sol-Re-La-Mi, si ottiene la terza maggiore ottimale di $5/4$ se l'iterazione di quattro quinte produce il rapporto $5/1$. Tenuto conto che l'addizione degli intervalli corrisponde al prodotto dei rapporti, si tratta quindi di attribuire alla quinta un rapporto pari a $\sqrt[4]{5}$, anziché $3/2$. Riportato il tutto nell'ottava-modello, fra note contigue si riscontrano, come nel sistema pitagorico, solo due tipi di intervallo: il semitono (diatonico) mesotonico, di rapporto $5\sqrt[4]{125}/16$, e il tono mesotonico, di rapporto $\sqrt{5}/2$. Il fatto che si potesse ottenere una polifonia tanto gradevole partendo da un rapporto addirittura irrazionale avrà sconcertato i pitagorici di stretta osservanza, ma il Rinascimento è anche l'epoca in cui si risolvono con radicali le equazioni di terzo e quarto grado, un'epoca matura per conferire dignità anche ai “numeri surdi”, ossia gli irrazionali.

Tutti i sistemi fin qui descritti condividono però un problema. Volendo trasportare il modello per partire da un'altra nota, diversa da Do, occorre aggiungere altre note. Questo equivale a salire di quinta anche dopo il Si e scendere di quinta sotto il Fa. Per altri cinque passi si ottengono suoni ben distinguibili da quelli già presenti. Per convenzione, i teorici tendono a fare due passi verso il basso e tre verso l'alto, ottenendo la sequenza per quinte ascendenti

Mi bemolle, Si bemolle, Fa, Do, Sol, Re, La, Mi, Si, Fa diesis,
Do diesis, Sol diesis

Riportata nell'ottava modello, questa produce una suddivisione dell'ottava in dodici parti

Do, Do diesis, Re, Mi bemolle, Mi, Fa, Fa diesis, Sol, Sol diesis, La, Si bemolle, Si, Do

che va poi proseguita indefinitamente in entrambe le direzioni. Dal sesto ulteriore passo in poi si pone però un problema spinoso, perché si ottengono sempre nuove note, ma assai poco diverse da quelle già presenti. Per esempio, proseguendo ancora verso l'alto si ottiene una nota, Re diesis, leggermente diversa da Mi bemolle. Più precisamente, Re diesis è leggermente più acuto di Mi bemolle nel sistema pitagorico e leggermente più grave in quello mesotonico. Occorre quindi scegliere se "chiudere" il sistema su questa divisione a 12 dell'ottava o tenerlo "aperto" alla possibilità di introdurre le ulteriori note che si rendessero necessarie.

Nel caso del canto e degli strumenti, quali gli archi, che consentono di graduare con continuità l'intonazione, la seconda soluzione è la più naturale. Ma strumenti, come l'organo e il clavicembalo, che hanno intonazione fissa obbligano in sostanza a perseguire la prima alternativa. In verità furono anche realizzati strumenti come l'arciorgano e l'archicembalo, con i tasti neri suddivisi a metà in modo da pervenire almeno a un sistema di 17 note nell'ambito dell'ottava, ma incontrarono scarso favore. Viceversa, limitando la suddivisione a 12 note, si impone l'identificazione, per esempio, di Re diesis con Mi bemolle. Questo comporta che sia accettato come "quinta giusta" un intervallo Sol diesis-Mi bemolle che comunque suona dissonante, soprattutto a un orecchio educato. Per la sua sgradevolezza, questo intervallo, accettato come quinta *obtorto collo*, fu chiamato "quinta del lupo".

Una geniale soluzione fu però trovata nella seconda metà del XVII secolo dal teorico e organista tedesco Andreas Werckmeister (1645-1706). Si è già detto che Re diesis è più acuto di Mi bemolle nel sistema pitagorico e più grave in quello mesotonico. Questo significa che, rispetto a un'ideale "chiusura" con Re diesis = Mi bemolle, il sistema pitagorico "arriva lungo", mentre quello mesotonico "arriva corto". È allora naturale domandarsi se una via di mezzo, con alcune quinte pitagoriche e altre mesotoniche, non centri la chiusura perfetta. Si può verificare che questo è impossibile, ma con quattro quinte mesotoniche e otto quinte pitagoriche l'errore, detto schisma, è veramente piccolissimo. Rimane la questione di scegliere come alternare i due tipi di quinte. È qui che Werckmeister ebbe un'idea geniale, che tiene conto del fatto che non tutte le tonalità erano adoperate con uguale frequenza nella pratica musicale del tempo. Egli

decise quindi di accordare come mesotoniche le quinte Do-Sol, Sol-Re, Re-La e Si-Fa diesis, come pitagoriche Mi bemolle-Si bemolle, Si bemolle-Fa, Fa-Do, La-Mi, Mi-Si, Fa diesis-Do diesis e Do diesis-Sol diesis. Rimane Sol diesis-Mi bemolle, che risulta essere quasi pitagorica (a meno dello schisma).

La suddivisione dell'ottava che ne segue è passata alla storia come “temperamento Werckmeister”. Un grande vantaggio pratico del temperamento Werckmeister, oltre alla (quasi) chiusura Re diesis = Mi bemolle, è che combina due sistemi che producono intervalli ben riconoscibili: la quinta di rapporto $3/2$ per il pitagorico e la terza maggiore di rapporto $5/4$ per il mesotonico. Sia pure con una certa fatica, è quindi utilizzabile per accordare uno strumento procedendo solo a orecchio.

Un'altra peculiarità è che, data la disuniforme ampiezza delle quinte, i possibili intervalli fra note contigue sono assai variegati. Già nell'ambito della suddivisione in sette parti dell'ottava Do-Do si riscontrano un tipo di semitono (diatonico) e tre tipi di tono. E prendendo un'altra nota come punto di partenza la suddivisione cambia leggermente, nonostante l'ausilio delle alterazioni. Ma proprio queste microdiversità costituiscono una cifra del temperamento Werckmeister, in cui ogni tonalità ha una sua “personalità”. È plausibile che Johann Sebastian Bach¹¹⁷ avesse in mente questo temperamento nel comporre il ciclo di preludi e fughe *Il clavicembalo ben temperato*. Nel passaggio dal XVIII al XIX secolo, si accentua la tendenza a modulare e a comporre brani musicali di ampiezza sempre maggiore. Diventa allora naturale andare a utilizzare tonalità in precedenza meno usate e si afferma progressivamente la convinzione che un temperamento “uniforme” sia preferibile a quello disuniforme di Werckmeister. In sostanza, l'obiettivo diventa di porre la chiusura perfetta Re diesis = Mi bemolle e di regolare l'ampiezza della quinta di conseguenza. Poiché fra il Mi bemolle e il Re diesis ci sono 12 quinte e circa 7 ottave, occorre che l'iterazione di 12 quinte produca esattamente il rapporto 128:1. Bisogna allora attribuire alla quinta un rapporto pari a $\sqrt[12]{128}$, ovvero a $\sqrt{2} \sqrt[12]{2}$. Come conseguenza, l'ottava viene suddivisa in 12 semitoni tutti uguali, corrispondenti al rapporto $\sqrt[12]{2}$, e tutti gli intervalli, tranne l'ottava e sue iterate, sono caratterizzati da numeri irrazionali. Questo sistema prende il nome di “temperamen-

¹¹⁷ Johann Sebastian Bach (Eisenach, 1685 – Lipsia, 1750), compositore e strumentista tedesco. Uno dei maggiori musicisti di tutti i tempi.

to equabile” ed è in realtà antichissimo, essendo stato descritto già da Aristosseno di Taranto¹¹⁸ nel IV secolo a.C. In epoca moderna era stato riscoperto da diversi studiosi, ma era rimasto un’ipotesi puramente teorica, perché l’assenza di intervalli corrispondenti a frazioni semplici lo rendeva di ardua realizzazione se si procede puramente a orecchio.

Solo l’avvento della meccanica di precisione prima e dell’elettronica poi ha consentito la sua effettiva realizzazione. Viceversa, proprio l’uniformità della suddivisione ha aperto la strada a nuove espressioni artistiche. Mentre l’alternanza di toni e semitoni comunica dei punti di riferimento all’ascoltatore, un uso non gerarchizzato di tutti i 12 suoni uniformemente distribuiti nell’ambito dell’ottava priva l’ascoltatore di ogni riferimento. È questo un aspetto programmatico della musica dodecafonica, teorizzata da Arnold Schönberg¹¹⁹. Un’altra possibilità è quella di utilizzare una scala fatta solo di toni interi, tutti di rapporto $\sqrt[6]{2}$, che divide l’ottava in 6 parti uguali. Questa scala esatonale, che pure non dà punti di riferimento all’ascoltatore, è stata utilizzata, per esempio, da Claude Debussy¹²⁰.

¹¹⁸ Aristosseno di Taranto, compositore e filosofo greco, vissuto alla fine del IV secolo a.C. È considerato il maggiore teorico musicale dell’antichità.

¹¹⁹ Arnold Schönberg (Vienna, 1874 – Los Angeles, 1951), compositore austriaco. È il principale teorico della dissoluzione della tonalità e della musica atonale e dodecafonica.

¹²⁰ Claude-Achille Debussy (Saint-Germain-en-Laye, 1862 – Parigi, 1918), compositore e pianista francese. Uno dei maggiori rappresentanti dell’impressionismo musicale.

GIOCHI E APPLICAZIONI

1. Alcune considerazioni sul gioco

Che cosa fanno i bambini non appena cominciano a interagire con il mondo esterno? Cominciano a giocare. Che cosa accomuna da sempre gli uomini, gli antichi come i contemporanei? La passione per lo sport e per le “storie” sia in forma di favola, sia nelle loro rappresentazioni, come a teatro. C’è un filo comune che lega le due cose, il gioco da una parte, le storie dall’altra. Il gioco per il bambino rappresenta un allenamento, l’allenamento indispensabile per affrontare la vita. Le fiabe, il teatro e il cinema invece offrono rappresentazioni simboliche della vita stessa, attraverso la narrazione di fatti e storie. L’uomo da sempre ha bisogno di simboli per esprimersi, esercitarsi, capire. Non a caso i bambini sono sempre molto seri, e molto concentrati, quando giocano.

Riflettendo un attimo, ci rendiamo conto che, in maniera semplificata, il gioco propone problematiche simili a quelle che ogni essere umano si trova ben presto ad affrontare nella vita. Pensiamo in particolare ai giochi svolti da più persone: la teoria si occupa soprattutto di questi. Come possono essere descritti in maniera astratta? Un gioco è caratterizzato dalla presenza di agenti, i giocatori, da regole, che determinano che cosa è possibile fare e che cosa invece è proibito, e da tanti possibili esiti finali, per ognuno dei quali è specificato che cosa ottengono i giocatori: chi vince, chi perde o quando c’è pareggio.

Naturalmente si presume che i giocatori preferiscano vincere piuttosto che perdere, e più generalmente che abbiano delle preferenze sui risultati possibili¹²¹. Ebbene, questa potrebbe essere la descrizione di quasi ogni possibile situazione in cui due o più individui si trovano a interagire.

¹²¹ Questo significa che sanno ordinare i vari esiti: cioè dire, per esempio, che l’esito A è meglio di B , il quale a sua volta è meglio di C ...

Gli uomini, come si dice, sono animali sociali: hanno bisogno gli uni degli altri, tutta la loro vita ruota attorno ai rapporti interpersonali. Questo vale in famiglia, a scuola o nel lavoro, nei rapporti con gli amici, in una coppia o quando compriamo un oggetto, insomma vale praticamente sempre. Ecco perché il gioco rappresenta un allenamento: giocare significa imparare a ragionare, a elaborare strategie, a inseguire scopi, a porsi obbiettivi e cercare di raggiungerli.

Dal momento che la scienza si propone di studiare l'uomo e tutto quello che lo circonda, e dal momento che il modo migliore per procedere è quello di analizzare esempi, per imparare da essi e operare generalizzazioni, si capisce bene perché il gioco sia efficace per descrivere le più svariate situazioni. Nonostante questo, nella storia della scienza, lo studio sistematico delle situazioni descrivibili come giochi è cominciato solo molto recentemente. La scienza da sempre si è occupata dello studio dei fenomeni naturali, mentre la ricerca sul comportamento umano, sulle sue motivazioni e i suoi fini è stata piuttosto delegata ad altre forme di pensiero, soprattutto la filosofia e la religione. Anche l'economia, scienza abbastanza antica, è sempre stata trattata da un punto di vista qualitativo, con una descrizione singolare dei fenomeni, senza analisi più quantitative e modelli rigorosamente matematici.

Nel '900 tutto questo è cambiato. Agli inizi del secolo nelle scienze c'era un'atmosfera di euforia: gli scienziati erano convinti che con i nuovi strumenti scientifici la conoscenza dell'uomo e il suo benessere sarebbero aumentati in maniera molto significativa. Naturalmente la visione era troppo ottimistica, tuttavia dobbiamo proprio a quel clima culturale la nascita di discipline nuove e lo studio più sistematico del comportamento razionale. Oggi siamo forse un po' disillusi rispetto agli obbiettivi troppo ambiziosi di allora, ma di quello spirito è rimasto il desiderio di sviluppare sempre più strumenti scientifici, in particolare matematici, per analizzare il comportamento umano.

Questo capitolo si propone di dare un'idea dei risultati di base della teoria dei giochi, quella parte della matematica che cerca di descrivere il comportamento umano nelle situazioni in cui più persone si trovano a interagire. In effetti, "teoria delle decisioni interattive" è il nome che oggi gli esperti utilizzano: un modo più appropriato, ma meno efficace, dal punto di vista comunicativo, per identificare questa parte della matematica. I giochi e i semplici esempi che saranno proposti sono talmente ingenui da sembrare del tutto irrealistici; tut-

tavia non è possibile fare diversamente, perché la descrizione delle situazioni di interazione tra agenti è molto spesso estremamente complicata e richiede strumenti matematici molto sofisticati, oltre a una capacità di calcolo che in genere è impraticabile “a mente”. Ma questo non è un problema, perché anche gli esempi più semplici possono mettere bene in luce le tematiche di cui si occupa questa teoria affascinante.

Osserviamo che occorre essere sempre ben consapevoli del valore predittivo di una teoria. In certi casi la scienza è in grado di descrivere fenomeni con precisione quasi assoluta: si pensi al fatto che già gli antichi non avevano problemi a prevedere le eclissi con grande accuratezza. In altri ambiti però (ne abbiamo già parlato nel capitolo “Modelli e previsioni”), le previsioni sono necessariamente meno accurate; in particolare, quando si applicano al comportamento umano, che ben difficilmente (e per fortuna!) può essere previsto con accuratezza. Occorre quindi fare attenzione a credere troppo fideisticamente ai risultati ottenuti.

Tuttavia avere una teoria di riferimento che spieghi, per esempio, quale sia il comportamento razionale in certi casi è sempre molto utile. Non solo: se è vero che il comportamento singolo spesso si allontana in maniera significativa da quanto la razionalità suggerirebbe, è altrettanto vero che lo studio di fenomeni più generali – per esempio, gli equilibri di mercato – si rivela di solito piuttosto affidabile, sebbene con numerose eccezioni.

2. Alcuni esempi

In questo paragrafo vediamo una carrellata di esempi significativi e famosi di casi in cui è stata applicata la teoria dei giochi nei suoi vari aspetti.

ESEMPIO 1: OSPEDALI E MEDICI INTERNI

Supponiamo che ci siano due gruppi di agenti, denotati con A e B , che devono essere abbinati. La letteratura scientifica ha studiato l'esempio in cui A è un gruppo di neolaureati in medicina, in cerca di posti da interni, disponibili nel gruppo B di ospedali. I medici hanno preferenze su quali ospedali richiedere, ma anche gli ospedali hanno preferenze sui neolaureati. Esiste un modo “ottimale” per abbinare medici e ospedali? E se ce n'è più d'uno, è possibile stabilire se un metodo è più favorevole di un altro per un determinato grup-

po? In questo esempio la cosa forse più interessante è il fatto che già l'idea di "soluzione", e cioè di abbinamento ottimale, non risulta ovvia. Altre applicazioni dello stesso modello sono state: come distribuire gli studenti fra le varie università, oppure anche la formazione di coppie tra ragazzi e ragazze che vogliono mettersi assieme.

