U Math

VOLUMI PUBBLICATI NELLA COLLANA U MATH

1. Maurizio Trombetta, *Calcolo combinatorio*
2. Carlo Càssola, *Geometria piana per le gare di matematica*
3. Salvatore Damantino, *Teoria dei numeri*
4. Sandro Campigotto, *I giochi matematici di PhiQuadro*
5. Carlo Càssola, *Geometria solida per le gare di matematica*
6. Samuele Maschio, *Tecniche dimostrative*
7. Salvatore Damantino, Emanuele Campeotto, *Aritmetica modulare*
8. Terence Tao, *Risolvere problemi matematici*
9. Luigi Amedeo Bianchi, *Probabilità*
10. Paolo Fiorini, *Manuale di allenamento per le gare di matematica*
11. Emanuele Callegari, *Combinatoria per problemi*

U MATH JUNIOR

1. S. Campigotto, P. Dall'Aglio, *Giocare con la matematica e il problem solving*

Luigi Amedeo Bianchi

Probabilità

Un'introduzione con esercizi

scienza express
edizioni

Prima edizione in *U Math* febbraio 2021

PROBABILITÀ

Grafica di copertina di Nicole Vascotto

ISBN 979-12-800-6809-5

La collana *U Math* è promossa dalla Sezione di Udine della Mathesis, ente che ha come scopo la diffusione delle scienze matematiche e fisiche, la valorizzazione e il progresso dell'insegnamento e della cultura scientifica. Tra le sue attività annovera la preparazione degli studenti alle competizioni matematiche, attraverso l'organizzazione di campus e stages di formazione, e gare di matematica a squadre.

Introduzione

La probabilità, come molte branche della matematica, nasce per affrontare problemi reali. Se però la geometria ha mosso i primi passi dalla necessità di rideterminare, a ogni piena del Nilo, i confini dei campi, a dare l'impeto iniziale alla probabilità sono stati il gioco d'azzardo e le scommesse. Questa origine meno nobile e tardiva, assieme a resistenze filosofiche e dogmatiche ("Dio non gioca a dadi", disse Einstein[1]), hanno penalizzato la probabilità e la sua diffusione come disciplina del sapere. Ancora oggi, nonostante l'importanza del ragionamento probabilistico e statistico sia accettata razionalmente, cerchiamo inconsciamente di sfuggire al caso e di trovare una causa per ogni avvenimento.

Questi sono alcuni dei motivi per cui la probabilità è stata storicamente trascurata, anche nella scuola e nei programmi universitari. È un peccato, perché la probabilità è bella, oltre che utile e molto presente nella vita quotidiana. Entra in gioco ogni volta che abbiamo informazioni incomplete o incertezza. In particolare è un ingrediente fondamentale del machine learning.

Per poterla apprezzare e usare servono però strumenti e confidenza e il modo migliore di ottenerli è mettersi alla prova con dei buoni problemi. In questo modo, possiamo acquisire familiarità con la probabilità senza fretta, avendo tempo per

[1] Quella di Einstein non era un'affermazione teologica, ma filosofica: era convinto che le leggi dell'Universo dovessero essere intrinsecamente deterministiche, fatto messo in discussione dalla nascita della meccanica quantistica, che ne suggeriva una natura probabilistica.

gustare i dettagli e costruendo una solida base all'intuizione, con la speranza di evitare alcune delle trappole che il nostro istinto, non ottimizzato per il calcolo aleatorio, ci tende in continuazione.

Questo libro prova ad andare proprio in questa direzione: offre una collezione di problemi su cui mettersi alla prova, accompagnati da una presentazione rigorosa dei fondamenti della teoria. In circolazione ci sono già molti libri di probabilità. Ci sono alcuni bellissimi libri di divulgazione che permettono di darle un primo sguardo, come ad esempio *La matematica dell'incertezza* [14] e *La passeggiata dell'ubriaco* [15]. Altri, pur mantenendo il fine divulgativo, ne approfondiscono l'applicazione in contesti specifici, come il gioco d'azzardo in *Fate il nostro gioco* [4]. In tutti questi, però, manca la partecipazione attiva: non ci sono esercizi. Cambiando prospettiva, troviamo libri che mettono al centro la risoluzione di problemi e offrono raccolte di esercizi svolti, come il meraviglioso *Fifty Challenging Problems in Probability with Solutions* [16] o *Probability and Expectation* [21]. In questi testi, però, sono presenti solo pochi cenni di teoria: chi non avesse già le basi potrebbe trovarsi spaesato o insoddisfatto. Un discorso a parte, poi, meritano i manuali universitari, nei quali la teoria trova decisamente più spazio, ma la parte iniziale, i concetti di base, sono spesso sacrificati ad un programma ambizioso. Questo libro copre in un centinaio di pagine quello che nei manuali costituisce il primo capitolo, con l'idea che, rafforzando con più calma le basi e la propria intuizione probabilistica, sia poi più facile affrontare argomenti più avanzati.

Vedendolo in questa collana, potreste pensare che sia un libro di preparazione alle Olimpiadi, ma, in senso stretto, non lo è. Alle Olimpiadi della Matematica infatti i problemi veramente di probabilità sono pochi e, in generale, i problemi che compaiono nelle gare non possono andare a toccare aspetti troppo tecnici. Nella Bibbia della preparazione olimpica, le *Schede Olimpiche* di Massimo Gobbino [10], alla probabilità sono dedicate due schede, che si basano su una "definizione" di probabilità molto semplificata e che trattano solo la probabilità finita. Lo stesso succede nelle due collezioni *Le Olimpiadi Della Matematica: problemi dalle gare italiane* [6, 2], anche se con qualche interessante eccezione, o nell'ottimo *La matematica delle Olimpiadi* [18].

Le Olimpiadi hanno tuttavia un forte influsso sul mio lavoro. L'idea di concentrarsi su esercizi significativi e diversi l'uno dall'altro, invece di insistere con esercizi ripetitivi, viene proprio delle gare di matematica: vogliamo imparare a risolvere problemi nuovi, non memorizzare uno stesso esercizio declinato in cento modi.

I problemi proposti sono di varia difficoltà, ma ognuno di essi merita un po' di tempo. Se non riuscite a risolverne uno al primo colpo, resistete alla tentazione di saltare subito a guardare la soluzione: meglio prendere fiato e riprovare una seconda volta, e poi una terza. In questo modo, anche se alla fine la vostra risposta non fosse completa, potrete apprezzare molto di più la soluzione, avendo nel frattempo conosciuto meglio i concetti coinvolti e avendo preso familiarità con l'esercizio stesso. Le soluzioni di tutti i problemi sono raccolte nel Capitolo 6. Un discorso leggermente diverso va fatto per gli esempi: alcuni sono un supporto alle definizioni, altri sono esercizi in incognito, usati per introdurre nuove idee o approfondire quelle già viste. Questi ultimi si riconoscono perché sono in un riquadro grigio, con una soluzione a seguire.

Voglio concludere questa introduzione con alcuni ringraziamenti. Innanzitutto a Carlo, Pino e Anna che, in modi diversi, mi hanno convinto a scrivere questo libro. A Francesco, che mi ha insegnato come divertirmi con la probabilità e al prof. Giorgio Letta, che mi ha parlato di tribù e mi ha ricordato l'importanza di essere precisi, in probabilità. Ai Rudi Mathematici e a Marco e Alberto, per i problemi (da tre birre) condivisi anni fa. A tutte le amiche e gli amici che ho incontrato attraverso le Olimpiadi, per quello che mi hanno insegnato. A Luigi, Daniele e Pino per gli appunti, i suggerimenti e gli errori segnalati in una versione preliminare del libro. A Carlo e Daniele, per un'esaustiva correzione delle bozze. Agli sviluppatori di TeXmacs [11], in cui è stata scritta la prima bozza di questo libro, di LaTeX, in cui la stesura è stata finalizzata, di PGF/TikZ, in cui sono state create la maggior parte delle figure, e di R, in cui sono state condotte le simulazioni.

A questo punto non mi resta che augurarvi buona lettura e buon divertimento!

Luigi Amedeo Bianchi

1. Combinatoria – Riscaldamento

Chi incontra per la prima volta la probabilità lo fa spesso attraverso la definizione "casi favorevoli su casi totali". Come vedremo, questa definizione, per quanto limitata, può funzionare. Nella pratica trasforma il problema di probabilità in uno di combinatoria.

Cos'è allora la combinatoria? In questo libro non abbiamo spazio per una risposta completa o una trattazione esaustiva della combinatoria: abbiamo appena iniziato e già stiamo divagando. Rimandiamo chi volesse saperne di più sulla combinatoria a libri dedicati, ad esempio *Calcolo combinatorio* [22] o *Combinatoria per problemi* [3] in questa stessa collana.

In questo libro chiameremo combinatoria il contare oggetti e studiare modi e metodi per farlo. La incontriamo subito in questo primo capitolo non solo perché è un primo modo di avvicinarsi alla probabilità, ma anche perché ci dà la possibilità di costruire esempi e problemi da usare nei prossimi capitoli per provare a passare a una definizione più appropriata di probabilità. Nel mondo delle Olimpiadi della Matematica, la probabilità compare come caso particolare della combinatoria, un po' per comodità nell'avere un numero limitato di etichette con le quali classificare i problemi (e in questo combinatoria è un po' il jolly, nel senso che tutto quello che non è algebra, geometria o teoria dei numeri viene classificato come combinatoria), ma anche perché spesso i problemi olimpici che parlano di probabilità considerano solamente casi elementari equiprobabili, concentrando la difficoltà del problema nel trovare un buon modo di contare i casi favorevoli.

Un'ultima cosa prima di continuare: in questo capitolo considereremo solamente insiemi con un numero finito di elementi. Per quelli infiniti dovremo aspettare i capitoli successivi. Alcuni richiami di teoria degli insiemi sono raccolti in Appendice.

1.1 I principi fondamentali della combinatoria

Abbiamo detto che la combinatoria è contare oggetti: vediamo allora come farlo. Per prima cosa andiamo a introdurre i tre principi fondamentali della combinatoria, a partire dai quali passeremo poi a definire metodi di conteggio più sofisticati.
Nel seguito di questo libro ci capiterà spesso di contare il numero di elementi di un insieme, cioè la sua cardinalità. Useremo il simbolo $\#A$ per indicare la cardinalità di un insieme A. Questa può essere un numero (naturale) finito, ma anche infinito, sia numerabile, denotato con $\aleph_0$, sia pari al continuo, denotato con $2^{\aleph_0}$. Senza entrare nei dettagli della teoria degli insiemi, ricordiamo che un insieme ha cardinalità numerabile se può essere messo in relazione biunivoca con l'insieme $\mathbb{N}$ dei numeri naturali (che per noi includono lo 0) e ha cardinalità del continuo se può essere messo in relazione biunivoca con l'insieme $\mathbb{R}$ dei numeri reali. Per qualche dettaglio in più, si veda l'Appendice.
Il *primo principio della combinatoria* sostituisce il conteggio degli elementi di un insieme con il conteggio degli elementi di una sua partizione, ossia con una rappresentazione dell'insieme come unione disgiunta di suoi sottoinsiemi[1]. È il principio che ci apre la via al paradigma del *divide et impera*: spezzare un problema in parti più piccole e mutualmente esclusive affrontandole separatamente e combinando alla fine i risultati.

Proposizione 1.1.1. *Siano A un insieme e $\{E_i\}_{i=1}^n$ una partizione di A. Allora*[2] $\#A = \sum_{i=1}^n \#E_i$.

Cosa c'entra questo con la combinatoria? Proviamo a fare un paio di esempi.

Esempio 1.1.1. Con un buono regalo possiamo decidere se avere o un film o un videogioco. Sapendo che ci sono 10 film e 6 videogiochi disponibili, in tutto abbiamo $10 + 6$ omaggi diversi tra cui scegliere quale portarci a casa. In questo caso A è l'insieme di tutti gli omaggi tra cui possiamo scegliere e la sua partizione è data da E_1, insieme dei film disponibili, ed E_2, insieme dei videogiochi disponibili.

---

[1] Chi avesse bisogno di un ripasso sulle partizioni trova un breve richiamo in Appendice.

[2] Il simbolo $\sum$, detto *sommatoria*, indica l'operazione di somma di un certo numero di elementi per cui sia definito un concetto di somma, ad esempio numeri reali. In particolare $\sum_{i=1}^n a_i = a_1 + a_2 + \cdots + a_{n-1} + a_n$.

Esempio 1.1.2. In una scuola ci sono 28 studentesse e studenti del primo anno, 25 del secondo, 21 del terzo, 26 del quarto e 26 del quinto. In tutto, nella scuola ci sono allora $28 + 25 + 21 + 26 + 26 = 126$ studentesse e studenti; infatti ognuno di loro non può che appartenere a uno e un solo anno di corso. Qui A è l'insieme di tutti gli studenti della scuola, ed E_i, per i da 1 a 5, l'insieme di quelli dell'i-esimo anno.

Per introdurre il *secondo principio della combinatoria*, ossia il principio del prodotto, dobbiamo prima richiamare brevemente il prodotto cartesiano di insiemi.

Definizione 1.1.2. Dati due insiemi A e B, il loro prodotto cartesiano, indicato con $A \times B$, è l'insieme delle coppie ordinate (a, b) tali che $a \in A$ e $b \in B$.

Notiamo l'aggettivo che compare nella precedente definizione: le coppie che consideriamo sono *ordinate*. Questo significa che una coppia non è determinata solamente dagli elementi che la compongono, ma anche dall'ordine in cui compaiono: le due coppie $(1, 3)$ e $(3, 1)$ sono coppie ordinate distinte. Vedremo che è importante non dimenticarsi se stiamo considerando coppie (o terne, o n-uple) ordinate o no.
Ora che sappiamo che cosa è il prodotto cartesiano, andiamo a vedere perché ci interessa in combinatoria. In questo caso vogliamo (poco sorprendentemente) contare le coppie ordinate.

Proposizione 1.1.3. *Dati due insiemi A e B e il loro prodotto cartesiano $A \times B$, vale la seguente uguaglianza: $\#(A \times B) = \#A \cdot \#B$.*

Dobbiamo fare attenzione: in generale i due insiemi $A \times B$ e $B \times A$, pur avendo la stessa cardinalità, sono diversi, perché formati da coppie ordinate diverse. D'altra parte, anche se gli insiemi sono distinti, possiamo mostrare una relazione biunivoca tra essi. In particolare la mappa che scambia le due componenti soddisfa questa condizione. In effetti, se pensiamo a quello che significano le varie quantità, stiamo dicendo che anche se i due insiemi hanno lo stesso numero di elementi, non necessariamente sono uguali.

Esempio 1.1.3. Per fare un esempio più che classico, pensiamo a un pasto in una mensa o in una tavola calda: il pasto consiste di un primo a scelta tra minestra, pasta e riso e di un secondo a scelta tra carne, pesce, formaggio, uova e sformato di verdure. In quanti modi diversi possiamo comporre un pasto?

Soluzione. Vogliamo contare le coppie ordinate in cui alla prima componente abbiamo un primo e alla seconda un secondo (molto appropriatamente). I modi che abbiamo sono in questo caso $3 \cdot 5$. Possiamo leggere il risultato così: per ogni scelta del primo tra i 3 disponibili, abbiamo 5 possibili secondi (e viceversa: visto che la moltiplicazione è commutativa, possiamo anche fissare prima il secondo, scegliendolo fra i 5 a nostra disposizione e, in seguito, determina uno dei 3 primi). $\square$

Possiamo definire in modo del tutto simile il prodotto cartesiano tra più di due insiemi, a patto che siano in numero finito, e vale un risultato analogo per la sua cardinalità.

Proposizione 1.1.4. *Data una famiglia finita di insiemi* $\{A_i\}_{i=1}^n$*, prendiamo il loro prodotto cartesiano* $A_1 \times \cdots \times A_n$ *che denotiamo anche con* $\bigotimes_{i=1}^n A_i$*. Vale la seguente uguaglianza*[3]*:* $\#\left(\bigotimes_{i=1}^n A_i\right) = \prod_{i=1}^n \#A_i$*.*

Esempio 1.1.4. Torniamo alla nostra mensa: è cambiata la gestione e ora, oltre a un primo e a un secondo come prima, possiamo scegliere anche un contorno, tra patate, carote, spinaci e piselli e un dessert tra budino, crème caramel e gelato. In quanti modi diversi possiamo ora comporre un pasto?

Soluzione. Ora vogliamo contare le 4-uple ordinate, in cui compaiono, nell'ordine, un primo, un secondo, un contorno e un dessert. Le scelte sono, nel medesimo ordine, 3, 5, 4 e 3, per un numero totale di $3 \cdot 5 \cdot 4 \cdot 3 = 180$ modi differenti di comporre un pasto. $\square$

Gli insiemi A_i che andiamo a moltiplicare non devono necessariamente essere disgiunti e, in realtà, nemmeno distinti. In particolare nulla ci impedisce di considerare n copie dello stesso insieme. In questo caso abbiamo semplicemente l'insieme $\bigotimes_{i=1}^n A = A^n$, la cui cardinalità è $\#(A^n) = (\#A)^n$.

Esempio 1.1.5. Quanti sono i possibili PIN a 6 cifre?

Soluzione. In questo caso abbiamo $A = \{0, 1, \ldots, 9\}$ come insieme nel quale peschiamo ciascuna delle 6 cifre del PIN. Stiamo quindi cercando la cardinalità dell'insieme A^6, cioè il numero 10^6. $\square$

Esempio 1.1.6. Se invece volessimo i PIN a 6 cifre in cui non ci sono cifre consecutive uguali?

---

[3] Il simbolo $\prod$, detto *produttoria*, è l'analogo per il prodotto della sommatoria $\sum$ già introdotta: indica dunque la moltiplicazione di un certo numero di elementi per cui sia definito un concetto di prodotto. In particolare $\prod_{i=1}^n a_i = a_1 \cdot a_2 \cdot \cdots \cdot a_{n-1} \cdot a_n$.

Soluzione. Come prima cosa notiamo che non siamo più nel caso precedente; in particolare ci aspettiamo di ottenere un numero più basso, visto che stiamo considerando un sottoinsieme di tutti i PIN possibili. Per la prima cifra[4] abbiamo 10 possibili valori (tutti i numeri tra 0 e 9). Quando passiamo alla seconda cifra, adiacente alla prima, uno dei valori non è più a nostra disposizione (quello scelto per la prima cifra). Ma solamente quel valore va escluso, quindi ci restano 9 scelte possibili. Similmente per le cifre successive, per un totale di $10 \cdot 9^5$ possibili PIN che soddisfano la nostra condizione. □

Osservazione 1.1.5. Riguardiamo ancora l'esempio precedente: potremmo pensare di procedere in un modo diverso, non necessariamente passando alle cifre vicine. Ad esempio potremmo cominciare scegliendo la prima, la terza e la quinta. Siccome esse non si toccano, possiamo scegliere ciascuna di esse in 10 modi. Quando però andiamo a considerare la seconda cifra del nostro PIN, dobbiamo distinguere due casi, per sapere quante scelte siano possibili: se la prima e la terza cifra sono uguali, allora la seconda può essere scelta in 9 modi. Se invece sono diverse tra loro, la seconda può essere scelta solamente in 8 modi. Questo modo di conteggiare, per quanto corretto e possibile, è quindi più a rischio per quanto riguarda gli errori di conto.
Nell'esempio, infatti, stiamo sfruttando un approccio (un algoritmo) che sfrutta una proprietà particolare: non ci interessa quale cifra estraiamo in un dato punto, perché stiamo considerando qualcosa di invariante rispetto alla scelta specifica della cifra, ossia la cardinalità delle cifre che ci restano da scegliere al passaggio successivo. È per questo che fissare cifre del PIN saltando qua e là non è altrettanto efficace: non abbiamo un invariante analogo.

È difficile definire in modo generale gli invarianti, ma possiamo pensarli come quantità (non sempre immediate) che rimangono costanti al variare di un sistema, ad esempio non cambiano qualunque sia la mossa che facciamo in un gioco. Alcuni esempi di invarianti sono molto semplici: se moltiplichiamo un numero per un numero positivo, il segno non cambia, se sommiamo a un numero intero un numero pari, la parità non cambia. Altri lo sono meno. Capita comunque molto spesso che coinvolgano le classi di resto[5]. Un contesto in cui sono spesso usati è per trovare

---

[4]La prima cifra che inseriamo. Infatti si può osservare abbastanza facilmente che il ragionamento non cambia se andiamo a scegliere per prima la cifra in una qualunque posizione (ad esempio in posizione 3), muovendoci poi in entrambe le direzioni passando ogni volta a una cifra adiacente a una già scelta.

[5]Questo argomento è trattato, ad esempio, nel Capitolo 1 di *Aritmetica modulare* [7], in questa stessa collana.

chi possa essere il vincitore in un gioco o per mostrare che non ci sono modi di fare qualcosa (ad esempio che non è possibile tassellare una scacchiera).
Nell'Esempio 1.1.6 l'invariante è il numero di possibili scelte per la cifra successiva, costante a ogni passo, indipendente dalla particolare cifra scelta. Altre volte gli invarianti sono più complicati, almeno finché non ci vengono mostrati. Vediamo ancora un esempio, anche questo abbastanza semplice. Per chi volesse approfondire il tema degli invarianti, un buon punto di partenza sono le *Schede Olimpiche* [10].

Esempio 1.1.7. Eracle viene mandato, nella sua seconda fatica, a uccidere l'Idra di Lerna. È un terrificante mostro con 50 teste, che muore solo quando rimane senza alcuna testa. Eracle, in un colpo, può tagliarne 3 (ma in questo caso ne ricrescono 6), 5 (e ne ricrescono 2), 13 (e ne ricrescono 16), oppure 20 (e ne ricrescono 8). Può riuscire Eracle a uccidere l'Idra?

Soluzione. Contrariamente al mito, in questo caso Eracle non può riuscire a uccidere l'Idra, almeno finché si limita a tagliarle le teste. Osserviamo che alcune "mosse" causano un aumento delle teste, mentre altre ne causano una diminuzione. Tuttavia in tutti i casi la variazione del numero di teste in modulo 3 non cambia: $-3+6 \equiv -5+2 \equiv -13+16 \equiv -20+8 \equiv 0 \pmod 3$. Quindi il numero di teste rimane costante, in modulo 3. Dal momento che all'inizio l'Idra ha 50 teste (congruo a 2 modulo 3), non potrà mai portarla a 0 teste e ucciderla. Questo almeno finché non si rende conto dell'invariante e chiama Iolao e la sua torcia in aiuto[6] per bruciare le teste troncate in modo che non ricrescano. □

Dopo questa breve divagazione torniamo alla combinatoria e, per concludere questa sezione, vediamo il *terzo principio della combinatoria*. Per farlo, ci mettiamo in una situazione simile a quella vista per il primo principio: vogliamo ottenere la cardinalità dell'unione di alcuni insiemi, lasciando però cadere l'ipotesi che siano disgiunti.

Esempio 1.1.8. Alcuni eventi della Coppa del Mondo di arrampicata hanno gare di due diverse specialità: *boulder* e *lead*. Sapendo che a uno di questi hanno partecipato 37 atleti nel *boulder*, 33 nel *lead* e 14 a entrambe le specialità, quanti erano gli atleti presenti all'evento?

Soluzione. In analogia a quanto visto per il primo principio, la prima idea che ci viene è quella di andare a sommare i partecipanti al *boulder* con quelli al *lead*,

[6] Per saperne di più sulle fatiche di Eracle, il testo che le racconta più nei dettagli è la *Biblioteca* di Pseudo-Apollodoro [1].

ottenendo $37+33=70$ atleti. Tuttavia sappiamo che il primo principio richiede che gli insiemi siano disgiunti, mentre qui sappiamo che questa ipotesi non è verificata. Cosa cambia? Pensiamo ai 14 atleti che hanno preso parte a entrambe le gare di specialità: li abbiamo contati due volte, sia nel *boulder*, sia nel *lead*, quindi per avere il numero totale di atleti presenti dobbiamo sottrarre 14 da 70, ottenendo in tutto 56 partecipanti. $\square$

In generale, possiamo enunciare il terzo principio come segue.

Proposizione 1.1.6. *Se abbiamo due insiemi A_1 e A_2, la cardinalità della loro unione sarà*

$$\#(A_1 \cup A_2) = \#A_1 + \#A_2 - \#(A_1 \cap A_2).$$

Dimostrazione. Possiamo dimostrare questo risultato riconducendoci al primo principio, scrivendo l'unione come unione disgiunta:

$$A_1 \cup A_2 = (A_1 \setminus A_2) \cup (A_2 \setminus A_1) \cup (A_1 \cap A_2).$$

Ora non ci resta che osservare che $A_1 = (A_1 \setminus A_2) \cup (A_1 \cap A_2)$ (e analogamente per A_2) e mettere assieme i vari pezzi per ottenere quanto cercato. $\square$

Anche qui, come in precedenza, nulla ci costringe a considerare solamente due insiemi. Se passiamo all'unione di tre insiemi A_1, A_2 e A_3, iniziamo come prima, sommando le cardinalità dei tre insiemi e togliendo le (tre) intersezioni degli insiemi a due a due. In questo modo abbiamo contato una sola volta tutti gli elementi, tranne quelli di un sottoinsieme: l'intersezione di A_1, A_2 e A_3. Guardiamo un elemento di questo sottoinsieme: lo abbiamo contato una volta in ciascuno dei tre insiemi, lo abbiamo poi tolto una volta per ciascuna delle tre intersezioni a due a due, col risultato che lo abbiamo contato zero volte. Dobbiamo quindi andare ad aggiungere l'intersezione a tre,

$$\begin{aligned}\#(A_1 \cup A_2 \cup A_3) = {} & \#A_1 + \#A_2 + \#A_3 \\ & - \#(A_1 \cap A_2) - \#(A_1 \cap A_3) - \#(A_2 \cap A_3) \\ & + \#(A_1 \cap A_2 \cap A_3).\end{aligned}$$

Come conseguenza di questo aggiungere e togliere elementi, il terzo principio prende anche il nome di *principio di inclusione-esclusione.*

Proposizione 1.1.7. *Con n insiemi $A_1, \dots, A_n$ abbiamo l'uguaglianza*

$$\# \bigcup_{i=1}^{n} A_i = \sum_{i=1}^{n} \#A_i - \sum_{i<j} \#(A_i \cap A_j) + \sum_{i<j<k} \#(A_i \cap A_j \cap A_k) - \dots + (-1)^{n+1} \# \bigcap_{i=1}^{n} A_i.$$

Osservazione 1.1.8. Osserviamo che se non consideriamo l'intera somma, ma solo i primi addendi, possiamo avere una stima del totale. È una stima dal basso nel caso in cui il primo termine della somma che ignoriamo ha segno positivo (cioè è in posizione dispari), dall'alto se ha segno negativo (ossia è in posizione pari).

1.2 Permutazioni e anagrammi

Pensiamo ora alla seguente situazione: abbiamo un insieme A che contiene n oggetti distinti. Ci chiediamo quante siano le permutazioni di questi oggetti, ossia i modi di disporli in fila.

Iniziamo dal primo oggetto della fila: lo possiamo scegliere a piacere tra tutti gli elementi di A, cioè abbiamo n modi per sceglierlo. Passiamo ora al secondo. Anche senza sapere quale elemento di A abbiamo messo al primo posto, sappiamo che ce ne sono rimasti altri $n-1$ tra cui scegliere: tutti gli elementi di A, tranne quello già usato. Possiamo continuare in questo modo: a ogni passo avanti nella fila di oggetti, avremo un elemento in meno tra cui scegliere, fino ad arrivare all'ultimo posto, per il quale non ci sarà rimasto che un solo elemento.

Scrivendo tutto questo abbiamo che le *permutazioni*, o *riordinamenti*, di A sono $n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 3 \cdot 2 \cdot 1$, cioè il prodotto di tutti i numeri interi positivi minori o uguali di n. Questo prodotto è talmente importante in matematica che viene denotato con un simbolo, $n!$, detto n fattoriale[7]. Per il caso limite $n = 0$, poniamo $0! = 1$, con l'idea che abbiamo un solo modo per ordinare l'insieme vuoto.

Come nel prodotto cartesiano, anche qui l'ordine è importante. E in un certo senso siamo ancora nel caso del prodotto cartesiano: semplicemente partiamo con l'insieme al completo e, a ogni passo, lo moltiplichiamo (cartesianamente) con una versione sempre più piccola, che ha perso l'elemento appena scelto. Anche se non sappiamo con precisione quale sia l'elemento che abbiamo scelto, a ogni passo ce ne sarà rimasto uno in meno rispetto a quelli che avevamo in precedenza.

Tra gli insiemi da riordinare un ruolo speciale è costituito dalle parole, intese come insiemi di lettere. In questo caso indichiamo i riordinamenti col nome *anagrammi*. Attenzione, vogliamo contare tutti gli anagrammi, non stiamo chiedendo che abbiano senso in qualche lingua.

Esempio 1.2.1. Prendiamo ora una parola, ad esempio "PRENDIAMO": quanti sono i suoi anagrammi?

[7] O n BUM, se preferite la terminologia del *Mago dei numeri* [9].

Soluzione. Siamo nello stesso caso visto sopra: il nostro insieme A è ora

$$A = \{\mathrm{P}, \mathrm{R}, \mathrm{E}, \mathrm{N}, \mathrm{D}, \mathrm{I}, \mathrm{A}, \mathrm{M}, \mathrm{O}\}$$

e in particolare ha 9 elementi, tutti distinti tra loro. I loro riordinamenti, cioè gli anagrammi di "PRENDIAMO", sono quindi $9! = 362\,880$. □

Prima di continuare con altri esempi di anagrammi, parliamo per un momento del fattoriale. Una delle prime cose che possiamo osservare incontrandolo è quanto velocemente cresce: nell'esempio precedente abbiamo visto che $9! = 362\,880$, mentre $10!$ è 10 volte più grande. Insomma, diventa rapidamente complicato scriverlo per esteso e conviene lasciarlo indicato con il suo simbolo finché si può. Non solo, nel momento in cui volessimo semplificarlo, ci conviene sfruttare la fattorizzazione naturale nascosta nella sua definizione, cioè la scrittura come prodotto dei primi n interi positivi, per semplificare tutto il semplificabile. Vedremo alcuni esempi di queste semplificazioni più avanti, perché il fattoriale salterà fuori spesso (cosa che rende ancora più comoda la notazione col punto esclamativo).
Consideriamo una variante della situazione precedente, molto comune quando stiamo anagrammando parole: cosa succede se abbiamo delle ripetizioni? Nel caso delle parole: cosa succede se una lettera compare più volte?