ESEMPIO 2: LA CONTRATTAZIONE

La mamma chiama i suoi figli, Tommaso e Andrea, e dice che darà loro 100 euro, da suddividersi come premio promozione, a condizione però che si accordino su come suddividersi la somma. Tommaso e Andrea si trovano dunque a contrattare: se non si mettono d'accordo non otterranno nulla. Non occorre elencare situazioni simili, nella vita quotidiana: in moltissime situazioni le persone contrattano, per i motivi più diversi. Probabilmente la nostra prima reazione, davanti a questo problema, è di pensare che sia banale, perché basta che si prendano 50 euro a testa. Però questa soluzione è troppo ingenua. Ben inteso, una spartizione in parti uguali sarà la soluzione ottimale in molti casi, come probabilmente in questo esempio, ma non sempre. Se vogliamo ottenere il massimo della soddisfazione complessiva (in qualche senso da rendere preciso), è possibile che uno dei due sia felice con poco, e che per l'altro invece sia molto importante avere la maggior parte della somma.

ESEMPIO 3: IL DILEMMA DEL PRIGIONIERO

La mamma chiama i suoi figli, Filippo e Niccolò, e dice loro: "Ognuno di voi deve dirmi, all'insaputa dell'altro, se preferisce che gli dia 10 euro, o che ne dia 100 al fratello". Come nell'esempio precedente, i due fratelli devono interagire, ma questa volta la situazione è un po' diversa: pur potendo parlare per mettersi d'accordo, Filippo e Niccolò sanno che eventuali accordi tra loro sono molto fragili: ognuno dei due potrebbe far credere all'altro di rinunciare ai 10 euro per sé, in favore dei 100 per l'altro, ma poi chi garantisce che mantengano il patto? Infatti, mantenerlo per loro non è favorevole. Ovviamente, in questo caso si presuppone che eventuali accordi tra i giocatori non siano vincolanti. Questa semplice storiella è una versione possibile del celeberrimo dilemma del prigioniero, paradigma di un numero incredibile di situazioni di tipo economico e psicologico, che è stato applicato anche per spiegare il comportamento di certe specie animali in determinate situazioni. Ne ripareremo.

ESEMPIO 4: LA PISTA DELL'AEROPORTO

Supponiamo che l'aeroporto di una città decida di costruire una nuova pista perché tre nuove compagnie hanno deciso di atterrare in città. La prima ha bisogno di una pista di 1 km, il cui costo è c_1 , la seconda di una pista di 2 km, il cui costo è c_2 , la terza infine di 3 km, il cui costo è c_3 . Naturalmente si ha che $c_1 < c_2 < c_3$, ma anche $c_2 < 2c_1$ ecc, perché il raddoppio della pista non implica l'automatico raddoppio dei costi, in quanto sono possibili le cosiddette economie di scala: per fare il secondo chilometro, per esempio, non c'è bisogno del trasporto di costosi macchinari, già sul posto per la costruzione del primo. L'affitto stesso dei macchinari poi non raddoppia con il raddoppio dei tempi, perché contratti a lungo termine sono unitariamente più economici. Dal momento che sembra del tutto ragionevole costruire una pista sola, di 3 km, come dovranno essere suddivisi i costi tra le tre compagnie?

ESEMPIO 5: LA BANCAROTTA

Una società, che ha tre creditori, va in bancarotta, perché possiede un capitale pari a x , ma la somma dei suoi debiti è ben superiore. Come suddividere x tra i tre creditori, che vantano crediti diversi?

ESEMPIO 6: LA SOCIETÀ PER AZIONI

Un problema che ha grande importanza è il seguente: data una società per azioni, come stabilire il potere dei singoli azionisti? Questo è cruciale non solo al momento di spartire eventuali utili, poiché un aspetto importantissimo riguarda la proprietà (eventualmente non dichiarata) della società. E ben noto che la legge spesso vuole impedire che una società sia proprietà di una sola persona: caso tipico quando le sia affidato un bene di utilità pubblica (per esempio, energia). Come stabilire se un dato azionista è il vero padrone della società? Certo, se ne possiede il 51% delle azioni..., ma ovviamente questo è un caso semplice. Se ne possiede invece il 10%? La risposta giusta è: dipende. Certamente spesso non è il padrone, ma in molti casi potrebbe accadere il contrario: per esempio, se il resto delle azioni è equamente distribuito tra migliaia e migliaia di piccolissimi azionisti. È ben noto che aggregare consenso tra moltissime persone è molto difficile, per cui in pratica non è infrequente che chi ha anche il 10% di una società abbia agli effetti pratici il potere di decidere da solo. Per un capitalista accorto potrebbe essere molto interessante avere il 10% di due società piuttosto che il 20% di una (a

parità di investimento), perché così potrebbe anche aumentare molto la sua forza sul mercato. La teoria dei giochi propone degli indici di potere, adatti a quantificare la forza degli agenti in situazioni simili.

ESEMPIO 7: LA VOTAZIONE

Tre politici devono decidere tra tre diverse alternative. Per esempio, se costruire un nuovo ponte (alternativa A), se raddoppiare un tratto di autostrada (alternativa B), o utilizzare il capitale per ridurre il debito pubblico (C). Il sistema di preferenze dei tre è il seguente:

- | | | | | | |
|----|-----|-----|-----|-----|-----|
| 1) | A | $>$ | B | $>$ | C |
| 2) | B | $>$ | C | $>$ | A |
| 3) | C | $>$ | A | $>$ | B |

Questo significa che il primo preferisce l'opzione A , poi eventualmente la B , infine la meno preferita è la C . E così via, per il secondo e il terzo. Supponiamo anche che nel caso votino tutti e tre diversamente, il voto del primo sia decisivo. Che provvedimento passerà? Non è sorprendente che la teoria non dia una risposta univoca a questa domanda. Sarebbe strano il contrario. Però può chiarire certi meccanismi, non del tutto ovvi. Per esempio, in questo caso dichiarare apertamente che si voterà il provvedimento preferito potrebbe non essere un'idea intelligente.

ESEMPIO 8: LA BATTAGLIA DEI SESSI

Questo esempio, che nella sua semplicità è molto efficace e altrettanto famoso, può essere descritto così. Ci sono due novelli sposi, diciamo Cesare e Ilaria, che devono decidere dove passare la serata: le alternative sono due, andare al cinema oppure a teatro. Entrambi vogliono stare assieme, ma Cesare preferisce il cinema, Ilaria invece il teatro. Possiamo esemplificare la situazione nella seguente tabella, che si chiama bimatrice:

$$\begin{pmatrix} (10, 0) & (-5, -5) \\ (-10, -10) & (0, 10) \end{pmatrix}$$

Come si legge? Una volta capite alcune convenzioni, non è difficile. Essa rappresenta un gioco tra due persone, e per convenzione il

primo sceglie una riga, il secondo una colonna. La scelta contemporanea dei due individua una coppia di numeri: il primo rappresenta l'utilità del giocatore di riga, il secondo quella del giocatore di colonna. Nel gioco di questo caso, la prima riga (colonna) rappresenta la scelta da parte di Cesare (Ilaria) di andare al cinema. Vediamo di capire il perché della scelta dei numeri che troviamo nella bimatrice. Consideriamo, per esempio, il caso prima riga-seconda colonna, che corrisponde alla scelta di Cesare di andare al cinema, e di Ilaria di andare a teatro. Questa situazione non è molto gradita ai due, che vorrebbero stare assieme; quantifichiamo il risultato assegnando utilità -5 a entrambi. D'altra parte, potrebbe essere meglio per loro stare da soli ma nel posto preferito, piuttosto che soli e nel posto dove non volevano andare! Quindi a quest'ultimo esito diamo utilità più bassa, diciamo -10 per entrambi. Infine diamo utilità 10 al giocatore che si trova in compagnia nel posto che preferisce, 0 se in compagnia ma nel posto preferito dall'altro. È importante osservare a questo proposito che la scelta dei numeri è abbastanza arbitraria: ciò che davvero conta è il confronto tra le utilità dei giocatori nei vari esiti possibili del gioco. Così qui è importante che l'utilità di Cesare sia massima quando l'esito è andare al cinema assieme, e che l'utilità di andare a teatro assieme, seppur minore di quella di andare entrambi al cinema, sia per lui superiore a quella delle situazioni in cui sono separati. Per questo avremmo potuto mettere 5 al posto di 10 o anche di 0 , senza alterare il modello¹²².

ESEMPIO 9: LA MORRA CINESE

I giochi "a somma zero" rappresentano una classe importante di giochi. Sono quelli strettamente competitivi, nei quali se il primo giocatore preferisce la situazione finale A , allora il secondo preferisce la B . Si chiamano a somma zero perché in pratica si può supporre che in ogni esito la somma delle utilità sia nulla. Nei casi più semplici il gioco prevede che alla vittoria dell'uno corrisponde la sconfitta dell'altro, oppure eventualmente il pareggio. In questi casi la bimatrice può essere sostituita da una matrice, tabella che riporta solo l'utilità del giocatore di riga: evidentemente quella del secondo è la stessa cambiata di segno.

¹²² Questa non è un'ipotesi del tutto ovvia. Stiamo infatti assumendo che non importa l'intensità del desiderio dei giocatori, ma solo l'ordine delle loro preferenze.

Ecco una matrice che rappresenta un gioco familiare alla gran parte dei bambini:

$$\begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}$$

Si tratta di una rappresentazione possibile del gioco della morra cinese, in cui i due giocatori mostrano contemporaneamente una mano che, a seconda di come sono le dita, può rappresentare un sasso, un foglio di carta o le forbici; il sasso spezza le forbici, che tagliano la carta, che avvolge il sasso: ogni mossa è vincente contro una mossa e perdente contro un'altra. Nella matrice di cui sopra, possiamo supporre che la prima riga (colonna) corrisponda alla scelta del sasso, la seconda delle forbici, la terza della carta. Così, per esempio, in corrispondenza della prima riga alla seconda colonna troviamo 1, che indica il fatto che il sasso batte le forbici, alla terza colonna -1 , perché la carta batte il sasso.

3. Giochi in forma estesa

Abbiamo visto nel paragrafo precedente un modo efficace di descrivere un gioco: si tratta di utilizzare tabelle, chiamate bimatrici, oppure matrici se il gioco è a somma zero. Non è certamente questa l'unica descrizione possibile del gioco: esiste una forma che per certi versi è ancora più efficace. Si tratta di utilizzare l'idea di "grafo"; per meglio capire questo concetto basta illustrarlo con un esempio semplice.

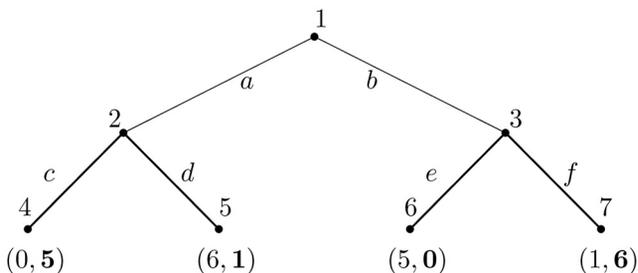
Prima però conviene mettere in evidenza quali sono gli ingredienti fondamentali di ogni gioco, e per farlo ci riferiamo agli scacchi, probabilmente il gioco più famoso al mondo. Poi vedremo un grafo che descrive la semplicissima situazione in cui ci sono due giocatori, che giocano in sequenza e hanno solo due mosse possibili. Per descrivere in maniera completa il gioco, occorre specificare:

- la situazione iniziale del gioco (la disposizione dei vari pezzi sulla scacchiera);
- le regole che determinano le possibili evoluzioni del gioco (chi deve fare la mossa, le mosse che ogni pezzo può fare, come mangiare un pezzo avversario e così via);
- i suoi stati finali (scacco matto, oppure tutte le situazioni che portano al pareggio);

- le preferenze dei giocatori sugli esiti possibili (in questo caso semplice si suppone, ovviamente, che il giocatore preferisca prima di tutto vincere, poi eventualmente pareggiare); come già osservato, le preferenze saranno ordinate per mezzo di numeri opportuni.

3.1 L'albero del gioco

Basta pensarci un secondo per capire che anche una versione semplificata degli scacchi richiede una struttura descrittiva molto complessa. Per questo illustriamo l'idea di grafo con un esempio molto semplice, in cui i giocatori sono Alberto e Laura e la situazione iniziale prevede che Alberto debba scegliere tra due opzioni, diciamo *essere simpatico*, oppure *essere antipatico*¹²³. Una volta che Alberto ha espresso la sua scelta, tocca a Laura fare una scelta simile. Ecco la descrizione per mezzo di un grafo.



Questo si chiama “albero del gioco”, e l'albero è un caso particolare di grafo. Abbiamo indicato con i numeri 1,..., 7 quelli che chiameremo i “nodi dell'albero”, e che rappresentano le varie situazioni del gioco: 1 rappresenta la situazione, nodo, iniziale, mentre 4, 5, 6, 7 rappresentano i nodi, situazioni, finali. Ogni nodo, esclusi quelli finali, è etichettato in modo da indicare quale giocatore deve fare la mossa a quel nodo: nell'albero di sopra per distinguere i giocatori sono disegnati con spessore diverso i rami uscenti dai nodi: Alberto ha uno spessore normale, Laura invece ha uno spessore maggiore. Inoltre, il ramo di sinistra indica un comportamento aggressivo mentre quello di destra un comportamento gentile, per entrambi. Ai nodi finali sono associate coppie di numeri, che indicano le utilità di Alberto e Laura in quella situazione. Se ambedue sono gen-

¹²³ Userò anche i termini “gentile” e “aggressivo”.

tili, Alberto ottiene 1 e Laura 6. E così via. Perché usiamo i termini aggressivo e gentile? Vediamo il caso di Alberto (per Laura vale lo stesso ragionamento): il suo comportamento gentile può far ottenere a Laura la massima soddisfazione possibile (6), ma a costo di accontentarsi di 1, pur sapendo che uno degli esiti possibili gli farebbe ottenere 6. Prima di provare a chiederci se possiamo prevedere l'esito del gioco, occorre fare due osservazioni molto importanti:

1. Quello che importa ai giocatori è solo quello che otterranno. In particolare, la soddisfazione dell'altro viene presa in considerazione solo per cercare di capire come l'altro si comporterà: nel nostro esempio, non dobbiamo essere sviati dal fatto che Laura potrebbe voler essere gentile con Alberto per altri motivi: ogni eventuale motivazione psicologica deve essere inglobata nell'utilità associata alle situazioni finali¹²⁴.
2. Nonostante le mosse a disposizione dei giocatori siano le stesse (al momento di decidere devono entrambi scegliere tra le due alternative: aggressivo/gentile), la loro situazione strategica è radicalmente diversa. Le regole del gioco infatti specificano che Laura decide sapendo che cosa ha deciso Alberto al momento precedente, e questo chiaramente cambia la sua situazione rispetto a quella di Alberto.

L'osservazione al punto 2 necessita di essere formalizzata, e l'idea è quella di utilizzare il fondamentale concetto di strategia: la strategia per Laura – come per ogni giocatore – è la specificazione di un'azione possibile in ogni nodo etichettato con il nome del giocatore stesso. Nel nostro esempio, dunque, Alberto ha 2 strategie possibili, che coincidono con le mosse aggressivo/gentile, visto che deve decidere solo alla radice, ma Laura ha 4 strategie possibili:

1. essere gentile se Alberto è gentile, essere gentile se Alberto è aggressivo;
2. essere gentile se Alberto è gentile, essere aggressiva se Alberto è aggressivo;

¹²⁴ Questo è un punto molto importante da capire: tutti amano avere più soldi che meno, in generale. Tuttavia in un gioco è ben possibile che io preferisca avere un po' meno, a favore di un altro giocatore, per i motivi più diversi. In questo caso assocerò al risultato con meno denaro un'utilità maggiore di quello in cui ottengo di più. Dunque è possibile osservare comportamenti "altruistici": il modello però prevede che questi debbano essere inclusi nelle preferenze dei giocatori.

3. essere aggressiva se Alberto è gentile, essere gentile se Alberto è aggressivo;
4. essere aggressiva se Alberto è gentile, essere aggressiva se Alberto è aggressivo.

In realtà, osservando l'albero del gioco, ci conviene scrivere più semplicemente (*df*, *cf*, *de*, *ce*), per elencarle tutte. Pur con queste premesse, non è evidente quale sarà l'esito del gioco. A prima vista Alberto potrebbe scegliere il ramo *a*, nella speranza che Laura lo porti alla situazione finale che gli dà la massima utilità (6). Ma è credibile che questo succeda? Vediamo un metodo efficace e convincente per scoprire come giocatori intelligenti e razionali analizzano giochi come quello appena descritto, e di conseguenza tutti quelli con struttura simile.