Esempio 1.2.2. Consideriamo la parola "ANAGRAMMI": quanti sono i suoi anagrammi?

Soluzione. Sicuramente sono al più $9!$, cioè tutte le permutazioni delle sue lettere. Però questo non tiene conto del fatto che abbiamo alcune lettere che si ripetono: A compare tre volte, M due. Se contassimo solamente le permutazioni, come fatto prima, staremmo contando come distinti due anagrammi ottenuti scambiando tra loro due lettere uguali (ad esempio le due M). Tuttavia questi sono indistinguibili tra loro:

$$\mathrm{ANAGRAM_1M_2I} = \mathrm{ANAGRAM_2M_1I}.$$

Dobbiamo allora contare in quanti modi possiamo permutare tra loro le lettere uguali. In questo esempio possiamo riordinare le M tra loro in $2! = 2$ modi e le A in $3! = 6$ modi. Dividiamo allora il fattoriale della lunghezza della parola per il numero di permutazioni di ciascun gruppo di lettere uguali, cioè per il fattoriale del numero delle loro occorrenze. In questo caso le permutazioni distinte di "ANAGRAMMI" sono

$$\frac{9!}{3! \cdot 2!} = \frac{362\,880}{12} = 30\,240.$$ □

Possiamo seguire questo approccio in generale, non solo per insiemi di lettere: se abbiamo un insieme A costituito da n elementi di m tipi diversi (necessariamente deve essere $m \leqslant n$), ciascun tipo $i \in \{1, \dots m\}$ presente in k_i copie, le permutazioni possibili di tutti gli elementi di A, non distinguendo elementi di uno stesso tipo, sono

$$\frac{n!}{k_1! \cdot k_2! \cdot \dots \cdot k_m!}.$$

Esempio 1.2.3. In una famiglia è consuetudine, per le festività invernali, decorare la ringhiera del balcone. Per farlo, mettono in fila palline luminose di tre colori: 8 sono rosse, 6 sono verdi e 4 sono azzurre. Ogni anno vogliono avere una decorazione diversa da quelle degli anni precedenti: dopo quanti anni dovranno necessariamente ripetersi?

Soluzione. Ci sono

$$\frac{18!}{8! \cdot 6! \cdot 4!} = \frac{9 \cdot 10 \cdot \dots \cdot 17 \cdot 18}{2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 2 \cdot 3 \cdot 4} = 11 \cdot 13 \cdot 14 \cdot 15 \cdot 17 \cdot 18 = 9\,189\,180$$

possibili anagrammi delle lampadine a loro disposizione, quindi molto probabilmente ci saremo già estinti da un po'. Prima di continuare, notiamo come abbiamo semplificato tutto quello che potevamo, prima di fare il conto conclusivo[8]. □

Esempio 1.2.4. Quanti sono gli anagrammi di "ANAGRAMMI" in cui le due "M" sono adiacenti?

Soluzione. Se le due "M" devono essere adiacenti, possiamo considerarle come un'unica lettera "X" e contare gli anagrammi della parola "ANAGRAXI".
A questo punto abbiamo $\frac{8!}{3!} = 6720$ anagrammi possibili, avendo 8 lettere di cui una, la "A", ripetuta 3 volte. □

Esempio 1.2.5. Goffredo ha recentemente avuto una delusione in amore, quindi odia tutto quello che gli ricorda il tema. Nel fare gli anagrammi di "ANAGRAMMI" esclude tutti quelli in cui compaiono le stringhe "AMA" o "AMI". Quanti anagrammi gli rimangono?

[8] Non dobbiamo pensare che siano semplificazioni inutili, anche quando abbiamo a portata di mano una calcolatrice o un computer: i fattoriali crescono talmente in fretta che, anche se il risultato finale è alla loro portata, gli strumenti di calcolo possono dare errori di approssimazione prima di arrivare in fondo.

Soluzione. Ci conviene contare quanti sono in tutto gli anagrammi, quanti sono quelli con una delle stringhe incriminate e sottrarre il secondo numero dal primo. Dobbiamo anche prestare attenzione al fatto che, essendoci più "A" e "M", potremmo avere più stringhe incriminate in un medesimo anagramma. Ci servirà allora il principio di inclusione-esclusione.

Cominciamo a codificare "X"="AMA" e "Y"="AMI". Contiamo gli anagrammi che contengono "AMA": sono i riordinamenti di "NGRAMIX", che sono 7!. Ce ne sono però alcuni che stiamo contando due volte: quelli in cui compare la stringa "AMAMA". Se la chiamiamo "W", per sapere quanti ne abbiamo contati di troppo, ci basta contare gli anagrammi di "NGRIW", che sono 5!.

Passiamo ora agli anagrammi di "ANAGRAMMI" che contengono "AMI": sono quelli di "NAGRAMY" cioè $\frac{7!}{2!}$. Anche in questo caso, però, ci sono alcuni anagrammi che contengono sia "AMA" sia "AMI" e che quindi abbiamo già contato in precedenza. Sono quelli della parola "NGRXY", 5!, ma anche quelli della parola "NGRAZ", in cui "Z"="AMAMI", anche questi 5!. Quindi gli anagrammi di "ANAGRAMMI" che contengono "AMA" o "AMI" sono

$$7! - 5! + \frac{7!}{2!} - 5! - 5! = 5! \cdot (6 \cdot 7 + 3 \cdot 7 - 3) = 5! \cdot 60.$$

Ora dobbiamo sottrarre questo numero, che ci dice quanti riordinamenti non vanno bene a Goffredo, dal numero di tutte le possibili permutazioni di "ANAGRAMMI", che sono $\frac{9!}{3!2!}$. La risposta è quindi

$$\frac{9!}{3!2!} - 5! \cdot 60 = 5!(7 \cdot 4 \cdot 9 - 60) = 120 \cdot 192 = 23\,040. \qquad \square$$

Torniamo ora al caso in cui tutti gli n elementi del nostro insieme sono distinti tra loro. Questa volta, però, vogliamo contare quante sono le possibili disposizioni di un numero $k \leqslant n$ di suoi elementi.

Esempio 1.2.6. Dodici amici hanno organizzato tra loro una lotteria, per la quale hanno 5 premi di valore decrescente. Quanti sono i diversi modi di distribuire i premi?

Soluzione. In un certo senso ci stiamo chiedendo nuovamente quanti siano i modi di mettere in fila i 12 amici (al primo della fila daremo il primo premio e così via), con la differenza che non ci interessa davvero sapere come sono disposti dalla sesta posizione in poi, perché uno scambio tra due persone oltre la quinta posizione non ha influenza sulla distribuzione dei premi. Abbiamo quindi, in questo caso, 12 scelte per il vincitore del primo premio, 11 per il secondo, fino a 8 scelte per il vincitore del quinto premio. La risposta è quindi $12 \cdot 11 \cdot 10 \cdot 9 \cdot 8 = 95\,040$.

Possiamo però osservare che questo è il prodotto dei numeri consecutivi da 8 a 12, una quantità che sappiamo esprimere come un rapporto di fattoriali:

$$\frac{12!}{7!} = 8 \cdot 9 \cdot 10 \cdot 11 \cdot 12.$$

Questo ci suggerisce un altro modo di vedere lo stesso risultato: le ultime 7 posizioni sono uguali tra loro, nel senso che sono tutte non vincenti, quindi stiamo contando le permutazioni di un insieme con 5 elementi tutti distinti tra loro e altri 7 tutti dello stesso tipo. Possiamo vederlo come l'insieme dei premi: primo, secondo, ..., quinto, niente, niente, ..., niente. □

In casi come questo si parla di *permutazioni incomplete* o *k-permutazioni*. Il numero di permutazioni di k elementi in un insieme di n elementi distinti è $\frac{n!}{(n-k)!}$.

Esempio 1.2.7. A ogni Gran Premio di Formula E partecipano 24 piloti. Al termine della gara vengono assegnati punti (diversi per ciascun piazzamento) ai primi 10 piloti. In quanti modi diversi è possibile assegnare i punteggi?

Soluzione. Abbiamo in tutto 24! riordinamenti possibili dei piloti. Ai fini della classifica, però, contano solamente le prime 10 posizioni, quindi non sono diversi tra loro quei riordinamenti che differiscono solo per permutazioni delle ultime 14 posizioni. Quindi la risposta è $\frac{24!}{14!} = 7\,117\,005\,772\,800$. □

Esempio 1.2.8. Nelle gare di Coppa del Mondo di arrampicata, vengono assegnati punti ai primi 30 classificati. Uomini e donne gareggiano in competizioni separate. Se alla gara di Garmisch-Partenkirchen hanno preso parte 50 uomini e 48 donne, quanti sono i modi diversi di assegnare i punteggi?

Soluzione. Cominciamo considerando separatamente la classifica maschile e quella femminile. Come visto nell'Esempio 1.2.7, abbiamo $\frac{50!}{(50-30)!}$ modi di assegnare punti nella gara maschile e $\frac{48!}{(48-30)!}$ modi per la gara femminile.
Dobbiamo ora combinare questi risultati, per avere il numero delle possibili classifiche dell'intero evento. Per il secondo principio della combinatoria, siccome i due ambiti sono distinti, dobbiamo moltiplicare i due risultati parziali, per avere quello totale: $\frac{50!}{20!} \cdot \frac{48!}{18!}$. Questo numero si può semplificare un po', ma è dell'ordine di 10^{91}. Sconsiglio di provare a calcolarlo. □

Resta per il momento in sospeso il caso delle k-permutazioni con ripetizioni. Le idee non sono molto diverse da quelle viste finora, ma diventano più semplici se viste sotto una lente diversa, quella delle combinazioni, argomento della prossima sezione.

1.3 Combinazioni e coefficiente binomiale

Passiamo a un problema diverso, anche se l'ambientazione è analoga a quella dell'Esempio 1.2.7.

Esempio 1.3.1. Le qualifiche di Formula E per stabilire l'ordine di partenza in un Gran Premio sono divise in due fasi: nella prima concorrono tutti i partecipanti, dopodiché i 6 più veloci nella prima fase competono tra loro nella Super Pole per determinare le prime 6 posizioni. In quanti modi diversi possiamo scegliere i 6 piloti (tra i 24 totali) che parteciperanno alla Super Pole?

Soluzione. Osserviamo che non siamo nella situazione già vista delle permutazioni incomplete, perché non ci interessa in che ordine siano i primi 6: tutti i risultati delle qualifiche (cioè tutti gli ordinamenti dei 24 piloti) che differiscono tra loro per riordinamenti dei primi 6 o degli ultimi 18 sono equivalenti. Quindi possiamo prendere tutti gli ordinamenti, dividerli per i riordinamenti degli ultimi 18, ottenendo le permutazioni incomplete viste prima, e dividere ancora una volta per i riarrangiamenti dei primi 6: abbiamo allora $\frac{24!}{18!\cdot 6!}$. □

Quello che abbiamo fatto in questo esempio è semplicemente contare i modi di scegliere 6 piloti tra 24. Possiamo generalizzarlo a n e k qualunque tra i numeri naturali, contando i modi di scegliere k oggetti tra n disponibili (ovviamente ci aspettiamo di farlo per $0 \leqslant k \leqslant n$) o, equivalentemente, di dividere gli n elementi di un insieme in k sottoinsiemi: essi prendono il nome di *combinazioni* di k oggetti scelti tra n. Per quanto appena detto, tali combinazioni saranno $\frac{n!}{k!(n-k)!}$, quantità per cui introduciamo la notazione $\binom{n}{k}$, detta *coefficiente binomiale*.

Esempio 1.3.2. Babbo Natale ha un grande sacco con tutti i regali per i bambini buoni del vicinato (che sono solo 13). La sua prima sosta è nella casa della famiglia Petruzzelli, in cui tutti e 3 i bambini sono stati buoni. In quanti modi può scegliere 3 regali da lasciar loro?

Soluzione. Babbo Natale deve scegliere 3 regali tra i 13 nel sacco. I modi possibili sono quindi $\binom{13}{3} = 286$. □

Fino a qui può sembrare che il coefficiente binomiale sia solo una comoda scrittura. Ma oltre a essere comodo è anche importante, perché tende a saltare fuori molto spesso nei problemi di combinatoria, anche più difficili di quelli appena visti. Prima di passare ad altri esempi più interessanti, tuttavia, vediamo alcune proprietà del coefficiente binomiale.

Proposizione 1.3.1. *Siano k e n numeri naturali tali che $0 \leqslant k \leqslant n$. Valgono le seguenti proprietà:*

i. $\binom{n}{k} = \binom{n}{n-k}$

ii. $\binom{n}{0} = \binom{n}{n} = 1$

iii. $\sum_{k=0}^{n} \binom{n}{k} = 2^n$

iv. $\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$.

Dimostrazione. Lasciata come Problema 4, alla fine di questo capitolo. □

Anche nel caso del coefficiente binomiale, come per il fattoriale e per la combinatoria in generale, non possiamo andare a fondo e studiare tutte le sue proprietà. Accenniamo solamente alla rappresentazione dei coefficienti binomiali in forma grafica, con il triangolo di Tartaglia[9] (o di Pascal[10]), del quale si può scoprire di più cercando online o consultando altri libri dedicati alla combinatoria[11].
Ci sono però problemi in cui il coefficiente binomiale entra in gioco in maniera non ovvia, come possiamo vedere nel prossimo esempio.

Esempio 1.3.3. Sul Lungarno a Pisa ci sono 18 palazzi, l'uno accanto all'altro. Il nuovo sindaco vuole ridipingerli in modo che siano soddisfatte le seguenti condizioni:

1. devono essere usati tutti e 7 i colori dell'arcobaleno;
2. tutti i palazzi del medesimo colore devono essere adiacenti.

In quanti modi diversi può farlo?

Soluzione. Cominciamo subito spezzando il problema in due parti: siccome tutti i palazzi del medesimo colore sono adiacenti, possiamo separare la scelta dell'ordine dei colori e i modi di colorare i palazzi una volta fissato l'ordine dei colori. In particolare nella soluzione comparirà un fattore 7! a contare i possibili riordinamenti dei colori.

[9] Niccolò Fontana detto Tartaglia (1499 circa – 1557).

[10] Blaise Pascal (1623 – 1662).

[11] Ad esempio il già citato *Calcolo combinatorio* [22], pag. 45.

Supponiamo ora fissato l'ordine dei colori. La seconda parte del problema è scegliere in quanti modi possiamo raggruppare i 18 palazzi in 7 sottoinsiemi, tenendo conto dei vincoli. Sentiamo puzza di coefficiente binomiale, ma non possiamo usarlo direttamente. Proviamo allora a cambiare punto di vista: mettiamoci sul Lungarno anche noi e guardiamo i palazzi che abbiamo accanto. Cominciamo a camminare: prima ne abbiamo un po' di un colore, poi passano al secondo, al terzo e così via, fino al passaggio dal sesto al settimo colore. Ehi! Abbiamo 6 cambi di colore, per via delle due condizioni. Quanti sono i posti in cui possiamo avere questi cambi di colore? Sono possibili a ogni confine tra due palazzi, quindi ne abbiamo uno dopo il primo palazzo, uno dopo il secondo e così via fino all'ultimo confine, dopo il penultimo (diciassettesimo) palazzo e prima dell'ultimo (diciottesimo). Quindi dobbiamo piazzare 6 cambi di colore in 17 posti possibili, per un contributo di $\binom{17}{6}$.
In generale, con p palazzi e c colori avremmo $\binom{p-1}{c-1}$ possibilità.
Mettendo assieme i due pezzi del problema, abbiamo allora $7! \cdot \binom{17}{6}$. □

1.4 Combinatoria e probabilità

Arriviamo finalmente a parlare di probabilità, grazie alla combinatoria. Se siamo in una situazione speciale in cui tutti i casi sono *equiprobabili*, possiamo calcolare la probabilità di qualcosa semplicemente contando tutti i casi favorevoli (cioè i casi in cui si verifica il qualcosa che cerchiamo) e dividere questo numero per quello di tutti i casi possibili.
È chiaro però che, se da un punto di vista intuitivo questa definizione ci può andare bene, da un punto di vista rigoroso lascia molto a desiderare: se non abbiamo ancora definito cosa significhi *probabilità*, come possiamo parlare di casi equiprobabili?
Lasciamo per il momento da parte questa perplessità e abbracciamo l'approccio intuitivo: possiamo comunque vedere numerosi esercizi ed esempi interessanti. Dobbiamo però controllare che le ipotesi di equiprobabilità siano verificate, altrimenti rischiamo di sbagliare, anche grossolanamente. Il classico controesempio alla formuletta mnemonica "casi favorevoli su casi totali" è quello della lotteria: ci sono due casi possibili, vincere e non vincere, di cui uno solo è a noi favorevole, quindi la probabilità di vittoria è $\frac{1}{2}$.
Ce ne sono però anche di più subdoli, in cui il risultato con un'interpretazione errata non è così lontano da quello corretto. In questi casi non possiamo sfruttare l'implausibilità della probabilità che otteniamo per accorgerci di aver sbagliato.

Esempio 1.4.1. Lanciamo due normali dadi a 6 facce. Qual è la probabilità di ottenere almeno un 4?

Soluzione. Quanti sono i possibili risultati, visti come coppie non ordinate? Ne abbiamo sei in cui compare almeno un 1, cinque in cui compare almeno un 2 e non compaiono 1 (abbiamo già contato $\{1, 2\}$) e così via. Le possibili coppie non ordinate sono $\frac{6 \cdot 7}{2} = 21$. Quelle in cui compare almeno un 4 sono sei. Quindi potremmo dire che la probabilità di vedere almeno un 4 sia $\frac{6}{21} = \frac{2}{7} \approx 29\%$.
Come però si può notare, queste coppie non ordinate non sono tra loro equiprobabili. E infatti se andiamo a contare le coppie ordinate, in cui il primo elemento rappresenta il risultato del primo dado e il secondo elemento quello del secondo dado, abbiamo 36 casi possibili, di cui 11 favorevoli, per una probabilità di vedere almeno un 4 uguale a $\frac{11}{36} \approx 31\%$. □

Alle volte è l'ambientazione del problema a complicare le cose, ad esempio quando ci presenta (come singoli) oggetti che siamo abituati a considerare a coppie.

Esempio 1.4.2. In una scarpiera, Andrea ha n paia di scarpe. Se prende a caso un numero pari di scarpe inferiore alla metà del totale, con che probabilità non avrà un paio completo?

Soluzione. Dobbiamo fare attenzione a non confondere scarpe e paia. Nella scarpiera ci sono $2n$ scarpe e Andrea ne prende $2s$, con $2s < n$. In quanti modi può farlo? È un coefficiente binomiale: può scegliere le scarpe in $\binom{2n}{2s}$ modi.
Passiamo allora al secondo conteggio, quello dei casi favorevoli[12]. Cosa vuol dire che non ha alcun paio completo? Significa che ha scelto al più una scarpa per ogni paio disponibile e, in particolare, ha scelto $2s$ *tipi* di scarpa (cioè tipi di paia) tra gli n disponibili e per ciascuno di essi (cioè per $2s$ volte) ha scelto una delle due scarpe. In altre parole lo può fare in $\binom{n}{2s} \cdot \binom{2}{1}^{2s} = \binom{n}{2s} \cdot 2^{2s}$ modi diversi. La probabilità cercata è allora

$$\frac{\binom{n}{2s} \cdot 2^{2s}}{\binom{2n}{2s}} = \frac{n!}{(2n)!} \cdot \frac{(2n-2s)!}{(n-2s)!} \cdot 2^{2s}.$$

Se questo ragionamento non ci convince del tutto, magari perché ci confondiamo nel passare da scarpe a paia, possiamo provare a calcolare il tutto in altri modi. Supponiamo che le scarpe siano tutte in fila e che quelle scelte da Andrea siano le prime $2s$. I casi totali sono allora $(2n)!$.
Passiamo allora ai casi favorevoli. Possiamo scegliere la prima scarpa come vogliamo, quindi abbiamo $2n$ modi di farlo. Per la seconda, non volendo avere paia complete, abbiamo $2n - 2$ scelte, per la terza $2n - 4$ e così via, fino alla scarpa in posizione $2s$

[12] In realtà per Andrea non sono molto favorevoli.

che possiamo scegliere in $2n - 2 \cdot (2s-1) = 2n - 4s + 2$ modi. A questo punto tutti i possibili ordinamenti delle successive $2n - 2s$ scarpe ci vanno bene, quindi abbiamo un fattore $(2n-2s)!$.
I casi favorevoli sono in tutto $2n \cdot (2n-2) \cdot (2n-4) \cdot \cdots \cdot [2n - 2 \cdot (2s-1)] \cdot (2n-2s)!$ e la probabilità cercata è

$$\frac{2n \cdot (2n-2) \cdot \cdots \cdot [2n - 2 \cdot (2s-1)] \cdot (2n-2s)!}{(2n)!} = \\ = 2^{2s} \cdot n \cdot (n-1) \cdot \cdots \cdot [n - (2s-1)] \cdot \frac{(2n-2s)!}{(2n)!},$$

cioè lo stesso risultato ottenuto prima (per fortuna).
E se volessimo ragionare per probabilità sulle singole scarpe, potremmo osservare che la prima ci va bene in $2n$ casi su $2n$, cioè con probabilità $\frac{2n}{2n} = 1$, la seconda con probabilità $\frac{2n-2}{2n-1}$, e così via fino alla scarpa numero $2s$ che ci va bene con probabilità $\frac{2n-4s+2}{2n-2s+1}$. Mettendo il tutto assieme,

$$\frac{2n}{2n} \cdot \frac{2n-2}{2n-1} \cdot \cdots \cdot \frac{2n - 2 \cdot (2s-1)}{2n - (2s-1)} = 2^{2s} \cdot \frac{n!}{(n-2s)!} \cdot \frac{(2n-2s)!}{(2n)!}. \qquad \square$$

Per i problemi delle Olimpiadi, possiamo quasi sempre fermarci qui: con un po' di creatività e di attenzione, sono tantissimi i problemi di probabilità che si possono scrivere in termini di casi equiprobabili. Ma come detto, la definizione data non ci soddisfa del tutto. Non solo, ci restringe a casi in cui possiamo contare cose, quindi in numero finito. E ci obbliga a prestare attenzione al fatto che tutti i casi sono equiprobabili (come possiamo ad esempio considerare una moneta truccata?). Non è impossibile, ma come vedremo nei prossimi capitoli, con poca fatica e un po' di astrazione in più, riusciremo affrontare problemi e situazioni molto più generali.

1.5 Esercizi

Problema 1

Il sindaco di Lucca vuole fare meglio del sindaco di Pisa visto nell'Esempio 1.3.3: si è messo in testa di dipingere tutti e 26 i palazzi che si affacciano su piazza Anfiteatro (una piazza a pianta ellittica completamente circondata da edifici, senza interruzioni) in modo che siano soddisfatte le due condizioni seguenti:

1. devono essere usati tutti gli 8 colori a disposizione, rigorosamente nell'ordine prefissato, in senso orario nella piazza;

2. tutti i palazzi del medesimo colore devono essere adiacenti.

In quanti modi diversi può farlo?

Problema 2
In quanti modi distinti possiamo scrivere 21 come somma di 5 interi positivi?

Problema 3
In quanti modi distinti possiamo scrivere 21 come somma di 5 numeri naturali?

Problema 4 *(Proposizione 1.3.1)*
Siano k e n numeri naturali tali che $0 \leqslant k \leqslant n$. Verificare le seguenti proprietà:

i. $\displaystyle\binom{n}{k} = \binom{n}{n-k}$

ii. $\displaystyle\binom{n}{0} = \binom{n}{n} = 1$

iii. $\displaystyle\sum_{k=0}^{n} \binom{n}{k} = 2^n$

iv. $\displaystyle\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$.

Problema 5
Una coppia di rapinatori ha svaligiato una banca ed è in fuga a piedi verso il proprio covo. La città ha una struttura tipicamente romana, come si vede in Figura 1.1. Quanti sono i percorsi di lunghezza minima che i ladri hanno a disposizione?

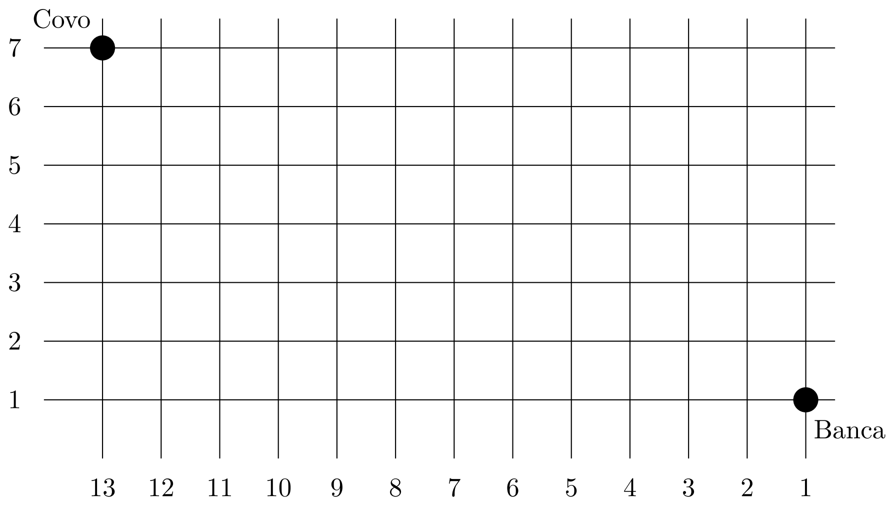

Figura 1.1: La fuga dei malfattori

Problema 6
Nelle stesse ipotesi del problema precedente, la polizia ha avuto una soffiata e ha piazzato due posti di blocco, come indicato in Figura 1.2: se i rapinatori passano di lì, vengono arrestati. Se i rapinatori scelgono uniformemente a caso tra tutti i percorsi di lunghezza minima, con che probabilità verranno catturati?

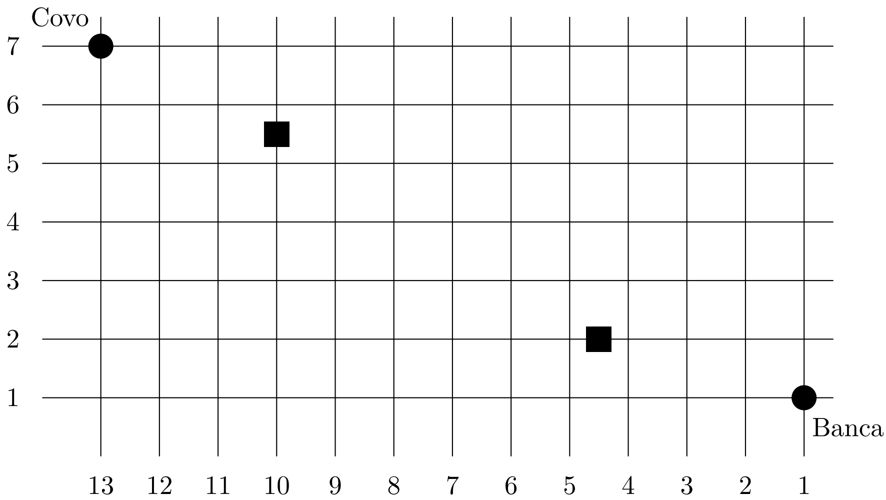

Figura 1.2: Fuga con posti di blocco

Problema 7
Quante sono le partizioni di un insieme di cardinalità $n = 10$?

Problema 8
Quante sono le funzioni iniettive da un insieme A a un insieme B?

Problema 9
Quante sono le funzioni suriettive da un insieme A a un insieme B?

2. La strada di Kolmogorov

In questo capitolo ci avviamo a costruire un modello matematico per esperimenti dei quali non sappiamo il risultato, in modo che sia un po' più ricco di quelli visti nel Capitolo 1. In particolare, in queste pagine consideriamo tale modello in modo astratto, come se fosse calato dall'alto, esaminandone le proprietà e studiandone alcuni esempi. Dobbiamo pazientare fino al Capitolo 3 per vedere strategie ed euristiche per costruire tali modelli per problemi specifici.

Gli assiomi di Kolmogorov[1], che vedremo in questo capitolo assieme alle loro conseguenze, sono per l'appunto assiomi: stabiliscono le regole del gioco e ci dicono come calcolare e manipolare la probabilità in modo coerente. Questo modo di procedere può sembrare noioso, ma ha una sua bellezza intrinseca. In ogni caso è necessario per evitare di essere fuorviati dall'istinto, che come vedremo gioca spesso brutti scherzi quando si parla di probabilità. Costruiremo la teoria assiomatica della probabilità mattone dopo mattone, partendo dalla teoria degli insiemi.

2.1 Esiti e universo

Vogliamo introdurre strumenti per studiare esperimenti dei quali non sappiamo a priori il risultato, e che possiamo quindi considerare casuali. La prima cosa di cui vogliamo tenere traccia sono i possibili risultati di un esperimento casuale (o aleatorio).

[1] Andrej Nikolaevič Kolmogorov (1903 – 1987).

Definizione 2.1.1. I possibili risultati di un esperimento casuale prendono il nome di *esiti*.
L'insieme di tutti gli esiti possibili si chiama *insieme universo* o *insieme degli esiti*[2] e si indica[3] solitamente con Ω o U.

Questo insieme Ω non ha necessariamente un numero finito di elementi: possono essere in numero finito o anche infiniti, numerabili o più che numerabili. Vediamo quindi una prima differenza rispetto al Capitolo 1: vogliamo poter considerare situazioni in cui il numero di esiti è infinito. Questo ci creerebbe qualche problema, se volessimo usare la definizione "casi favorevoli su casi totali", perché dovremmo confrontare cardinali infiniti, ma per essi non vale la legge di cancellazione. Per risolvere questo problema abbiamo bisogno di rendere più robusta la nostra teoria. Prima di proseguire, vediamo alcuni esempi di esperimenti casuali e dei corrispondenti insiemi degli esiti.

- Il lancio di una moneta: in questo caso abbiamo $\Omega = \{\text{testa}, \text{croce}\}$ che è un insieme finito.
- Il numero di tentativi prima di colpire il centro a freccette: qui $\Omega = \mathbb{N}$ e ha cardinalità numerabile. Infatti non è possibile stabilire a priori quanti lanci saranno sufficienti (e dare quindi un limite superiore a Ω).
- Le possibili lunghezze di un segmento contenuto nell'intervallo reale $[0, 1]$: il corrispondente insieme degli esiti è $\Omega = (0, 1]$, un intervallo di cardinalità pari al continuo.

2.2 Algebre e tribù di insiemi

Quando pensiamo ai risultati dell'esperimento, non necessariamente siamo interessati a un solo esito. Ci possono essere occasioni (che costituiranno la maggior parte dei casi) in cui non ci interessa un risultato specifico, ma vogliamo tener conto di tutti quei risultati che soddisfano certe caratteristiche e, quindi, formino un insieme, in particolare un sottoinsieme di Ω. Inoltre possiamo osservare che, se vogliamo pensare per insiemi, gli esiti sono in corrispondenza biunivoca (e quindi identificabili) con i singoletti[4] in Ω. Avere una teoria delle probabilità sugli insiemi, quindi, ci dà come caso speciale la probabilità sugli esiti.

---

[2] Si parla anche di popolazione o di spazio campionario, soprattutto tra gli statistici.

[3] Ω è una lettera dell'alfabeto greco. Si legge *omega*.

[4] Un *singoletto* è un insieme con un solo elemento.

Volgiamo allora la nostra attenzione ai sottoinsiemi di Ω. Come prima cosa ci chiediamo quanti sono.

Proposizione 2.2.1. *Dato un insieme Ω di cardinalità eventualmente infinita (anche più che numerabile) l'insieme delle parti di Ω (o insieme potenza di Ω) $\mathcal{P}(\Omega)$ ha cardinalità $\#\mathcal{P}(\Omega) = 2^{\#\Omega}$.*

La dimostrazione di questo risultato è molto semplice e va vista almeno una volta nella vita. Chi non l'avesse ancora incontrata, la può trovare in Appendice.
Guardando il risultato, possiamo vedere che, anche quando Ω è un insieme finito, il numero dei suoi sottoinsiemi cresce molto in fretta al crescere della cardinalità di Ω.

Esempio 2.2.1. L'insieme $\Omega = \{1, 2, 3\}$ ha $2^3 = 8$ sottoinsiemi. L'insieme $\Omega = \{1, \dots, 10\}$ ne ha 1024. L'insieme $\Omega = \{1, \dots, 20\}$ ne ha oltre un milione.

Quando poi passiamo a considerare un insieme Ω di cardinalità infinita, le cose si complicano ulteriormente, visto che ci sono infiniti "infiniti".

Proposizione 2.2.2. *L'insieme $\mathcal{P}(\mathbb{N})$ ha cardinalità uguale a quella dell'insieme dei numeri reali.*

E qui ci va ancora bene, perché i reali li conosciamo, ma se andiamo un passo oltre siamo persi.

Proposizione 2.2.3. *L'insieme $\mathcal{P}(\mathbb{R})$ ha cardinalità $2^{(2^{\aleph_0})}$, in particolare ha cardinalità maggiore di quella di $\mathbb{R}$.*

Le dimostrazioni delle Proposizioni 2.2.2 e 2.2.3 si possono trovare in Appendice.
Avendo visto questi risultati, sorge spontaneo un pensiero: sarebbe bello poter considerare solo una parte dei sottoinsiemi, qualora non ci interessassero proprio tutti. Ad esempio, se stessimo scegliendo un numero tra tutti i naturali, ma ci interessasse solo sapere se il numero è pari o no, ci farebbe comodo poter considerare solo i due sottoinsiemi "numeri pari" e "numeri dispari", invece che tutti i sottoinsiemi di $\mathbb{N}$.
Pensiamo infatti al nostro obiettivo: vogliamo definire una probabilità, ma vorremmo farlo solo su alcuni insiemi, quelli che ci interessano, e non necessariamente su tutti quanti, perché sarebbe un po' uno spreco. Possiamo pensare che definire la probabilità di un evento abbia un costo non trascurabile, dal momento che, come vedremo, dobbiamo scegliere tale probabilità con attenzione in modo che soddisfi certe importanti proprietà. È un prezzo che non vogliamo pagare inutilmente.
Il nostro piano è quindi quello di considerare in generale una famiglia $\mathcal{F}$ di sottoinsiemi, quindi $\mathcal{F} \subseteq \mathcal{P}(\Omega)$, ma non necessariamente tutto $\mathcal{P}(\Omega)$. Ancora una volta,

pensiamo al nostro traguardo a lungo termine: definire una probabilità su questi sottoinsiemi. Abbiamo quindi bisogno che questa famiglia sia, in un qualche senso, "stabile".
Per capire meglio cosa intendiamo, pensiamo di nuovo al nostro obiettivo: vogliamo definire una probabilità in modo sensato e vogliamo definirla su questa collezione di insiemi. Vorremmo in particolare che questa collezione contenesse l'insieme Ω e che fosse chiusa rispetto alle operazioni di unione, intersezione e complementare. In altre parole: se due insiemi appartengono alla collezione, vorremmo che ci appartenessero anche la loro unione, la loro intersezione e i loro complementari.

Definizione 2.2.4. Una famiglia $\mathcal{F}$ di sottoinsiemi di un insieme Ω è un'*algebra* se valgono tutte le seguenti proprietà:

i. $\Omega \in \mathcal{F}$;

ii. se $A \in \mathcal{F}$, allora anche il suo complementare $A^c \in \mathcal{F}$;

iii-finita. se $A, B \in \mathcal{F}$, allora $A \cup B \in \mathcal{F}$.

Osserviamo che, per come è scritta, la proprietà **iii-finita** della Definizione 2.2.4 dovrebbe essere chiamata **iii-binaria**. Possiamo però estenderla al caso più generale dell'unione finita: se abbiamo una famiglia finita $(A_i)_{i=1}^n$ di sottoinsiemi di Ω tali che $(A_i)_{i=1}^n \subseteq \mathcal{F}$, allora per la proprietà associativa dell'unione $\bigcup_{i=1}^n A_i \in \mathcal{F}$.
Qualche riga più in alto parlavamo di avere una collezione chiusa anche rispetto all'intersezione. La Definizione 2.2.4 non menziona esplicitamente l'intersezione e parla solo di unione e complementare. Tuttavia ci garantisce anche che un'algebra sia chiusa rispetto all'intersezione, assieme ad altre proprietà, come mostrato nel seguente risultato.

Proposizione 2.2.5. *Data un'algebra $\mathcal{F}$ su Ω, valgono le seguenti proprietà:*

1. $\varnothing \in \mathcal{F}$*;*

2. se A*,* $B \in \mathcal{F}$*, allora* $A \cap B \in \mathcal{F}$*;*

3. se $(A_i)_{i=1}^n \subseteq \mathcal{F}$*, allora* $\bigcap_{i=1}^n A_i \in \mathcal{F}$*;*

4. se A*,* $B \in \mathcal{F}$*, allora* $A \setminus B \in \mathcal{F}$*;*

5. se A*,* $B \in \mathcal{F}$*, allora*[5] $A \triangle B \in \mathcal{F}$*.*

---

[5]Con la notazione $A \triangle B$ indichiamo la differenza simmetrica tra gli insiemi A e B. Per ulteriori dettagli rimandiamo all'Appendice.

Dimostrazione. Procediamo in ordine.

1. Sappiamo che $\Omega \in \mathcal{F}$, per la prima proprietà, e che anche il suo complementare appartiene a $\mathcal{F}$, per la seconda. Ma $\Omega^c = \varnothing$, che quindi appartiene a $\mathcal{F}$.

2. Osserviamo che $A \cap B = (A^c \cup B^c)^c$. Ora, sia A^c sia B^c appartengono a $\mathcal{F}$, dunque anche $A^c \cup B^c$ e il suo complementare.

3. Possiamo iterare il ragionamento visto al punto precedente, sfruttando l'associatività dell'intersezione.

4. Anche qui il trucco è riscrivere l'insieme in una forma più comoda della precedente: $A \setminus B = A \cap B^c$. A questo punto ci basta usare la seconda proprietà di algebra e la chiusura rispetto all'intersezione mostrata sopra.

5. Riscriviamo $A \triangle B = (A \cup B) \cap (A^c \cup B^c)$ e concludiamo usando le proprietà già mostrate (i dettagli sono in Appendice). □

Esempio 2.2.2. Prendiamo $\Omega = \{0, 1, 2\}$. Allora $\mathcal{F} = \{\varnothing, \{0\}, \{1, 2\}, \Omega\}$ è un'algebra su Ω. In particolare quest'algebra è diversa dall'insieme potenza $\mathcal{P}(\Omega)$.

Questo ci suggerisce in particolare che, dato un insieme Ω (con almeno due elementi), esiste più di un'algebra su di esso. Quante ne possiamo avere? Nel caso di Ω finito, le algebre sono tante quante le partizioni di Ω, cioè $B_{\#\Omega}$, come abbiamo visto nel Problema 7.

Esempio 2.2.3. Prendiamo ora $\Omega = \{a, b, c, d, e, f, g\}$. Le seguenti famiglie di insiemi non sono algebre:

- $\mathcal{F}_1 = \{\varnothing, \{a, b, c, d, e\}, \Omega\}$. Infatti manca il complementare di $\{a, b, c, d, e\}$; il completamento di $\mathcal{F}_1$ a un'algebra è $\{\varnothing, \{a, b, c, d, e\}, \{f, g\}, \Omega\}$.

- $\mathcal{F}_2 = \{\{a\}, \{b, c, d\}, \{e, f, g\}, \Omega\}$, poiché manca un complementare, l'insieme vuoto.

- $\mathcal{F}_3 = \{\varnothing, \{a\}, \{b\}, \{b, c, d, e, f, g\}, \{a, c, d, e, f, g\}, \Omega\}$, siccome mancano i due insiemi $\{a, b\}$, $\{c, d, e, f, g\}$, l'unione di $\{a\}$ e $\{b\}$ e il suo complementare.

Nella Definizione 2.2.4 la terza proprietà è chiamata "iii-finita" e non "iii": questo potrebbe farci sospettare che esistano famiglie di sottoinsiemi per cui la proprietà è sostituita da una sua variante infinita. Così è, ma è un infinito "controllato".

Definizione 2.2.6. Una famiglia $\mathcal{F}$ di sottoinsiemi di un insieme Ω è una *tribù* (o *σ-algebra*[6]) se valgono tutte le seguenti proprietà:

i. $\Omega \in \mathcal{F}$;

ii. per ogni $A \subseteq \Omega$, se $A \in \mathcal{F}$, allora $A^c \in \mathcal{F}$;

iii. per ogni famiglia numerabile $(A_i)_{i=1}^{+\infty}$ di insiemi di Ω, se tutti gli insiemi A_i della famiglia appartengono a $\mathcal{F}$, allora $\bigcup_{i=1}^{+\infty} A_i \in \mathcal{F}$.

Esempio 2.2.4. L'insieme delle parti di Ω è esso stesso una tribù. Possiamo osservare, infatti, che soddisfa tutte le proprietà richieste. Dal momento che include tutti i possibili sottoinsiemi di Ω, contiene Ω stesso, il complementare di ogni sottoinsieme di Ω e anche ogni unione numerabile di sottoinsiemi di Ω.

Rispetto alla definizione di algebra, stiamo chiedendo che anche l'unione numerabile sia un'operazione interna. Osserviamo inoltre che, se $\mathcal{F}$ è una tribù, è in particolare un'algebra, ma il viceversa non è vero in generale. Vale tuttavia il risultato seguente.

Proposizione 2.2.7. *Sia $\mathcal{F}$ un'algebra finita su un insieme Ω. Allora $\mathcal{F}$ è una tribù.*

Dimostrazione. La differenza tra un'algebra e una tribù sta nella proprietà **iii** della Definizione 2.2.6: dobbiamo mostrare che ogni unione numerabile di elementi di $\mathcal{F}$ sta in $\mathcal{F}$. Siccome $\mathcal{F}$ è finita, contiene solamente un numero finito di elementi, cioè di sottoinsiemi di Ω. Di conseguenza ogni unione numerabile di elementi di $\mathcal{F}$ sarà in realtà un'unione finita, dal momento che abbiamo solo un numero finito di possibili elementi. Tale unione finita appartiene a $\mathcal{F}$, poiché $\mathcal{F}$ è un'algebra. $\square$

Quindi, finché abbiamo a che fare con insiemi finiti, non abbiamo davvero bisogno di parlare di tribù: ci basta controllare che la nostra famiglia di sottoinsiemi sia un'algebra. Questa proprietà ci dà anche un'interessante condizione necessaria affinché una famiglia finita di sottoinsiemi sia una tribù: deve avere un numero di elementi uguale a una potenza di 2. Lasciamo da parte la dimostrazione (che si può fare per induzione, con qualche accortezza), ma osserviamo che grazie a questa condizione abbiamo un modo rapido per dire che una famiglia di sottoinsiemi non è una tribù. Infatti, se una collezione di insiemi ha cardinalità diversa da una potenza di 2, sappiamo che sicuramente non può essere una tribù.
Proseguiamo con altre proprietà di algebre e tribù.

[6] In realtà questo termine non è del tutto corretto: bisognerebbe parlare di *σ-algebre (o σ-campi) di insiemi*, che sono un particolare caso di σ-algebre booleane. Nella pratica tra i probabilisti il termine σ-algebra è sdoganato.

Proposizione 2.2.8. *Date su Ω due algebre $\mathcal{F}_1$ ed $\mathcal{F}_2$, la loro intersezione $\mathcal{F}_1 \cap \mathcal{F}_2$ è a sua volta un'algebra su Ω. Lo stesso vale se sostituiamo "algebra" con "tribù".*

Dimostrazione. Dimostriamo questa proposizione per due tribù: in questo modo abbiamo il risultato anche per le algebre.
Dobbiamo far vedere che $\mathcal{F}_1 \cap \mathcal{F}_2$ soddisfa le proprietà di una tribù. Procediamo punto per punto.

i. Siccome $\mathcal{F}_1$ ed $\mathcal{F}_2$ sono tribù, $\Omega \in \mathcal{F}_1$ e $\Omega \in \mathcal{F}_2$, quindi $\Omega \in \mathcal{F}_1 \cap \mathcal{F}_2$.

ii. Sia $E \in \mathcal{F}_1 \cap \mathcal{F}_2$, allora $E \in \mathcal{F}_1$ ed $E \in \mathcal{F}_2$. Siccome $\mathcal{F}_1$ ed $\mathcal{F}_2$ sono due tribù, $E^c \in \mathcal{F}_1$ ed $E^c \in \mathcal{F}_2$, quindi $E^c \in \mathcal{F}_1 \cap \mathcal{F}_2$.

iii. Prendiamo $(E_i)_{i=1}^{+\infty} \subset \mathcal{F}_1 \cap \mathcal{F}_2$. Allora la successione sarà in entrambe le tribù, $(E_i)_{i=1}^{+\infty} \subset \mathcal{F}_1$ ed $(E_i)_{i=1}^{+\infty} \subset \mathcal{F}_2$. Di conseguenza, $\bigcup_{i=1}^{+\infty} E_i \in \mathcal{F}_1$ e $\bigcup_{i=1}^{+\infty} E_i \in \mathcal{F}_2$ e quindi anche $\bigcup_{i=1}^{+\infty} E_i \in \mathcal{F}_1 \cap \mathcal{F}_2$.

Questo conclude la dimostrazione. □

Avendo preso familiarità con le tribù, ripensiamo al motivo per cui le abbiamo definite. Dato un insieme Ω e una tribù $\mathcal{F}$ su di esso, ci interessano gli elementi di $\mathcal{F}$. Andiamo quindi a dar loro un nome.

Definizione 2.2.9. Sia $\mathcal{F}$ una tribù su Ω. Ogni elemento $E \in \mathcal{F}$ prende il nome di *evento*. I singoletti in $\mathcal{F}$ prendono il nome di *eventi elementari*. Si dice che un evento E si *verifica* se il risultato osservato dell'esperimento casuale è un esito appartenente a E.

Un evento è un elemento di una famiglia di insiemi, quindi è lui stesso un insieme. Questo può alle volte causare un po' di confusione di terminologia con uno scontro tra elementi e insiemi. È per questo che chiamiamo esiti gli elementi di Ω, eventi gli elementi di $\mathcal{F}$ e universo l'insieme Ω.

Esempio 2.2.5. Prendiamo un insieme $\Omega = \{a, b, c, d\}$, e su di esso la tribù $\mathcal{F} = \{\varnothing, \{a\}, \{d\}, \{a, d\}, \{b, c\}, \{a, b, c\}, \{b, c, d\}, \Omega\}$. Allora $A = \{a\}$ è un evento, in particolare un evento elementare, ma è anche un sottoinsieme di Ω, quindi un insieme, e un elemento di $\mathcal{F}$. A sua volta $E = \{a, b, c\}$ è un evento, ma non elementare, mentre $N = \{b\}$ non è un evento, poiché non compare in $\mathcal{F}$.

Non sempre, come vedremo, viene assegnata esplicitamente una tribù e non sempre abbiamo una sola scelta possibile, anche quando sappiamo quali eventi vogliamo che siano al suo interno. In questi casi può venir comoda la seguente definizione.

Definizione 2.2.10. Data una famiglia $\mathcal{G}$ di sottoinsiemi di Ω, definiamo $\sigma(\mathcal{G})$, detta *tribù generata da* $\mathcal{G}$, la più piccola tribù che contiene $\mathcal{G}$, cioè

$$\sigma(\mathcal{G}) = \bigcap \{\mathcal{F} \,:\, \mathcal{F} \text{ è una tribù e } \mathcal{G} \subseteq \mathcal{F}\}\,.$$

Se vogliamo, questo è un modo per semplificarci la vita: sappiamo quali sono gli eventi che vogliamo avere e generiamo a partire da essi una famiglia che li contenga e che sia anche una tribù. Per farlo possiamo pensare di aggiungere a $\mathcal{G}$ i complementari di insiemi di $\mathcal{G}$, poi unioni numerabili, poi ancora complementari e così via. Prendiamo la più piccola possibile perché non vogliamo doverci occupare di più eventi di quanto non sia strettamente necessario. Il perché di questo essere un po' avari, già menzionato in precedenza, sarà chiaro nella prossima sezione.

2.3 Spazi di probabilità

Abbiamo fatto tutto questo lavoro di teoria degli insiemi per poter introdurre le prossime tre definizioni sulla probabilità. Cominciamo mettendo assieme due oggetti che abbiamo già definito.

Definizione 2.3.1. Dati un insieme Ω e una[7] tribù $\mathcal{F}$ su di esso, la coppia $(\Omega, \mathcal{F})$ prende il nome di *spazio probabilizzabile*.

Il nome ci suggerisce che siamo quasi arrivati al nostro obiettivo: abbiamo le fondamenta su cui costruire o definire la probabilità, anche se siamo ancora a una probabilità "in potenza". Ricordiamo che vogliamo far sì che ogni evento abbia una probabilità, quindi dobbiamo definire una funzione che abbia come dominio $\mathcal{F}$. Qui entra in gioco Kolmogorov, che ci dice quali sono le proprietà che deve soddisfare una funzione per essere accettabile come funzione di probabilità.

Definizione 2.3.2. Assegnato uno spazio probabilizzabile $(\Omega, \mathcal{F})$, una funzione $P : \mathcal{F} \to \mathbb{R}$ si dice *funzione o misura*[8] *di probabilità* se soddisfa le seguenti proprietà (dette *assiomi di Kolmogorov*):

1. per ogni evento E, $P(E) \geqslant 0$ (*non negatività*);

[7] Abbiamo già visto, ma lo sottolineiamo ancora una volta, che dato Ω, in genere $\mathcal{F}$ non è unica. Scegliere una particolare tribù tra quelle disponibili è una scelta di modello: a priori non esiste una scelta giusta, dipende dal problema che stiamo considerando. Di volta in volta, sceglieremo $\mathcal{F}$ in modo che sia adatta ai nostri scopi. Vedremo meglio cosa intendiamo nel Capitolo 3.

[8] Il nome *misura* viene dal fatto che questa funzione misura la grandezza dell'evento in termini di probabilità. Vedremo nel Capitolo 3 che prendendo come Ω l'intervallo $[0, 1]$, una particolare misura di probabilità è quella che restituisce la lunghezza dei segmenti, cioè la loro misura.

2. $P(\Omega) = 1$ (*normalizzazione*);

3. data una famiglia numerabile $(E_i)_{i=1}^{+\infty}$ di eventi a due a due disgiunti (cioè $E_i \cap E_j = \varnothing$ se $i \neq j$) allora $P\left(\bigcup_{i=1}^{+\infty} E_i\right) = \sum_{i=1}^{+\infty} P(E_i)$ (σ-*additività*).

Il valore $P(E)$ della funzione in un evento E si dice *probabilità di* E.

Possiamo considerare una versione finita del terzo assioma: se abbiamo una famiglia finita di eventi disgiunti $(E_i)_{i=1}^{n}$, allora $P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i)$. Chiaramente il terzo assioma implica questa versione finita, ma in genere non vale il viceversa, a meno che $\mathcal{F}$ non sia un'algebra finita, cosa che sappiamo essere vera ogni volta che Ω è finito.
Possiamo ora dare la definizione cui stavamo puntando dall'inizio di questo capitolo.

Definizione 2.3.3. Siano Ω un insieme, $\mathcal{F}$ una tribù su Ω e P una funzione di probabilità su $\mathcal{F}$. La tripla $(\Omega, \mathcal{F}, P)$ prende il nome di *spazio di probabilità*.

Vediamo alcuni esempi di spazi di probabilità.

Esempio 2.3.1. Se prendiamo $\Omega = \{0, 1\}$, $\mathcal{F} = \{\varnothing, \{0\}, \{1\}, \{0, 1\}\} = \mathcal{P}(\Omega)$ e P tale che

$$P(\varnothing) = 0, \qquad P(\{0\}) = P(\{1\}) = \frac{1}{2}, \qquad P(\{0, 1\}) = 1,$$

abbiamo uno spazio di probabilità. Possiamo mostrare che tutte le proprietà sono soddisfatte: $\mathcal{F}$ è una tribù, $P(E) \geqslant 0$ per ogni $E \in \mathcal{F}$, $P(\Omega) = 1$, e $P(\{0\}) + P(\{1\}) = 1 = P(\{0, 1\})$.
In particolare se identifichiamo 0 con "testa" e 1 con "croce", questo è un modo di rappresentare il lancio di una moneta bilanciata come spazio di probabilità.

Esempio 2.3.2. Prendiamo ora $\Omega = \{\clubsuit, \diamondsuit, \heartsuit, \spadesuit\}$ e $\mathcal{F} = \mathcal{P}(\Omega)$. Della probabilità P sappiamo quanto segue:

$$\begin{array}{lll} P(\varnothing) = 0 & P(\{\clubsuit\}) = P(\{\diamondsuit\}) = \dfrac{1}{3} & P(\{\clubsuit, \diamondsuit, \spadesuit\}) = \dfrac{7}{9} \\ P(\{\spadesuit\}) = q & P(\{\heartsuit\}) = p. & \end{array}$$

Possiamo determinare p e q tali per cui P può essere una probabilità?

Soluzione. Se vogliamo che P sia una probabilità, $P(\Omega) = 1$ e quindi

$$1 = P(\Omega) = P(\{\clubsuit, \diamondsuit, \spadesuit\} \cup \{\heartsuit\}) = P(\{\clubsuit, \diamondsuit, \spadesuit\}) + P(\{\heartsuit\}) = \frac{7}{9} + p,$$

da cui $p = \frac{2}{9}$. A questo punto possiamo ricavare q, in modo del tutto simile,

$$\begin{aligned} 1 = P(\Omega) &= P(\{\clubsuit\} \cup \{\diamondsuit\} \cup \{\heartsuit\} \cup \{\spadesuit\}) \\ &= P(\{\clubsuit\}) + P(\{\diamondsuit\}) + P(\{\heartsuit\}) + P(\{\spadesuit\}) \\ &= \frac{1}{3} + \frac{1}{3} + \frac{2}{9} + q \end{aligned}$$

da cui $q = \frac{1}{3} - \frac{2}{9} = \frac{1}{9}$.
Perché questo ci dice che P *può* essere una probabilità (e non che P *è* una probabilità)? Perché non sappiamo quanto valga, ad esempio, $P(\{\clubsuit, \diamondsuit\})$. Se fosse $P(\{\clubsuit, \diamondsuit\}) \neq \frac{2}{3}$, P non potrebbe essere una probabilità, perché avremmo una contraddizione con il terzo assioma. □

Prima di continuare, vale la pena fare un'osservazione: gli assiomi di Kolmogorov non ci dicono come definire *la* probabilità sul nostro spazio probabilizzabile, ma ci permettono di dire se una funzione definita su $(\Omega, \mathcal{F})$ sia o meno una misura di probabilità. C'è un buon motivo per cui gli assiomi non ci garantiscono l'unicità della probabilità: questa unicità non c'è! Una volta fissato lo spazio probabilizzabile, possiamo definire più probabilità non equivalenti tra loro.

Esempio 2.3.3. Prendiamo $\Omega = \{0, 1\}$, $\mathcal{F} = \mathcal{P}(\Omega)$, cioè lo stesso spazio probabilizzabile visto nell'Esempio 2.3.1. Possiamo definire $Q : \mathcal{F} \to [0, 1]$ come segue:

$$Q(\varnothing) = 0, \quad Q(\{0\}) = \frac{3}{5}, \quad Q(\{1\}) = \frac{2}{5}, \quad Q(\{0, 1\}) = 1.$$

Anche la funzione Q appena definita è una probabilità, ma è diversa dalla probabilità P vista prima. In particolare, possiamo vedere questo spazio di probabilità come un modello matematico per una moneta sbilanciata in cui la "testa" è una volta e mezza più probabile della "croce".

Quello che abbiamo visto in quest'ultimo esempio non è un caso isolato: vedremo più avanti, nel Capitolo 3, come costruire probabilità in modo che soddisfino gli assiomi di Kolmogorov, ma siano anche buoni modelli per i problemi che considereremo volta per volta.

2.4 Proprietà della (misura di) probabilità

Le proprietà viste sopra sono quelle essenziali per caratterizzare una probabilità. Tuttavia ce ne sono molte altre, che possiamo dedurre da quelle enunciate nella Definizione 2.3.2 e dalle proprietà delle tribù. Nelle prossime pagine ne vedremo un po', alcune ovvie, altre meno. Tutte quante, però, importanti per manipolare le probabilità, come vedremo negli esempi.

Proposizione 2.4.1. *La probabilità dell'evento $\varnothing$ è sempre uguale a 0.*

Dimostrazione. Osserviamo che $\Omega \cup \varnothing = \Omega$ e che allo stesso tempo $\Omega \cap \varnothing = \varnothing$. Allora abbiamo

$$1 = P(\Omega) = P(\Omega \cup \varnothing) = P(\Omega) + P(\varnothing) = 1 + P(\varnothing),$$

in cui la prima e la quarta uguaglianza seguono dal secondo assioma nella Definizione 2.3.2 e la terza identità dal terzo assioma in versione finita. Da questa identità ricaviamo $P(\varnothing) = 0$. □

Proposizione 2.4.2. *Se $E \in \mathcal{F}$, la probabilità del suo complementare E^c è $P(E^c) = 1 - P(E)$.*

Dimostrazione. Come prima cosa, sappiamo che P è definita in E^c, poiché $\mathcal{F}$ è una tribù ed è chiusa rispetto all'operazione di complementare. Inoltre, possiamo osservare che $E \cup E^c = \Omega$ e che $E \cap E^c = \varnothing$, quindi

$$1 = P(\Omega) = P(E \cup E^c) = P(E) + P(E^c),$$

in cui l'ultima uguaglianza segue dal terzo assioma (in versione finita) della Definizione 2.3.2. □

Questa è forse la proprietà della probabilità che sfrutteremo più di tutte nello svolgere esercizi e problemi: molte volte infatti ci verranno forniti dati incompleti, che potremo ricostruire in questo modo. Capiterà spesso che il calcolo diretto della probabilità di un evento sia molto complicato (ad esempio perché ci sono parecchi casi possibili), mentre passando al complementare i conti si semplificano notevolmente.

Esempio 2.4.1. In un "Gratta e vinci"[9] ci sono premi di prima e seconda fascia. La probabilità di vincere un premio di prima fascia è $\frac{1}{1\,000\,000}$, quella di vincere un premio di seconda fascia è $\frac{1}{100}$. Con che probabilità, giocando, non si vince nulla?

---

[9]Le probabilità usate in questo esempio non sono quelle vere, principalmente perché "Gratta e vinci" comprende un'ampia famiglia di lotterie istantanee, che cambia spesso e con premi in numero e taglia variabile. Sono comunque probabilità di un ordine di grandezza non dissimile da quello vero. Per chi volesse approfondire, le coordinate di riferimento sono quelle del sito dell'agenzia Dogane e Monopoli, dove per legge sono mostrate le probabilità dei vari premi nelle varie lotterie: https://www.adm.gov.it/portale/monopoli/giochi/lotterie/lotterie_istantanee/lot_ist_note. Sempre su questo tema e in generale su quello dei giochi d'azzardo e della probabilità a essi collegata, una lettura divertente e interessante è *Fate il nostro gioco* [4].