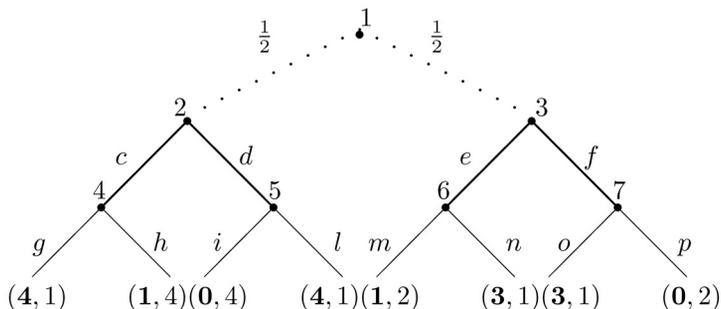
L'idea consiste nell'analizzare che cosa decideranno i giocatori in ogni nodo in cui la scelta porta a una situazione finale: nel nostro caso semplice, nei nodi 2 e 3. In 2, Laura sceglie il ramo *c* che le dà utilità **5**, piuttosto che *d*, che per lei vale solo **1**. Analogamente, a destra, sceglie *f* (infatti **6** > **0**). L'analisi è resa ovvia dal fatto che, portando tali nodi a un risultato finale, l'agente coinvolto è solo uno (Laura), e la sua scelta è immediatamente chiara a tutti: sceglie il ramo che le porta utilità maggiore. Un passaggio semplice, dunque, ma anche cruciale. L'ipotesi di intelligenza dei giocatori, che è sottintesa in tutta la teoria, implica che Alberto sappia benissimo che cosa sceglierebbe Laura in ogni circostanza (le utilità dei giocatori sono note a tutti) e quindi che sia consapevole che se sceglie il ramo *a* ottiene 0, mentre se porta il gioco al nodo 3 ottiene 1. Dunque sceglie di essere simpatico, e lo stesso farà Laura¹²⁵.

La procedura utilizzata si chiama di induzione a ritroso. È chiaro che questo metodo ci permette di prevedere a priori l'esito di giochi di questo tipo, ed è anche un risultato intuitivo, almeno in parte, perché tutti abbiamo la sensazione che certi giochi debbano finire sempre esattamente allo stesso modo; un po' di esperienza ci convince, per esempio, che nel famoso gioco del tris i due giocatori pareggiano sempre.

¹²⁵ Questo è quel che ci interessa, ma occorre precisare che la strategia di Laura prevede di essere aggressiva nel nodo 2 e simpatica nel nodo 4. Non ha poi bisogno di comportarsi aggressivamente, nel caso decidano davvero di giocare, perché il nodo 2 non viene raggiunto. Tuttavia Alberto sa che lei lo farebbe, e questo alla fine determina il fatto che Laura ottenga un risultato migliore di Alberto.

Il passaggio successivo naturalmente riguarda il capire quali sono le caratteristiche di un gioco che ci permettono di utilizzare l'induzione a ritroso. Dal punto di vista della descrizione a parole, si capisce che possiamo applicare tale metodo in ogni gioco in cui ai giocatori sono note tutte le informazioni rilevanti. In altre parole, ogni giocatore al momento di decidere sa quale è stata la storia precedente del gioco, e quali sono tutti i suoi possibili sviluppi futuri. Dal punto di vista matematico, la cosa si esprime dicendo che a partire da un nodo qualunque e guardando solo quello che segue da quel nodo, la struttura ad albero del gioco non cambia.

Osserviamo che è anche possibile che sia coinvolto qualche evento probabilistico; la cosa importante è che il risultato dell'evento dovrà essere noto a tutti. Per esempio, all'inizio il gioco potrebbe prevedere l'estrazione di una carta da un mazzo e avere poi sviluppi diversi a seconda che essa sia rossa oppure nera. Ecco un albero che presenta una situazione simile.



L'analisi del gioco non cambia molto dall'esempio in cui non ci sono eventi casuali. Possiamo osservare, per esempio, che il giocatore che tira per secondo sceglie il ramo di destra al nodo 4, quello di sinistra al nodo 5, e così via.

Facendo i calcoli, l'induzione a ritroso mostra che nel caso la sorte privilegi il ramo di sinistra, il primo giocatore avrà utilità pari a 1, il secondo pari a 4, mentre nel caso del ramo di destra, le utilità saranno rispettivamente di 1 e 2. Le strategie dei giocatori sono *ce* per il primo giocatore e *himp* per il secondo. La cosa più interessante è valutare le utilità dei due, una volta che usano tali strategie: il primo prende 1 nei due casi, e quindi la sua utilità è 1, mentre il secondo prende 4 a sinistra, 2 a destra, quindi la sua utilità è pari a 3. Questo non deve suonare paradossale: stiamo assumendo l'ipotesi, in certi

casi non del tutto realistica ma certamente molto comune e forse inevitabile, del calcolo dell'utilità in senso atteso: guadagnare 2 o 4 con eguale probabilità è la stessa cosa che guadagnare 3 con certezza. In questo senso abbiamo abolito ogni incertezza: questo gioco, come il precedente, è determinato, nel senso che il suo esito è prevedibile a priori; giocatori intelligenti faranno sempre la stessa cosa, ottenendo inevitabilmente lo stesso risultato¹²⁶.

Concludiamo questo paragrafo sui giochi in forma estesa accennando solo al fatto che quanto affermato precedentemente, e cioè che l'induzione a ritroso "risolve" questi giochi, ha una conseguenza molto sorprendente: dal punto di vista di questa teoria un gioco come gli scacchi non è molto interessante (come il tris) perché teoricamente due giocatori terminerebbero le loro partite sempre con lo stesso risultato. È chiaro che questo risultato è teoricamente interessante, ma è anche ovvio che i giocatori della teoria sono, almeno in questo caso, molto idealizzati: non esiste e non esisterà mai un essere umano che possa disegnare l'albero del gioco degli scacchi per poi applicare l'induzione a ritroso¹²⁷!

La teoria precedente non può evidentemente essere applicata ai giochi in cui sono previste mosse contemporanee¹²⁸. E dunque necessario introdurre concetti nuovi. Si arriva così all'idea di gioco in forma strategica, e si divide la teoria in due filoni principali: la non cooperativa e la cooperativa. Nei prossimi paragrafi vedremo alcuni aspetti di entrambe.

4. La teoria non cooperativa

Le prime applicazioni della teoria non cooperativa riguardano i giochi competitivi, o a somma zero: la morra cinese (esempio 9, a p. 151), la dama, gli scacchi, il tris ne sono esempi tipici. Sono i giochi più semplici da analizzare, perché gli interessi dei giocatori sono totalmente contrapposti, e quindi in un certo senso è chiaro che ottenere il meglio per sé è equivalente a cercare che l'avversario ottenga il meno possibile, e questo, come vedremo, semplifica l'analisi.

¹²⁶ Che ottengano sempre lo stesso risultato sembra un po' difficile da digerire, visto che il guadagno finale dipende da una scelta iniziale casuale. Ma ricordo che questo va inteso in senso atteso.

¹²⁷ Oppure che possa farselo disegnare dal più potente dei computer.

¹²⁸ Possiamo considerare come contemporanee anche mosse che pur avvenendo in tempi diversi non sono note a tutti i giocatori: nel gioco della morra cinese Matteo può anche tirare per primo, l'importante è che Marina non sappia che cosa ha tirato!

Quando invece, e sono le situazioni più comuni, le utilità dei giocatori sono per tutti più alte in certi esiti del gioco che in altri (come nell'esempio 3 sul dilemma del prigioniero, a p. 148), possono nascere situazioni anche paradossali, più difficili in generale da trattare.

Dal momento che nel caso competitivo ogni forma di collaborazione è preclusa, perché gli interessi dei giocatori sono opposti, il primo concetto da tenere in considerazione è quello dei valori conservativi dei giocatori (detti anche i loro livelli di sicurezza). Proviamo a capirne l'idea su un gioco in forma di matrice:

$$\begin{pmatrix} 8 & -1 & 1 \\ -4 & 0 & 2 \\ 4 & 4 & 3 \end{pmatrix}$$

Supponiamo che Emanuele debba scegliere una riga e Francesca una colonna; ricordiamo che il coefficiente nella casella individuata è quanto Francesca paga a Emanuele (o riceve in valore assoluto, se il numero è negativo). Emanuele sa che giocando la prima riga potrebbe dover pagare 1 (se Francesca scegliesse la seconda colonna), con la seconda riga invece rischia di dover pagare 4, con la terza si garantisce un guadagno minimo di 3. Quindi con la terza riga può ottenere almeno 3, che rappresenta il valore conservativo per lui. Mettiamoci ora dal punto di vista di Francesca. Con la prima colonna rischia di dover pagare 8, con la seconda 4, con la terza 3; giocando giudiziosamente dunque si può garantire di pagare non più di 3. Riassumiamo la situazione. Emanuele sa di poter garantirsi almeno 3, Francesca sa di avere il modo di pagare non più di 3: è dunque evidente che 3 rappresenta esattamente quel che Emanuele (che gioca la terza riga) riceve da Francesca (che gioca la terza colonna) e, ogni volta che due giocatori si troveranno a giocare la matrice precedente, il risultato sarà quello. Dunque il gioco è determinato e Francesca ed Emanuele in pratica non lo giocano mai, perché non c'è nessun divertimento a giocare un gioco che finisce sempre allo stesso modo!

Per capire quello che abbiamo fatto, riscriviamo in termini un po' diversi i calcoli del gioco:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

Come abbiamo determinato la soluzione del gioco precedente? Dapprima abbiamo isolato i valori a_{12} , a_{21} , a_{33} , che rappresentano ri-

spettivamente i valori minimi nelle righe 1, 2, 3; abbiamo poi selezionato il più grande di questi valori (a_{33}), che rappresenta il livello di sicurezza di Emanuele (cioè il minimo che è in grado di garantirsi, con un comportamento razionale), e abbiamo fatto la stessa cosa per Francesca, con l'unica differenza che lei paga, e quindi dobbiamo tenere conto di un cambiamento di segno. Il valore conservativo per Francesca coincide con quello di Emanuele, e il coefficiente a_{33} , che rappresenta la soluzione del gioco, verifica la proprietà:

$$a_{i3} \leq a_{33} \leq a_{3j}$$

con gli indici i, j che possono assumere valore 1, 2, 3. In generale, se indichiamo con x una strategia del primo giocatore, con y una strategia del secondo e con $f(x, y)$ il pagamento del secondo al primo, una soluzione del gioco è una coppia di strategie (\bar{x}, \bar{y})

$$f(x, \bar{y}) \leq f(\bar{x}, \bar{y}) \leq f(\bar{x}, y) \quad (1)$$

per ogni possibile strategia x del primo e y del secondo.

Riprendiamo l'esempio 9 sulla morra cinese, a p. 151. Se calcoliamo i valori conservativi dei due giocatori, ci accorgiamo che sono differenti: -1 per il primo, 1 per il secondo. Questo significa che nel gioco della morra cinese entrambi i giocatori non possono garantirsi niente di più che . . . perdere! In generale, si può dimostrare che il valore conservativo del primo è sempre minore o uguale al valore conservativo del secondo¹²⁹. Dunque nei giochi in cui i due valori sono diversi, sembra che ci sia un problema. Ma per fortuna è così! Sarebbe andato veramente troppo contro la nostra intuizione scoprire che anche nella morra cinese esista un modo infallibile per vincere¹³⁰! È più ragionevole aspettarsi che non tutti i giochi che prevedono mosse contemporanee siano strettamente determinati. Tuttavia la teoria non si ferma qui, perché riesce a dire qualcosa di interessante anche per quei giochi che prevedono mosse contemporanee. Visto che non esiste un modo infallibile per garantirsi il risultato, ci si può

¹²⁹ Il risultato è perfettamente naturale. Supponiamo che in un gioco il valore conservativo del primo giocatore sia 5. Potrebbe essere il valore del secondo minore di 5? Evidentemente no, perché il primo è in grado di garantirsi almeno 5, qualunque cosa faccia l'altro, perciò il secondo non può avere una strategia che gli permette di pagare di meno, e quindi il suo valore conservativo, cioè il pagamento minimo che è in grado di garantirsi, deve essere non minore di 5.

¹³⁰ Oppure che i giocatori pareggino sempre.

chiedere se comunque si può provare a definire un comportamento ottimale anche in questi ultimi casi.

Un modo per affrontare il problema è di immaginare di giocare più volte contro lo stesso giocatore.

Mi spiego nel caso della morra cinese. Se uno dei due decidesse di non giocare mai carta, allora dopo un po' l'altro giocatore se ne accorgerebbe e comincerebbe a giocare sempre sasso, con indubbio guadagno¹³¹. Dunque esistono modi più o meno razionali di giocare un tale gioco. Come individuarli? L'idea consiste nell'assegnare (a priori) una certa probabilità con cui le varie strategie verranno giocate, e poi giocare in accordo con i risultati di un esperimento casuale che rispetti le probabilità stabilite. La funzione di utilità dei giocatori si calcola adesso come valore atteso, come abbiamo già visto per il gioco in forma estesa con la presenza di un evento casuale: guadagnare 100 o 0 con uguale probabilità equivale a guadagnare 50 con certezza. Si parla in questo caso di strategie miste per i giocatori, in contrapposizione alle strategie pure, che sono le scelte fatte con certezza. Un esito razionale del gioco diventa allora una coppia che verifica la relazione dell'equazione (1)¹³².

Il celebre teorema di von Neumann¹³³, enunciato nel 1928, asserisce che "Ogni gioco finito a somma zero ammette equilibrio in strategie miste". Questo non significa che ogni singola partita abbia esito scontato, ma che se, per esempio, due giocatori razionali giocano ripetutamente alla morra cinese, il risultato sarà in media un pareggio¹³⁴. Da questo punto di vista possiamo dunque dire che il teorema di von Neumann afferma che anche i giochi come la morra cinese sono strettamente determinati, sia pure nel mondo allargato delle strategie miste e delle utilità attese. È importante osservare come questo risultato sia comunque un progresso della teoria (tra l'altro all'epoca in cui von Neumann l'ha dimostrato c'era chi pensava che un risultato simile non fosse vero), sebbene evidentemente la risposta non possa essere decisiva come nel caso in cui esistono equilibri in strategie pure.

¹³¹ Infatti non perderebbe mai, e vincerebbe un certo numero di partite.

¹³² Dove x, y sono le strategie miste dei giocatori 1 e 2 rispettivamente, e $f(x, y)$ rappresenta quanto paga il secondo giocatore al primo come valore atteso se viene giocata la coppia (x, y) .

¹³³ John von Neumann (Budapest, 1903 – Washington, 1957), matematico e informatico ungherese naturalizzato statunitense. Uno dei più importanti scienziati del XX secolo, a cui si devono fondamentali contributi nel campo della teoria degli insiemi, dell'analisi funzionale, della topologia, della fisica quantistica, dell'economia, dell'informatica, della teoria dei giochi, della fluidodinamica e in molti altri settori della matematica.

¹³⁴ Infatti il valore atteso nel gioco esteso diventa 0 per entrambi i giocatori.

La teoria dei giochi strettamente competitivi ha a questo punto una forma completa. Certamente sono possibili molte generalizzazioni, ma i risultati ottenuti sono del tutto soddisfacenti. Tuttavia, come già accennato, i casi più interessanti si verificano quando i giocatori non hanno necessariamente interessi contrapposti in ogni situazione. Il dilemma del prigioniero (esempio 3 a p. 148) ne è l'esempio più noto ed efficace: Filippo e Niccolò, a seconda delle scelte che fanno, possono avere 10 euro a testa, oppure 100, due situazioni ben diverse.

D'altra parte, si può intuire come in questi casi la situazione si complichino, dal momento che nei giochi a somma zero un giocatore può trovare le sue strategie di equilibrio senza bisogno di conoscere che cosa fa l'altro giocatore. Inoltre nel caso di molteplici equilibri i giocatori sono indifferenti su quale equilibrio scegliere, perché la loro utilità è sempre la stessa (il loro valore conservativo). È naturale ipotizzare che questo non succeda se il gioco non è strettamente competitivo. Così nel caso del gioco della battaglia dei sessi (esempio 8, p. 150) è intuibile come i comportamenti razionali siano di andare assieme a teatro oppure al cinema, però sembra difficile privilegiare una delle due soluzioni, e d'altra parte Ilaria preferisce andare a teatro, Cesare andare al cinema; quindi i due hanno idee diverse sugli equilibri; inoltre, se non si mettono d'accordo su dove andare, rischiano di passare la serata soli e tristi; in definitiva, è per loro necessario coordinare le azioni per giungere a un equilibrio.

Storicamente la teoria ha avuto un momento di stasi e non è riuscita a trovare un'idea brillante per procedere, finché fu pubblicata nel 1944 un'opera fondamentale di von Neumann e Morgenstern¹³⁵, *Theory of games and economic behaviour* (Teoria dei giochi e comportamento economico). La loro idea guida è che essendo possibile che in certe situazioni i giocatori stiano meglio che in altre, questo li porterà naturalmente a cercare di stringere coalizioni: è la nascita della teoria cooperativa.

Ma di questo parleremo nel prossimo paragrafo, ora invece vediamo le idee fondamentali del modello non cooperativo di John Nash¹³⁶.

¹³⁵ Oskar Morgenstern (Görlitz, 1902 – Princeton, 1977), economista austriaco. È coautore, con John von Neumann, del libro *Theory of games and economic behaviour*, opera fondamentale, ritenuta l'atto ufficiale di nascita della teoria dei giochi.

¹³⁶ John F. Nash (Bluefield, 1928), matematico ed economista statunitense. Ha rivoluzionato l'economia con i suoi studi di matematica applicata alla teoria dei giochi, vincendo il premio Nobel per l'economia nel 1994. Ha anche ottenuto risultati fondamentali in matematica pura, tra cui uno storico risultato di regolarità sulle equazioni differenziali alle derivate parziali, dimostrato indipendentemente anche da Ennio De Giorgi. A lui è dedicato il film *A beautiful mind*.

Gli oggetti primitivi del modello non cooperativo sono gli spazi delle strategie dei giocatori e le loro funzioni di utilità, che dipendono dalle scelte congiunte di tutti i giocatori. Un equilibrio di Nash allora, nel caso di due giocatori, è una coppia di strategie (\bar{x}, \bar{y}) tali che:

- fissata la strategia \bar{y} del secondo giocatore, il primo ottiene il massimo possibile nella sua funzione utilità con la strategia $x = \bar{x}$;
- fissata la strategia \bar{x} del primo giocatore, il secondo ottiene il massimo possibile nella sua funzione utilità con la strategia $y = \bar{y}$.

Nel caso di un gioco bimatrice, le strategie dei giocatori sono rispettivamente le righe e le colonne, e se vogliamo considerare anche le strategie miste allora gli insiemi X e Y sono tutte le distribuzioni di probabilità possibili sulle righe/colonne (rispettivamente).

Come spesso succede, un esempio dovrebbe chiarire le idee. Consideriamo il gioco seguente:

$$\begin{pmatrix} (4, 5) & (3, 6) \\ (6, 1) & (2, 1) \end{pmatrix}$$

e proviamo a verificare che l'esito $(3, 6)$ proviene da un equilibrio di Nash. Il primo giocatore deve guardare la seconda colonna (che rappresenta la scelta del secondo giocatore) e osservare se la scelta della prima riga sia a lui favorevole o se non gli conviene cambiare, ma se cambiasse otterrebbe 2 invece di 3: non gli interessa. Analogamente, il secondo sa che cambiando otterrebbe 5 invece di 6: non gli conviene. Nessuno dei due ha interesse a cambiare: siamo in una situazione di equilibrio.

Ritornando all'equazione (1) a p. 159, che caratterizza gli equilibri nei giochi a somma zero, si vede che questi effettivamente sono equilibri di Nash. Tuttavia, al di fuori del caso a somma zero, in generale i valori conservativi non hanno molta importanza¹³⁷.

Abbiamo visto in precedenza che i giochi possono anche essere matematicamente formalizzati per mezzo di un albero, e per trovarne un esito razionale abbiamo utilizzato il metodo dell'induzione a ritroso; se riscriviamo il gioco descritto dall'albero in forma strategi-

¹³⁷ I valori conservativi nel caso non a somma zero possono essere molto poco significativi, proprio perché in certe situazioni i giocatori non hanno interesse a prendere una decisione ostile all'altro. Nel gioco:

$$\begin{pmatrix} (1, 1) & (-1, -1) \\ (0, 0) & (10, 10) \end{pmatrix},$$

i valori conservativi di entrambi sono zero, ma l'esito $(0, 0)$ chiaramente non interessa a nessuno.

ca, un teorema garantisce che le soluzioni determinate dall'induzione a ritroso sono equilibri di Nash.

Questo modello astratto ha in più la caratteristica di mettere bene in evidenza che cosa sia in generale un comportamento razionale da parte dei giocatori: ogni giocatore massimizza la propria utilità, prendendo per buono che l'altro utilizzi la strategia proposta. D'altra parte, prendere per buono che l'altro si comporti così è consistente, perché anche l'altro non ha interesse a cambiare la strategia a lui proposta. Non è il caso qui di esaminare oltre questa idea, che Nash ha sviluppato nella sua tesi di dottorato e che quaranta anni dopo gli ha procurato il premio Nobel per l'economia. È però interessante aggiungere almeno che l'idea di equilibrio sviluppata da Nash è servita anche per studiare l'evoluzione di due specie animali in competizione su un territorio. Questo è un modello diverso (ma altrettanto efficace) rispetto a quello che descriveremo nel capitolo "La matematica nelle scienze umane e nella vita". Non solo, si tratta del primo esempio di applicazione della teoria a giocatori non umani.

La conclusione è che l'idea di equilibrio sviluppata da Nash è il concetto centrale della teoria non cooperativa, e anche i raffinamenti successivi, quali per esempio l'idea di equilibrio correlato, si basano sostanzialmente sul concetto introdotto da Nash.

Va comunque messo in evidenza che, pur così naturale, non risolve i problemi in caso di non unicità, e soprattutto non sembra fornire una soluzione "interessante" nei giochi del tipo dilemma del prigioniero. Riprendiamolo un attimo, per raccontarlo nella sua forma originale (l'esempio 3, p. 148, ne propone una formulazione equivalente). Un giudice convoca due persone sospettate di essere complici di un grave crimine, e fa loro il discorso seguente: "Se uno dei due confessa e l'altro no, chi confessa è libero per aver assicurato un irriducibile alla giustizia, l'altro prende 10 anni di galera. Se entrambi confessate, la pena prevista è di 7 anni di galera. Se nessuno confessa, le prove sono insufficienti per provare la vostra colpevolezza riguardo al crimine, ma troverò il modo di condannare entrambi alla pena detentiva di un anno, per un reato più lieve". Ecco la bimatrice:

$$\begin{pmatrix} (-7, -7) & (0, -10) \\ (-10, 0) & (-1, -1) \end{pmatrix}$$

in cui il valore -7 , per esempio, è associato al fatto che confessando entrambi si faranno 7 anni di galera. È facile accorgersi che l'u-

nico equilibrio di Nash sta nel confessare: una conclusione molto deludente, visto che i due giocatori, accordandosi, potrebbero fare solo 1 anno di galera, invece di 7. Ma c'è di più: per ogni giocatore è più conveniente confessare, qualunque strategia decida di utilizzare l'altro; in questo caso si parla di strategia dominante, e chiaramente le strategie dominanti sono le uniche di Nash¹³⁸.

È evidente il fatto che questa soluzione appare molto deludente dal punto di vista dell'efficienza. D'altra parte, la teoria non è stata capace di elaborare un concetto ragionevole che induca alla collaborazione in un esempio come questo. Tuttavia uno sviluppo molto importante si ottiene considerando il gioco ripetuto più volte; un'ipotesi naturale da assumere, visto che la ripetizione rende "l'esperimento" interessante. In tal caso un risultato profondo ma difficile da spiegare nei dettagli dice sostanzialmente che in certe ipotesi il gioco ripetuto prevede, tra i molti equilibri di Nash, anche la collaborazione tra i giocatori: in altre parole, due persone possono decidere razionalmente e coerentemente di collaborare, e questo accordo può reggere. Tuttavia non è detto che il meccanismo funzioni sempre. Tutta la nostra vita sociale è condizionata dallo stesso dilemma: l'uomo si dà delle regole di convivenza civile nella convinzione che queste migliorino le condizioni della collettività, e quindi di tutti, ma chi le viola ne ha spesso un vantaggio immediato; ne vediamo tutti i giorni esempi: viola i patti, per esempio, chi non paga le tasse o chi guidando salta continuamente di corsia in autostrada.

5. La teoria cooperativa

Come accennato in precedenza, l'idea di gioco cooperativo è stata introdotta da von Neumann e Morgenstern. Il contributo del loro libro sulla teoria dei giochi e del comportamento economico è fondamentale perché ha reso lo studio dei giochi una disciplina sistematica e ha proposto un cambiamento radicale nel modo di studiare i problemi dell'economia, delle scienze politiche e di quelle sociali. Il metodo proposto consiste nel tradurre i problemi in giochi opportuni, nel trovare le soluzioni di questi giochi con le tecniche

¹³⁸ Strategia dominata invece è quella per cui esiste un'altra strategia che procura al giocatore maggiore utilità, qualunque cosa faccia l'altro. Una strategia dominata non può far parte di un equilibrio di Nash.

sviluppate dalla teoria, e nel ritradurre le soluzioni trovate in termini di comportamenti economici ottimali.

L'idea di gioco cooperativo nasce, come abbiamo accennato, dall'esigenza di analizzare il comportamento razionale di agenti che interagiscono in situazioni non strettamente competitive: in tal caso è ragionevole pensare che i giocatori possano fare alleanze, formare coalizioni ecc.; ogni coalizione sarà in grado poi di garantire una certa distribuzione di utilità all'interno dei suoi membri.

Che cosa distingue il gioco cooperativo da quello non cooperativo? Il fatto che si ipotizzi la nascita delle coalizioni non significa che si supponga che i giocatori siano meno egoisti; le coalizioni sono uno strumento possibile per ottenere migliori risultati individuali. La differenza nei due approcci sta in un'altra cosa: secondo J. Harsanyi¹³⁹, premio Nobel per l'economia insieme con Nash, un gioco è cooperativo se gli accordi tra i giocatori sono vincolanti. In caso contrario, il gioco è non cooperativo.

All'interno dei giochi cooperativi, la teoria distingue fra quelli TU (*a utilità trasferibile*) e quelli NTU (*a utilità non trasferibile*). Qui ci limitiamo a 3 esempi di gioco TU, già sufficienti a introdurre le idee principali di questo approccio.

Per definire un gioco cooperativo abbiamo bisogno dell'insieme $N = \{1, \dots, n\}$ dei giocatori, del dato per ogni $A \subseteq N$, e di un numero reale denotato con $v(A)$. Ogni $A \subseteq N$ rappresenta una possibile coalizione e $v(A)$ rappresenta l'utilità, o in altri casi un costo, che la coalizione è in grado di garantirsi se i giocatori di A si alleano; v è detta la funzione caratteristica del gioco. Il modo migliore di capire l'idea sottostante questa definizione è di illustrarla con esempi.

ESEMPIO 10: DUE COMPRATORI E UN VENDITORE

Due persone sono interessate a un bene che è in possesso di una terza persona. Il giocatore 1, che possiede il bene, lo valuta meno di chi lo vuole comprare (altrimenti non ci sarebbe situazione di interazione tra i tre). Fissiamo, per esempio, a 100 il valore che il possessore assegna al bene. Gli altri due, che chiamiamo rispettivamente 2 e 3, valutano il bene 200 e 300. Possiamo allora definire il gioco come $N = \{1, 2, 3\}$, e le coalizioni sono otto:

¹³⁹ John Charles Harsanyi (Budapest, 1920 – Berkeley, 2000), economista statunitense di origine ungherese, vincitore del premio Nobel per l'economia nel 1994, insieme a John Nash e Reinhard Selten. Ha portato contributi fondamentali alla teoria dei giochi.

$$\{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\} = N\}^{140}$$

Possiamo inoltre porre

$$v(\{1\}) = 100, v(\{2\}) = v(\{3\}) = v(\{2, 3\}) = 0, v(\{1, 2\}) = 200, \\ v(\{1, 3\}) = v(N) = 300^{141}.$$

ESEMPIO 11: DUE VENDITORI E UN COMPRATORE

Consideriamo invece il caso di un compratore (giocatore 1) e due venditori dello stesso bene; la situazione può essere descritta efficacemente ponendo $v(A) = 1$ se $A = \{1, 2\}, \{1, 3\}, \{1, 2, 3\}$, altrimenti 0. In questo caso, quando la funzione caratteristica v assume solo valori 0 e 1, il gioco si chiama semplice, e v assume più il significato di indice di forza della coalizione (A è coalizione vincente se e solo se $v(A) = 1$). Il gioco non cambia se al posto di 1 mettiamo un altro numero positivo.

ESEMPIO 12: LA PISTA DELL'AEROPORTO, LA BANCAROTTA, LA SOCIETÀ PER AZIONI

Gli esempi 4, 5 e 6 (p. 149) sono anch'essi descrivibili come giochi cooperativi.

Nel caso della pista dell'aeroporto, v rappresenta un costo e non un'utilità. È naturale pensare che a una coalizione venga assegnato il costo della pista più lunga necessaria per le compagnie che formano la coalizione. Dunque si ha $v(\{1\}) = c_1, v(\{2\}) = c_2, v(\{3\}) = c_3, v(\{1, 2\}) = c_2, v(\{1, 3\}) = v(\{2, 3\}) = v(N) = c_3$. Il caso della bancarotta, anche se si intuisce facilmente che è un problema analogo a quello dell'aeroporto, è un pochino più complicato, perché non è chiaro a priori che cosa una coalizione possa garantire per sé; una stima molto prudente potrebbe essere quello che rimane dopo che tutti gli altri creditori sono stati pagati. Nel caso della società per azioni, siamo in presenza di un gioco semplice, e daremo valore 1 a quelle coalizioni in grado di avere la maggioranza dei voti necessaria nei vari tipi di votazioni (semplice, qualificata ecc.).

¹⁴⁰ \emptyset rappresenta l'insieme vuoto, cioè la coalizione che non contiene giocatori. Anche se può sembrare inutile, è invece opportuno tenerla in considerazione; qualunque sia v , si assume che $v(\emptyset) = 0$.

¹⁴¹ Perché abbiamo definito in questo modo il gioco? Vediamo un paio di casi. Per esempio, $v(\{2, 3\}) = 0$ perché la coalizione $\{2, 3\}$ non possiede il bene, $v(\{1, 3\}) = 300$ perché la coalizione $\{1, 3\}$ possiede il bene, che valuta 300 (infatti non se ne priva per meno).

Una generica soluzione di un gioco cooperativo con $N = \{1, 2, \dots, n\}$ come insieme di giocatori è un vettore a n componenti, ciascuna delle quali è un numero reale. Il significato dovrebbe essere chiaro: se (x_1, x_2, \dots, x_n) è tale vettore, allora x_i è l'utilità assegnata (o il costo, se v rappresenta dei costi) al giocatore i . Per fare un esempio, nel caso dei due compratori e un venditore, se proponessimo come soluzione $(100, 100, 100)$ ciò significherebbe che l'esito del gioco prevede un'utilità di 100 a testa per i tre¹⁴². Un concetto di soluzione invece rappresenta un modo per trovare vettori che soddisfano particolari proprietà. A un gioco la soluzione può associare un insieme grande di vettori, a un altro nessun vettore, ad altri ancora un solo vettore. È bene osservare che la soluzione in genere non è interessata a quanto viene assegnato alle coalizioni, ma solo a quel che è dato ai giocatori: ancora una volta va ricordato che le coalizioni sono solo un mezzo che gli individui utilizzano per ottenere il meglio per sé. L'idea di gioco cooperativo è così generale da rendere necessaria l'introduzione di molti concetti di soluzione: qui accenniamo rapidamente ad alcuni fra i più importanti. Una soluzione deve per prima cosa essere un'imputazione, cioè un vettore (x_1, \dots, x_n) tale che:

- $x_i \geq v(\{i\})$ per ogni i ;
- $x_1 + x_2 + \dots + x_n = v(N)$ ¹⁴³

Si richiede cioè a ogni soluzione di godere delle proprietà di razionalità individuale e di efficienza collettiva: ogni giocatore deve ricavare almeno quel che è in grado di garantirsi da solo (altrimenti esce dal gioco), e tutto l'utile disponibile va distribuito (e ovviamente non di più)¹⁴⁴.