Soluzione. Le due fasce cui appartengono i premi sono distinte tra loro, quindi la probabilità di vincere qualcosa è la somma delle due probabilità assegnate, cioè $\frac{1}{1\,000\,000} + \frac{1}{100} = \frac{10\,001}{1\,000\,000}$. Allo stesso tempo, non vincere nulla è l'evento complementare al vincere qualcosa, quindi la sua probabilità è

$$1 - \frac{10\,001}{1\,000\,000} = \frac{989\,999}{1\,000\,000} \approx 99\%.$$ □

Introduciamo la prossima proprietà della probabilità con un esempio. È particolare: non ci sono conti da fare e non sono assegnate probabilità di eventi. Vale però anche in questo caso il consiglio di provare a dare una risposta, prima di proseguire la lettura.

Esempio 2.4.2. Linda è una brillante trentunenne, laureata con lode in filosofia. È molto schietta e, da studentessa, si è interessata ai problemi di discriminazione e di giustizia sociale, partecipando anche a manifestazioni contro la corruzione. Sapendo questo, quale delle seguenti è più probabile?

1. Linda è impiegata in una banca italiana.
2. Linda è responsabile delle pari opportunità in Banca Etica.

Prima di discutere questo esempio, vediamo un'altra proprietà, che prende il nome di *monotonia* della probabilità.

Proposizione 2.4.3. *Siano E, F due eventi in $\mathcal{F}$ tali che $E \subseteq F$. Allora vale la disuguaglianza $P(E) \leqslant P(F)$.*

Dimostrazione. Possiamo riscrivere F come

$$F = (E \cap F) \cup (E^c \cap F) = E \cup (E^c \cap F),$$

che è un'unione disgiunta. A questo punto

$$P(F) = P(E) + P(E^c \cap F) \geqslant P(E),$$

dove per l'uguaglianza sfruttiamo il terzo assioma (in versione finita) della Definizione 2.3.2, per la disuguaglianza la non negatività del primo assioma. □

Soluzione all'Esempio 2.4.2. A questo punto la risposta all'esempio di Linda è ovvia: siccome essere una responsabile delle pari opportunità è un caso particolare (un sottoinsieme) dell'essere un'impiegata e Banca Etica è una banca italiana, nonostante il modo in cui Linda ci è stata presentata, è più probabile che sia impiegata in una banca italiana. Questo è un esempio classico, che magari qualcuno ha già visto, introdotto da Kahneman[10] e Tversky[11] per studiare (e mostrare) quanto per noi esseri umani la *narrativa* sia più sedimentata e istintiva rispetto al ragionamento logico o probabilistico[12]. Leggendo il passato di Linda, ci è difficile immaginarla come impiegata in una banca, ma aggiungere che non si tratta di un impiego qualunque, ma uno allineato con i suoi valori, per di più in una banca particolare, ci presenta una storia più credibile. Anche se meno probabile. □

Tornando alle proprietà più matematiche, una conseguenza della monotonia della probabilità è la seguente.

Corollario 2.4.4. *L'immagine della funzione di probabilità è contenuta nell'intervallo unitario* $[0, 1]$.

Dimostrazione. Segue immediatamente dal fatto che, per ogni evento $E \in \mathcal{F}$, $\varnothing \subseteq E \subseteq \Omega$ e dalla monotonia. □

Vediamo ora qualcosa che apparentemente abbiamo già incontrato: la probabilità dell'unione di due eventi.

Proposizione 2.4.5. *Siano E, F due eventi in $\mathcal{F}$, allora la probabilità della loro unione è*

$$P(E \cup F) = P(E) + P(F) - P(E \cap F).$$

Dimostrazione. Come nelle dimostrazioni precedenti vogliamo andare a riscrivere questo insieme come unione disgiunta. Per farlo, osserviamo che $E \cup F = E \setminus F \cup F$ e che $E \setminus F \cap F = \varnothing$. Allora

$$P(E \cup F) = P(E \setminus F) + P(F).$$

Tuttavia, non sappiamo quale sia il valore[13] di $P(E \setminus F)$. Possiamo però riscrivere E come $E = (E \cap F) \cup (E \setminus F)$, notando che si tratta di un'unione disgiunta, quindi $P(E) = P(E \cap F) + P(E \setminus F)$. A questo punto dobbiamo solo andare a sostituire per ottenere la tesi. □

Confrontiamo quanto visto ora e l'enunciato in versione finita del terzo assioma: in quest'ultimo la probabilità dell'unione era la probabilità che accadesse esattamente uno dei due eventi (poiché erano mutualmente esclusivi). Qui invece abbiamo

[10] Daniel Kahneman (1934 –).

[11] Amos Tversky (1937 – 1996).

[12] Per una presentazione divulgativa di questo aspetto della ricerca di Kahneman e Tversky, vedi anche *Pensieri lenti e veloci* [13].

[13] Anche se sappiamo che è definita, poiché $E \setminus F = E \cap F^c \in \mathcal{F}$.

una vera unione: stiamo chiedendo che almeno uno degli eventi si sia verificato e contempliamo anche la possibilità che si siano verificati entrambi. In analogia con il principio di inclusione ed esclusione, togliamo la probabilità dell'evento intersezione, cioè "sono avvenuti entrambi", dal totale, per non contarla due volte.

Esempio 2.4.3. In un videogioco, la probabilità di trovare un oggetto raro in uno dei contenitori posti in giro è del 4%, mentre quella di trovare un oggetto magico è del 12%. La probabilità di trovare un oggetto raro che sia anche magico è dell'1%. Qual è la probabilità di trovare un oggetto che sia magico o raro?

Soluzione. Dobbiamo sommare la probabilità di avere un oggetto magico e quella di avere un oggetto raro, per un totale del 16%. Tuttavia, abbiamo contato due volte la probabilità di avere un oggetto che sia contemporaneamente magico e raro, uguale all'1%. Dobbiamo quindi sottrarre, ottenendo 15%. □

Esempio 2.4.4. In una scuola, la probabilità che una studentessa o uno studente abbia in pagella un'insufficienza in matematica è $\frac{17}{24}$, che ne abbia una in inglese è $\frac{5}{6}$. Quanto vale, come minimo, la probabilità di avere un'insufficienza in entrambe le materie?

Soluzione. Questo problema sembra diverso da quello precedente, ma in realtà possiamo risolverlo in modo molto simile. Cominciamo sommando le due probabilità che conosciamo: $\frac{17}{24} + \frac{5}{6} = \frac{37}{24}$. Osserviamo che questa quantità è maggiore di 1. Non abbiamo ancora scritto nulla che coinvolga la probabilità dell'intersezione, che chiamiamo p ed è la quantità che vogliamo calcolare. Sappiamo però che la probabilità di avere un'insufficienza in almeno una materia è $\frac{17}{24} + \frac{5}{6} - p = \frac{37}{24} - p$. Affinché sia una probabilità, questa quantità deve essere minore o uguale a 1, cioè $\frac{37}{24} - p \leqslant 1$, da cui $\frac{37}{24} - 1 \leqslant p$, quindi $p \geqslant \frac{13}{24}$. □

Abbiamo una conseguenza immediata della Proposizione 2.4.5.

Corollario 2.4.6. *Possiamo maggiorare la probabilità dell'unione di due eventi con la somma delle probabilità dei due eventi:*

$$P(E \cup F) \leqslant P(E) + P(F).$$

Questa proprietà prende il nome di sub-additività.

Riguardiamo la Proposizione 2.4.5 e il parallelo fatto col principio di inclusione-esclusione. Non ci sorprende, a questo punto, che come il principio combinatorio vale per un qualunque numero finito di insiemi, la Proposizione 2.4.5 possa essere estesa a un generico numero n di eventi.

Proposizione 2.4.7. *Sia* $(E_i)_{i=1}^n$ *una famiglia finita di eventi. Allora*

$$P\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n P(E_i) - \sum_{i<j} P(E_i \cap E_j) + \cdots + (-1)^{n+1} P\left(\bigcap_{i=1}^n E_i\right). \tag{2.1}$$

Possiamo generalizzare a questo caso il Corollario 2.4.6.

Corollario 2.4.8. *È possibile maggiorare la probabilità di un'unione finita di eventi con la somma delle probabilità:*

$$P\left(\bigcup_{i=1}^n E_i\right) \leqslant \sum_{i=1}^n P(E_i).$$

Possiamo in realtà dare un risultato più raffinato, ripensando ancora una volta a quanto detto per il principio di inclusione-esclusione.

Proposizione 2.4.9. *La probabilità dell'unione di un numero finito di eventi può essere stimata dall'alto troncando il secondo membro della* (2.1) *in modo che il primo termine che tralasciamo sia di segno negativo, oppure dal basso, se il primo termine che ignoriamo è di segno positivo. In particolare*

$$\sum_{i=1}^n P(E_i) - \sum_{i<j} P(E_i \cap E_j) \leqslant P\left(\bigcup_{i=1}^n E_i\right) \leqslant \sum_{i=1}^n P(E_i).$$

Queste disuguaglianze prendono il nome di disuguaglianze di Bonferroni[14].

A differenza di quanto visto per la combinatoria, però, per la probabilità abbiamo anche il caso delle unioni numerabili: abbiamo stabilito nella definizione di tribù che tali unioni di eventi fossero esse stesse eventi. Un risultato elementare in questo contesto è la generalizzazione del Corollario 2.4.8 al caso numerabile, detta anche disuguaglianza di Boole[15].

Proposizione 2.4.10. *Data una famiglia numerabile di eventi* $(E_i)_{i=1}^{+\infty}$*, possiamo stimare dall'alto la probabilità della sua unione con la somma delle probabilità dei singoli eventi:*

$$P\left(\bigcup_{i=1}^{+\infty} E_i\right) \leqslant \sum_{i=1}^{+\infty} P(E_i).$$

Questo significa che la probabilità è σ*-sub-additiva.*

[14] Carlo Emilio Bonferroni (1892 – 1960).

[15] George Boole (1815 – 1864).

Dimostrazione. Per l'unione numerabile al momento abbiamo solo l'assioma 3, quindi dobbiamo trovare un modo di riscrivere il problema in termini di unione di eventi disgiunti. Possiamo farlo nel modo seguente:

$$\begin{cases} F_1 = E_1 \\ F_k = E_k \setminus \bigcup_{i=1}^{k-1} F_i, \quad k \geqslant 2. \end{cases}$$

In questo modo gli eventi F_i sono a due a due disgiunti e la loro unione coincide con l'unione degli E_i. In più, per ogni $k \in \mathbb{N}$, $F_k \subseteq E_k$, quindi possiamo sfruttare la monotonia:

$$P\left(\bigcup_{i=1}^{+\infty} E_i\right) = P\left(\bigcup_{i=1}^{+\infty} F_i\right) = \sum_{i=1}^{+\infty} P(F_i) \leqslant \sum_{i=1}^{+\infty} P(E_i),$$

in cui abbiamo usato per seconda uguaglianza il terzo assioma nella Definizione 2.3.2 e per la disuguaglianza la Proposizione 2.4.3. □

2.5 Esercizi

Problema 10

Lanciando un dado a 12 facce in cui ogni faccia pari esce con probabilità $\frac{1}{18}$ e ogni faccia dispari con probabilità $\frac{1}{9}$, con che probabilità esce un multiplo di 3 o di 7?

Problema 11

In una particolare estrazione del Lotto matematico su tutti i numeri naturali, gli infiniti numeri non escono tutti con la medesima probabilità. Sui numeri dispari abbiamo un po' di informazioni: 1 esce con probabilità $\frac{1}{3}$, 3 con probabilità $\frac{1}{9}$, 5 con probabilità $\frac{1}{27}$ e così via. In generale il k-esimo numero dispari esce con probabilità $\frac{1}{3^k}$. La probabilità che esca un numero pari, poi, è doppia rispetto alla probabilità che esca un numero pari positivo. Con che probabilità esce 0?

Problema 12

Un'urna contiene 16 biglie bianche e 11 nere. Pescandone 4 assieme, qual è la probabilità che non siano tutte del medesimo colore?

Problema 13

A una festa ciascuna delle n invitate porta un regalo, chiuso in un pacchetto e incartato. Questi pacchetti, tutti della stessa dimensione e con la stessa carta, vengono messi su un tavolo e, nel momento clou della festa, ridistribuiti tra le partecipanti. Qual è la probabilità che almeno un'invitata riceva il regalo che ha portato?

Problema 14

Dati un insieme Ω e due algebre $\mathcal{A}$ e $\mathcal{B}$ su di esso, l'unione $\mathcal{A} \cup \mathcal{B}$ è anch'essa un'algebra su Ω?

Problema 15

Dato un insieme Ω, è vero che $\mathcal{A} = \{E \subseteq \Omega \, : E \text{ è finito}\}$ è sempre un'algebra?

3. Costruire spazi di probabilità

In questo capitolo introdurremo alcuni modi per costruire spazi di probabilità che soddisfino le proprietà di tribù e gli assiomi di Kolmogorov visti nel capitolo precedente. Come vedremo, non abbiamo a disposizione algoritmi che ci dicono esattamente cosa fare in ogni caso o che ci danno il miglior modello possibile. Dovremo sporcarci le mani e dare il nostro contributo, spesso in maniera ovvia, ma alle volte pensando in modo creativo.

Un problema nasce nel momento in cui vogliamo usare la probabilità come strumento per descrivere matematicamente un fenomeno reale. Non è detto che possiamo misurare le probabilità di tutti gli eventi che ci interessano: potremmo non avere a disposizione i dati che ci occorrono.

In realtà c'è anche un aspetto filosofico non trascurabile, al quale però accenniamo solamente: come facciamo ad assegnare una probabilità a qualcosa? Proviamo a pensarci un momento: cosa significa dire che un certo evento ha probabilità p? Cosa vuol dire che la probabilità di fare 6 in un dado non truccato è $\frac{1}{6}$? Una possibile interpretazione, detta *frequentista*[1], sostiene che potendo ripetere il nostro esperimento infinite volte, la probabilità di un evento altro non è che la frequenza con cui l'evento si realizza in queste prove.

È un'interpretazione sensata, ma presta il fianco a due domande insidiose: come possiamo ripetere un esperimento infinite volte e come facciamo ad assegnare delle

[1] Tra i primi frequentisti ricordiamo Richard Edler von Mises (1883 – 1953).

probabilità a un esperimento che non può essere ripetuto? Il tipico esempio in questo caso è quello delle elezioni: cosa vuol dire che in una certa sfida elettorale la candidata F ha una probabilità p di vincere e il suo sfidante Q ha una probabilità $1 - p$ di vincere? Non possiamo pensarla come frequenza, dato che questa elezione è unica: non si ripeterà mai allo stesso modo (anche se si dovessero ripresentare gli stessi sfidanti, l'istante nel tempo sarà cambiato e quindi anche i votanti e le loro convinzioni potrebbero essere diversi). Da un punto di vista frequentista, non è possibile in questo caso parlare di probabilità e men che meno assegnarle dei valori.

Il fatto stesso che in occasione delle elezioni vengano dichiarate delle probabilità ci suggerisce che esistano altri punti di vista. Uno di questi è il punto di vista *soggettivista bayesiano*[2], nel quale la probabilità di un evento rappresenta il grado di convinzione o di certezza che un soggetto assegna all'evento: in altre parole quanto il soggetto sarebbe disponibile a scommettere (in euro, dollari, altra unità a piacere, supponendole infinitamente divisibili) per avere 1 (euro, dollaro, unità a piacere) se l'evento si verifica e 0 altrimenti.

Anche questo punto di vista ha i suoi pregi e i suoi difetti. Una critica che i frequentisti muovono ai bayesiani è che individui diversi potrebbero assegnare probabilità diverse e che quindi la probabilità non può più essere considerata una proprietà intrinseca dell'evento. Inoltre non è chiaro *come* gli individui possano misurare il loro grado di certezza: per i bayesiani i metodi possono dipendere dal particolare esperimento casuale che si considera, passando da esperimenti di laboratorio a sondaggi popolari a interrogazioni di esperti o un misto[3] di questi.

La contrapposizione, che abbiamo visto in maniera estremamente semplificata, tra frequentisti e bayesiani non esaurisce la varietà delle interpretazioni della probabilità. Abbiamo appena sfiorato la superficie di una discussione filosofica sicuramente affascinante, ma approfondirla ci porterebbe fuori strada rispetto agli obiettivi di questo libro. In un certo senso da un punto di vista astratto ci importa poco: finché non assegniamo un valore specifico alla probabilità di un evento, questo dibattito non entra in gioco e per molti esperimenti semplici le due filosofie danno il medesimo valore di probabilità. Inoltre, come già detto, nell'ambito della teoria di Kolmogorov il risultato è vero indipendentemente da come viene assegnata la probabilità.

---

[2] Tra i primi soggettivisti ricordiamo Bruno De Finetti (1906 – 1985) e Frank Plumpton Ramsey (1903 – 1930).

[3] Discuteremo ulteriormente nel Capitolo 4 di come mettere insieme valutazioni diverse della probabilità di un evento, in particolare nel contesto del Teorema di Bayes (Thomas Bayes, 1701 – 1761), dal quale i bayesiani prendono il nome.

3.1 Spazi finiti o numerabili

Cominciamo esaminando un caso semplice. Supponiamo di aver individuato lo spazio degli esiti Ω e di aver visto che esso è un insieme finito o numerabile. Come prima cosa prendiamo $\mathcal{F} = \mathcal{P}(\Omega)$, cioè l'insieme delle parti di Ω. In altre parole vogliamo che tutti i possibili sottoinsiemi di Ω siano eventi. Se siamo alla ricerca di un algoritmo generale, questa è una buona idea, perché qualunque insieme ci capiti di avere in Ω, esso potrà avere una probabilità.
Come abbiamo visto, la cardinalità di $\mathcal{F}$ è $2^{\#\Omega}$. Nel caso finito non è un grave problema, ma nel caso numerabile dovremo andare ad assegnare una probabilità a tanti eventi quanti sono i numeri reali (non solo infiniti, ma più che numerabili). Questo può sembrare un problema, dal momento che non lo possiamo fare ricorsivamente, a differenza di quanto accade nel caso di una quantità numerabile di oggetti.
Ma proprio qui sta il trucco: andiamo ad assegnare una probabilità a ciascun singoletto in Ω, in modo che per ogni $\omega \in \Omega$, $P(\{\omega\}) \geqslant 0$ e $\sum_{\omega\in\Omega} P(\{\omega\}) = 1$.
In pratica quello che facciamo è scegliere una funzione che soddisfi queste due proprietà, più eventuali altre condizioni imposte dal problema specifico, e a questo punto siamo a posto. Infatti per ogni $E \in \mathcal{F}$

$$P(E) := \sum_{\omega\in E} P(\{\omega\}),$$

dove abbiamo usato una proprietà che volevamo soddisfare, ossia che la probabilità di un'unione disgiunta sia la somma delle probabilità. Inoltre possiamo osservare che la somma si svolge su una quantità di indici al più numerabile, quindi non stiamo commettendo alcun abuso di notazione[4].
La difficoltà più grande in questo caso è individuare una funzione P definita su Ω che soddisfi le due proprietà enunciate sopra, cioè la non negatività e la somma a 1, e che al contempo catturi le proprietà del particolare problema che stiamo considerando.

Esempio 3.1.1. Tre amici si sfidano abitualmente nella corsa, sempre sullo stesso percorso. Prisca arriva per prima il doppio delle volte di Carlo, Daniele arriva primo la metà delle volte di Carlo. Qual è la probabilità che, in un giorno qualunque, Carlo sia il più veloce?

[4] Un vero abuso che si vede spesso è il seguente: si lasciano cadere le parentesi graffe e si identifica il singoletto di ω, un evento, con ω stesso, un esito. Anche se il desiderio di alleggerire la notazione è condivisibile, si tratta di una scelta pericolosa, perché genera ambiguità.

Soluzione. Indichiamo con d la frequenza con cui Daniele vince. Dai dati del problema sappiamo che Carlo vince con frequenza $2d$ e Prisca con frequenza $2 \cdot 2d = 4d$. Sappiamo anche che, dal momento che i concorrenti sono solo loro tre, $1 = 4d + 2d + d = 7d$, cioè Carlo arriva primo con probabilità $\frac{2}{7}$. $\square$

Non sempre, però, abbiamo le informazioni per dare una probabilità esplicita a ogni esito, come vedremo nel prossimo esempio. In questo caso abbiamo due possibilità: accontentarci di assegnare la probabilità solo su una tribù di eventi, oppure cambiare l'insieme Ω in modo che gli "eventi indivisibili" diventino esiti nella nuova rappresentazione.

Esempio 3.1.2. Sull'isola dei matematici applicati c'è una particolare lotteria, in cui viene estratto un numero naturale a caso. Tuttavia, non tutti i numeri hanno la medesima probabilità di uscire: ciascun numero pari ha la stessa probabilità di uscire, il 7 esce con probabilità $\frac{1}{2}$, l'evento $\{1, 2, 3, 5\}$ ha probabilità $\frac{1}{3}$, mentre gli eventi $\{9\}$, $\{9, 11\}$ e $\{n : n \geqslant 9\}$ hanno la stessa probabilità.
Possiamo iniziare osservando che i numeri pari possono avere solamente probabilità 0: se così non fosse, avremmo una probabilità totale maggiore di 1, dal momento che i numeri naturali soddisfano la proprietà archimedea. Questo ci dice anche che $P(\{1, 2, 3, 5\}) = P(\{1, 3, 5\}) = \frac{1}{3}$. Con le stesse idee possiamo anche mostrare che ogni numero naturale strettamente maggiore di 9 ha probabilità 0. A questo punto sappiamo che

$$1 = P(\Omega) = P(\{1, 3, 5\}) + P(7) + P(9) + P(\{0, 2, 4, 6, 8\}) + P(\{n : n > 9\}),$$

quindi $P(\{9\}) = \frac{1}{6}$. Osserviamo che, con i dati forniti, non siamo in grado di dire quali siano le probabilità degli eventi $\{1\}$, $\{3\}$, $\{5\}$, $\{1, 3\}$, $\{1, 5\}$, $\{1, 7\}$...
Possiamo considerare solo eventi in cui $\{1, 3, 5\}$ sia un blocco unico.
Un modo per ricondursi a quanto visto prima è scegliere un Ω diverso. In questo caso prendiamo, per esempio, Ω che ha per elementi l'insieme dei naturali pari, l'insieme dei naturali dispari maggiori di 5 e l'insieme $\{1, 3, 5\}$.

Nel caso numerabile, dato che ci sono infiniti singoletti, potremmo aspettarci che un numero infinito di essi dovrà necessariamente avere probabilità zero. Questo è falso, come possiamo vedere nel seguente esempio.

Esempio 3.1.3. Anche sull'isola dei matematici puri c'è una lotteria infinita, su tutti i numeri naturali, in cui ogni numero ha il doppio della probabilità di essere estratto rispetto al suo successore.

In questo caso abbiamo bisogno di sfruttare la serie geometrica[5]. Sappiamo infatti che, posta z la probabilità di estrarre 0, la probabilità di estrarre n è $2^{-n} \cdot z$, ma anche che

$$1 = \sum_{n=0}^{+\infty} 2^{-n} \cdot z = z \cdot \sum_{n=0}^{+\infty} 2^{-n} = z \cdot 2,$$

da cui abbiamo che lo zero esce con probabilità $\frac{1}{2}$ e che in generale un numero naturale n esce con probabilità $2^{-(n+1)}$. In particolare, nessun numero naturale ha probabilità 0 di uscire.

3.2 Lo spazio dei numeri reali

Consideriamo ora il caso in cui Ω è l'intervallo di numeri reali $[0,1]$. Dobbiamo scegliere la tribù e valutare come definire una misura di probabilità. Cominciamo dalla tribù.

Come già accennato in precedenza, potremmo prendere come tribù l'insieme delle parti di $[0,1]$, ma questo ha cardinalità pari all'insieme potenza di $\mathbb{R}$, cioè $2^{(2^{\aleph_0})}$, che è un po' grande per i nostri gusti, visto che poi a ogni elemento della tribù andrà assegnata una probabilità[6]. Consideriamo insiemi di numeri reali. Quelli che ci possono venire in mente di solito[7] sono punti singoli, segmenti, semirette e loro combinazioni (unioni finite o numerabili, differenze e così via). Dato che per il momento ci stiamo interessando solamente all'intervallo $[0,1]$, intersecheremo quest'ultimo con gli insiemi visti sopra. Dentro alla nostra tribù dovranno esserci insiemi di questo tipo, perché è di questi che vogliamo calcolare la probabilità.

In altre parole, vogliamo la tribù generata da punti isolati, intervalli (aperti, chiusi, semiaperti a destra e a sinistra) e loro unioni numerabili, cioè la più piccola tribù che contiene tutti questi insiemi. Con un po' di teoria degli insiemi possiamo osservare che a partire dai soli intervalli chiusi in $[0,1]$ possiamo ottenere, attraverso il passaggio al complementare e all'unione numerabile:

- gli intervalli aperti (a,b), definendo per ogni $n \in \mathbb{N} \setminus \{0\}$, $I_n = \left[a + \frac{1}{n}, b - \frac{1}{n}\right]$, intervallo chiuso, e prendendone l'unione $\bigcup_{n\in\mathbb{N}} I_n = (a,b)$;
- gli intervalli semiaperti della forma $[a,b)$ e $(a,b]$, in maniera analoga;

[5] Per un brevissimo ripasso sulla serie geometrica, vedere l'Appendice.

[6] Ci sono anche altri motivi per non scegliere l'insieme delle parti: non possiamo farlo, se vogliamo definire una probabilità che soddisfi alcune ragionevoli condizioni. Discutere di questo, però, ci porterebbe un po' troppo fuori strada, verso la teoria della misura.

[7] Qualcuno potrebbe pensare immediatamente a casi patologici come l'insieme di Vitali (Giuseppe Vitali, 1875 – 1932), ma probabilmente non ha bisogno di questo libro.

- i singoletti;
- le intersezioni...

Quindi se vogliamo avere una tribù che contenga tutti questi insiemi, possiamo generarla a partire dai soli intervalli chiusi, dal momento che unioni numerabili e complementari di elementi di una tribù sono essi stessi nella tribù.
In realtà è possibile usare come generatori gli intervalli semiaperti a sinistra, ossia della forma $(a, b]$. Come vedremo tra poco, questo modo di procedere è anche più comodo. In modo analogo a quanto fatto sopra, a partire dagli intervalli semiaperti a sinistra possiamo ottenere (con unioni numerabili e passaggi al complementare) gli intervalli chiusi e, di conseguenza, tutti gli altri insiemi che ci interessano. Questo è il Problema 17, al termine di questo capitolo.
Insomma, sia usando gli intervalli chiusi, sia usando gli intervalli semichiusi a destra (cioè semiaperti a sinistra), generiamo una tribù che soddisfa le nostre richieste, poiché contiene gli insiemi che consideriamo interessanti. Essa prende il nome di *tribù dei Boreliani*[8] (su $[0, 1]$) e viene indicata con $\mathcal{B}([0, 1])$. La sua cardinalità è quella del continuo, cosa che non dimostreremo qui (si fa per induzione transfinita).
Ora che abbiamo Ω e $\mathcal{F}$, dobbiamo solo scegliere una misura di probabilità. Anche in questo caso, come in quello degli spazi finiti o numerabili, non esiste un'unica scelta: il modo in cui definiamo la probabilità dipende dal problema che stiamo considerando. Tuttavia possiamo stabilire una procedura per definire misure di probabilità valide: dal momento che dobbiamo assegnare una probabilità a ciascun evento, cioè a ciascun elemento della tribù dei Boreliani, cominciamo assegnando una probabilità a ciascun intervallo utilizzato per generare la tribù[9]. Vogliamo farlo in modo che la probabilità dipenda solo dai due estremi dell'intervallo, senza dimenticare che anche le altre proprietà devono essere soddisfatte.

Esempio 3.2.1. Una possibile scelta di misura di probabilità sull'intervallo unitario $[0, 1]$ è la seguente: interpretiamo ogni intervallo $[a, b]$ contenuto in $[0, 1]$ come un segmento e gli assegniamo come probabilità la sua lunghezza. Abbiamo allora $P([a, b]) = b - a$.
A partire da questo, possiamo calcolare le probabilità degli altri elementi della tribù, anche di forma diversa da $[a, b]$, come ad esempio $[a, b)$, sfruttando gli assiomi

---

[8] Émile Borel (1871 – 1956).

[9] Il fatto che questo sia sufficiente a definire una probabilità su tutta la tribù dei Boreliani, anche se è intuitivo, non è un fatto banale. Esiste però un risultato, il Teorema di Carathéodory (Constantin Carathéodory, 1873 – 1950), che ce lo garantisce, ma che in questo libro è relegato, senza dimostrazione, in Appendice.

di Kolmogorov. Infatti, preso c tale che $b \leqslant c \leqslant 1$, abbiamo $[a, c] = [a, b) \cup [b, c]$, in cui l'unione è disgiunta. Allora

$$c - a = P([a, c]) = P([a, b)) + P([b, c]) = P([a, b)) + (c - b),$$

da cui $P([a, b)) = c - a - (c - b) = b - a$. In modo analogo possiamo calcolare la probabilità degli altri elementi della tribù, ad esempio quella dei singoletti. Questa è *una* possibile scelta di probabilità sull'intervallo $[0, 1]$, che prende anche il nome di *probabilità uniforme* o *misura di Lebesgue*[10], ma non è l'unica.

Se vogliamo che la probabilità di un intervallo dipenda solo dai suoi estremi, possiamo considerare una funzione $F : [0, 1] \to \mathbb{R}$, tale che $P((a, b]) = F(b) - F(a)$. Da questo punto di vista la probabilità nell'Esempio 3.2.1 è stata ottenuta scegliendo $F(x) = x$ (la funzione identità) nell'intervallo $[0, 1]$. Chiaramente ci sono altre scelte possibili, come vedremo ora.

Esempio 3.2.2. Prendiamo la funzione $F : [0, 1] \to \mathbb{R}$ definita da

$$F(x) = \begin{cases} x & 0 \leqslant x < \frac{1}{2} \\ \frac{1}{2}(x+1) & \frac{1}{2} \leqslant x \leqslant 1 \end{cases} .$$

La probabilità definita sulla tribù dei Boreliani a partire da $P((a, b]) = F(b) - F(a)$ non è più quella uniforme vista nell'Esempio 3.2.1. Per certi intervalli (e quindi per certi eventi) le due probabilità coincidono, ma possiamo vedere facilmente che su alcuni intervalli, come ad esempio $\left(\frac{1}{4}, \frac{3}{4}\right]$, esse assumono valori diversi. Inoltre, in questo caso, non è sempre vero che $P((a, b]) = P((a, b))$. Infatti $P\left(\left(\frac{1}{4}, \frac{1}{2}\right]\right) = \frac{3}{4} - \frac{1}{4} = \frac{1}{2}$, mentre

$$\begin{aligned} P\left(\left(\tfrac{1}{4}, \tfrac{1}{2}\right)\right) &= P\left(\bigcup_{n \in \mathbb{N}} \left(\tfrac{1}{4}, \tfrac{1}{2} - \tfrac{1}{n}\right]\right) \\ &= \lim_{n \to +\infty} P\left(\left(\tfrac{1}{4}, \tfrac{1}{2} - \tfrac{1}{n}\right]\right) \\ &= \lim_{n \to +\infty} F\left(\tfrac{1}{2} - \tfrac{1}{n}\right) - F\left(\tfrac{1}{4}\right) \\ &= \tfrac{1}{2} - \tfrac{1}{4} = \tfrac{1}{4}, \end{aligned}$$

in cui abbiamo dovuto scomodare il passaggio al limite[11] per n che tende a $+\infty$.

[10] Henri Léon Lebesgue (1875 – 1941).

[11] Per chi non conoscesse il concetto di limite, il suggerimento è quello di tracciare il grafico della funzione F e vedere cosa succede in $\frac{1}{2}$: se ci avviciniamo sempre di più a $\frac{1}{2}$ da sinistra, lungo il grafico, senza raggiungerlo, il valore della funzione F è sempre più vicino a $\frac{1}{2}$ anch'esso, pur non essendo questo il valore che la funzione assume per $x = \frac{1}{2}$.

Di conseguenza abbiamo anche che $P\left(\left\{\frac{1}{2}\right\}\right) = P\left(\left(\frac{1}{4}, \frac{1}{2}\right]\right) - P\left(\left(\frac{1}{4}, \frac{1}{2}\right)\right) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$, mentre si può verificare (con l'ausilio dei limiti) che per ogni $x \neq \frac{1}{2}$ in $[0, 1]$, $P(\{x\}) = 0$.

Non tutte le funzioni F vanno bene, però: non dobbiamo dimenticare che stiamo cercando delle probabilità, quindi gli assiomi dovranno essere soddisfatti. Abbiamo visto, nell'Esempio 3.2.2, che F non deve necessariamente essere continua. Tuttavia deve essere monotona debolmente crescente, poiché per $0 \leqslant a < b < c \leqslant 1$,

$$F(b) - F(a) = P((a, b]) \leqslant P((a, c]) = F(c) - F(a),$$

per la Proposizione 2.4.3, quindi $F(c) \geqslant F(b)$. Questo ancora non basta: per vedere le altre proprietà di queste funzioni, conviene però passare al caso in cui Ω è l'intera retta reale e considerare $[0, 1]$ come un caso speciale.
Se vogliamo lavorare sull'intera retta dei numeri reali $\mathbb{R}$, dobbiamo come prima cosa definire nuovamente la tribù che consideriamo. I Boreliani su $[0, 1]$ non sono più sufficienti, ma basterà modificarli un po' per estenderli a tutto $\mathbb{R}$.
Quali sono queste modifiche? Per comodità, al posto dei segmenti prenderemo le semirette come mattoni base della nostra costruzione. In particolare, sostituiamo gli intervalli semiaperti $(a, b]$ con le semirette sinistre chiuse, cioè della forma $(-\infty, b]$, con $b \in \mathbb{R}$ (e non più limitato al solo intervallo $[0, 1]$).
A partire da queste semirette possiamo generare, con le solite operazioni di unione numerabile e passaggio al complementare, gli intervalli (aperti, chiusi e semiaperti), i singoletti, le semirette sinistre aperte e le semirette destre aperte e chiuse, nonché tutte le loro unioni: abbiamo quindi dei buoni generatori. La tribù generata dalle semirette sinistre chiuse prende il nome di *tribù dei Boreliani* (su $\mathbb{R}$) e viene indicata con $\mathcal{B}(\mathbb{R})$ (o brevemente con $\mathcal{B}$)[12]. La sua cardinalità è anche in questo caso quella del continuo.
Per definire una probabilità sullo spazio probabilizzabile $(\mathbb{R}, \mathcal{B})$, sfruttiamo la medesima idea vista per l'intervallo unitario: la faremo dipendere solamente dagli estremi. In questo caso però abbiamo un solo estremo "agibile": il secondo. Allora definiamo la probabilità della semiretta $(-\infty, b]$ come funzione del solo estremo b, mediante un'opportuna funzione F definita su tutti i reali: $P((-\infty, b]) = F(b)$. Questo è del tutto compatibile con quanto visto prima: per differenza di insiemi abbiamo infatti che $P((a, b]) = P((-\infty, b]) - P((-\infty, a]) = F(b) - F(a)$.
Non tutte le funzioni F vanno bene, tuttavia. Abbiamo già visto che F deve essere monotona non decrescente, ma ora non possiamo più avere come probabilità la

[12] Si può ottenere la stessa tribù anche usando altri generatori, ma come vedremo questa scelta è particolarmente comoda per definire le probabilità.

lunghezza dei segmenti, ossia F uguale all'identità: dal momento che le semirette hanno lunghezza infinita, non potremmo più rispettare gli assiomi di Kolmogorov[13]. Abbiamo però una buona caratterizzazione delle funzioni ammesse: sono quelle funzioni $F : \mathbb{R} \to \mathbb{R}$ tali che

- F è non decrescente (o debolmente crescente);
- esiste il limite di $F(x)$ per x che tende a $+\infty$ e vale $\lim_{x\to+\infty} F(x) = 1$;
- esiste il limite di $F(x)$ per x che tende a $-\infty$ e vale $\lim_{x\to-\infty} F(x) = 0$;
- in ogni punto x_0 la funzione F è continua a destra, cioè $\lim_{x\to x_0^+} F(x) = F(x_0)$ e limitata a sinistra, ossia $\lim_{x\to x_0^-} F(x) \leqslant F(x_0)$.

La prima di queste proprietà ci è familiare e segue dalla monotonia della probabilità. Le due successive seguono dal fatto che $P(\Omega) = P(\mathbb{R}) = 1$. L'ultima proprietà (o meglio, le due proprietà all'ultimo punto) possono apparire più sorprendenti. In realtà servono per darci la possibilità di assegnare a un punto una probabilità diversa da 0:

$$\begin{aligned} P(\{x_0\}) &= P((-\infty, x_0]) - P\left(\bigcup_{n\in\mathbb{N}} (-\infty, x_0 - \tfrac{1}{n}]\right) \\ &= F(x_0) - \lim_{n\to+\infty} F\left(x_0 - \tfrac{1}{n}\right) \\ &= F(x_0) - \lim_{x\to x_0^-} F(x). \end{aligned}$$

Allo stesso tempo ci garantiscono che la probabilità si comporta bene anche in tali punti e, in particolare,

$$\lim_{n\to+\infty} F\left(b + \tfrac{1}{n}\right) - F(a) = \lim_{n\to+\infty} P\left(\left(a, b + \tfrac{1}{n}\right]\right) = P((a, b]) = F(b) - F(a).$$

Insomma, ci basta definire una funzione F di questo tipo per avere una probabilità sulla retta reale[14].

Esempio 3.2.3. Consideriamo la funzione $F : \mathbb{R} \to \mathbb{R}$ definita da

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-x} & x \geqslant 0 \end{cases}.$$

[13]Non possiamo nemmeno prendere una funzione che sia proporzionale alla lunghezza dei segmenti, perché avremmo il medesimo problema.

[14]Stiamo ancora imbrogliando, perché stiamo sfruttando in silenzio il Teorema di Carathéodory già nominato in precedenza.

Questa funzione soddisfa le proprietà viste sopra (è addirittura continua in ogni punto), quindi definisce una probabilità. Possiamo in particolare vedere che ogni intervallo non vuoto nei reali positivi ha una probabilità strettamente positiva:

$$P((a,b]) = F(b) - F(a) = 1 - e^{-b} - 1 + e^{-a} = e^{-a} - e^{-b},$$

mentre ogni singoletto ha probabilità 0 (conseguenza del fatto che F è continua).

Non possiamo davvero apprezzarlo in questo libro, ma imparare a costruire una misura di probabilità (e quindi uno spazio di probabilità) sui reali è un ottimo investimento, se non addirittura il migliore che possiamo fare. È infatti possibile trasformare ogni esperimento aleatorio in uno equivalente in cui lo spazio probabilizzabile sia $(\mathbb{R}, \mathcal{B})$ e tutte le caratteristiche peculiari del problema siano codificate dalla probabilità P (cioè dalla funzione F, che prende il nome di *funzione di ripartizione*). Questo è reso possibile dalla nozione di variabile aleatoria o casuale[15].

3.3 Spazi prodotto

In probabilità succede spesso che qualcosa possa essere visto come una combinazione di più fenomeni aleatori. Quando questi sono distinti e non si influenzano a vicenda (un concetto che approfondiremo meglio parlando di indipendenza, nel Capitolo 4), possiamo descriverli tutti assieme come spazio prodotto, portandoci dietro quello che sappiamo sulle varie componenti. Per farci un'idea, vediamo qualche esempio.

Esempio 3.3.1. Se lanciamo un dado a 4 facce e una moneta, possiamo scrivere gli esiti come coppie ordinate in cui la prima componente è l'esito del lancio del dado e la seconda l'esito del lancio della moneta. In altre parole, $\Omega = \{(1,T),(2,T),(3,T),(4,T),(1,C),(2,C),(3,C),(4,C)\}$. Come insieme, questo è il prodotto cartesiano dei due insiemi universo $\Omega_1 = \{1,2,3,4\}$ e $\Omega_2 = \{T,C\}$, cioè $\Omega = \Omega_1 \times \Omega_2$. Se assumiamo che il dado e la moneta non si influenzino, possiamo definire una probabilità su questo spazio a partire dalle probabilità del dado e della moneta, ovviamente su un'opportuna tribù.

Esempio 3.3.2. Prendiamo ora n monete tutte uguali tra loro e lanciamole (o in alternativa prendiamo una sola moneta e lanciamola n volte). In questo caso uno spazio naturale per descrivere il fenomeno è quello delle n-uple ordinate di elementi di $\Omega_1 = \{T,C\}$, cioè $\Omega = (\Omega_1)^n$. E se pensassimo di lanciare la moneta

[15] Incontreremo di nuovo le variabili aleatorie parlando di valore atteso nella Sezione 4.3, ma sempre di sfuggita: purtroppo non possiamo dedicare loro spazio in questo libro.

infinite volte? Avremmo che un esito è una successione di elementi di Ω_1, cioè avremmo $\Omega = (\Omega_1)^{\mathbb{N}}$. In entrambi i casi, però, per poter calcolare probabilità di eventi abbiamo bisogno di definire una tribù $\mathcal{F}$ e una funzione di probabilità P. Nel secondo caso possiamo identificare Ω con $\{0,1\}^{\mathbb{N}}$, che a sua volta possiamo identificare coi numeri reali in $[0,1]$: potremmo allora usare quanto visto nella Sezione 3.2. Questo maschererebbe però la struttura di "esperimento ripetuto" che invece è più facilmente riconoscibile nella rappresentazione come prodotto (infinito).

Come primo caso, consideriamo il prodotto di due esperimenti aleatori descritti rispettivamente dagli spazi di probabilità $(\Omega_1, \mathcal{F}_1, P_1)$ e $(\Omega_2, \mathcal{F}_2, P_2)$. Vogliamo costruire uno spazio di probabilità $(\Omega, \mathcal{F}, P)$ che descriva la coppia di esperimenti. Iniziamo dallo spazio degli esiti: come abbiamo già detto nell'Esempio 3.3.1, è ragionevole prendere il prodotto cartesiano $\Omega = \Omega_1 \times \Omega_2$.

Passiamo allora alla tribù: $\mathcal{F}$ sarà generata dai prodotti di elementi delle due tribù $\mathcal{F}_1$ e $\mathcal{F}_2$, quindi

$$\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma\left(\{E_1 \times E_2 \,:\, E_1 \in \mathcal{F}_1, E_2 \in \mathcal{F}_2\}\right),$$

cioè $\mathcal{F}$ è la tribù generata dai rettangoli in cui la prima coordinata è data dal primo esperimento e la seconda coordinata dal secondo. Non ci fermiamo alla famiglia dei rettangoli, ma ne prendiamo la tribù generata perché vogliamo essere sicuri di avere una famiglia di insiemi che sia una tribù. Usando l'analogia geometrica, vogliamo che nella tribù ci siano anche altre figure (triangoli, cerchi...), che costruiamo come unione numerabile di rettangoli (o complementari).

Come ultimo passo, dobbiamo parlare della probabilità P. Ancora una volta vogliamo mettere in evidenza che si tratta di una combinazione di esperimenti, quindi vorremmo che la proiezione su ogni coordinata fosse la probabilità del corrispondente esperimento singolo, cioè che la probabilità di ogni esperimento di Ω_1 fosse inalterata nel prodotto con Ω_2 e viceversa.

Possiamo ottenere una probabilità P con le proprietà richieste se la definiamo, per ogni rettangolo $E_1 \times E_2$ in $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$, come

$$P(E_1 \times E_2) = P_1(E_1) \cdot P_2(E_2). \tag{3.1}$$

Questo giustifica anche la notazione $P = P_1 \otimes P_2$. Si può obiettare che la (3.1) non definisce da sola una probabilità per ogni elemento di $\mathcal{F}$ se, come abbiamo detto, non ogni elemento di $\mathcal{F}$ è un rettangolo. Tuttavia, potendo scrivere ogni elemento di $\mathcal{F}$ a partire da rettangoli, mediante unione e complementare, e sapendo come si comporta la probabilità rispetto all'unione (disgiunta) e al complementare, possiamo limitarci a definirla sui rettangoli e l'estensione sarà unica[16].

Esempio 3.3.3. Tornando all'Esempio 3.3.1, osserviamo che nella tribù $\mathcal{F}$ non ci sono solo i prodotti di elementi delle due tribù $\mathcal{F}_1$ e $\mathcal{F}_2$: infatti il complementare di $\{1\}\times\{C\}$ non può essere scritto come prodotto (in particolare non è $\{2,3,4\} \times \{T\}$, che non contiene la coppia $(1,T)$, che appartiene al complementare di $\{1\} \times \{C\}$). Dobbiamo prendere la tribù generata, che contiene anche tutti i complementari e le unioni di rettangoli (cioè di prodotti di elementi di $\mathcal{F}_1$ e $\mathcal{F}_2$).
La probabilità dell'evento $\{(1,T)\}$, supponendo il dado equilibrato e la moneta non truccata, sarà $P(\{(1,T)\}) = P_1(\{1\}) \cdot P_2(\{T\}) = \frac{1}{8}$, mentre quella dell'evento $(\{1,2\} \times \{C\})^c$ sarà

$$P((\{1,2\} \times \{C\})^c) = 1 - P(\{1,2\} \times \{C\}) = 1 - P_1(\{1,2\}) \cdot P_2(\{C\}) = 1 - \frac{1}{4} = \frac{3}{4}.$$

In modo analogo possiamo fare per un numero finito di esperimenti distinti quello che abbiamo mostrato per due esperimenti. Possiamo anche passare a una quantità numerabile, ma vedremo i dettagli solamente in un caso speciale: quello degli esperimenti ripetuti.
Parliamo di esperimenti ripetuti quando tutti gli esperimenti sono copie identiche del medesimo esperimento, cioè possono essere tutti descritti con lo stesso spazio di probabilità $(\Omega_S, \mathcal{F}_S, P_S)$. Ne abbiamo visti due nell'Esempio 3.3.2: il lancio di n monete uguali o quello di infinite (numerabili) monete uguali.
Nel caso di un numero finito di ripetizioni, abbiamo una versione semplificata di quanto visto per il prodotto di esperimenti qualunque: abbiamo infatti (considerando ad esempio due sole ripetizioni) che $\Omega = \Omega_1 \times \Omega_2 = \Omega_S^2$, perché i due spazi dei singoli elementi coincidono; abbiamo inoltre che la tribù

$$\mathcal{F} = \mathcal{F}_S \otimes \mathcal{F}_S = \mathcal{F}_S^2 = \sigma(\{E_1 \times E_2 \,:\, E_1 \in \mathcal{F}_S, E_2 \in \mathcal{F}_S\})$$

e che la probabilità $P = P_S^2$.

Esempio 3.3.4. Francesco lancia 6 volte una moneta truccata (o una volta sei monete truccate identiche tra loro), che dà testa con probabilità p e croce con probabilità $1-p$. Con che probabilità i primi due lanci sono entrambi testa? Con che probabilità i primi tre lanci non sono tutti uguali tra loro?

---

[16] Ancora una volta stiamo facendo le cose più facili di quanto non siano in realtà: anche qui viene in aiuto il Teorema di Carathéodory, che garantisce che tale estensione è unica.

Soluzione. Come spazio Ω abbiamo $\{T, C\}^6$ o $\{0,1\}^6$. La tribù è quella generata dai rettangoli, mentre la probabilità su ciascuna componente vale 0 sull'insieme vuoto, p su $\{T\}$, $1-p$ su $\{C\}$ e 1 su $\Omega_S = \{T, C\}$. Il primo evento cui siamo interessati, "i primi due lanci sono entrambi testa", è $\{T\} \times \{T\} \times \Omega_S^4$, la cui probabilità è

$$P(\{T\} \times \{T\} \times \Omega_S^4) = P(\{T\})^2 P(\Omega_S)^4 = p^2 1^4 = p^2.$$

Il secondo evento è un po' più complicato: lo possiamo scrivere come unione di rettangoli, oppure in modo più semplice come complementare di unione di rettangoli,

$$E = (\{T\} \times \{T\} \times \{T\} \times \Omega_S^3 \cup \{C\} \times \{C\} \times \{C\} \times \Omega_S^3)^c.$$

Per quanto riguarda la probabilità abbiamo allora

$$\begin{aligned} P(E) &= 1 - P(\{T\} \times \{T\} \times \{T\} \times \Omega_S^3 \cup \{C\} \times \{C\} \times \{C\} \times \Omega_S^3) \\ &= 1 - P(\{T\} \times \{T\} \times \{T\} \times \Omega_S^3) - P(\{C\} \times \{C\} \times \{C\} \times \Omega_S^3) \\ &= 1 - p^3 - (1-p)^3 \\ &= 3p - 3p^2, \end{aligned}$$

dove nel secondo passaggio abbiamo usato che i due eventi $\{T\} \times \{T\} \times \{T\} \times \Omega_S^3$ e $\{C\} \times \{C\} \times \{C\} \times \Omega_S^3$ sono disgiunti, dal momento che le sestuple nei due insiemi hanno sicuramente le prime tre coordinate distinte e sono quindi diverse. □

Passiamo al caso di infinite ripetizioni di uno stesso esperimento $(\Omega_S, \mathcal{F}_S, P_S)$: siamo alla ricerca di un unico spazio $(\Omega, \mathcal{F}, P)$ che le descriva tutte assieme. Cominciamo come sempre dallo spazio Ω: esso sarà costituito da successioni di elementi di Ω_S, quindi $\Omega = \Omega_S^{\mathbb{N}}$. Fin qui nulla di difficile.

Ci dedichiamo ora alla tribù $\mathcal{F}$. Qui, almeno in apparenza, quando le ripetizioni sono infinite le cose si complicano: ci sono troppe componenti da controllare. Proviamo dunque a sfruttare le idee viste prima e a concentrarci solo sulla ricerca dei generatori della tribù. Non solo, cerchiamo anche di imparare da quanto visto nella Sezione 3.2 per $\mathbb{R}$.

Una cosa che possiamo fare è fissare un numero naturale n e mettere in un unico insieme tutti gli elementi $\omega \in \Omega$ che hanno in comune le prime n coordinate. Possiamo farlo per ogni numero naturale n, considerando per ciascun n tutte le possibili n-uple di elementi di Ω_S. Questi insiemi, al variare di n, prendono il nome di *n-cilindri*, perché come i cilindri geometrici sono caratterizzati dall'avere una sezione fissata (le prime n componenti).

Prendiamo allora la collezione $\mathcal{C}$ di tutti gli n-cilindri al variare di n: la chiamiamo *famiglia degli insiemi cilindrici*. Analogamente a quanto abbiamo visto per i rettangoli, la famiglia dei cilindri in generale non è una tribù. Possiamo però usarla per generarne una: $\mathcal{F} = \sigma(\mathcal{C})$, che è una tribù su $\Omega^{\mathbb{N}}$.
Avendo costruito spazio e tribù, non resta che l'ultimo passo, la probabilità. Per definirla usiamo la forma dei cilindri che generano la tribù $\mathcal{F}$ e il fatto che stiamo parlando di esperimenti ripetuti: su ogni cilindro definiamo la probabilità come il prodotto della probabilità P_S su ciascuna delle n componenti del cilindro e di fattori 1 per tutte le altre (in sostanza le stiamo ignorando). In questo modo abbiamo una probabilità che generalizza al caso infinito quanto già visto per il caso del prodotto finito: per una successione di eventi $E_i \in \mathcal{F}_S$ abbiamo che la probabilità dell'evento $\bigotimes_{i=1}^{+\infty} E_i \in \mathcal{F}$ è

$$P\left(\bigotimes_{i=1}^{+\infty} E_i\right) = \prod_{i=1}^{+\infty} P_S(E_i).$$

In realtà, stiamo tacendo molti dettagli: non abbiamo la pretesa di essere precisi e nemmeno lo spazio o i prerequisiti per poterlo fare, ma vogliamo solo farci un'idea. Per vedere a fondo tutti i dettagli, ancora una volta, è necessario prendere in mano un libro di testo avanzato o seguire un corso universitario di probabilità o di teoria della misura.
Un'ultima osservazione, prima di passare a qualche esempio: quello che abbiamo fatto per la ripetizione infinita di un esperimento può essere adattato al caso del prodotto infinito di esperimenti non necessariamente uguali tra loro. Anche in tal caso possiamo definire dei cilindri, in cui però le componenti devono essere "pescate" dagli spazi corrispondenti alla coordinata in questione. Questo appesantisce la notazione, ma non cambia la sostanza.

Esempio 3.3.5. Federico ha infinite monete identiche tra loro, ciascuna delle quali dà testa con probabilità p e croce con probabilità $1 - p$. Come sempre possiamo pensare che in realtà ne abbia una sola e la lanci infinite volte. Con che probabilità Federico ottiene la prima testa al k-esimo lancio?

Soluzione. Osserviamo che in questo esempio non possiamo fissare a priori un numero massimo di lanci (o di monete), perché qualunque sia questo numero, potremmo avere croci in tutti questi lanci (improbabile, al crescere del numero dei lanci, ma mai con probabilità identicamente zero). Ha allora senso considerare una ripetizione infinita dell'esperimento "lancio di una moneta"[17]. Qual è l'evento

[17]Questo esperimento costituito da una ripetizione infinita del lancio di una moneta si chiama

del quale vogliamo calcolare la probabilità? È un k-cilindro le cui prime $k-1$ componenti sono C e la cui k-esima componente è T. Delle successive non ci interessa. Sappiamo che i cilindri stanno nella tribù, dal momento che ne sono i generatori. La probabilità di questo cilindro è

$$P_S(\{\mathrm{C}\})^{k-1} P_S(\{\mathrm{T}\}) \prod_{i=k+1}^{+\infty} 1 = (1-p)^{k-1} p. \qquad \square$$

3.4 Farsi le ossa

Nelle sezioni precedenti abbiamo visto alcuni modi per costruire spazi di probabilità. Non è però garantito che siano i migliori per i particolari problemi che incontreremo, né che in ogni problema avremo tutte le informazioni necessarie per costruirli nei modi visti, pur avendo magari tutto quello che ci serve per arrivare a una soluzione.

Esempio 3.4.1. Supponiamo di avere un dado a 6 facce, di cui sappiamo che $P(1) = \frac{1}{6}$. Se volessimo procedere come visto nella Sezione 3.1 e ci concentrassimo su $\Omega = \{1, \ldots, 6\}$, dovremmo assegnare una probabilità a tutte le facce del dado. Però non possiamo farlo, perché non abbiamo alcuna informazione sulle altre facce. Potremmo assumere che il dado sia bilanciato e che quindi tutte le facce escano con la medesima probabilità, ma il risultato che otterremmo sarebbe vero solo se questa ipotesi fosse soddisfatta, cosa che non abbiamo la possibilità di controllare.

Alle volte la formulazione del problema ci suggerisce una costruzione diversa da quella standard. Come possiamo accorgercene? È un'attività creativa e non meccanica: dobbiamo fare apprendistato, il solo modo di allenare l'occhio e la mano è fare tanti esercizi. Dobbiamo cercare soluzioni diverse alle quali ispirarci in futuro per affrontare altri problemi, in particolare quelli in cui i metodi standard non funzioneranno. Per questo stesso motivo, uno dei metodi migliori per allenarsi a risolvere problemi è risolvere altri problemi e confrontare le proprie soluzioni con quelle altrui. Pólya[18] nel suo libro *Come risolvere i problemi di matematica* [19] indica tra le diverse euristiche per trovare una soluzione a un problema quella di cercare un altro problema, analogo o simile, del quale ci sia nota una soluzione, per poi cercare di adattare quest'ultima al problema corrente. Parla però di euristiche, non di teoremi, perché questa somiglianza non è definita in modo formale, poiché esula dagli scopi del testo: sta a noi individuarla e sfruttarla.

---

anche *processo di Bernoulli* (Jakob Bernoulli 1654 – 1705). Il modello che descrive il primo istante di successo in un processo di Bernoulli si chiama *geometrico*.

[18] György Pólya (1887 – 1985).

Nel caso dei problemi di probabilità, dobbiamo imparare a estrarre gli oggetti giusti dal testo che abbiamo: non solo la probabilità, ma prima di essa lo spazio degli esiti e la tribù. In particolare è un ottimo esercizio, soprattutto all'inizio, essere molto precisi (quasi noiosi) nello scrivere esplicitamente cosa scegliamo come spazio degli esiti e come tribù, perché quest'accortezza ci eviterà di prendere dei granchi, come ad esempio definire una probabilità su coppie ordinate, quando gli oggetti su cui stiamo lavorando magari sono coppie non ordinate.

In questa sezione esamineremo alcuni problemi, abbastanza classici e famosi, e le loro soluzioni, per iniziare a costruire la nostra cassetta degli attrezzi di esempi e idee per risolvere i prossimi problemi che incontreremo, come quelli del Capitolo 5.

Esempio 3.4.2. In una calda serata di inizio maggio, due amici si danno appuntamento a Cesenatico sotto il grattacielo per passare assieme la serata. Sanno che durante la cena in albergo e nel primo dopocena ci potrebbero essere ritardi e imprevisti e che i loro telefoni sono scarichi, quindi si mettono d'accordo nel modo seguente. Ciascuno di loro si impegna ad arrivare al grattacielo in un orario compreso tra le 22 e le 23 e a fermarsi ad aspettare per 5 minuti. Passati questi 5 minuti (o allo scoccare delle 23) se ne andrà. Con che probabilità i due amici si incontreranno?

Soluzione. La prima volta che si affronta un problema di questo tipo, la tentazione più forte è quella di discretizzare in minuti o secondi. In questo caso, però, il tempo va considerato una quantità continua. Concentriamoci allora su uno dei due amici. Con che probabilità arriverà nei primi dieci minuti dell'ora? Il segmento favorevole è lungo $\frac{1}{6}$ del segmento totale (10 minuti su 60), quindi la probabilità che arrivi in quei dieci minuti è proprio $\frac{1}{6}$. Analogamente, la probabilità che il secondo amico arrivi tra le 22.21 e le 22.41 è $\frac{1}{3}$, poiché c'è un intervallo lungo 20 (minuti) favorevole su un intervallo totale lungo 60 (minuti).

In questo ragionamento, però, stiamo considerando i due amici separatamente e stiamo trascurando il fatto che sono disposti ad aspettare. Se sapessimo che il primo amico arriva alle 22.13, allora la probabilità che i due si incontrino sarebbe uguale alla probabilità che il secondo arrivi nei 5 minuti precedenti alle 22.13 o nei 5 minuti successivi, cioè $\frac{1}{6}$.

È arrivato il momento di passare dai segmenti ai quadrati. Mettiamo sull'asse delle ascisse l'orario di arrivo del primo amico e su quello delle ordinate quello del secondo. Le coordinate interne al quadrato rappresentano le combinazioni di arrivi dei due amici. I due si incontrano se le due coordinate non differiscono più di 5. Ma geometricamente questo cosa significa?