Questa richiesta è quindi da ritenere minimale. In realtà, visto che le coalizioni sono possibili, sembra naturale richiedere che esse stesse gradiscano una distribuzione di utilità, altrimenti una parte dei giocatori potrebbe ritirarsi. Si arriva così a uno dei concetti fonamen-

¹⁴² Per il momento, non ci poniamo il problema se la suddivisione di utili proposta sia ragionevole. Vogliamo semplicemente capire che cosa significa soluzione in questo contesto.

¹⁴³ Per esempio, sono imputazioni i vettori $(100, 100, 100)$, nel gioco dei due compratori e un venditore (esempio 10), e $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ nel gioco dei due venditori e un compratore (esempio 11), mentre in quest'ultimo esempio non lo sono $(0, 0, 0)$ e $(1, -1, 1)$.

¹⁴⁴ Anche se non è presupposta esplicitamente, l'ipotesi che $v(N) \geq v(A)$ per ogni $A \subseteq N$ è verificata in quasi tutti i giochi interessanti. Anzi, spesso i giochi verificano l'ipotesi detta di superadditività, che cioè $v(A \cup B) \geq v(A) + v(B)$ se $A \cap B = \emptyset$, che stabilisce che l'unione fa la forza. Questo fa sì che sia ragionevole assumere che i giocatori si metteranno d'accordo per spartirsi tutta la quantità $v(N)$.

tali di soluzione: il nucleo del gioco v è l'insieme di quelle distribuzioni di utilità che nessuna coalizione ha interesse a rifiutare. D'altra parte, la coalizione A rifiuta quel che le viene proposto se la somma delle utilità proposte ai suoi giocatori è inferiore al valore $v(A)$ che, come detto, rappresenta quel che lei è complessivamente in grado di procurarsi. Per capire meglio l'idea vediamo di caratterizzare il nucleo in un esempio semplice: quello dei due venditori e un compratore (esempio 11, a p. 166): un elemento del nucleo è un vettore x fatto da tre elementi, scriviamo $x = (x_1, x_2, x_3)$. Ora scriviamo i vincoli che questo vettore deve soddisfare:

$$\begin{cases} x_1 \geq 0, x_2 \geq 0, x_3 \geq 0 \\ x_1 + x_2 \geq 1 \\ x_1 + x_3 \geq 1 \\ x_2 + x_3 \geq 0 \\ x_1 + x_2 + x_3 = 1 \end{cases}$$

La prima riga impone le disequazioni relative alle coalizioni fatte dai singoli individui: essi non accettano meno di 0. La seconda riga riguarda il vincolo imposto dalla coalizione $\{1, 2\}$; essa è in grado di garantirsi 1, quindi la somma di quel che viene proposto ai giocatori 1 e 2, cioè $x_1 + x_2$ deve essere maggiore o uguale a 1, e così via, fino all'ultima coalizione $N = \{1, 2, 3\}$. Ora, confrontando l'ultima equazione con la seconda si vede che deve essere $x_3 \leq 0$, ma la prima dice $x_3 \geq 0$, quindi $x_3 = 0$. Analogamente $x_2 = 0$. Poiché la somma delle utilità deve essere 1, allora $x_1 = 1$. Quindi il nucleo consiste del solo vettore $(1, 0, 0)$.

Vediamo ora che cosa ci propone il nucleo in alcuni dei giochi introdotti in precedenza. Nel gioco dei due compratori e un venditore (esempio 10, a p. 165), la soluzione proposta dal nucleo è che il primo vende l'oggetto al terzo (che lo valuta di più rispetto al secondo), a un prezzo che può variare fra 200 e i 300 euro, quindi il nucleo propone in questo caso più spartizioni possibili. Nel gioco invece in cui ci sono un compratore e due venditori dello stesso bene, come abbiamo visto il nucleo consiste nell'unico vettore $(1, 0, 0)$, il che significa che il compratore ottiene il bene per nulla.

È interessante notare che, nel caso dei due compratori e un venditore, il ruolo del secondo giocatore, che pure alla fine non fa nulla, è messo in evidenza dal fatto che il prezzo di vendita è influenzato dalla sua presenza. D'altra parte questo è logico: se il terzo facesse

un'offerta minore di 200 euro, allora il secondo potrebbe a sua volta fare un'offerta superiore, fino a un massimo di 200 euro. In questo caso il nucleo propone tante soluzioni possibili. Nel caso di un compratore e due venditori ciò che indica il nucleo è un fatto ben noto in economia: l'eccesso di offerta mette i venditori in balia del compratore. Infatti nel nucleo sta solo il vettore che assegna tutto al compratore, nulla ai venditori. Altre soluzioni, come vedremo, propongono una soluzione diversa, che tiene conto del fatto che in qualche modo i due venditori non sono del tutto inutili.

Un esempio ancora più interessante di come il nucleo possa proporre soluzioni bizzarre è il famoso gioco delle paia di guanti, di cui esistono infinite varianti: una versione che ne mette bene in luce la stranezza è quando si hanno 4 giocatori; il primo e il secondo possiedono 1 e 2 guanti sinistri, rispettivamente, mentre il terzo e quarto 1 guanto destro ciascuno. Lo scopo del gioco consiste nel formare paia di guanti. In questo caso il nucleo è costituito dal solo vettore $(0, 0, 1, 1)$, il che significa che i possessori di un guanto sinistro (guanti che sono in eccedenza) devono cedere il loro guanto per nulla. Risultato che appare ancora più bizzarro se si pensa che il secondo giocatore potrebbe cambiare la situazione semplicemente eliminando un guanto in suo possesso. A dispetto del fatto che a volte le soluzioni proposte dal nucleo sembrano controintuitive, esso rappresenta un concetto di soluzione molto importante, soprattutto in applicazioni economiche. Tuttavia, come concetto di soluzione, il nucleo presenta un altro problema: è facile verificare che per certi giochi può essere vuoto. L'esempio più semplice è quando tre giocatori si devono spartire a maggioranza una somma fissata (possiamo porre l'utilità pari a 1). In tal caso, le coalizioni A con almeno due giocatori sono vincenti ($v(A) = 1$), mentre per le altre $v(A) = 0$, e un calcolo immediato mostra che il nucleo è vuoto¹⁴⁵. Il che rende indispensabile la definizione di altre soluzioni, che possano suggerire possibili spartizioni anche nel caso in cui almeno una coalizione non sia soddisfatta della spartizione proposta.

Una soluzione, che qui illustriamo solo a parole, considera, per ogni possibile imputazione, il grado di insoddisfazione $e(A, x)$ della coa-

¹⁴⁵ Supponiamo (x_1, x_2, x_3) sia un vettore del nucleo. Le condizioni $x_1 + x_2 \geq 1$, $x_1 + x_3 \geq 1$, $x_2 + x_3 \geq 1$, imposte dalle coalizioni formate da due giocatori implicano, prendendo la loro somma, $2(x_1 + x_2 + x_3) \geq 3$, che è in contraddizione con la condizione di efficienza $x_1 + x_2 + x_3 = 1$. Quindi il nucleo è vuoto.

lizione A per la distribuzione dell'imputazione x : $e(A, x) = v(A) - \sum_{i \in A} x_i$. L'imputazione x sta nel nucleo, per esempio, se e solo se $e(A, x) \leq 0$ per ogni A , cioè se nessuna coalizione si lamenta. Se però il nucleo è vuoto, allora qualunque sia la distribuzione proposta c'è almeno una coalizione che si lamenta. Che fare in questo caso? Un'idea intelligente è di considerare, per ogni imputazione x , il lamento della coalizione più sfavorita (cioè di quella che si lamenta maggiormente), e poi scegliere quella distribuzione di utilità efficiente che minimizza questo lamento massimo. Se poi sono molte le distribuzioni che hanno questa proprietà, si possono scegliere quelle che minimizzano il secondo massimo lamento, e così via. Si dimostra che in questo modo si arriva a un'unica distribuzione di utilità, che viene chiamata il nucleolo del gioco.

Nel gioco dei compratori, esposto in precedenza, il prezzo di vendita è 250, e cioè il prezzo intermedio fra quello minimo e quello massimo proposti dal nucleo; nel gioco di maggioranza a tre giocatori il nucleo propone l'imputazione $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$: in questo caso ogni coalizione di due giocatori si lamenta $\frac{1}{3}$, e non è difficile verificare che ogni distribuzione di utilità diversa farebbe lamentare di più una coalizione. I risultati precedenti non sono sorprendenti, dal momento che il nucleolo è soluzione che gode di forti proprietà di simmetria; purtroppo però anche il nucleolo può dare risultati bizzarri: per esempio, siccome appartiene al nucleo, purché naturalmente questo non sia vuoto, nel gioco dei due venditori e un compratore il nucleolo assegna tutto al compratore.

Passiamo al terzo concetto di soluzione che qui consideriamo: si chiama indice di Shapley¹⁴⁶. La sua formula è un po' complicata, a una prima lettura, ma non bisogna spaventarsi. Se poi non si capiscono i dettagli, questo non impedisce a chi vuole di capire lo stesso le idee, come ha scritto Nash nella sua celebre tesi.

Dunque, intanto va osservato che questa soluzione, come il nucleolo, ha l'interessante proprietà di assegnare un'unica distribuzione di utilità a ogni giocatore. La indichiamo con S , in onore di Shapley. Risulta così definita, per un qualunque gioco v ¹⁴⁷

$$S_i(v) = \sum_{i \in A \subseteq N} \frac{(a-1)!(n-a)!}{n!} [v(A) - v(A \setminus \{i\})]$$

¹⁴⁶ Lloyd Stowell Shapley (Cambridge, Massachusetts, 1923), matematico statunitense, noto per i suoi contributi fondamentali nell'ambito della teoria dei giochi.

¹⁴⁷ Data una coalizione A , indicheremo con a la sua cardinalità, cioè il numero dei giocatori che formano la coalizione A .

L'indice di Shapley associa al giocatore i i contributi marginali¹⁴⁸ che esso porta a ogni coalizione, pesati secondo un certo coefficiente (per la coalizione $A \setminus \{i\}$ esso è $\frac{(a-1)!(n-a)!}{n!}$). Tale coefficiente ha un'interpretazione probabilistica interessante: supponendo che i giocatori decidano di trovarsi per giocare in un certo luogo e a una data ora, il coefficiente $\frac{(a-1)!(n-a)!}{n!}$ rappresenta la probabilità che i al suo arrivo trovi gli altri giocatori della coalizione A , e solo loro¹⁴⁹. Nel gioco di maggioranza semplice fra 3 giocatori, l'indice di Shapley propone $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, come il nucleolo.

Nel gioco dei guanti, invece, la soluzione è $(\frac{1}{4}, \frac{7}{12}, \frac{7}{12}, \frac{7}{12})$, vettore che presenta caratteristiche interessanti: tiene conto del fatto che c'è un eccesso di offerta di guanti sinistri, il che rende un po' più debole degli altri il giocatore 1; il secondo ne risente relativamente, perché sfrutta il fatto di poter soddisfare da solo la domanda dei giocatori con il guanto destro. Questo mostra che il valore tiene conto di altri aspetti, ignorati dal nucleo.

L'indice di Shapley ha applicazioni importanti anche nei giochi semplici. L'indice di Shapley è quindi uno dei modi possibili per valutare il potere dei giocatori in un gioco. Per concludere, ecco la risposta che dà al problema di come suddividere le spese per la costruzione della pista dell'aeroporto (esempi 4 e 12, pp. 149 e 166): il primo paga $\frac{1}{3}c_1$, il secondo $\frac{1}{2}c_2 - \frac{1}{6}c_1$, il terzo $c_3 - \frac{1}{6}c_1 - \frac{1}{2}c_2$.

Detto così non sembra molto significativo, ma per prima cosa è utile osservare che la somma dei 3 pagamenti fa proprio c_3 , il che mostra su un esempio quel che è vero sempre, e cioè che l'indice è efficiente. Inoltre, e questo è molto interessante, il risultato ha la seguente interpretazione molto naturale: il primo, che da solo spenderebbe c_1 , divide questa spesa equamente con gli altri due, che usufruiscono dello stesso servizio. Il secondo chilometro porta un costo aggiuntivo di $c_2 - c_1$: questa spesa viene equamente divisa tra gli altri due che utilizzano la pista. Il resto che manca ($c_3 - c_2$) infine è pagato dall'unico utente che ha bisogno del terzo chilometro.

Concludiamo questo paragrafo riprendendo un concetto già espresso: il fatto che esistano tante soluzioni per i giochi cooperativi non deve essere considerato segno di confusione. La varietà di situazioni che sono descritte come gioco cooperativo impone, in un certo sen-

¹⁴⁸ Il contributo marginale che il giocatore i porta alla coalizione C è la quantità $v(C \cup \{i\}) - v(C)$. Chiaramente può essere interpretato come l'apporto che il giocatore porta alla coalizione.

¹⁴⁹ Assumendo equiprobabile l'ordine d'arrivo dei giocatori.

so, che si considerino diverse soluzioni possibili. Sta a chi utilizza questi modelli scegliere quella più adatta. E nessuna soluzione è adatta a ogni gioco: per esempio, l'indice di Shapley per il gioco del venditore e dei due compratori è $(\frac{650}{3}, \frac{50}{3}, \frac{200}{3})$, a cui sembra difficile dare un significato sensato. Per questo le varie soluzioni sono caratterizzate da proprietà che servono a descriverle: abbiamo, per esempio, ricordato che l'indice di Shapley e il nucleolo godono di proprietà di simmetria, il che significa che non privilegiano alcuni giocatori rispetto ad altri.

6. Conclusioni

Naturalmente il capitolo di un libro può contenere solo una parte molto limitata delle varie idee sviluppate nell'ambito di una disciplina; nonostante questo il lettore può essersi fatto un'idea del tipo di ragionamenti alla base della teoria dei giochi. Per concludere il nostro discorso, riprendiamo due esempi introdotti all'inizio, per fare emergere qualche altra questione interessante.

Cominciamo dall'esempio dei due gruppi di agenti che devono abbinarsi, avendo ogni elemento di un gruppo un ordine di preferenze sull'altro gruppo (esempio 1 a p. 147). Il primo problema che si pone è: che cosa significa dare una soluzione a questo problema? La risposta degli esperti è stata che si deve impedire che persone abbinate in un certo modo possano ottenere un abbinamento migliore. Come sempre, un esempio può aiutare a capire. Supponiamo di aver deciso di abbinare Laura con Emanuele e Alberto con Francesca. Se però Emanuele preferisse Francesca a Laura, e allo stesso tempo Francesca preferisse Emanuele ad Alberto, la nostra proposta non sarebbe ottimale perché Francesca e Emanuele non obbediranno alla nostra raccomandazione¹⁵⁰.

Stabilito questo, la domanda successiva diventa: è vero che, qualunque sia il numero degli elementi dei due gruppi, e qualsiasi siano le loro preferenze, è possibile formare coppie in maniera ottimale? Un teorema non difficile da dimostrare ci garantisce che è proprio così. Ma non ci basta: dal momento che di solito il sistema di abbinamenti ottimali non è unico, sarebbe interessante sapere se esistono procedure che favoriscono un gruppo piuttosto che un altro. E anche in

¹⁵⁰ Questo è vero qualunque siano le preferenze di Alberto e Laura.

questo caso la risposta è positiva! Come conclusione, possiamo dire che, pur se il modello è molto semplificato (per esempio, non tiene conto che le preferenze degli individui potrebbero avere un'evoluzione nel tempo), le sue risposte sono molto interessanti.

Anche l'esempio 2 (p. 148) che parla della contrattazione merita due parole, non foss'altro perché la prima e più celebre soluzione è stata proposta da un giovanotto di genio, poco più che ventenne, di cui abbiamo già parlato, John Nash, che nel suo primo lavoro, propose un modello matematico per trattare ogni problema possibile di contrattazione e si chiese quali proprietà deve avere una soluzione ottimale di tali problemi. Dopo aver fatto un elenco di poche proprietà ragionevoli, dimostrò il notevole teorema che una sola funzione soddisfa i requisiti richiesti dall'insieme di tutti i problemi di contrattazione. Il che significa, almeno in linea teorica, aver risolto ogni problema di contrattazione.

Incidentalmente, possiamo dire che la proposta di Nash consiste nel dire ai giocatori di massimizzare il prodotto delle loro utilità, tra tutte quelle rese possibili da una distribuzione qualunque della somma in palio.