Se arrivano insieme, si incontrano. Questi punti sono la diagonale del quadrato. Ma non sono, come detto, la loro unica possibilità di incontrarsi[19]. Possiamo spostarci orizzontalmente o verticalmente di 5 minuti, rispetto alla diagonale, cioè considerare la diagonale ingrassata, colorata in grigio chiaro nella Figura 3.1. Questa superficie rappresenta tutte le coppie di orario d'arrivo per cui i due amici si incontrano. Per calcolare la probabilità richiesta dobbiamo considerare il rapporto tra quest'area e quella totale, che rappresenta tutte le possibili coppie di tempi d'arrivo dei due amici. Questo rapporto è $\frac{11}{36}$.

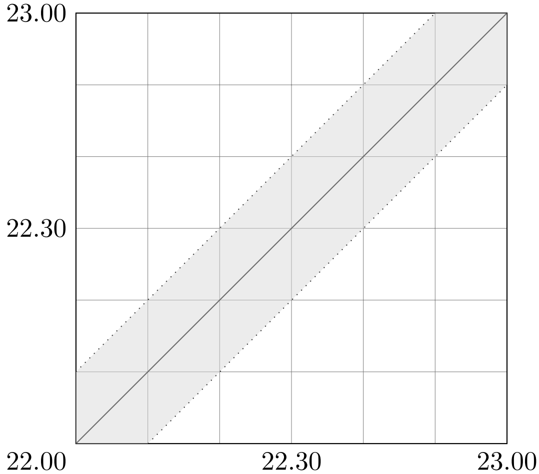

Figura 3.1: Incontro sotto il Grattacielo

Possiamo esaminare questo stesso esercizio sotto la lente più formale introdotta nella prima parte di questo capitolo. Quello che cambia è solamente il linguaggio, non l'idea sottostante, né tanto meno il risultato. Giusto per dare uno spunto: abbiamo considerato per ciascuno dei due amici uno spazio di probabilità in cui $\Omega = [0, 60]$ e $\mathcal{F} = \mathcal{B}([0, 60])$ (cioè la tribù dei Boreliani generata dagli intervalli semiaperti in $[0, 60]$, il che equivale a prendere la tribù dei Boreliani su $\mathbb{R}$ intersecata con l'intervallo che ci interessa). Per quanto riguarda P stiamo prendendo la lunghezza dei segmenti riscalata (in modo che $P([0, 60]) = 1$), cioè $P([a, b]) = \frac{b-a}{60}$. Possiamo vedere la stessa probabilità come generata dalla funzione

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{60} & 0 \leqslant x \leqslant 60 \\ 1 & x > 60 \end{cases} .$$

[19]In realtà la probabilità che arrivino insieme è 0, come possiamo vedere calcolando il rapporto tra l'area della diagonale (nulla) e quella del quadrato.

Quando poi passiamo a considerare insieme i due amici, siamo in uno spazio prodotto (in realtà il quadrato dello stesso spazio), con la misura prodotto, che è l'area delle porzioni del quadrato (il nostro Ω^2), rinormalizzata dividendo per $60 \cdot 60 = 3600$, in modo da avere una probabilità. $\square$

Osservazione 3.4.1. Un dettaglio interessante, anche se non necessario per il problema appena esaminato, è il seguente: possiamo calcolare la probabilità anche di eventi che nello spazio bidimensionale non sono rettangoli, ma che si ottengono come unione (eventualmente numerabile) di rettangoli, come ad esempio triangoli, poligoni, cerchi o altre figure convesse.

Esempio 3.4.3. Prendendo a caso due punti su un segmento, lo si divide in tre parti. Con che probabilità questi tre segmenti possono formare un triangolo?

Soluzione. Come prima cosa, fissiamo uguale a 1 la lunghezza del segmento iniziale[20]. Consideriamo questo segmento unitario in un sistema cartesiano dove l'origine coincide con il primo estremo e chiamiamo P e Q i due punti presi a caso su di esso. Ciascuno di essi è univocamente identificato dalla sua distanza dall'origine, che indichiamo rispettivamente con p e q.
Osserviamo che l'evento in cui P e Q coincidono (e dunque $p = q$) è un punto e ha quindi probabilità nulla[21] di accadere, per le proprietà della probabilità uniforme sul segmento $[0, 1]$. Possiamo quindi trascurarlo e abbiamo così due casi possibili: $p < q$ e $p > q$. Data la simmetria del problema, possiamo studiare solamente uno dei due casi (a patto di ricordarcene nel momento in cui consideriamo Ω o di moltiplicare per 2 il risultato, se Ω non tiene conto della simmetria).
Se $p < q$, allora i tre segmenti hanno lunghezza p, $q - p$ e $1 - q$ (quello che resta a destra del secondo punto). Affinché possano essere le lunghezze dei lati di un triangolo, devono soddisfare le disuguaglianze triangolari, cioè ogni lunghezza deve essere minore della somma delle altre due:

$$\begin{cases} p < q - p + 1 - q = 1 - p \\ q - p < p + 1 - q \\ 1 - q < p + q - p = q. \end{cases}$$

[20] In sostanza stiamo assumendo come unità di misura "la lunghezza di questo segmento".

[21] Abbiamo già visto un fenomeno simile nell'Esempio 3.4.2: in entrambi i casi abbiamo che, con una distribuzione uniforme di probabilità, oggetti geometrici di dimensione più bassa (come i punti in un segmento o i segmenti in una superficie) hanno misura nulla.

In ciascuna di queste disuguaglianze sommiamo a entrambi i membri quanto compare a primo membro, ottenendo

$$\begin{cases} 2p < 1 \\ 2(q-p) < 1 \\ 2(1-q) < 1, \end{cases}$$

da cui risulta che le tre lunghezze p, $q-p$ e $1-q$ devono tutte essere minori di $\frac{1}{2}$. Anche in questo problema, come nel precedente, abbiamo però un continuo di valori possibili per p e q. Rappresentiamoli anche in questo caso in due dimensioni: siccome abbiamo assunto $p < q$, il nostro Ω sarà il solo triangolo sopra la diagonale, colorato in grigio chiaro. Quali sono in questo triangolo le coppie (p, q) che vanno bene? Cominciamo scartando tutti i punti in cui $p > \frac{1}{2}$ o $q < \frac{1}{2}$. Rimane solamente un vincolo da considerare: $q - p < \frac{1}{2}$, cioè scartiamo i punti che distano più di $\frac{1}{2}$ dalla diagonale. Rimane il triangolo colorato in grigio scuro nella Figura 3.2, che è rettangolo isoscele di cateto $\frac{1}{2}$ e area $\frac{1}{8}$. La probabilità cercata è il rapporto tra quest'area e l'area dell'intero triangolo sopra la diagonale, che vale $\frac{1}{2}$. La probabilità cercata è allora $\frac{1}{4}$.

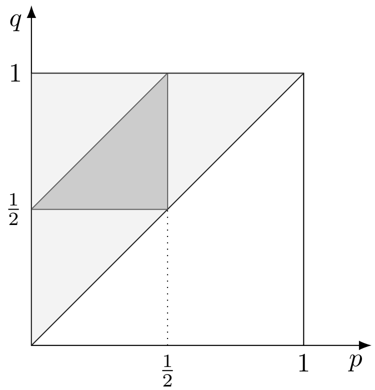

Figura 3.2: Spezzare un segmento per avere triangoli, prima soluzione

Per evitare di calcolare le aree, possiamo osservare che il triangolo più grande è diviso in quattro triangolini equivalenti dalle linee che abbiamo tracciato, di cui solo uno, quello più scuro, è costituito da punti che rappresentano casi favorevoli. Se non siamo convinti che sia sufficiente considerare solo il caso $p < q$, possiamo guardare i casi favorevoli all'interno dell'intero quadrato, considerando anche il caso simmetrico in cui $p > q$.

Vediamo ora una seconda soluzione. Mettiamoci in un sistema di riferimento cartesiano tridimensionale, con i tre assi che rappresentano le lunghezze dei tre segmenti. Le terne possibili (cioè quelle nel primo ottante in cui la somma delle tre coordinate è uguale a 1) giacciono tutte su un piano e, in particolare, sulla superficie di un triangolo (all'interno del cubo unitario) di vertici $(1, 0, 0), (0, 1, 0), (0, 0, 1)$, colorato in grigio nella Figura 3.3 a sinistra.

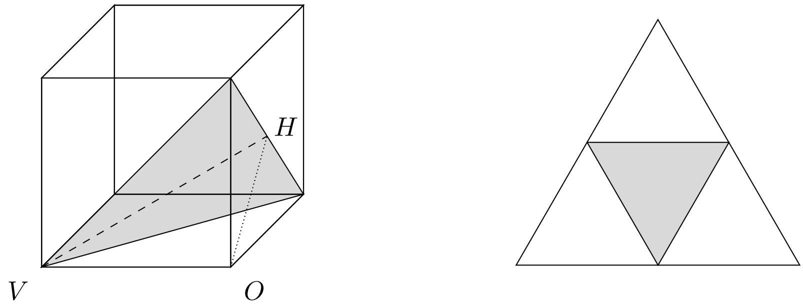

Figura 3.3: Spezzare un segmento per avere triangoli, seconda soluzione

Possiamo proiettare i tre assi cartesiani sul triangolo, in modo che siano le altezze rispetto ai tre lati (ad esempio, nella Figura 3.3 a sinistra proiettiamo l'asse OV sull'altezza VH): in altre parole, possiamo scegliere di rappresentare le lunghezze dei tre segmenti come le tre altezze del triangolo, che consideriamo dunque tutte di lunghezza 1. Il Teorema di Viviani[22] ci dice che i punti del triangolo sono precisamente i punti in cui la somma delle distanze dai lati è uguale alla lunghezza di una delle altezze e quindi a 1. I punti interni al triangolo sono quindi tutti i modi possibili in cui un segmento unitario può essere spezzato in tre parti, il nostro Ω. Ora dobbiamo rappresentare le altre condizioni, già viste nella prima soluzione: nessuno dei tre segmenti può essere più lungo di $\frac{1}{2}$. L'intersezione di queste tre condizioni è il triangolino centrale in Figura 3.3 a destra, la cui superficie è $\frac{1}{4}$ della superficie totale. □

Qualche commento prima di proseguire con un altro esempio. Nella seconda soluzione dell'Esempio 3.4.3 abbiamo trasformato il nostro problema di probabilità in un problema di geometria. E in un certo senso avevamo fatto la stessa cosa anche nella prima soluzione e in quella dell'Esempio 3.4.2. Ciò accade perché, se proviamo a suddividere la matematica in compartimenti stagni risolvendo problemi di probabilità o di geometria o di teoria dei numeri, la matematica si mostrerà comunque come un tutt'uno, in cui lo stesso problema può (e alle volte deve) essere affrontato con tecniche diverse che arrivano da ambiti apparentemente distinti.

[22] Vincenzo Viviani (1622 – 1703). Per approfondimenti su questo teorema, consultare *Geometria piana per le gare di matematica* [5], pagina 13.

Inoltre sembra che ci siano spesso più strade che portano al medesimo risultato. Anche questa è una verità più generale della matematica e non solo dei problemi di probabilità. Questa osservazione, però, ci suggerisce un buon esercizio: cercare nuove soluzioni a un problema già visto, che magari potranno tornare utili per altri problemi che incontreremo in futuro.

È importante che le diverse soluzioni portino davvero al medesimo risultato: guadagneremo così un po' di confidenza nella correttezza di quanto abbiamo ottenuto. Se i risultati invece saranno tutti diversi, avremo la certezza che almeno uno di quelli trovati sia sbagliato. Infatti in probabilità può capitare, se non si fa attenzione a scrivere tutto nei dettagli, di trovarsi in situazioni paradossali, come quella presentata nel prossimo esempio. Data la peculiarità di questo esercizio, è consigliabile, ancor più del solito, cercarne una o più soluzioni, prima di leggere quelle proposte.

Esempio 3.4.4. Consideriamo un triangolo equilatero e la circonferenza a esso circoscritta. Prendiamo casualmente una corda della circonferenza. Qual è la probabilità che la corda scelta sia più lunga del lato del triangolo?

Prima soluzione. Possiamo associare a ogni corda il punto P in cui interseca il raggio a essa ortogonale (il suo asse). In questo modo ogni corda è univocamente identificata da questo punto.

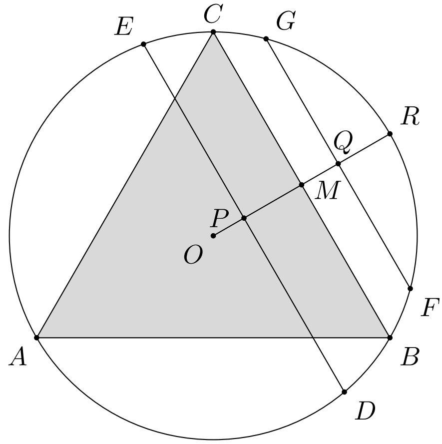

Figura 3.4: Corde e triangoli, prima soluzione

Contemporaneamente possiamo osservare che il problema è invariante per rotazioni, cioè il comportamento sarà lo stesso su ogni raggio. Prendiamo il raggio OR perpendicolare a uno dei lati del triangolo, come indicato nella Figura 3.4 (oppure, equivalentemente, ruotiamo il triangolo in modo che un suo lato sia ortogonale al raggio fissato).

Poiché il triangolo è equilatero, il raggio e il lato si bisecano a vicenda: in figura M è punto medio sia di OR sia di BC. Tutte le corde perpendicolari al raggio che lo intersecano in un punto del segmento OM sono più lunghe del lato del triangolo, come ad esempio DE, che interseca OM in P. Viceversa, tutte le corde perpendicolari al raggio OR che lo intersecano in un punto di MR sono più corte del lato, come FG che passa per il punto Q. Dal momento che i segmenti OM ed MR hanno la stessa lunghezza e che a ogni loro punto corrisponde una corda, la probabilità cercata è $\frac{1}{2}$. Grazie all'invarianza per rotazioni del problema questa probabilità è uguale per tutti i raggi ed è quindi la risposta cercata. □

Seconda soluzione. In alternativa, osserviamo che ogni corda è identificata dai suoi due estremi. Grazie all'invarianza per rotazione possiamo considerare fissato uno di questi estremi. In particolare, possiamo far coincidere l'estremo fissato con uno dei vertici del triangolo, ad esempio A (Figura 3.5), oppure, in modo equivalente, fissare un punto sulla circonferenza e ruotare il triangolo in modo che uno dei suoi vertici vada a sovrapporsi al punto fissato.

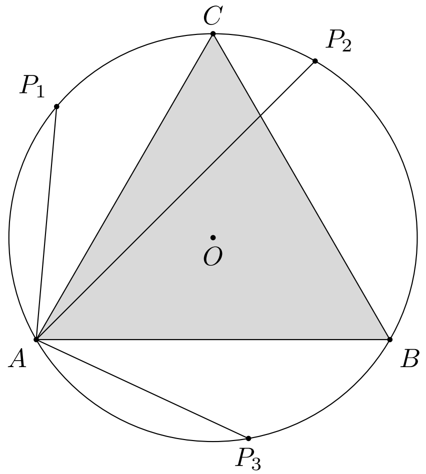

Figura 3.5: Corde e triangoli, seconda soluzione

Ora ogni corda è determinata scegliendo come suo secondo estremo un punto casuale sulla circonferenza. Osserviamo che se questo punto cade tra il vertice fissato e uno degli altri due, cioè negli archi AB o AC, la corda sarà più corta del lato, come nel caso delle corde AP_1 e AP_3 in figura. Se invece il secondo punto cade sull'arco di circonferenza BC, delimitato dai vertici del triangolo distinti da quello fissato, allora la corda, AP_2 in figura, è più lunga del lato del triangolo. Questo arco "favorevole" è lungo $\frac{1}{3}$ della circonferenza, quindi la probabilità cercata è uguale a $\frac{1}{3}$. □

Terza soluzione. Abbiamo un terzo modo di individuare una corda in modo univoco: prendere il suo punto medio. Per ricostruire la corda a partire dal punto, procediamo in questo modo: tracciamo il raggio passante per il punto, poi la retta perpendicolare al raggio per il medesimo punto. Essa intersecherà la circonferenza in due punti che sono gli estremi della corda.
La corda, come visto nella prima soluzione, ha lunghezza maggiore del lato del triangolo se il suo punto medio non dista dal centro della circonferenza più del lato, ossia se esso giace all'interno della circonferenza inscritta al triangolo, concentrica alla circonferenza circoscritta e di raggio lungo la metà. Ad esempio, in Figura 3.6 il punto M giace all'interno della circonferenza inscritta e la corda che ha M come punto medio è più lunga del lato del triangolo. Viceversa, il punto N è esterno alla circonferenza inscritta al triangolo e la corda che ha N come punto medio è più corta del lato del triangolo.

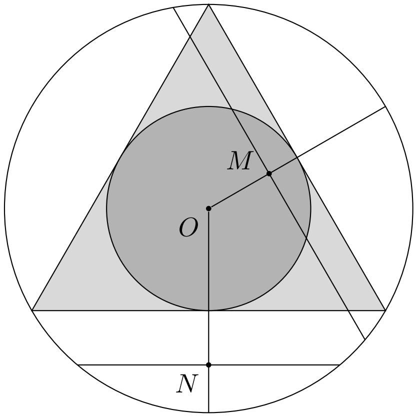

Figura 3.6: Corde e triangoli, terza soluzione

Possiamo allora calcolare la probabilità che la corda abbia lunghezza maggiore del lato come rapporto tra la superficie del cerchio più piccolo (i punti medi favorevoli), in grigio scuro in Figura 3.6, e quella del cerchio più grande (tutti i possibili punti medi). Tale rapporto è il quadrato del rapporto tra i raggi, ossia $\frac{1}{4}$. □

Cosa è successo? Dove stiamo sbagliando? Quello che abbiamo appena visto prende il nome di paradosso di Bertrand[23]. Il punto chiave è che la frase “prendiamo casualmente una corda” non è precisa e può essere interpretata in tutti e tre i modi visti, che però non sono tra loro equivalenti. Le tre soluzioni proposte sono tre soluzioni valide di tre problemi distinti, in cui scegliamo uniformemente un raggio e un punto su esso, due punti sulla circonferenza e un punto del cerchio,

[23] Joseph Bertrand (1822 – 1900).

rispettivamente. Nella simulazione in Figura 3.7 possiamo vedere la distribuzione, all'interno del cerchio unitario, dei punti medi delle corde ottenuti nei tre modi. Nel primo caso essi tendono ad accumularsi attorno al centro, nel secondo attorno al centro e vicino al bordo, mentre nel terzo sono distribuiti uniformemente in tutto il cerchio.

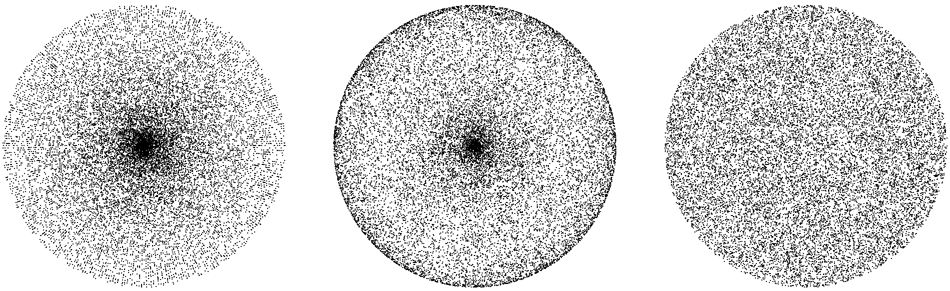

Figura 3.7: Simulazioni della distribuzione dei punti medi delle corde nei tre casi

Ora lasciamoci alle spalle questo esempio "imbroglione": di seguito troviamo qualche altro esercizio. Questo non è che un assaggio: più avanti ci aspetta il Capitolo 5 dedicato esclusivamente a esercizi di vario tipo.

3.5 Esercizi

Problema 16

In un $(2n+1)$-gono regolare si scelgono a caso 3 dei $2n+1$ vertici. Con che probabilità il triangolo individuato da questi punti contiene il circocentro del poligono?

Problema 17

In che modo è possibile generare un intervallo chiuso $[a,b]$ (in $[0,1]$ o in $\mathbb{R}$) a partire da intervalli semiaperti della forma $(c,d]$, usando solamente unioni numerabili e passaggi al complementare?

Problema 18

Torniamo dal nostro amico Francesco, visto nell'Esempio 3.3.4. Questa volta Francesco lancia 8 monete identiche tra loro, ciascuna delle quali dà testa con probabilità p e croce con probabilità $1-p$. Qual è la probabilità che Francesco ottenga esattamente 5 croci?

4. Sapendo che. . .

Gli assiomi di Kolmogorov e la teoria della probabilità che se ne ricava non ci dicono nulla su come calcolare e assegnare le probabilità stesse, problema che abbiamo affrontato nel Capitolo 3, né su cosa significhi dire che un evento ha una determinata probabilità. Un possibile punto di vista, come accennato, è interpretare la probabilità di un evento come misura della nostra incertezza o ignoranza riguardo a esso: non siamo sicuri che avverrà, ma in qualche modo ci sentiamo di dire, ad esempio, che accadrà due volte su tre, se potessimo ripetere questo "esperimento" più volte, oppure che siamo disposti a scommettere che accada, purché in tal caso riceviamo almeno una volta e mezzo quello che abbiamo puntato.

In questo modo la probabilità diventa in un certo senso una misura di informazione, ed è quindi naturale pensare che, se accumuliamo nuovi dati riguardo a un evento, possiamo e dobbiamo aggiornare la probabilità che gli assegniamo. Questo punto di vista è molto vicino al metodo scientifico: non possiamo mai avere certezza di nulla, ma una serie di risultati sperimentali favorevoli a una nostra teoria farà aumentare la nostra confidenza nel fatto che tale teoria possa dare una buona spiegazione.

4.1 Condizionamenti

Supponiamo, in un certo spazio di probabilità, di venire a sapere che un certo evento F si è verificato. Se a questo punto vogliamo valutare di nuovo la probabilità di un altro evento E, vorremo tener conto delle informazioni in più che abbiamo, ossia che è successo F. Parliamo in questo caso di probabilità condizionata.

Definizione 4.1.1. Dato uno spazio di probabilità $(\Omega, \mathcal{F}, P)$ e due eventi E ed F in $\mathcal{F}$, con $P(F) \neq 0$, definiamo la *probabilità di E condizionata a F* come

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)}.$$

Se guardiamo il numeratore della definizione, stiamo considerando solo gli esiti in E che possono verificarsi in un mondo nel quale sappiamo che F non è più solo una possibilità, ma un dato di fatto. Per quanto riguarda il denominatore, dividiamo per $P(F)$ perché il nostro universo è ora il solo F e, dal momento che vogliamo avere di nuovo una probabilità, dobbiamo rinormalizzare opportunamente. Nella Figura 4.1 possiamo vedere un'illustrazione in termini di insiemi della definizione.

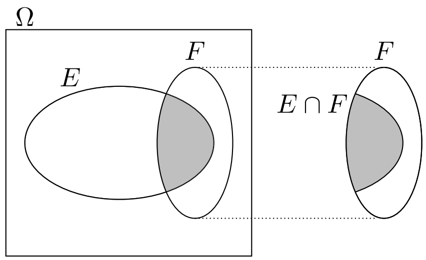

Figura 4.1: Nel condizionamento F è il "nuovo" universo

Esempio 4.1.1. Rosalia lancia un normale dado a 6 facce. Come spesso accade, il dado cade a terra e Rosalia non vede cos'è uscito. Stefano, che vede il risultato del dado, le dice che è uscito un numero dispari. Qual è la probabilità che Rosalia abbia fatto 3? E qual è la probabilità che non abbia fatto 1?

Soluzione. Stiamo considerando il lancio di un dado a 6 facce. Abbiamo quindi come possibile scelta dello spazio degli esiti $\Omega = \{1, 2, 3, 4, 5, 6\}$, per l'algebra $\mathcal{F} = \mathcal{P}(\Omega)$ e per funzione di probabilità quella che dà peso $\frac{1}{6}$ a ogni singoletto. L'informazione fornita da Stefano è che si è verificato l'evento $F = \{1, 3, 5\}$. A questo punto possiamo usare la definizione per calcolare le quantità richieste.
La probabilità che sia uscito 3 è

$$P(\{3\} \mid F) = \frac{P(\{3\} \cap \{1,3,5\})}{P(\{1,3,5\})} = \frac{P(\{3\})}{P(\{1,3,5\})} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}.$$

La probabilità che non sia uscito 1 è

$$P(\{1\}^c \mid F) = \frac{P(\{2,3,4,5,6\} \cap \{1,3,5\})}{P(\{1,3,5\})} = \frac{P(\{3,5\})}{P(\{1,3,5\})} = \frac{\frac{2}{6}}{\frac{1}{2}} = \frac{2}{3}. \qquad \square$$

Esempio 4.1.2. In una sperduta isola dell'arcipelago matematico i traghetti passano in modo abbastanza casuale. Gli abitanti sono interessati a studiare le probabilità del tempo passato ad aspettare un traghetto, misurato in ore. Qual è la probabilità di aspettare almeno 2 ore? E quella di aspettare almeno 4 ore, se dopo due ore il traghetto non è ancora passato?

Soluzione. Prima di poter risolvere questo problema, abbiamo bisogno di qualche informazione in più. Lo spazio Ω è l'insieme dei reali positivi o, in alternativa, quello di tutti i reali, che abbiamo già trattato nella Sezione 3.2, facendo poi attenzione che la probabilità nei negativi sia nulla. Nella Sezione 3.2 abbiamo visto che la tribù che scegliamo in questo caso è quella dei Boreliani, $\mathcal{B}$. Non resta che scegliere la probabilità P: in questo caso scegliamo quella indotta dalla funzione F descritta nell'Esempio 3.2.3, per cui $P((-\infty, a]) = 1 - e^{-a}$ per ogni $a \geqslant 0$ ed è zero per $a < 0$, come richiesto[1].
Qual è l'evento che ci interessa, nel primo caso? Che il tempo sia maggiore di 2, quindi nella semiretta $(2, +\infty)$, che è il complementare di $(-\infty, 2]$. Quindi

$$P((2, +\infty)) = 1 - P((-\infty, 2]) = 1 - (1 - e^{-2}) = e^{-2} \approx 13{,}5\%.$$

E nel secondo? Vogliamo calcolare $P((4, +\infty) \mid (2, +\infty))$, cioè

$$P((4, +\infty) \mid (2, +\infty)) = \frac{P((4, +\infty) \cap (2, +\infty))}{P((2, +\infty))} = \frac{P((4, +\infty))}{P((2, +\infty))} = \frac{e^{-4}}{e^{-2}} = e^{-2}.$$

In altre parole, se un abitante ha aspettato invano per 2 ore, la probabilità che ne debba aspettare ancora 2 (e quindi 4 in totale) è uguale a quella di aspettarne 2: con la misura di probabilità che abbiamo scelto, il traghetto non ha memoria del proprio ritardo! □

Osservazione 4.1.2. Se fissiamo un evento F di probabilità non nulla, allora la funzione $P_F : \mathcal{F} \to \mathbb{R}$ definita per ogni $E \in \mathcal{F}$ da $P_F(E) = P(E \mid F)$ è una funzione di probabilità, poiché soddisfa tutti gli assiomi visti. Abbiamo, infatti, che $P_F(E) \geqslant 0$ per qualunque evento E in $\mathcal{F}$, dal momento che stiamo facendo il rapporto tra una quantità non negativa e una positiva. Allo stesso tempo, $P_F(\Omega) = \frac{P(\Omega \cap F)}{P(F)} = 1$. Non resta che verificare che anche il terzo assioma sia soddisfatto: prendiamo una famiglia numerabile $(E_i)_{i=1}^{+\infty}$ di eventi a due a due

[1] È importante sottolineare che questa è una scelta per P, non l'unica possibile. In generale, la scelta della probabilità è una scelta di modello, fatta tenendo conto delle caratteristiche della situazione che si vuole descrivere, come abbiamo già discusso nel Capitolo 3.

disgiunti e scriviamone la probabilità condizionata a F dell'unione,

$$
\begin{aligned}
P_F\left(\bigcup_{i=1}^{+\infty} E_i\right) &= \frac{P\left[\left(\bigcup_{i=1}^{+\infty} E_i\right)\cap F\right]}{P(F)} = \frac{P\left[\bigcup_{i=1}^{+\infty} (E_i \cap F)\right]}{P(F)} \\
&= \frac{\sum_{i=1}^{+\infty} P(E_i \cap F)}{P(F)} = \sum_{i=1}^{+\infty} P_F(E_i),
\end{aligned}
$$

in cui abbiamo usato la distributività dell'intersezione rispetto all'unione.
Allora anche per P_F valgono le proprietà viste per una qualunque misura di probabilità P. Tutto questo non ci sorprende: abbiamo dato la definizione di probabilità condizionata proprio con l'idea di avere alla fine una misura di probabilità.

Esempio 4.1.3. Edoardo ama molto correre in montagna quindi, se le previsioni sono buone, la probabilità che passi la domenica sui monti è del 70%. Se le previsioni sono buone, con che probabilità rimane a casa?

Soluzione. Siano M l'evento "correre in montagna" e S l'evento "buone previsioni":

$$P(M \mid S) + P(M^c \mid S) = P(M \cup M^c \mid S) = P(\Omega \mid S) = 1,$$

quindi $P(M^c \mid S) = 1 - P(M \mid S) = 30\%$. □

Nel definire la probabilità condizionata, il nostro scopo era quantificare l'effetto di un evento su un altro, in termini di probabilità. Tuttavia, il bello delle identità è che possiamo rigirarle un po' per mettere in evidenza altri aspetti. In particolare, dalla definizione di probabilità condizionata possiamo ricavare un modo (anzi, due) per scrivere la probabilità dell'intersezione tra due eventi:

$$P(E \cap F) = P(E \mid F) \cdot P(F) = P(F \mid E) \cdot P(E). \tag{4.1}$$

La doppia identità (4.1) prende anche il nome di *teorema (o regola) del prodotto*.
Osserviamo che ella (4.1) siamo stati un po' imprecisi: non abbiamo specificato che $P(E) \neq 0 \neq P(F)$. Tuttavia, se anche $P(E)$ o $P(F)$ fossero nulli, avremmo che $P(E \cap F) = 0$, perché l'intersezione $E \cap F$ è un evento contenuto in un evento di probabilità nulla (E o F): qualunque valore (finito) assegniamo a $P(E \mid F)$ (o $P(F \mid E)$), lo annulleremo moltiplicandolo per 0.
Potrebbe essere interessante caratterizzare quegli eventi che non interagiscono tra loro, quelli che intuitivamente chiameremmo eventi indipendenti. Come prossimo passo vogliamo quindi dare una definizione matematica di indipendenza tra eventi, per poi vedere come essa si sposi con l'idea intuitiva di eventi indipendenti.

Definizione 4.1.3. In uno spazio di probabilità $(\Omega, \mathcal{F}, P)$, due eventi E ed F in $\mathcal{F}$ si dicono *indipendenti* se vale l'uguaglianza $P(E \cap F) = P(E) \cdot P(F)$.

Questa definizione, a un primo sguardo, ci sorprende un po': com'è che parliamo di indipendenza tra eventi e ci ritroviamo con una "formula" per la probabilità dell'intersezione? In realtà grazie al legame tra probabilità dell'intersezione e probabilità condizionata possiamo dare un altro punto di vista sull'indipendenza appena definita. Infatti, se due eventi E ed F sono indipendenti, abbiamo

$$P(E) \cdot P(F) = P(E \cap F) = P(E \mid F) \cdot P(F) \quad \text{e} \quad P(E) \cdot P(F) = P(F \mid E) \cdot P(E),$$

cioè, supponendo $P(E) \neq 0$ e $P(F) \neq 0$ (ne discuteremo nella Sezione 4.4),

$$P(E \mid F) = P(E) \quad \text{e} \quad P(F \mid E) = P(F).$$

Quindi sapere che è accaduto F non cambia quello che sappiamo della probabilità di E e, viceversa. Inoltre, avendo due catene di uguaglianze possiamo invertire il ragionamento: sapere che E non ci dà informazioni su F ed F non ci dà informazioni su E implica che E ed F sono indipendenti, per la definizione di indipendenza data sopra. La definizione data caratterizza proprio quello che ci aspettavamo e il termine usato è giustificato.
Se siamo invece interessati alla probabilità dell'intersezione di due eventi, sappiamo che essa è uguale al prodotto delle probabilità dei due eventi se questi ultimi sono tra loro indipendenti. Se non abbiamo questa informazione, dobbiamo usare la regola del prodotto (4.1) vista sopra, oppure l'identità incontrata nel Capitolo 2:

$$P(E \cap F) = P(E) + P(F) - P(E \cup F).$$

Esempio 4.1.4. Gaia sa che le sue professoresse di Storia e di Agiografia[2] interrogano in ognuna delle due materie a sorteggio tra coloro che ancora non hanno un voto. Sapendo che nella classe, formata da 25 tra studentesse e studenti, nessuno è ancora stato interrogato in Storia e 6 persone (ma non Gaia) hanno un voto in Agiografia, con che probabilità Gaia verrà interrogata domani?

Soluzione. Gaia ha una probabilità di essere interrogata in Storia uguale a $P(S) = \frac{1}{25}$, come tutte le sue compagne e i suoi compagni di classe, e $P(A) = \frac{1}{19}$ di essere interrogata in Agiografia. Le due estrazioni vengono fatte da liste (o

[2] Lo studio delle vite dei santi.

contenitori) diversi, quindi non si influenzano l'una con l'altra: possiamo allora considerare le due interrogazioni come indipendenti. La probabilità di interrogazione di Gaia è allora

$$P(S \cup A) = P(S) + P(A) - P(S \cap A) = \frac{1}{25} + \frac{1}{19} - \frac{1}{25 \cdot 19} = \frac{43}{475},$$

in cui abbiamo usato l'identità per la probabilità dell'unione vista nella Proposizione 2.4.5 e, per valutare $P(S \cap A)$, l'indipendenza. □

Esempio 4.1.5. Nicolò possiede un'auto sportiva gialla. Un giorno, in un parcheggio, vede che l'auto in sosta accanto alla sua è anch'essa una sportiva gialla. Mentre rientra verso casa, si chiede quanto sia probabile che un'auto sia una sportiva gialla. Da una rapida ricerca online scopre che solamente 1 auto ogni 100 è un'auto sportiva e che solo 1 auto su 200 è gialla. Ne conclude quindi che la probabilità che un'auto sia una sportiva gialla è $\frac{1}{20\,000}$.
In realtà questo ragionamento non è corretto, perché nulla garantisce che i due eventi "auto sportiva" e "auto gialla" siano indipendenti. In effetti, con un po' di attenzione, Nicolò scopre poco dopo che tra le auto gialle, 1 su 3 è un'auto sportiva, quindi la probabilità che un'auto a caso sia gialla e sportiva è

$$P(S \cap G) = P(S \mid G) \cdot P(G) = \frac{1}{3} \cdot \frac{1}{200} = \frac{1}{600} \neq \frac{1}{20\,000} = P(S) \cdot P(G).$$

Qui l'errore non ha gravi conseguenze, ma una svista simile ha contribuito alla condanna, poi annullata, di Malcolm Ricardo Collins[3].

Dalla riscrittura in termini di probabilità condizionata dell'indipendenza, abbiamo che, se due eventi E ed F sono indipendenti, allora $P(F \mid E) = P(F)$, ma anche $P(F^c \mid E) = P(F^c)$. Ma che succede se abbiamo il complementare dall'altro lato del condizionamento?

Esempio 4.1.6. Prendiamo, in uno spazio di probabilità $(\Omega, \mathcal{F}, P)$, due eventi E ed F tali che $0 < P(F) < 1$ e $P(E \mid F) = P(E \mid F^c)$. Possiamo dire che gli eventi E ed F sono indipendenti?

Soluzione. Potremmo sospettare un trabocchetto, quindi andiamo a scriverci con attenzione le quantità:

$$\frac{P(E \cap F)}{P(F)} = P(E \mid F) = P(E \mid F^c) = \frac{P(E \cap F^c)}{P(F^c)} = \frac{P(E \cap F^c)}{1 - P(F)}.$$

[3] È un caso giudiziario realmente accaduto negli anni Sessanta in California. Una discussione più dettagliata di questo e di altri esempi reali di errori matematici in ambito processuale si trova nel libro *Math on Trial* [20].

Ora prendiamo il primo e l'ultimo termine e moltiplichiamoli per $P(F)(1 - P(F))$

$$P(E \cap F)(1 - P(F)) = P(E \cap F^c)P(F)$$

e continuiamo raccogliendo i termini moltiplicati per $P(F)$ a secondo membro,

$$P(E \cap F) = (P(E \cap F) + P(E \cap F^c))P(F).$$

A questo punto possiamo osservare che i due eventi $E \cap F$ ed $E \cap F^c$ sono disgiunti e la loro unione è E, quindi, siccome la probabilità dell'unione disgiunta è la somma delle probabilità, abbiamo $P(E \cap F) = P(E)P(F)$, ossia l'indipendenza.
Torniamo allora a guardare il testo iniziale e proviamo a rileggere quello che c'è scritto. La condizione $P(E \mid F) = P(E \mid F^c)$ ci dice che sapere che F sia accaduto o no non dà alcuna informazione su E; infatti non modifica la sua probabilità. □

4.2 Formula di fattorizzazione

Nell'Esempio 4.1.6 abbiamo incontrato un'idea interessante: abbiamo scritto un evento dividendolo in due pezzi disgiunti, che però esaurissero tutte le possibilità. In realtà non c'è nulla di speciale nel fatto che siano due eventi complementari: le caratteristiche fondamentali sono che gli eventi siano tutti disgiunti, ma che allo stesso tempo coprano tutto lo spazio, cioè ne siano una partizione. Andare a riscrivere la probabilità di un evento in termini delle sue probabilità condizionate a una partizione di eventi è una tecnica molto importante che prende il nome di formula di fattorizzazione (o legge delle probabilità totali). La sua validità è garantita dal seguente teorema.

Teorema 4.2.1. *Dato uno spazio di probabilità $(\Omega, \mathcal{F}, P)$, consideriamo una famiglia al più numerabile di eventi disgiunti $(E_i)_{i \in I}$ che sia anche una partizione di Ω. Supponiamo che ogni evento nella partizione abbia probabilità non nulla. Allora per ogni evento $F \in \mathcal{F}$,*

$$P(F) = \sum_{i \in I} P(F \cap E_i) = \sum_{i \in I} P(F \mid E_i) \cdot P(E_i).$$

Dimostrazione. Osserviamo che la seconda uguaglianza deriva, addendo per addendo, dalla definizione di probabilità condizionata. Per quanto riguarda la prima, basta osservare che

$$P(F) = P(F \cap \Omega) = P\left(F \cap \left(\bigcup_{i \in I} E_i\right)\right) = P\left(\bigcup_{i \in I} (F \cap E_i)\right)$$

e che l'unione è necessariamente disgiunta, dal momento che per ogni i risulta $F \cap E_i \subset E_i$. □

Osserviamo che in realtà la richiesta che gli eventi nella partizione non abbiano misura nulla non è cruciale: se è vero che non sappiamo determinare il valore di $P(F \mid E_i)$ per tali eventi, sappiamo che comunque è una probabilità, quindi un numero compreso tra 0 e 1. Questo numero compare moltiplicato per $P(E_i)$, cioè per 0, e ciò risolve i nostri problemi.
La formula di fattorizzazione va a estendere al mondo della probabilità quello che è il principio della somma nella combinatoria: ci permette di dividere il problema in sotto-problemi (auspicabilmente) più facili, dandoci un modo per combinarli alla fine. Di nuovo la strategia del *divide et impera*. Come accennato in precedenza, questo risultato è di grandissima utilità pratica, perché ci permette di spezzare il calcolo della probabilità in più sotto-casi, scelti opportunamente, spesso semplificando enormemente i conti.

Esempio 4.2.1. A un allenamento per la gara di matematica a squadre partecipano solo tre giocatori della squadra: il capitano, il consegnatore e una studentessa del primo anno. Il capitano è un esperto di problemi di combinatoria e ne risolve sei su sette, gli altri due preferiscono entrambi geometria e risolvono gli esercizi di combinatoria solo una volta su quattro. Oggi lavorano indipendentemente su tre problemi, assegnati a caso, di cui solo uno di combinatoria. Qual è la probabilità che alla fine dell'allenamento la squadra abbia risolto questo problema?

Soluzione. Cosa sappiamo? Chiamiamo C l'evento "il problema di combinatoria viene assegnato al capitano" e R l'evento "il problema di combinatoria viene risolto". Allora il testo ci dice che

$$P(R \mid C) = \frac{6}{7}, \qquad P(R \mid C^c) = \frac{1}{4}, \qquad P(C) = \frac{1}{3}.$$

Grazie alla formula di fattorizzazione possiamo riscrivere la probabilità cercata come

$$P(R) = P(R \mid C) \cdot P(C) + P(R \mid C^c) \cdot P(C^c) = \frac{6}{7} \cdot \frac{1}{3} + \frac{1}{4} \cdot \frac{2}{3} = \frac{19}{42}. \qquad \square$$

Esempio 4.2.2. Da un recente sondaggio svolto nell'Arcipelago delle Tre Isole è emerso che nell'isola di Idilos 2 abitanti su 15 sono matematici, nell'isola di Iremun è matematico 1 abitante su 5, mentre sulla terza isola, Erettel, sono 3 su 25. Qual è la probabilità che un qualunque abitante dell'arcipelago sia un matematico, se il 30% vive su Idilos, il 45% su Iremun e il 25% su Erettel?

Soluzione. Indicando con M l'essere un matematico e con S, N ed E l'essere abitante dell'isola di Idilos, Iremun ed Erettel rispettivamente, abbiamo

$$\begin{aligned} P(M) &= P(M \mid S) \cdot P(S) + P(M \mid N) \cdot P(N) + P(M \mid E) \cdot P(E) \\ &= \frac{2}{15} \cdot \frac{30}{100} + \frac{1}{5} \cdot \frac{45}{100} + \frac{3}{25} \cdot \frac{25}{200} \\ &= 16\%. \end{aligned}$$

□

In pratica quello che stiamo facendo è prendere la media delle probabilità dell'evento che ci interessa (essere matematici) condizionata ai casi disgiunti (vivere in una specifica isola), pesando questa media con le probabilità dei casi stessi.

4.3 Intermezzo - Il valore atteso

Prima di continuare, riguardiamo per un momento quest'ultima osservazione. Abbiamo visto che possiamo interpretare la formula di fattorizzazione come una media delle probabilità condizionate, pesata con le probabilità degli eventi che costituiscono la partizione. Possiamo fare qualcosa di analogo in un problema leggermente diverso.

Consideriamo il caso di un esperimento aleatorio in cui gli esiti siano un insieme finito o numerabile di numeri, ad esempio il lancio di un dado. Può avere senso cercare un valore da cui il risultato, qualunque esso sia, non si discosterà più di tanto. Se vogliamo scommettere con un amico con la regola che vince non necessariamente chi indovina il numero corretto, ma chi ci va più vicino, scegliere un numero che sia abbastanza vicino a qualunque risultato possibile potrebbe essere una buona strategia. Ma qual è questo valore "centrale" o "medio"?

La risposta è meno ovvia di quanto non sembri: una possibilità è infatti fare la media aritmetica degli esiti. Nel caso del lancio di un dado avremmo allora $\frac{1+2+3+4+5+6}{6} = \frac{7}{2}$. Ma se sapessimo che il dado è truccato, come possiamo tenerne conto? E poi, con tutte le medie che ci sono, perché proprio quella aritmetica e non, per esempio, quella geometrica? Non abbiamo una buona giustificazione da dare: in particolare, la scelta che abbiamo fatto non ha particolari riferimenti al contesto in cui siamo.

Un modo per includere le informazioni che abbiamo (cioè la probabilità) è il seguente: scegliamo come valore centrale la media dei possibili valori (cioè degli esiti) pesati con le loro rispettive probabilità[4]. Come strategia sembra promettente:

[4] A voler essere pignoli, le probabilità dei singoletti, visto che non assegniamo probabilità agli esiti, ma agli eventi.

i pesi hanno somma 1 (poiché gli esiti sono una partizione dello spazio Ω e la probabilità di Ω uguale a 1). Non solo, visto che stiamo considerando gli esiti del nostro esperimento aleatorio e le loro probabilità, stiamo usando informazioni sul nostro problema, perché gli eventi più probabili hanno un peso maggiore.
Questa quantità appena introdotta prende il nome di *valore atteso*, *media* o *speranza matematica*. Ha un ruolo importantissimo in probabilità, perché permette di riassumere in un numero alcune caratteristiche di un esperimento.

Esempio 4.3.1. Se scommettiamo con un amico 1 euro lanciando una moneta bilanciata, qual è in media il guadagno che ci aspettiamo? In altre parole, qual è il valore atteso del guadagno?

Soluzione. La moneta è bilanciata, quindi con probabilità $\frac{1}{2}$ paghiamo 1 euro per giocare, vinciamo e incassiamo 2 euro, per un guadagno di 1 euro. Con probabilità $\frac{1}{2}$ paghiamo 1 euro per giocare, perdiamo e incassiamo 0 euro, per un guadagno di -1 euro. Il guadagno medio è allora $\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot (-1) = 0$. □

Nel caso di una scommessa, se il valore medio è nullo si dice che è *equa*. La cosa potrebbe sembrare noiosa: vediamo allora un altro esempio in cui la speranza ha un valore diverso.

Esempio 4.3.2. Quando Marianna gioca a carte contro suo nonno ha una probabilità di vincere uguale a 0,25. Se in una settimana giocano a carte 12 volte, quante partite vincerà in media Marianna, supponendo che i risultati delle partite siano tra loro indipendenti?

Soluzione. Cerchiamo di risolvere il problema più in generale, chiamando p la probabilità che Marianna vinca una partita ed n il numero di partite giocate. Cominciamo col chiederci quale sia la probabilità che Marianna non ne vinca nemmeno una: vuol dire che in tutte le n partite giocate ha vinto il nonno, cosa che ha probabilità $(1-p)^n$. Qual è invece la probabilità che Marianna ne vinca esattamente una? Dobbiamo scegliere qual è la partita che vince, tra le n giocate, cosa che possiamo fare in $\binom{n}{1} = n$ modi. Ciascuno di questi casi ha una probabilità uguale a $p(1-p)^{n-1}$.
A questo punto non è difficile passare al caso più generale di un numero k (compreso tra 0 e n) di partite vinte da Marianna: la probabilità sarà $q_k = \binom{n}{k} p^k (1-p)^{n-k}$. Osserviamo che se sommiamo questa quantità al variare di tutti i valori di k possibili (cioè per k da 0 a n) otteniamo 1. Non può essere altrimenti, perché sono tutti i casi possibili, disgiunti tra loro a due a due, quindi la loro unione dà l'intero spazio e la somma delle loro probabilità è 1.

Il problema, però, riguarda la media, che chiameremo M. Andiamo quindi a pesare i possibili esiti, ossia il numero k di partite vinte da Marianna, ciascuno con la propria probabilità q_k. Abbiamo

$$\begin{aligned}
M &= \sum_{k=0}^{n} k \cdot \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n!}{(n-k)!(k-1)!} \cdot p^k \cdot (1-p)^{n-k} \\
&= \sum_{k=1}^{n} n \cdot \frac{(n-1)!}{(n-1-(k-1))!(k-1)!} \cdot p \cdot p^{k-1} \cdot (1-p)^{n-1-(k-1)}.
\end{aligned}$$

A questo punto introduciamo due nuove variabili: $h = k-1$ e $m = n-1$. L'identità per M diventa allora

$$\begin{aligned}
M &= \sum_{h=0}^{n-1} n \cdot p \cdot \frac{(n-1)!}{(n-1-h)!h!} \cdot p^h \cdot (1-p)^{n-1-h} \\
&= \sum_{h=0}^{m} n \cdot p \cdot \frac{m!}{(m-h)!h!} \cdot p^h \cdot (1-p)^{m-h} \\
&= n \cdot p \cdot \sum_{h=0}^{m} \binom{m}{h} \cdot p^h \cdot (1-p)^{m-h} \\
&= n \cdot p,
\end{aligned}$$

dove abbiamo usato il fatto che la somma nella penultima riga è uguale a 1 (per lo stesso motivo per il quale lo era quella dei q_k prima[5]). Possiamo vedere in questa identità quello che si intende con "Marianna vince una volta su k": fissiamo $M = 1$ e cerchiamo k per cui valga $1 = k \cdot p$, cioè $p = \frac{1}{k}$.
Nel caso particolare che ci interessa, la probabilità di vittoria è $p = 0{,}25$ (cioè Marianna vince una volta su 4) e il numero delle partite è $n = 12$. Il numero medio di partite che Marianna vince è dunque $m = 12 \cdot 0{,}25 = 3$. □

Come abbiamo detto, la definizione di media richiede che gli esiti siano numeri, altrimenti non potremmo fare il conto. In realtà si può mostrare, mediante il concetto di variabile aleatoria, che questa non è una restrizione molto forte: dobbiamo solo trovare come associare a ogni evento un numero, in un modo compatibile con gli assiomi della probabilità, dal momento che dovremo trasferire la probabilità dai nostri eventi alle loro immagini mediante questa trasformazione.

[5]Questo è in una veste diversa lo stesso modello binomiale visto nel Problema 18, di cui adesso sappiamo calcolare la media.

Queste trasformazioni, che sono a tutti gli effetti delle particolari funzioni dallo spazio di probabilità considerato alla retta dei numeri reali, prendono il nome di *variabili aleatorie*. Usando questo formalismo è possibile, tra le altre cose, lasciar cadere nella definizione di media la richiesta che gli esiti siano al più un insieme numerabile, al costo di qualche accortezza aggiuntiva, tra cui sostituire integrali a somme per fare la media pesata. Questo piccolo assaggio, però, è tutto quello che vedremo delle variabili aleatorie in questo libro. Per chi fosse stato stuzzicato è sufficiente prendere un qualunque testo universitario per un corso di probabilità o, meglio ancora, seguire il corso.

Tornando alla speranza matematica, possiamo osservare che per definizione essa è determinata dalla misura di probabilità. Quindi, dal momento che la probabilità condizionata a un evento F è essa stessa una probabilità (come abbiamo visto nell'Osservazione 4.1.2), indurrà un particolare valore atteso, che prende il nome di *speranza condizionata a* F. Questa altro non è che la speranza matematica nel mondo in cui l'evento F si è verificato.

Esempio 4.3.3. Luigi lancia un dado a 12 facce. Qual è la speranza del risultato del dado condizionata al fatto che non sono usciti né 1, né 2, né 12?

Soluzione. In generale, la media di un dado a 12 facce è $M = \frac{13}{2} = 6{,}5$. Ma abbiamo un'informazione in più: sappiamo che non sono usciti alcuni valori, cosa che rappresentiamo con l'evento F. Vogliamo allora calcolare il valore medio condizionato a F. Perciò facciamo la media pesata di tutti i possibili esiti del dado, con i pesi dati dalla probabilità condizionata a F:

$$M_F = \sum_{i=1}^{12} i \cdot P(\{i\} \mid F) = \sum_{i=3}^{11} i \cdot \frac{1}{9} + 0 \cdot (1 + 2 + 12) = \frac{63}{9} = 7. \qquad \square$$

4.4 Variazioni sull'indipendenza

Facciamo un passo indietro e torniamo all'indipendenza: l'abbiamo definita a partire dalla probabilità dell'intersezione e siamo poi passati al legame con la probabilità condizionata, che ci ha dato una caratterizzazione molto più intuitiva dell'indipendenza stessa. Perché allora non abbiamo usato direttamente la probabilità condizionata per dare la definizione?

Torniamo per un momento alla questione, lasciata in sospeso, del caso in cui abbiamo un evento di probabilità nulla. Cosa succede? Supponiamo che sia $P(E) = 0$. Allora, per monotonia, $P(E \cap F) \leqslant P(E) = 0$, cioè $P(E \cap F) = 0 = P(E) \cdot P(F)$,

ossia un evento di probabilità nulla è indipendente rispetto a ogni evento, usando la definizione data. Cosa succede se andiamo a considerare le probabilità condizionate? Quella che vogliamo guardare è $P(F \mid E)$, che però non è definita: questo ci obbligherebbe a dare una definizione più macchinosa di indipendenza, specificando a parte il caso in cui un evento ha probabilità nulla. Osserviamo che questo non è davvero influente, perché $P(F \mid E)$ compare moltiplicato per $P(E)$: $P(F \mid E)$ è una probabilità e ha un valore compreso tra 0 e 1, quindi anche se non ne conosciamo il valore, sappiamo che il prodotto varrà zero.
A questo punto abbiamo la curiosità di capire quanto possa valere $P(F \mid E)$ se $P(E) = 0$. Quando andiamo ad analizzare i dettagli, però, ci accorgiamo che non ha un valore univoco. Se $E \subset F$, allora sapere che è avvenuto E ci dice automaticamente che è avvenuto F, con probabilità 1. Matematicamente questo torna (con qualche equilibrismo), perché $E \cap F = E$ e quindi abbiamo che i due termini uguali "si semplificano". Se invece $E \cap F = \varnothing$, cioè $E \subset F^c$, sapere che è avvenuto E assegna automaticamente probabilità 0 a F, cosa che possiamo immaginare vedendo la probabilità dell'evento nullo "più nulla" di tutte le altre[6]. Fin qui sembra andare tutto bene, a parte la seccatura di dover distinguere queste due possibilità. Purtroppo però questi non sono i soli casi possibili: infatti un evento di probabilità nulla può avere intersezione non vuota e differenza non vuota con un altro evento e, in questo caso, non sappiamo assegnare un valore sensato alla probabilità condizionata.

Esempio 4.4.1. Consideriamo $\Omega = [0, 1]$ con la tribù dei Boreliani e la misura di Lebesgue vista nell'Esempio 3.2.1, cioè la probabilità che misura la lunghezza dei segmenti. Prendiamo come F l'evento $\omega \leqslant \frac{1}{2}$ e come E l'evento $\left\{\frac{1}{3}, \frac{2}{3}\right\}$.
In questo caso potremmo essere tentati di dire che $P(F \mid E)$ vale $\frac{1}{2}$, ma cosa succede se abbiamo come E tutti i razionali in Ω? Viene sempre 0,5 per simmetria? E se i valori non fossero distribuiti simmetricamente? E se fossero sicuramente di probabilità nulla, ma non facilmente caratterizzabili?
Insomma, come sempre lavorare con lo zero (e con l'infinito) richiede molte attenzioni per non cadere in contraddizioni e assurdi. Per ora quindi è meglio lasciar perdere il caso di condizionamento a un evento di probabilità nulla.

Proseguiamo ora con altre proprietà interessanti del condizionamento.

Esempio 4.4.2. Lanciamo per l'n-esima volta un dado a 6 facce. Lo spazio probabilizzabile che consideriamo è dunque $\Omega = \{1, 2, 3, 4, 5, 6\}$ e $\mathcal{F} = \mathcal{P}(\Omega)$. Prendiamo i due eventi $E = \{2, 4, 6\}$ ed $F = \{3, 6\}$.

---

[6]Vedi anche gli uguali più uguali degli altri nella *Fattoria degli animali* [17].

Supponiamo che il dado sia bilanciato, quindi ogni faccia del dado (ogni singoletto) ha probabilità $P(\{i\}) = \frac{1}{6}$, per ogni $i = 1, \dots, 6$. Allora

$$P(E) = \frac{1}{2}, \quad P(F) = \frac{1}{3} \quad \text{e} \quad P(E \cap F) = P(\{6\}) = \frac{1}{6} = P(E) \cdot P(F),$$

cioè i due eventi sono indipendenti.
Supponiamo invece che il dado sia truccato: allora abbiamo una nuova probabilità $\widetilde{P}$ tale che

$$\widetilde{P}(\{1\}) = \widetilde{P}(\{2\}) = \widetilde{P}(\{3\}) = \widetilde{P}(\{4\}) = \frac{1}{12}, \quad \widetilde{P}(\{5\}) = \widetilde{P}(\{6\}) = \frac{1}{3}.$$

Dopo aver verificato che si tratta effettivamente di una probabilità, andiamo a calcolare

$$\widetilde{P}(E) = \frac{1}{2}, \quad \widetilde{P}(F) = \frac{5}{12} \quad \text{e} \quad \widetilde{P}(E \cap F) = \widetilde{P}(\{6\}) = \frac{1}{3} \neq \frac{5}{24} = \widetilde{P}(E) \cdot \widetilde{P}(F),$$

ossia sotto questa probabilità i due eventi non sono indipendenti.

Grazie a quest'ultimo esempio, notiamo che l'indipendenza tra due eventi non è una proprietà intrinseca degli eventi stessi, ma dipende dall'intero spazio di probabilità scelto e, in particolare, dalla misura di probabilità. Se consideriamo sullo stesso spazio probabilizzabile due probabilità distinte, può succedere che con una di esse due eventi siano indipendenti e con l'altra no.
C'è un caso particolare che ci interessa, arrivati a questo punto: mettiamo assieme il concetto di indipendenza e una particolare misura di probabilità, la probabilità condizionata. Se fissiamo nel nostro spazio di probabilità $(\Omega, \mathcal{F}, P)$ un evento $F \in \mathcal{F}$ di probabilità non nulla, abbiamo visto che la funzione $P_F : \mathcal{F} \to \mathbb{R}$ definita per ogni $E \in \mathcal{F}$ da $P_F(E) = P(E \mid F)$ è una probabilità e possiamo quindi considerare l'indipendenza tra eventi rispetto a essa. Ne nasce la seguente definizione.

Definizione 4.4.1. In uno spazio di probabilità $(\Omega, \mathcal{F}, P)$, fissato un evento F tale che $P(F) \neq 0$, due eventi E_1 ed E_2 si dicono *indipendenti condizionalmente a* F se

$$P(E_1 \cap E_2 \mid F) = P(E_1 \mid F) \cdot P(E_2 \mid F).$$

L'indipendenza condizionale è distinta dall'indipendenza, come vediamo nei due esempi successivi.

Esempio 4.4.3. Marko ha due cassetti nel suo armadio, nei quali tiene i suoi calzini. In uno ci sono solo calzini invernali lunghi, nell'altro ci sono calzini estivi sia lunghi sia corti (metà e metà). Marko pesca contemporaneamente due calzini da uno dei due cassetti.

Se chiamiamo S l'evento "Marko pesca dal secondo cassetto", gli eventi L_1: "il primo calzino pescato è lungo" e C_2: "il secondo calzino pescato è corto" sono indipendenti condizionalmente a S. Infatti $P(L_1 \mid S) = 0{,}5 = P(C_2 \mid S)$, quindi $P(L_1 \cap C_2 \mid S) = 0{,}25 = P(L_1 \mid S)P(C_2 \mid S)$.
In generale, tuttavia, se supponiamo che Marko scelga con uguale probabilità dai due cassetti, i due eventi non sono indipendenti. Infatti abbiamo

$$\begin{aligned}
P(L_1) &= P(L_1 \mid S)P(S) + P(L_1 \mid S^c)(1 - P(S)) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4} \\
P(C_2) &= P(C_2 \mid S)P(S) + P(C_2 \mid S^c)(1 - P(S)) = \frac{1}{4} + 0 = \frac{1}{4} \\
P(L_1 \cap C_2) &= P(L_1 \cap C_2 \mid S)P(S) + P(L_1 \cap C_2 \mid S^c)(1 - P(S)) = \frac{1}{8} \\
P(L_1) \cdot P(C_2) &= \frac{3}{16} \neq \frac{1}{8}.
\end{aligned}$$

Esempio 4.4.4. Consideriamo ancora una volta il lancio di due dadi a 6 facce. I due eventi D_2: "il primo dado ha come risultato 2" ed E_5: "il secondo dado ha come risultato 5" sono tra loro indipendenti. Tuttavia se condizioniamo rispetto all'evento S_8: "la somma dei dadi è 8", vediamo che D_2 ed E_5 non sono indipendenti condizionalmente a S_8. Infatti

$$P(D_2 \cap E_5 \mid S_8) = 0 \neq \frac{1}{25} = P(D_2 \mid S_8) \cdot P(E_5 \mid S_8),$$

poiché per avere la somma dei due dadi uguale a 8, ciascuno dei due può prendere uno dei 5 valori tra 2 e 6, quindi entrambi i fattori a ultimo membro sono $\frac{1}{5}$.