Occorre comunque osservare che l'approccio di Nash semplifica eccessivamente certi aspetti del problema: per esempio, dà per scontato che le funzioni di utilità dei giocatori siano di conoscenza comune. Questo è davvero poco realistico in moltissimi casi, tanto è vero che la gente contratta e continuerà a farlo, e uno degli accorgimenti più utilizzati (e più ovvi) è proprio quello di cercare di tenere nascosto all'altro quanto interessa l'oggetto della contrattazione.

Un'ultima parola sull'esempio 7 (p. 150). Tra gli equilibri di Nash, c'è quello che prevede che passi il procedimento *C*. La procedura per ottenerlo è che il primo dichiari di giocare *A* (che per lui è strategia debolmente dominante), e gli altri eliminino le loro ultime scelte (strategie debolmente dominate)¹⁵¹. In questo caso il gioco si riduce a una matrice 2×2 in cui gli esiti sono *A* o *C*. Poiché entrambi i giocatori 2 e 3 preferiscono *C* ad *A*, il risultato è *C*. Quindi il primo giocatore, che in teoria ha più potere degli altri, vede passare il provvedimento a lui più invisio: ha sbagliato a dichiarare in pubblico l'uso di una strategia solo debolmente dominante.

¹⁵¹ Strategia *debolmente dominata* (o *debolmente dominante*) significa che in certi casi il giocatore può essere (a seconda delle scelte degli altri) indifferente tra questa e un'altra. Certi equilibri possono prevedere l'uso di strategie debolmente dominate (ma non dominate).

Arrivati in fondo, citiamo una frase di un libro che ricorda la motivazione profonda che sta alla base dello studio e dello sviluppo della teoria dei giochi. Si riferisce agli esseri umani, ma potrebbe essere estesa a tutti i viventi. È tratta dal romanzo di Amos Oz, *Storia d'amore e di tenebra*, e recita così:

“Nessun uomo è un'isola, piuttosto siamo tutti delle penisole, circondate quasi interamente da un'acqua nera, ma comunque collegate alle altre penisole.”

LA MATEMATICA NELLE SCIENZE UMANE E DELLA VITA

Questo capitolo si colloca per certi versi in una posizione intermedia rispetto ad alcune tematiche già trattate. Da un lato l'oggetto dello studio è costituito dal mondo della vita, incluso l'uomo, il che lo avvicina al capitolo su "Giochi e applicazioni". Dall'altro, da un punto di vista puramente matematico, gli strumenti impiegati sono più classici di quelli richiesti nel capitolo precedente e rimandano a quanto già incontrato nel capitolo su "Modelli e previsioni". Lo scopo è quello di descrivere, e in qualche modo predire, l'evoluzione nel tempo di certe specie viventi, incluso l'uomo, quando sono poste in determinate condizioni.

1. La crescita delle popolazioni

Supponiamo di voler descrivere e predire la crescita nel tempo di una popolazione. Storicamente la questione fu affrontata per la prima volta in termini quantitativi dall'inglese Thomas Robert Malthus¹⁵², che, in qualità di demografo, intendeva studiare la dinamica delle popolazioni. In seguito il suo modello si è rivelato applicabile, entro certi limiti, anche a "popolazioni" di ben altro tipo, come per esempio determinate specie viventi in un certo habitat oppure i batteri presenti in un organismo. Di conseguenza, le applicazioni dei modelli sull'evoluzione delle popolazioni spaziano dalla demografia all'ecologia fino alla medicina.

Più precisamente, denotiamo con $N(t)$ la numerosità della popolazione in questione all'istante di tempo t . Partendo dall'osservazione dell'andamento demografico nelle colonie inglesi della Nuova Inghilterra, Malthus arrivò a formulare un'equazione del tipo

¹⁵² Thomas Robert Malthus (Roocherry, 1766 – Bath, 1834), economista e demografo inglese. Nel 1798 pubblica il *Saggio sul principio della popolazione*, di grande impatto sugli studi economici successivi e non solo, avendo anche influenzato la teoria dell'evoluzione di Darwin.

$$N(t + \tau) - N(t) = \rho N(t) \quad (1)$$

da intendersi soddisfatta per ogni t . La costante $\tau > 0$ denota il tempo necessario a che i meccanismi riproduttivi diano il loro effetto. L'idea è che l'incremento di popolazione, dopo un intervallo di tempo di ampiezza τ , sarà proporzionale, secondo il fattore ρ , al numero di individui all'inizio dell'intervallo stesso. In effetti risulta più significativo esprimere la costante ρ in un modo leggermente diverso. Supponiamo per esempio di avere una prima popolazione che parte con $N_1(0) = 100$ ed è caratterizzata da $\tau_1 = 6$ mesi e $\rho_1 = 1/5$. Poiché

$$N(t + \tau) = (1 + \rho)N(t)$$

misurando il tempo in mesi risulta $N_1(6) = 120$, quindi $N_1(12) = 144$. Supponiamo ora di voler paragonare questa crescita con quella di un'altra popolazione che parte pure con $N_2(0) = 100$, ma è caratterizzata da $\tau_2 = 12$ mesi. Se immaginiamo di avere $\rho_2 = \rho_1 = 1/5$, risulta $N_2(12) = 120$, che è molto diverso da $N_1(12) = 144$. In altre parole, con τ diversi, i coefficienti ρ si paragonano male fra di loro. Se invece fosse $\rho_2 = 2/5$, si avrebbe $N_2(12) = 140$, che è paragonabile con $N_1(12)$.

Questo induce a ritenere che una buona pietra di paragone possa essere costituita dal rapporto ρ/τ : se τ si raddoppia, ottengo effetti paragonabili raddoppiando anche ρ . Denotiamo con r il rapporto ρ/τ , che si chiama tasso (*rate* in inglese) di crescita. Con questa riformulazione, la (1) diventa

$$N(t + \tau) - N(t) = r\tau N(t) \quad (2)$$

Beninteso, l'equazione (2) è assolutamente equivalente alla (1). L'unica differenza è che propone l'uso di un'altra costante, r invece di ρ , che è più adatta ai confronti, allorché vengono coinvolte crescite con τ diversi. D'altra parte molte informazioni in ambito scientifico, e non solo, vengono ricavate dai paragoni, per cui è conveniente disporsi in modo che questi siano il più trasparenti possibile. Naturalmente il valore $r = 0$ comporta $N(t + \tau) = N(t)$ e corrisponde alla crescita zero.

Assegnata la numerosità a un certo istante, diciamo per comodità 0, è possibile determinare la numerosità a un tempo successivo t ,

corrispondente a n iterazioni del tempo τ ? In altre parole, noto $N(0)$, è possibile determinare $N(t)$ con $t = n\tau$?

Tanto per cominciare, la (2) si può riscrivere nella forma

$$N(t + \tau) = (1 + r\tau)N(t)$$

da cui, scegliendo $t = 0$,

$$N(\tau) = (1 + r\tau)N(0)$$

Scegliendo invece $t = \tau$ e tenendo conto di quanto appena ottenuto, risulta

$$N(2\tau) = (1 + r\tau)N(\tau) = (1 + r\tau)^2 N(0)$$

Proseguendo con la scelta $t = 2\tau$, si perviene a

$$N(3\tau) = (1 + r\tau)N(2\tau) = (1 + r\tau)^3 N(0)$$

A questo punto non è difficile immaginare che la formula generale sia

$$N(n\tau) = (1 + r\tau)^n N(0)$$

ovvero, ricordando che $t = n\tau$,

$$N(t) = (1 + r\tau)^{t/\tau} N(0) \quad (3)$$

Sulla base della (3), Malthus formulò la previsione, essenzialmente corretta, che in presenza di risorse praticamente illimitate – questa era la situazione nelle colonie inglesi della Nuova Inghilterra – la popolazione sarebbe cresciuta in progressione geometrica. Lo stesso tipo di evoluzione si riscontra, se $N(0)$ e t non sono così grandi da far emergere il problema della limitatezza delle risorse, in molti altri casi: per esempio, nella riproduzione delle cellule di un embrione nel primo periodo di vita.

Ritornando alla (3), si può osservare che il rapporto t/τ fa entrare in gioco il confronto fra due diverse scale di tempi. In molti casi il tempo t , a cui si vuole effettuare una previsione, è nettamente maggiore del tempo τ . In altre parole τ risulta trascurabile rispetto a t . Ci si può allora domandare se la formula (3) non sia approssimabile effi-

cacemente con una formula più semplice e maneggevole. Ricordando che per ogni $x > 0$ si ha

$$x = e^{\ln x}$$

per le proprietà del logaritmo risulta

$$(1 + r\tau)^{t/\tau} = e^{\ln[(1+r\tau)^{t/\tau}]} = e^{\frac{t}{\tau} \ln(1+r\tau)} = e^{\frac{\ln(1+r\tau)(\tau/t)}{\tau/t}}$$

Quando τ/t si avvicina a 0, il valore di

$$\frac{\ln(1 + r\tau(\tau/t))}{\tau/t}$$

tende a confondersi con $r\tau$. Detto in modo più preciso, risulta

$$\lim_{x \rightarrow 0} \frac{\ln(1 + rtx)}{x} = r\tau$$

Allora diventa plausibile approssimare la (3) con

$$N(t) = e^{rt} N(0) \tag{4}$$

Quest'ultima formula è meno elementare della (3) sia per la presenza del numero di Nepero¹⁵³ sia per il fatto che adesso l'esponente rt non ha motivo di essere ritenuto intero, ma è più maneggevole, anche perché contiene un parametro in meno (è sparito τ). Un'ulteriore approssimazione consiste nell'immaginare che $N(t)$ possa assumere valori reali positivi, anziché soltanto interi. Se $N(t)$ è un numero abbastanza grande, il passaggio regge, perché, per esempio, un valore come $N(t) = 100,4$ può essere assimilato a 100, con un errore percentuale piccolo. Invece un valore come $N(t) = 1,4$ sarebbe problematico, perché fra 1,4 e 1 lo scarto percentuale è grande.

Anche il passaggio dai numeri naturali a quelli reali positivi segna l'abbandono di una struttura elementare, a favore di una più maneggevole (nell'ambito dei numeri reali positivi si può sempre fare la divisione e molte altre cose). Non di rado, in matematica, elementa-

¹⁵³ Giovanni Nepero, nome italianizzato di John Napier (Merchiston Castle, 1550 – Edimburgo, 1617), matematico, astronomo e fisico scozzese. Nel 1614 pubblicò la *Mirifici logarithmorum canonis descriptio* (*Descrizione della regola meravigliosa dei logaritmi*), in cui tra l'altro introduceva il numero e e il logaritmo naturale, che semplificano diverse formule riguardanti limiti e sviluppi in serie.

re e maneggevole non sono affatto sinonimi: ciò che è elementare non è maneggevole, mentre la maneggevolezza si ottiene con costrutti matematici non elementari. È anche conveniente collegare r con il cosiddetto tempo di raddoppiamento T_2 , caratterizzato dalla proprietà che $N(T_2) = 2N(0)$. Risulta

$$2N(0) = N(T_2) = e^{rT_2} N(0)$$

da cui

$$r = \frac{\ln 2}{T_2}$$

relazione assai utile per ricavare r dai dati sperimentali, allorché il modello è applicabile.

Ricordando che la formula (3) proveniva dalla risoluzione dell'equazione (2), ci si può domandare se la formula approssimante (4) non rappresenti essa stessa la soluzione di una qualche equazione, a sua volta approssimante la (2). Per capirlo, riscriviamo la (2) nella forma

$$\frac{N(t + \tau) - N(t)}{\tau} = rN(t) \quad (5)$$

Così come la (4) era stata ottenuta dalla (3) attraverso un processo di limite allorché τ/t si avvicinava a zero, in questo caso è naturale ipotizzare, come approssimazione della (2), un'equazione (differenziale) del tipo

$$N'(t) = rN(t) \quad (6)$$

in cui $N'(t)$ rappresenta il valore con cui può essere confuso il primo membro della (5), allorché τ è vicino a 0. Detto in modo più preciso,

$$N'(t) = \lim_{\tau \rightarrow 0} \frac{N(t + \tau) - N(t)}{\tau}$$

Quest'ultimo passaggio consente di “quadrare il cerchio”, nel senso che la funzione definita dalla (4) è effettivamente la soluzione dell'equazione (6) che all'istante 0 assume il valore $N(0)$. Data la forma della soluzione espressa nella (4), il modello di Malthus si chiama anche “modello esponenziale”.

Siccome le equazioni differenziali come la (6) sono meno elementari della (2), in quanto occorrono limiti e derivate, ma molto più maneggevoli, è ormai diventata prassi descrivere i fenomeni di crescita con modelli differenziali in cui la numerosità $N(t)$ può assumere valori reali. Si tratta di un caso, molto frequente, di approssimazione del discreto per mezzo del continuo. Va infine osservato che, nella molteplicità di situazioni in cui il modello trova applicazione, esistono anche casi nei quali è plausibile considerare tassi di crescita negativi, diciamo $r = -\mu$ con $\mu > 0$. Rientra in tale categoria il decadimento radioattivo, per esempio, del carbonio-14. In tale contesto è naturale collegare r , ovvero μ , non al tempo di raddoppiamento, ma al tempo di dimezzamento $T_{1/2}$. Risulta

$$\frac{1}{2} N(0) = N(T_{1/2}) = e^{-\mu T_{1/2}} N(0)$$

da cui

$$\mu = \frac{\ln 2}{T_{1/2}}$$

2. Popolazioni e ambiente

Il modello di Malthus, come si è detto, è plausibile in presenza di risorse praticamente illimitate e descrive correttamente molte situazioni, finché si trovano nella fase iniziale. Quando la numerosità della popolazione ha però raggiunto un certo livello, si rende necessario un modello un po' più complesso, affinché la descrizione sia un minimo plausibile. Il primo contributo in questa direzione è stato dato dallo statistico belga P.-R. Verhulst¹⁵⁴. Egli ha proposto di descrivere l'evoluzione nel tempo di $N(t)$ attraverso l'equazione differenziale

$$N'(t) = r \left(1 - \frac{N(t)}{M} \right) N(t) \quad (7)$$

in cui $r > 0$ e $M > 0$ è un ulteriore parametro. Anzitutto si dà per scontato che il modello sia di tipo differenziale e che la funzione $N(t)$

¹⁵⁴ Pierre-François Verhulst (Bruxelles, 1804 – Bruxelles, 1849), matematico e statistico belga. Nel 1838 pubblicò il saggio *Informazione sulla legge di crescita della popolazione*.

sia a valori reali positivi. Pertanto la (7) va paragonata con la (6). In sostanza il tasso di crescita costante r viene sostituito con la funzione

$$r \left(1 - \frac{N(t)}{M} \right)$$

che per valori piccoli di $N(t)$ si confonde con r , ma al crescere di $N(t)$ diminuisce e, oltrepassato il valore M , detto “popolazione massima sostenibile”, diventa addirittura negativo. Questo significa che, se la numerosità si trova al di sopra del valore M , le risorse non sono sufficienti e la popolazione decresce. L'equazione (7) si chiama equazione differenziale logistica e, se $N(0) > 0$, la soluzione al tempo $t \geq 0$ è

$$N(t) = \frac{e^{rt} N(0)}{1 + (N(0)/M)(e^{rt} - 1)} \tag{8}$$

In Figura 1 sono rappresentati, nel caso $r = 1/2$, $M = 500$ e $N(0) = 50$, gli andamenti di $N(t)$ secondo il modello esponenziale di Malthus e quello logistico di Verhulst. È anche indicato il valore $M = 500$, a cui tende asintoticamente $N(t)$ nel modello logistico. Dopo una prima fase in cui le due previsioni differiscono di poco, il modello esponenziale contempla una crescita sempre più accentuata, mentre in quello logistico si tende a una stabilizzazione verso il valore di popolazione massima sostenibile.

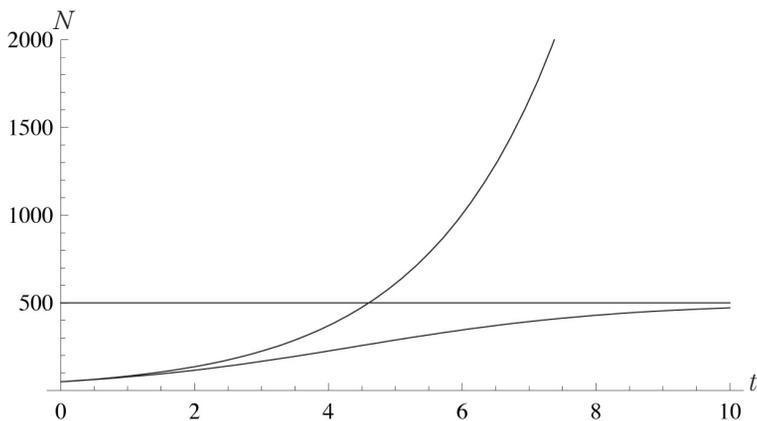


Figura 1. Gli andamenti previsti dal modello esponenziale e da quello logistico.

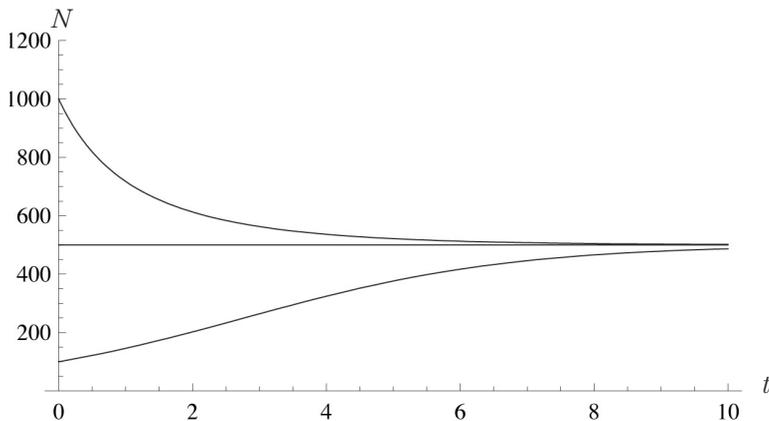


Figura 2. Alcune soluzioni dell'equazione logistica.

In Figura 2 sono rappresentate, sempre nel caso $r = \frac{1}{2}$ e $M = 500$, le soluzioni dell'equazione logistica corrispondenti a $N(0) = 100$, $N(0) = 500$ e $N(0) = 1000$. La soluzione con $N(0) = M$ è una costante a cui si avvicinano asintoticamente tutte le soluzioni con $N(0) > 0$.

In realtà il modello logistico fornisce previsioni attendibili per tempi maggiori rispetto a quello esponenziale, ma non sempre per il lungo periodo.

Il problema principale è che nel modello logistico il valore M di popolazione massima sostenibile è presupposto come costante, mentre nella realtà è una funzione del tempo che può variare anche non di poco. Per esempio, le previsioni sulla popolazione massima sostenibile per l'intera Terra, basate sul modello logistico, si sono rivelate finora sempre errate per difetto.

Ciononostante, il modello logistico costituisce un utile strumento di analisi e rappresenta il punto di partenza per l'elaborazione di modelli più sofisticati.

Come si è visto, ogni soluzione con $N(0) > 0$ cresce verso il valore M . Si possono immaginare situazioni più complesse in cui, oltre alla limitatezza delle risorse, la popolazione debba anche fronteggiare una presenza di predatori, o di malattie, tale da portare all'estinzione, se la consistenza numerica della popolazione stessa è troppo bassa.

Questo ulteriore effetto, detto "depensazione" o "effetto Allee", è tenuto in considerazione nell'equazione

$$N'(t) = r \left(1 - \frac{N(t)}{M} \right) \left(\frac{N(t)}{m} - 1 \right) N(t) \tag{9}$$

Si intende che $r > 0$ e $0 < m < M$. Il tasso di crescita è la funzione

$$r \left(1 - \frac{N(t)}{M} \right) \left(\frac{N(t)}{m} - 1 \right)$$

che è positiva se $m < N(t) < M$. Ma se $N(t) > M$, il tasso diventa negativo per l'effetto della limitatezza di risorse già considerato. Se $N(t) < m$ – e questa è la novità – il tasso è ugualmente negativo, perché l'azione dei predatori (o delle malattie) prevale sulla capacità di riproduzione della popolazione.

Nella Figura 3 sono rappresentate, nel caso $r = 2/5$, $m = 80$ e $M = 500$, le soluzioni corrispondenti a $N(0) = 60, 80, 100, 500, 600$. La soluzione con $N(0) = M$ è di nuovo una costante a cui si avvicinano asintoticamente tutte le soluzioni con $N(0) > m$. Invece le soluzioni con $N(0) < m$ decrescono e asintoticamente prefigurano un'estinzione della popolazione. La soluzione con $N(0) = m$ è pure una costante, ma è asintoticamente instabile, perché piccolissime variazioni di $N(0)$, da poco sotto a poco sopra m , inducono nel lungo periodo conseguenze drasticamente diverse, dall'estinzione alla stabilizzazione sul valore di popolazione massima sostenibile.

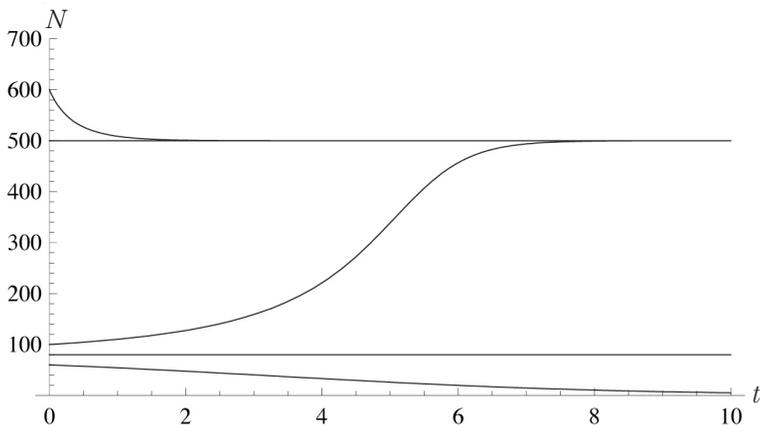


Figura 3. Alcune soluzioni in presenza di effetto Allee.

3. Cinetica dei farmaci

Nei due precedenti paragrafi abbiamo considerato situazioni descritte per mezzo di un'unica popolazione. Vediamo ora un altro problema, di origine medica, riconducibile a un modello con due popolazioni. Si tratta di descrivere la presenza nel corpo umano di un farmaco assunto periodicamente per via orale. Più precisamente, il farmaco viene assunto secondo il dosaggio $D(t)$ ed è rilasciato nel tratto gastrointestinale, da dove passa nel flusso sanguigno, nel quale viene metabolizzato e infine eliminato. Le due concentrazioni $I(t)$ e $S(t)$, con cui il farmaco è presente nel tratto gastrointestinale e nel sangue, costituiscono le due "popolazioni" di cui si vuole descrivere l'evoluzione in funzione del dosaggio $D(t)$. Un semplice modello consiste nel sistema di equazioni differenziali

$$\begin{cases} I'(t) = -\mu_I I(t) + D(t) \\ S'(t) = -\mu_S S(t) + \mu_I I(t) \end{cases} \quad (10)$$

in cui μ_I e μ_S sono due costanti positive. Si tratta sempre di una variante del modello di Malthus (6).

La prima equazione dice che il farmaco presente nel tratto gastrointestinale tende naturalmente a diminuire con tasso costante μ_I , dal momento che viene assorbito nel flusso sanguigno. Tale effetto è però in qualche misura controbilanciato dal fatto che dall'esterno viene assunto altro farmaco con il dosaggio $D(t)$. La seconda equazione dice che anche il farmaco presente nel flusso sanguigno tende naturalmente a diminuire con tasso costante μ_S , dal momento che viene metabolizzato ed eliminato. Tuttavia tale diminuzione va contenuta con l'apporto di farmaco che il flusso sanguigno riceve dal tratto gastrointestinale. Qui si immagina che il farmaco che lascia il tratto gastrointestinale vada tutto nel flusso sanguigno, per cui lo stesso termine $\mu_I I(t)$ compare, ovviamente con segni opposti, nelle due equazioni.

Supponiamo che il farmaco sia assunto ogni 6 ore e ogni volta in quantità unitaria, e che, una volta ingerito, sia tutto rilasciato nel tratto gastrointestinale in modo uniforme nell'arco di 2 ore. Tutto questo può essere descritto con il grafico di Figura 4. In sostanza si alternano 2 ore di rilascio costante e 4 ore senza alcun rilascio. L'area di ciascun rettangolo è unitaria, a rappresentare il fatto che ogni

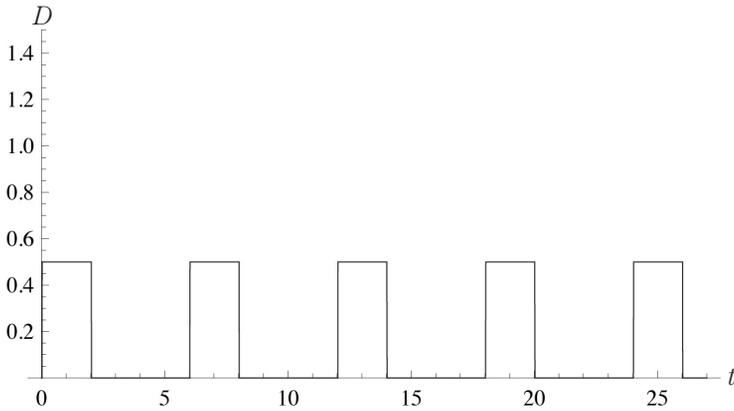


Figura 4. Il grafico di un possibile dosaggio.

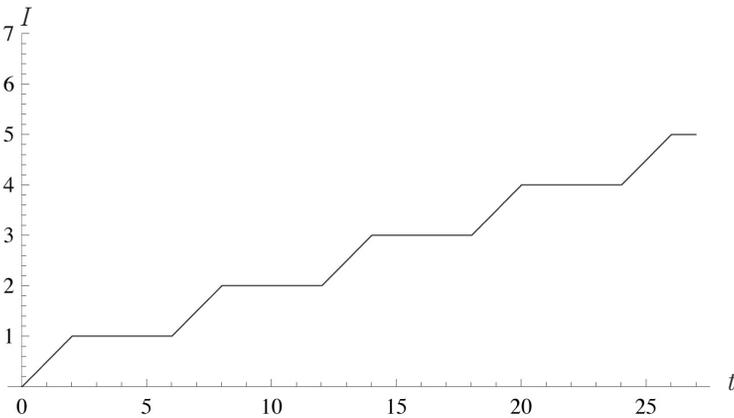


Figura 5. L'effetto della somministrazione, se ci fosse puro accumulo.

somministrazione è in quantità unitaria. Se nel tratto gastrointestinale ci fosse puro accumulo, ossia se si avesse $\mu_I = 0$, l'andamento di $I(t)$ con $I(0) = 0$ sarebbe quello indicato in Figura 5 in cui, al termine di ogni rilascio, il valore di $I(t)$ cresce di 1. Nelle Figure 6 e 7 sono invece indicati gli andamenti di $I(t)$ e $S(t)$ nel caso

$$\mu_I = 2 \ln 2, \quad \mu_S = \frac{\ln 2}{5}, \quad I(0) = 0, \quad S(0) = 0$$

mentre nella Figura 8 i tre grafici di $D(t)$, $I(t)$ e $S(t)$ sono sovrapposti. Si può notare come l'andamento di $I(t)$ sia collegato a quello di $D(t)$. Nella 2 ore di rilascio del farmaco $I(t)$ cresce, ma secondo una curva con concavità verso il basso anziché secondo una retta, perché immediatamente una parte del farmaco è assorbita nel flusso sanguigno. Per lo stesso motivo $I(t)$ non raggiunge il valore 1, come avverrebbe se non ci fosse passaggio nel sangue. Nelle 3 ore successive, $I(t)$ decade con tempo di dimezzamento di mezz'ora, dato il valore di μ_I , perché il farmaco viene quasi tutto assorbito. Nella sesta ora il farmaco è

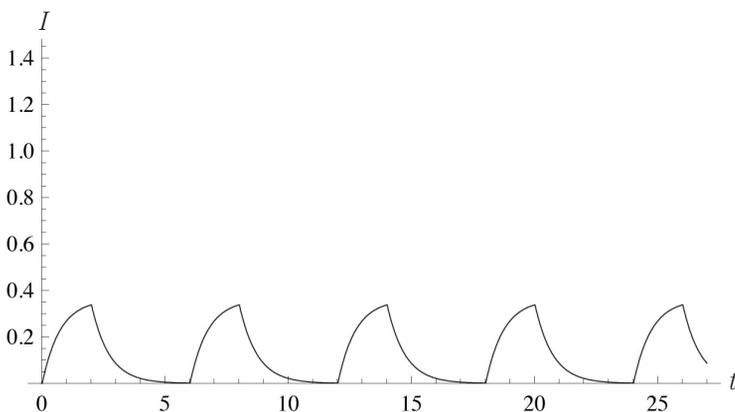


Figura 6. La concentrazione del farmaco nel tratto gastrointestinale.

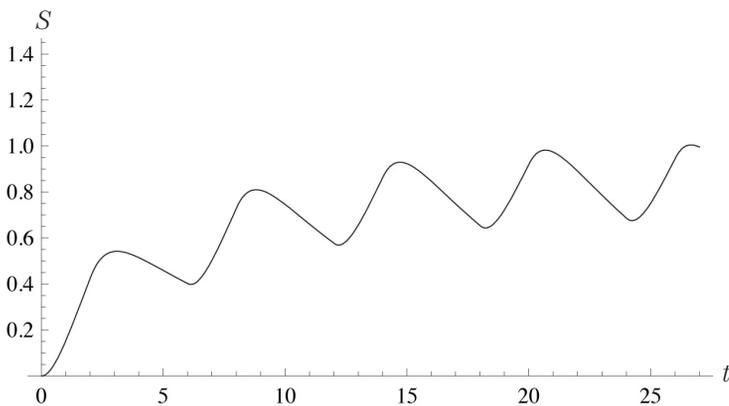


Figura 7. La concentrazione del farmaco nel sangue.

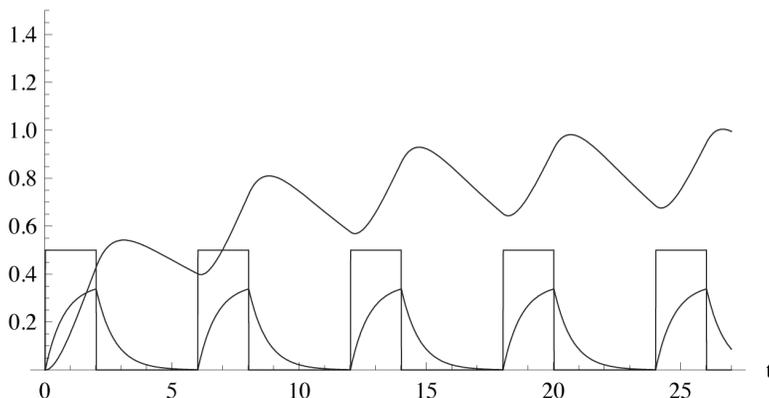


Figura 8. I grafici di $D(t)$, $I(t)$ e $S(t)$ sovrapposti.

praticamente assente nel tratto gastrointestinale, dopodiché parte un altro rilascio. Il valore scelto per μ_S comporta invece che nel sangue la metabolizzazione e l'eliminazione producano un decadimento, in assenza di apporti, con un tempo di dimezzamento di 5 ore. Di conseguenza il comportamento di $S(t)$, dopo una fase transitoria iniziale, dovuta al fatto che si parte con $S(0) = 0$, tende a stabilizzarsi su un andamento oscillatorio con un valore minimo ben positivo. Il fatto che μ_S sia abbastanza piccolo fa sì che una certa quantità di farmaco sia sempre presente nel sangue. L'ideale, in molti casi, è ottenere una funzione $S(t)$ che si stabilizzi su un valore quasi costante.

4. Predatori e prede

Nel secondo paragrafo, a p. 182, abbiamo contemplato la possibilità che la crescita della popolazione sia condizionata dalla presenza di predatori. Nel modello proposto, non era però previsto alcun effetto di ritorno sui predatori stessi. Uno sviluppo interessante, dovuto indipendentemente a Alfred J. Lotka¹⁵⁵ e Vito Volterra¹⁵⁶, si ottiene immaginando di dover descrivere contemporaneamente sia

¹⁵⁵ Alfred James Lotka (Lemberg, 1880 – New York, 1949), matematico, statistico e chimico fisico statunitense. È noto per i suoi lavori sulla dinamica delle popolazioni e l'energetica.

¹⁵⁶ Vito Volterra (Ancona, 1860 – Roma, 1940), matematico e fisico italiano. Fu uno dei fondatori dell'analisi funzionale e della connessa teoria delle equazioni integrali. Il suo nome è noto soprattutto per i suoi contributi alla biologia matematica.

la popolazione dei predatori, sia quella iniziale, diciamo delle prede. Denotata con $P(t)$ la numerosità dei predatori e con $N(t)$ quella delle prede, l'evoluzione è caratterizzata dal sistema di equazioni differenziali

$$\begin{cases} N'(t) = r_N N(t) - i_N N(t)P(t) \\ P'(t) = -\mu_P P(t) + i_P N(t)P(t) \end{cases} \quad (11)$$

in cui r_N , i_N , μ_P e i_P sono costanti positive. A questo livello, alcune drastiche semplificazioni sono state messe in atto. Tanto per cominciare, s'immagina che le prede dispongano di risorse illimitate, per cui la prima equazione appare come una variante del modello di Malthus (6) in cui è aggiunto un termine negativo a indicare che le prede soffrono dell'interazione prede-predatori quantificata dal prodotto $N(t)P(t)$.

A loro volta i predatori sono ritenuti assai più famelici che prolifici, tant'è che la loro capacità riproduttiva viene considerata trascurabile, mentre viene esaltato, attraverso il coefficiente negativo $-\mu_P$, il fatto che soffrono la competizione per le risorse.

L'unica loro speranza di sussistenza è costituita dall'interazione con le prede, rappresentata di nuovo dal termine $N(t)P(t)$, che nella seconda equazione interviene quindi con coefficiente positivo. Il sistema (11) non è risolvibile esplicitamente, ma è possibile scrivere una interessante relazione implicita che collega, per ogni t , $N(0)$, $P(0)$, $N(t)$ e $P(t)$. Risulta infatti

$$\begin{aligned} i_P [N(t) - N(0)] - \mu_P \ln \frac{N(t)}{N(0)} + \\ + i_N [P(t) - P(0)] - r_N \ln \frac{P(t)}{P(0)} = 0 \end{aligned} \quad (12)$$

Si parla in questo caso di descrizione nel piano delle fasi. Nella Figura 9 sono rappresentate le coppie $(N(t), P(t))$ che soddisfano la (12) nel caso

$$\begin{aligned} r_N = \frac{2}{5}, \quad i_N = \frac{1}{125}, \quad \mu_P = 2, \quad i_P = \frac{1}{50}, \\ N(0) = 100, \quad P(0) = 10 \end{aligned}$$

Di conseguenza, l'ascissa denota la numerosità delle prede e l'ordinata quella dei predatori. L'evoluzione nel tempo della coppia $(N(t), P(t))$ può essere seguita percorrendo la curva in senso antiorario. All'istante $t = 0$ ci sono 100 prede e 10 predatori, il che corrisponde al punto più in basso della curva. In queste condizioni, le prede crescono in numero, essendovi pochi predatori, i quali pure aumentano di numero, avendo prede a sufficienza. In questo modo però aumenta anche l'interazione, che avvantaggia i predatori e penalizza le prede, finché non si arriva a un certo momento in cui i predatori raggiungono il valore

$$P = \frac{r_N}{i_N} = 50$$

corrispondente a quasi 170 prede (il punto più a destra della curva). Qui c'è una prima inversione di tendenza, perché i predatori continuano a crescere, data l'abbondanza di prede, mentre queste incominciano a diminuire per un eccesso di predatori. Si prosegue così finché il numero delle prede non è calato al valore

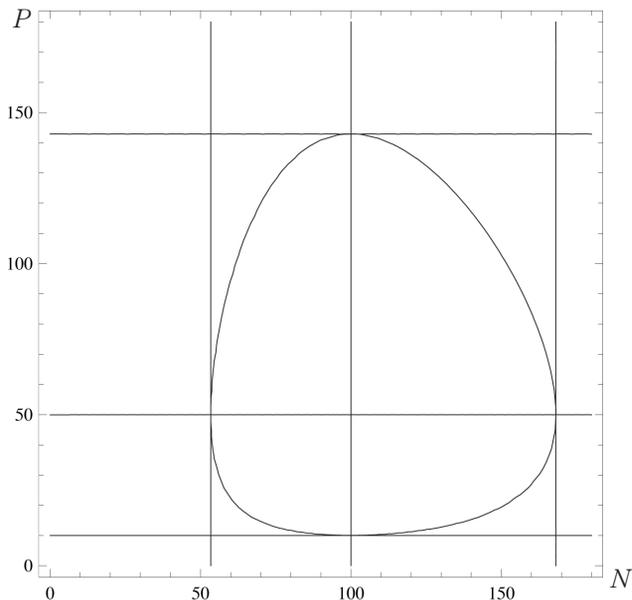


Figura 9. Una rappresentazione dell'insieme delle coppie $(N(t), P(t))$ per il sistema di Lotka-Volterra.

$$N = \frac{\mu_P}{i_P} = 100$$

corrispondente a un po' più di 140 predatori (il punto più in alto della curva). Qui si realizza una seconda inversione di tendenza, perché con poche prede e molti predatori le prede continuano a diminuire perché vittime dei predatori, che però iniziano a loro volta a calare per insufficienza di prede. Il fenomeno prosegue finché i predatori non sono ridiscesi al livello

$$P = \frac{r_N}{i_N} = 50$$

corrispondente questa volta a poco più di 50 prede (il punto più a sinistra della curva). È questo il momento in cui, con una terza inversione di tendenza, i predatori continuano a diminuire per scarsità di prede, mentre queste ultime possono ricominciare a crescere, essendosi ridimensionato il numero dei predatori. Si continua finché le prede non sono risalite al livello

$$N = \frac{\mu_P}{i_P} = 100$$

corrispondente questa volta a 10 predatori (il punto più in basso della curva). Con una quarta inversione di tendenza, le prede continuano ad aumentare per scarsità di predatori, che però possono riprendere a crescere, essendovi ora prede a sufficienza. In effetti, in questa quarta inversione di tendenza si riproducono esattamente i valori iniziali (100 prede e 10 predatori), per cui l'evoluzione prosegue ripercorrendo indefinitamente le fasi già descritte.

Anche il modello di Lotka-Volterra risulta essere estremamente semplificato. Tuttavia rappresenta il punto di partenza per l'elaborazione di altre analisi più sofisticate.

BIBLIOGRAFIA

Costantini D. e Monari P. (a cura di), *Le regole matematiche del gioco d'azzardo. Perché il banco non perde mai?*, Muzzio, 1996.

Feynman R., *La legge fisica*, Bollati Boringhieri, Torino, 1993.

Gamba E. e Montebelli V., *Piero della Francesca matematico*, Le Scienze n. 331 marzo 1996.

Lucchetti R., *Di duelli, scacchi e dilemmi: la teoria matematica dei giochi*, Bruno Mondadori, Milano 2008.

Naldi G., Pareschi L. e Russo G., *Introduzione al Calcolo scientifico*, McGraw-Hill, Milano, 2001.

Ossermann R., *Poesia dell'universo*, Longanesi, Milano, 2000.

Quarteroni A. e Saleri F., *Introduzione al calcolo scientifico*, Springer, Milano, 2004.

Trudeau R., *La rivoluzione non euclidea*, Bollati Boringhieri, Torino, 2004.

Vulpiani A., *Determinismo e caos*, Carocci, 2004.

ELENCO DEI SIMBOLI

\mathbb{N}	insieme dei numeri naturali
π	pi greco
\mathbb{R}	insieme dei numeri reali
g_{ij}	tensore metrico
\cdot	derivazione rispetto al tempo
$\ddot{\cdot}$	doppia derivazione rispetto al tempo
grad	gradiente
Δ	laplaciano
$B A$	evento B dato l'evento A
∇	simbolo alternativo per denotare il gradiente
div	divergenza
φ	fi, numero legato al rapporto aureo
\circ	composizione fra trasformazioni
e	numero di Nepero

INDICE ANALITICO

- Accordi vincolanti 148, 165
 Albero del gioco 153
 Algoritmo 19
 Anisotropia 109-110
 Aritmetizzazione dell'analisi 17
 Asintoticamente instabile 183
 Assi cartesiani 20
 Assiomi di Peano 17
 Astragalo 65
 Automi cellulari 114
- Bancarotta 149
 Battaglia dei sessi 150, 161
 Bimatrice 150
 Birapporto 130
 Bolle di sapone 99
- Calcolo delle variazioni 103
 Calcolo numerico 97, 110
 Calcolo simbolico 97
 Canali semicircolari 19
 Coalizione 165
 Comma sintonico o di Didimo 139
 Comportamento razionale 146
 Computer 93
 Concetto di soluzione (di gioco cooperativo) 167
 Continuità della retta reale 23
 Continuum spazio-temporale 113
- Contrattazione 148, 173
 Contributo marginale 171
 Convessità 109
 Curva di von Koch 56
 Curvatura gaussiana 30
 Curvatura media 100
 Curvature principali 31, 102
- Depensazione 182
 Dilemma del prigioniero 148, 161
 Dimensione 19
 Discretizzazione 111
- Economie di scala 149
 Effetto Allee 182
 Efficienza collettiva 167
 Elaborazione di immagini 115
 Equazione di Eulero 47
 Equazione di Eulero-Lagrange 104
 Equazione di Navier-Stokes 50, 106
 Equazione differenziale logistica 181
 Equazioni alle derivate parziali 48, 113
 Equazioni di Lotka-Volterra 188
 Equilibrio di Nash 162
 Errore 35

- Errore di arrotondamento 94
 Errore di discretizzazione 111
 Eventi indipendenti 88
 Evento 74

 Film di animazione 121
 Focchi di neve 107
 Formula di Bayes 89
 Frazioni 14
 Frequenza relativa 70
 Funzione caratteristica 165

 Geodetiche 28
 Geometria iperbolica 31, 133
 Geometria sferica 31
 Giochi a somma zero 151
 Giochi NTU 165
 Giochi strettamente
 competitivi 161
 Giochi TU 165
 Gioco dei guanti 169, 171
 Gioco del tris 155
 Gioco determinato 157
 Gioco semplice 166
 Grafo 152

 Imputazione 167
 Indice di Shapley 171-172
 Indici di potere 150
 Induzione a ritroso 155, 157
 Infinito 11
 Infinito attuale 16
 Intervallo di ottava 137
 Intervallo di quarta 139
 Intervallo di quinta 137
 Intervallo di terza maggiore
 139
 Intervallo di terza minore 139
 Intorno 98
 Inviluppo convesso 109

 Linux 121
 Livelli di sicurezza 159

 Matrice 151
 Meccanica newtoniana 37
 Misurazione 36
 Modello 36
 Modello discreto 111
 Modello esponenziale 179
 Modello logistico 181
 Morra cinese (gioco della) 151,
 157
 Musica dodecafonica 144

 Nodi dell'albero 153
 Notazione scientifica 93
 Nucleo 168-170
 Nucleolo 171
 Nucleolo del gioco 170
 Numero algebrico 23
 Numero decimale 17
 Numero naturale 14
 Numero razionale 14
 Numero reale 16
 Numero trascendente 23

 Operazioni elementari 93
 Ospedali e medici interni 147
 Ottimizzazione 97

 Piano delle fasi 188
 Pista dell'aeroporto 149, 166
 Popolazione massima
 sostenibile 181
 Probabilità classica 67
 Probabilità condizionata 87
 Probabilità frequentista 70
 Probabilità soggettiva 83
 Problema di Plateau 100
 Proiezione di Fuller 28

- Proiezione di Mercatore 28
 Prospettiva 127
 Punti di Steiner 105
 Punti stazionari 98
 Punto singolare 103

 Rami (dell'albero) 153
 Ray tracing 120
 Razionalità individuale 167
 Relatività generale 24
 Restauro di immagini 115
 Rettangolo aureo 123
 Riconoscimento di immagini 115

 Scacchi 152, 157
 Scuola pitagorica 22, 137
 Segmentazione di immagini 115
 Semitono pitagorico 139
 Simmetria 130
 Sintesi di immagini 117
 Sistema di riferimento 20
 Sistema mesotonico 141
 Sistema pitagorico 139
 Società per azioni 149, 166
 Software libero 118
 Software open source 118
 Soluzione
 (di gioco cooperativo) 163
 Soluzione approssimata 110
 Statistica 77
 Strategia 154
 Strategia dominante 164

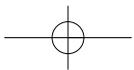
 Strategia dominata 164
 Strategie debolmente dominate/dominanti 173
 Strategie miste 160
 Strategie pure 160
 Superadditività 167
 Superfici minime 100

 Tassellatura di Penrose 132
 Tasso di crescita 176
 Temperamento equabile 144
 Temperamento Werckmeister 143
 Tempo di dimezzamento 180
 Tempo di raddoppiamento 179
 Tensione superficiale 99, 108
 Tensore metrico 30
 Teorema di Talete 14
 Teorema di von Neumann 160
 Teoria degli insiemi 16
 Tono pitagorico 139
 Transizione di fase 107

 Utilità in senso atteso 157

 Valore atteso 160
 Valore conservativo 158, 161-162
 Varietà bidimensionale 29
 Votazione 150

 Zara (gioco della) 65



INDICE DEI NOMI

- | | |
|---|--|
| Alberti, Leon Battista 129 | Euclide 21 |
| Aristosseno 144 | Eulero 46-47, 50, 53, 104 |
| Bach, Johann Sebastian 143 | Fëdorov, Evgraf 132 |
| Bayes, Thomas 89-92 | Fermat, Pierre de 68-69, 73 |
| Bernoulli, Jakob 70 | Fourier, Jean-Baptiste 45, 53,
61, 137 |
| Bessel, Friedrich 45 | |
| Buckminster Fuller, Richard 28 | |
| | Galilei, Galileo 35 |
| Caccioppoli, Renato 99 | Gauss, Carl 30-31, 82 |
| Cantor, Georg 16, 23 | |
| Cardano, Girolamo 67 | Harsanyi, John Charles 165 |
| Carneade 64 | Hausdorff, Felix 57-58 |
| Chasles, Michel 130 | Huygens, Christiaan 70 |
| Claudio (imperatore) 64 | |
| Coxeter, Harold 132 | Kolmogorov, Andrej 86 |
| | Kronecker, Leopold 17 |
| Dalí, Salvador 134 | |
| De Finetti, Bruno 84 | Lagrange, Joseph-Louis 45, 104 |
| De Giorgi, Ennio 99 | Laplace, Pierre-Simon 46, 52,
55, 82 |
| de Moivre, Abraham 82 | Leibniz, Gottfried 43, 81 |
| Debussy, Claude 144 | Leonardo da Vinci 125, 129 |
| Dedekind, Richard 16 | Ljapunov, Aleksandr 82 |
| Desargues, Girard 130 | Lorenz, Edward 53-55, 58 |
| Dürer, Albrecht 129 | Lotka, Alfred 187, 190 |
| | |
| Einstein, Albert 24, 26, 30-31,
133 | Malthus, Thomas 175, 177,
179-181, 184, 188 |
| Eratostene 25 | Mercator, Gerardus 28 |
| Escher, Maurits Cornelis 32,
132-133 | Mondrian, Piet 127 |

- Morgenstern, Oskar 161, 164
Nash, John 161-165, 170, 173
Navier, Claude-Louis 39, 50,
52-53, 106
Nepero, Giovanni 178
Neumann, John von 160-161,
164
Newton, Isaac 41-44, 46-47, 50,
54, 81
- Ovidio 64
- Pacioli, Luca 66-67, 69, 89
Pappo 130
Pascal, Blaise 67-69, 73
Peano, Giuseppe 17
Penrose, Roger 120, 132
Phong, Bui Tuong 120
Piero della Francesca 127-128
Pitagora 14, 22, 136
Plateau, Joseph 100, 103, 107
Poincaré, Henri 32-33, 133
- Riemann, Bernhard 26, 30-31
- Schönberg, Arnold 144
Shapley, Lloyd 170-172
Stefan, Jožef 107
Steiner, Jakob 105
Stokes, George 39, 50, 52-53,
106
Svetonio 64
- Taletto 14
Taylor, Brook 44-45
- Verhulst, Pierre 180-181
Volterra, Vito 187, 190
- Weierstrass, Karl 16, 56
Werckmeister, Andreas
142-143
- Zarlino, Gioseffo 140
Zenone 22