Possiamo prendere una variante dell'esempio precedente e osservare un altro aspetto.

Esempio 4.4.5. Siano D_2, E_5 come nell'esempio precedente e S_7: "la somma dei dadi è 7". Osserviamo che non solo D_2 ed E_5 sono indipendenti tra loro, ma ciascuno di loro è anche indipendente da S_7:

$$\begin{aligned}
P(D_2 \mid S_7) &= \frac{1}{6} = P(D_2) & \qquad P(S_7 \mid D_2) &= \frac{1}{6} = P(S_7) \\
P(E_5 \mid S_7) &= \frac{1}{6} = P(E_5) & \qquad P(S_7 \mid E_5) &= \frac{1}{6} = P(S_7).
\end{aligned}$$

Questo però non ci basta per dire che sono tutti e tre indipendenti tra loro, infatti

$$P(D_2 \cap E_5 \cap S_7) = \frac{1}{36} \neq \frac{1}{216} = P(D_2) \cdot P(E_5) \cdot P(S_7).$$

Concludiamo queste divagazioni sull'indipendenza con un ultimo esempio, in cui abbiamo indipendenza condizionale rispetto a una partizione.

Esempio 4.4.6. Prendiamo ora tre eventi D, E ed F sul nostro spazio di probabilità $(\Omega, \mathcal{F}, P)$, con $0 < P(F) < 1$ e supponiamo che D ed E siano indipendenti tra loro condizionalmente a F, ma anche a F^c. Possiamo dire che D ed E sono necessariamente indipendenti tra loro in senso stretto?

Soluzione. Da un lato avremmo la tentazione di rispondere affermativamente: sono indipendenti in ciascuna delle due possibilità determinate da F (sia con F vero, sia con F falso), quindi lo saranno anche globalmente. Allo stesso tempo, però, gli esempi precedenti ci hanno insegnato un po' di prudenza.
Proviamo allora a vedere se ci sono condizioni da soddisfare affinché questa indipendenza sia vera e, allo stesso tempo, se possiamo costruire un controesempio. Dalla formula di fattorizzazione abbiamo le seguenti identità:

$$\begin{aligned} P(D) &= P(D \mid F) \cdot P(F) + P(D \mid F^c) \cdot (1 - P(F)) \\ &= (P(D \mid F) - P(D \mid F^c)) \cdot P(F) + P(D \mid F^c) \end{aligned}$$

$$\begin{aligned} P(E) &= P(E \mid F) \cdot P(F) + P(E \mid F^c) \cdot (1 - P(F)) \\ &= (P(E \mid F) - P(E \mid F^c)) \cdot P(F) + P(E \mid F^c), \end{aligned}$$

cioè, chiamando per semplicità $d = P(D \mid F)$, $d' = P(D \mid F^c)$, $e = P(E \mid F)$, $e' = P(E \mid F^c)$ e anche $a = P(D)$, $b = P(E)$, $c = P(F)$,

$$\begin{aligned} a &= dc + d'(1 - c) = (d - d')c + d' \\ b &= ec + e'(1 - c) = (e - e')c + e'. \end{aligned}$$

Allo stesso tempo abbiamo anche, grazie all'indipendenza di D ed E condizionalmente a F ed F^c,

$$\begin{aligned} P(D \cap E) &= P(D \cap E \mid F) \cdot P(F) + P(D \cap E \mid F^c) \cdot P(F^c) \\ &= P(D \mid F) \cdot P(E \mid F) \cdot P(F) + P(D \mid F^c) \cdot P(E \mid F^c) \cdot P(F^c) \\ &= dec + d'e'(1 - c). \end{aligned}$$

Avremmo l'indipendenza se valesse $P(D \cap E) = P(D) \cdot P(E)$, cioè, con la nuova notazione, $dec + d'e'(1 - c) = ab$. Studiamo allora questa identità.

$$\begin{aligned} dec + d'e' - d'e'c = ab& \\ &= (dc + d' - d'c)(ec + e' - e'c) \\ &= dec^2 + de'c - de'c^2 + d'ec \\ &\quad + d'e' - d'e'c - d'ec^2 - d'e'c + d'e'c^2. \end{aligned}$$

Possiamo semplificare un po' di termini, arrivando a

$$dec^2 + de'c - de'c^2 + d'ec - d'ec^2 - d'e'c + d'e'c^2 - dec = 0$$

che possiamo riscrivere, raccogliendo più volte i fattori in comune, come

$$c(c-1)(d-d')(e-e') = 0,$$

o, tornando esplicitamente alle probabilità,

$$P(F) \cdot P(F^c) \cdot (P(D \mid F) - P(D \mid F^c)) \cdot (P(E \mid F) - P(E \mid F^c)) = 0.$$

I primi due fattori per ipotesi non possono essere 0 (altrimenti non potremmo parlare di probabilità condizionali), quindi ci sono due possibilità: o la probabilità di D non cambia nelle due parti F ed F^c, o quella di E non cambia.

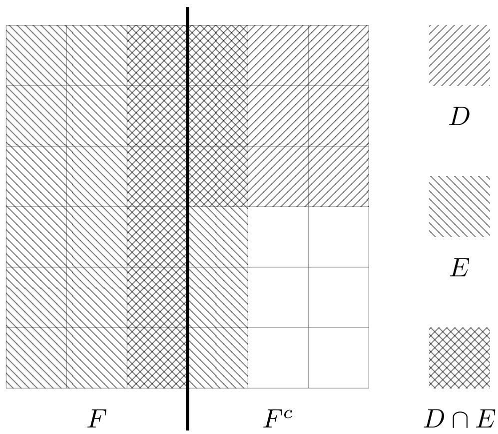

Figura 4.2: Un controesempio

Abbiamo allora tutti gli ingredienti per costruire un controesempio, rappresentato in Figura 4.2. In questo esempio abbiamo $P(F) = P(F^c) = \frac{1}{2}$. La probabilità di ciascun evento è data dalla sua area in quadratini divisa per l'area totale (sempre in quadratini). In F abbiamo $P(D \mid F) = \frac{1}{3}$, $P(E \mid F) = 1$ e $P(D \cap E \mid F) = \frac{1}{3}$, mentre in F^c valgono $P(D \mid F^c) = \frac{1}{2}$, $P(E \mid F^c) = \frac{1}{3}$ e $P(D \cap E \mid F^c) = \frac{1}{6}$. Allora, condizionalmente a F e F^c, D ed E sono indipendenti.
Se guardiamo però le probabilità di D ed E, vediamo $P(D) = \frac{5}{12}$ e $P(E) = \frac{2}{3}$, dunque $P(D) \cdot P(E) = \frac{5}{18}$, mentre $P(D \cap E) = \frac{1}{4}$, quindi D ed E non sono indipendenti. $\square$

4.5 Teorema di Bayes

La probabilità condizionata non è simmetrica: in generale $P(E \mid F) \neq P(F \mid E)$. Da un punto di vista matematico la cosa è immediata: basta guardare la definizione e osservare che non è simmetrica nei due insiemi considerati. Tuttavia, se andiamo a considerare l'uso della probabilità nella vita di tutti i giorni, ci accorgiamo che questo è uno degli errori (o fallacie) più frequenti.

Esempio 4.5.1. Da una recente indagine[7] sui vaccini per l'influenza stagionale, in Italia la copertura vaccinale per le persone di età maggiore o uguale a 65 anni è del 53,1%. Nella popolazione generale la copertura si riduce al 15,8%. Questo non significa che, scegliendo un vaccinato a caso, la probabilità che abbia almeno 65 anni sia il 53,1%. Infatti gli italiani con almeno 65 anni sono circa 14 milioni, di cui circa 7,5 milioni sono vaccinati. Al tempo stesso la popolazione italiana è costituita da 60 milioni di persone circa, di cui 9,5 milioni vaccinati. Tra i vaccinati, gli over 65 sono quasi il 79%. In termini di probabilità condizionate abbiamo

$$P\,(\text{vaccinato} \mid \text{over } 65) = 53{,}1\% \neq P\,(\text{over } 65 \mid \text{vaccinato}) = 78{,}9\%.$$

Purtroppo nel momento in cui ci si allontana dal contesto esplicitamente matematico, capita spesso che le due probabilità condizionate vengano confuse. Vediamo alcuni tipici esempi.

- "Se la maggior parte dei criminali appartiene a un certo gruppo, allora è altamente probabile che un generico membro del gruppo sia un criminale." Falso: tra i condannati per omicidio in Italia, oltre il 95% sono di sesso maschile, ma non ci verrebbe mai in mente di pensare che quasi tutti i maschi italiani siano assassini.

- "Se la probabilità che un imputato abbia indizi contro di lui pur essendo innocente è molto bassa, allora deve essere molto bassa anche la probabilità che sia innocente se ci sono indizi contro di lui." Falso: questo argomento prende il nome di *fallacia del procuratore* ed è stato ingrediente di molti casi di cattiva giustizia, con condanne annullate in fase di revisione dei processi, ad esempio il già citato caso Collins, ma anche con assoluzioni forse non meritate, come nel caso O. J. Simpson.

- "Se la maggior parte dei recenti attacchi terroristici in Europa è stata portata a termine da musulmani, allora la proporzione di musulmani che sono terroristi

[7] Fonte: Ministero della Salute-ISS per la stagione 2018/19.

è molto alta.” Falso anche questa volta: in realtà la probabilità che un musulmano europeo sia un terrorista è dell'ordine di $4 \cdot 10^{-6}$, cento volte più piccola della probabilità di essere colpiti da un fulmine nel corso della propria vita.

Pur non essendoci simmetria, le due probabilità condizionate $P(E \mid F)$ e $P(F \mid E)$ non sono completamente scollegate tra loro, come vediamo nel prossimo esempio.

Esempio 4.5.2. Tra i concorrenti delle Olimpiadi della Matematica[8], il 43% è del biennio, il rimanente 57% del triennio. Tra i concorrenti del biennio, il 51% sono ragazze, tra quelli del triennio tale percentuale scende al 23%. Se Giulietta è una concorrente, qual è la probabilità che sia una studentessa del biennio?

Soluzione. Indichiamo con B l'evento “concorrente del biennio” e con ♀ l'evento “concorrente è una ragazza”. Allora dai dati del problema abbiamo:

$$P(B) = 0{,}43, \quad P(B^c) = 0{,}57, \quad P(♀ \mid B) = 0{,}51, \quad P(♀ \mid B^c) = 0{,}23.$$

Noi però vorremmo calcolare $P(B \mid ♀)$, poiché Giulietta è una ragazza. Cominciamo a calcolare qualcosa di diverso: $P(B \cap ♀)$, cioè la probabilità che la persona presa sia del biennio e sia una ragazza. Lo facciamo perché per definizione $P(B \mid ♀) = \frac{P(B \cap ♀)}{P(♀)}$ e stiamo in questo modo calcolando il numeratore. Dalla definizione di probabilità condizionata otteniamo che

$$P(B \cap ♀) = P(♀ \mid B) \cdot P(B)$$

dove le due quantità a secondo membro sono note. Possiamo allora calcolare esplicitamente $P(B \cap ♀) = 0{,}51 \cdot 0{,}43 = 0{,}2193$.
Per calcolare $P(B \mid ♀)$, la quantità che cerchiamo, non resta che calcolare $P(♀)$, cosa che possiamo fare aiutandoci con la formula di fattorizzazione,

$$P(♀) = P(♀ \mid B) \cdot P(B) + P(♀ \mid B^c) \cdot P(B^c).$$

Anche in questo caso tutte le quantità sono note (e addirittura abbiamo già calcolato il primo prodotto), quindi abbiamo

$$P(♀) = 0{,}2193 + 0{,}23 \cdot 0{,}57 = 0{,}3504.$$

[8] Un sottoinsieme ben determinato degli studenti di scuola secondaria di secondo grado.

Mettendo assieme il tutto, abbiamo che quanto cerchiamo, cioè la probabilità che Giulietta sia del biennio, è

$$P(B \mid \female) = \frac{P(B \cap \female)}{P(\female)} = \frac{P(\female \mid B) \cdot P(B)}{P(\female)} = \frac{0{,}2193}{0{,}3504} \approx 0{,}6259. \qquad \square$$

Nell'esempio precedente abbiamo fatto qualcosa di interessante, che va oltre la risoluzione del problema assegnato: abbiamo calcolato una probabilità condizionata in funzione della sua speculare, ossia $P(E \mid F)$ a partire da $P(F \mid E)$. Possiamo fare la stessa cosa in generale, come mostrato dal seguente risultato.

Teorema 4.5.1 (Teorema di Bayes)**.** *Sia* $(\Omega, \mathcal{F}, P)$ *uno spazio di probabilità e siano* E, F *due eventi, entrambi di probabilità non nulla. Allora*

$$P(E \mid F) = \frac{P(F \mid E)}{P(F)} \cdot P(E).$$

Dimostrazione. Dalla definizione di probabilità condizionata abbiamo la seguente catena di uguaglianze:

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)} = \frac{P(E \cap F)}{P(E)} \cdot \frac{P(E)}{P(F)} = \frac{P(F \mid E) \cdot P(E)}{P(F)}.$$

$\square$

Possiamo poi combinare il Teorema di Bayes con il teorema delle probabilità totali, ricavando il seguente risultato.

Corollario 4.5.2. *Data una partizione di* Ω *in eventi disgiunti di probabilità non nulla, se* F *è un evento in* $\mathcal{F}$*, allora per ogni evento* E

$$P(E \mid F) = \frac{P(F \mid E) \cdot P(E)}{\sum_{i \in I} P(F \mid E_i) \cdot P(E_i)}. \tag{4.2}$$

Un trucco di pigrizia: se scegliamo la partizione in modo che E ne faccia parte, il prodotto al numeratore sulla destra compare anche nella somma a denominatore, quindi dobbiamo calcolare il valore di un addendo in meno.

Esempio 4.5.3. Un laboratorio propone un nuovo test per determinare la positività (o negatività) al virus SARS-CoV-2. La proporzione di infetti che risultano positivi al test (detta anche sensibilità) è il 99,9%, mentre la proporzione di sani che sono negativi al test (detta anche specificità) è il 99,7%. In Italia il virus contagia 5 persone su 1000. Jacopo si sottopone a questo test. Se il test è positivo, con che probabilità Jacopo è davvero infetto?

Soluzione. La prima tentazione è di rispondere 99,9%. Tuttavia, avendo visto che il condizionamento non è simmetrico, sappiamo distinguere tra $P(M \mid +)$ e $P(+ \mid M)$, dove M è l'evento "Jacopo è malato" e $+$ l'evento "Jacopo è positivo". Il dato del problema sulla sensibilità è $P(+ \mid M)$, mentre il problema ci chiede $P(M \mid +)$. Il Teorema di Bayes, però, ci suggerisce la strada da prendere:

$$\begin{aligned} P(M \mid +) &= \frac{P(+ \mid M) \cdot P(M)}{P(+)} \\ &= \frac{P(+ \mid M) \cdot P(M)}{P(+ \mid M) \cdot P(M) + P(+ \mid M^c) \cdot P(M^c)}, \end{aligned} \tag{4.3}$$

dove abbiamo usato anche l'identità (4.2) e la formula di fattorizzazione. Sostituiamo i valori disponibili, che abbiamo già come dati o che ricaviamo facilmente:

$$P(M) = 0{,}005, \quad P(M^c) = 1 - P(M) = 0{,}995, \quad P(+ \mid M) = 0{,}999$$

e anche

$$P(+ \mid M^c) = 1 - P(- \mid M^c) = 0{,}003.$$

Tornando alla (4.3), abbiamo allora

$$\begin{aligned} P(M \mid +) &= \frac{0{,}999 \cdot 0{,}005}{0{,}999 \cdot 0{,}005 + 0{,}003 \cdot 0{,}995} \\ &= \frac{0{,}004995}{0{,}00798} \approx 63\%. \end{aligned}$$

Questa probabilità, per quanto non trascurabile, è comunque inferiore rispetto a quella che ci aveva tentato inizialmente.
Spendiamo due parole per spiegare, per quanto in modo non approfondito, il motivo di questa discrepanza. Concentriamoci su quello che sappiamo: Jacopo è positivo al test. Quando succede questo? Se una persona è veramente malata, nel 99,9% dei casi il test sarà positivo, tuttavia l'incidenza della malattia, ossia la proporzione di persone effettivamente malate, è molto piccola. Allo stesso tempo, raramente (nello 0,3% dei casi) il test segnalerà come positivo qualcuno che è sano. Tuttavia la proporzione di persone sane è molto alta, quindi tra i positivi al test i falsi positivi sono una parte non trascurabile: più di un terzo. □

L'esempio precedente, oltre a essere un buon esercizio, ci mostra anche quanto sia importante il Teorema di Bayes nella vita reale. Il cervello umano non è portato intuitivamente al ragionamento probabilistico ed è quindi facile incappare in errori. Il Teorema di Bayes è uno degli strumenti che ci permettono di aggirare ed evitare questi errori. Una delle sue applicazioni, in sintonia con il metodo scientifico, consiste nello spingerci ad aggiornare le nostre convinzioni.

Cosa vogliamo dire con questo? Ci aspettiamo di fare ipotesi e metterle alla prova con opportuni esperimenti. Facciamo entrare in gioco anche la probabilità, usandola come misura del livello di convinzione nella nostra ipotesi.

Ad esempio, Maestra Rita potrebbe supporre che la probabilità che Pierino non abbia studiato la lezione sia del 70%. In questo caso il fenomeno d'interesse è "lo studio da parte degli scolari" (in particolare da parte di Pierino) e abbiamo come ipotesi "Pierino non ha studiato". Maestra Rita non è sicura di questa ipotesi: Pierino potrebbe finalmente aver capito ed essersi messo sui libri, ma se la maestra dovesse scommettere darebbe fiducia a Pierino solo al 30%. Maestra Rita però può mettere alla prova la sua ipotesi con un esperimento: interrogando Pierino ha modo di verificare se abbia studiato o no.

Scriviamo queste cose con la notazione della probabilità: H è la nostra ipotesi, (Pierino non ha studiato) che supponiamo vera con probabilità $P(H)$ (70% nell'esempio). Con E indichiamo il risultato di un esperimento (Pierino non sa rispondere alla domanda).

Prima di effettuare l'esperimento, possiamo assegnare le probabilità relative all'esperimento: $P(E \mid H)$ nel caso in cui H sia vera e $P(E \mid H^c)$ nel caso in cui H sia falsa. Nel caso di Pierino, Maestra Rita stima che $P(E \mid H) = 90\%$: se Pierino non ha studiato è probabile che non sappia rispondere, ma potrebbe avere fortuna e azzeccare la risposta. Viceversa, valuta $P(E \mid H^c) = 5\%$: se Pierino ha studiato, potrebbe comunque non rispondere correttamente, per qualche motivo, anche se è poco probabile.

Assegniamo queste probabilità condizionate prima di vedere l'effettivo risultato dell'esperimento. Quando però sappiamo cosa è successo, possiamo usare gli ingredienti che abbiamo preparato per vedere come cambia la nostra confidenza nell'ipotesi dopo aver visto il verificarsi di E. In altre parole siamo interessati a $P(H \mid E)$: quanto è convinta Maestra Rita che Pierino non abbia studiato se non ha saputo rispondere alla domanda che gli ha fatto?

Per il Teorema di Bayes,

$$P(H \mid E) = \frac{P(E \mid H)}{P(E \mid H)P(H) + P(E \mid H^c)P(H^c)} \cdot P(H)$$

e Maestra Rita, che prima pensava che ci fosse un 30% di possibilità che Pierino per una volta avesse studiato la lezione, dopo la scena muta aggiorna questa sua convinzione,

$$\frac{90}{100} \cdot \frac{10\,000}{90 \cdot 70 + 5 \cdot 30} \cdot \frac{70}{100} \approx 97{,}7\%$$

e ha quasi la certezza che Pierino non si sia preparato.

Tornando al caso generale, $P(H)$ è la probabilità che diamo alla verità di H prima di effettuare l'esperimento e prende quindi il nome di *probabilità a priori* (o *prior*). D'altra parte, $P(H \mid E)$ è la probabilità di H aggiornata dopo aver visto il risultato E dell'esperimento: prende il nome di *probabilità a posteriori* o *posterior*.
È allora più chiaro il parallelo col ragionamento scientifico. Nello studiare un fenomeno, facciamo un'ipotesi H di cui siamo convinti a un livello $P(H)$, per precedenti osservazioni o per altri motivi. Pianifichiamo un esperimento e, prima di effettuarlo, valutiamo con cura quali sono i possibili risultati e quanto li riteniamo plausibili in un mondo in cui H è vera e in uno in cui H è falsa, dando dei valori a $P(E \mid H)$ e $P(E \mid H^c)$, rispettivamente, per ogni possibile esito E dell'esperimento. A questo punto facciamo l'esperimento e ne osserviamo il risultato E. Possiamo poi aggiornare la nostra convinzione che H sia vera, con l'informazione in più raccolta con l'esperimento, calcolando $P(H \mid E)$ col Teorema di Bayes.
Nell'Esempio 4.5.3, prima di sottoporsi al test, Jacopo poteva stimare la probabilità di essere malato allo 0,5%; dopo il risultato positivo del test, rivaluta questa probabilità al 63%. Il modo in cui ha scelto la sua probabilità a priori di essere malato è di considerarsi un individuo qualunque della popolazione, all'interno della quale l'incidenza è 5 su 1000. Chiaramente altri fattori sarebbero potuti entrare in gioco: ad esempio se avesse avuto sintomi, magari avrebbe valutato diversamente la probabilità a priori.
Ci sono pochi vincoli sulla prior: deve essere una probabilità, quindi soddisfare le proprietà che ormai conosciamo (in particolare quella di monotonia). In più, se vogliamo poter usare il Teorema di Bayes in modo fruttuoso, non possiamo assegnare mai le probabilità 0 e 1.
Infatti se assegniamo a un evento probabilità 1, diciamo $P(H) = 1$, per quanti esperimenti contrari facciamo non potremo mai discostarci da quel valore:

$$P(H \mid E) = \frac{P(E \mid H)P(H)}{P(E \mid H)P(H) + P(E \mid H^c)P(H^c)} = \frac{P(E \mid H)}{P(E \mid H)} = 1$$

e analogamente per il caso $P(H) = 0$.
Questo ha senso, da un punto di vista astratto: se siamo certi di qualcosa nulla ci farà cambiare idea. In generale, però, quando dichiariamo di essere certi di qualcosa in un contesto sperimentale, questo significa che per cambiare idea avremo bisogno di una notevole quantità di evidenza contraria alla nostra convinzione precedente. Questo almeno finché vogliamo agire in modo razionale. I valori 0 e 1 sono quindi da evitare.
Quanto appena detto vale anche per i risultati degli esperimenti: anche se possono sembrare controesempi alla nostra ipotesi, dobbiamo tenerci un po' di margine (da valutare) che tenga conto di possibili errori nell'esperimento, ad esempio una

lettura sbagliata da parte dello strumento. Quindi non raggiungeremo mai certezze: per la gioia degli scienziati sperimentali possiamo continuare a fare esperimenti all'infinito!
E se continuiamo a fare esperimenti, vorremo combinare i risultati osservati in ciascuno di essi per aggiornare la nostra $P(H)$. Come primo passo, vediamo il caso di due esperimenti: abbiamo due esiti E_1 ed E_2 e vogliamo capire quanto vale $P(H \mid E_1 \cap E_2)$, la probabilità a posteriori della nostra ipotesi dopo entrambi gli esperimenti. Facciamo un esempio.

Esempio 4.5.4. Torniamo al caso di Jacopo, incontrato all'Esempio 4.5.3. Se Jacopo si sottoponesse di nuovo al test e questo risultasse nuovamente positivo, quale sarebbe la probabilità che sia effettivamente malato?

Soluzione. L'impostazione del problema è simile a quella dell'Esempio 4.5.3, solo che ora abbiamo due eventi rispetto ai quali condizioniamo. Allora

$$\begin{aligned} P(M \mid +_2 \cap +_1) &= \frac{P(+_2 \mid M \cap +_1)P(M \mid +_1)}{P(+_2 \mid M \cap +_1)P(M \mid +_1) + P(+_2 \mid M^c \cap +_1)P(M^c \mid +_1)} \\ &= \frac{P(+_2 \mid M \cap +_1)}{P(+_2 \mid M \cap +_1)P(M \mid +_1) + P(+_2 \mid M^c \cap +_1)P(M^c \mid +_1)} \cdot \\ &\quad \cdot \frac{P(+_1 \mid M)P(M)}{P(+_1 \mid M)P(M) + P(+_1 \mid M^c)P(M^c)}, \end{aligned} \tag{4.4}$$

in cui abbiamo messo in evidenza una specie di iterazione. Tuttavia non è facile semplificare ulteriormente questa espressione, a meno di non fare alcune ipotesi di indipendenza tra i due test. In particolare, prendiamo come ipotesi il fatto che i due test, ossia gli eventi $+_1$ e $+_2$, siano indipendenti *condizionatamente a* M ed M^c. In altre parole, quando sappiamo se Jacopo è malato o no, la positività dei due test è indipendente[9].
Se ora torniamo alla (4.4) abbiamo, sfruttando l'indipendenza condizionata,

$$\begin{aligned} P(M \mid +_1 \cap +_2) = &\frac{P(+_2 \mid M)}{P(+_2 \mid M)P(M \mid +_1) + P(+_2 \mid M^c)P(M^c \mid +_1)} \cdot \\ &\cdot \frac{P(+_1 \mid M)P(M)}{P(+_1 \mid M)P(M) + P(+_1 \mid M^c)P(M^c)} \end{aligned}$$

in cui il secondo fattore nel membro di destra è esattamente $P(M \mid +_1)$. In sostanza stiamo facendo esattamente la medesima cosa vista all'Esempio 4.5.3, solo che al

[9]Questo non significa che i due test siano indipendenti tra loro, nonostante stiamo considerandoli indipendenti se condizionati a un evento e al suo complementare, come abbiamo visto nell'Esempio 4.4.6.

secondo passaggio è cambiata la probabilità di partenza: non è più $P(M)$, bensì $P(M \mid +_1)$, perché dobbiamo tenere conto del primo test fatto.
Possiamo a questo punto sostituire nell'espressione i valori che conosciamo e ricavare la probabilità cercata: $P(M \mid +_1 \cap +_2) \approx 99{,}8\%$. Avere un secondo test positivo ci ha portati (quasi) alla certezza, nonostante quanto osservato prima (Esempio 4.5.3) sul fatto che in prima battuta i falsi positivi non sono trascurabili.
Può essere interessante notare, a margine di questo esempio, cosa succederebbe qualora il secondo test fosse negativo. La risposta di pancia potrebbe essere che il test positivo e quello negativo si "annullano" a vicenda, quindi che la probabilità che Jacopo sia malato ritorni a essere il valore di base 0,005. Le cose però non stanno proprio così: abbiamo

$$P(M \mid +_1 \cap -_2) = \frac{P(-_2 \mid M)}{P(-_2 \mid M)P(M \mid +_1) + P(-_2 \mid M^c)P(M^c \mid +_1)} \cdot \\ \cdot \frac{P(+_1 \mid M)P(M)}{P(+_1 \mid M)P(M) + P(+_1 \mid M^c)P(M^c)}$$

e, sostituendo i valori che conosciamo, otteniamo $P(M \mid +_1 \cap -_2) \approx 0{,}002$. Come mai? Il motivo è questo: se da un lato i falsi positivi non sono infrequenti, i falsi negativi lo sono molto meno, dato che complessivamente gli infetti sono una piccola parte della popolazione. □

Ispirati dall'Esempio 4.5.4, possiamo fare alcune osservazioni generali.

Osservazione 4.5.3. Ripetere un esperimento non è inutile, nemmeno se dà nuovamente il medesimo risultato: la nostra valutazione della probabilità cambierà ulteriormente dopo la seconda osservazione. Lo stesso vale per due esperimenti con risultato opposto: in generale vederne i risultati non ci riporta al punto di partenza, ma ci lascia comunque delle informazioni aggiuntive, codificate dentro la probabilità.

Osservazione 4.5.4. La grandezza dell'effetto delle due osservazioni sulla probabilità non è la stessa. Nell'esempio, il risultato del primo test porta la probabilità di malattia da 0,005 a 0,63, con un aumento di 0,625. Il secondo esperimento la fa crescere "solo" di 0,368. Questo potrebbe sorprenderci: la seconda osservazione non è in sé diversa dalla prima. Il fenomeno, che prende il nome di *diminuzione dei ritorni marginali*, è del tutto naturale: ogni successivo esperimento con il medesimo risultato ha un impatto sempre minore sulla probabilità. Può sembrare controintuitivo, ma ciò accade solo perché le informazioni che raccogliamo interagiscono con la probabilità attraverso una moltiplicazione e non una somma, come il nostro cervello preferirebbe. Possiamo vedere una traccia di questo comportamento moltiplicativo nell'identità (4.4).

Se siamo interessati alla probabilità che H sia vera, contrapposta alla probabilità che sia falsa, possiamo considerarne il rapporto $\frac{P(H)}{P(H^c)}$. Questa quantità prende il nome di *odd*[10] *a favore di* H.
È un modo diverso di vedere la probabilità: non consideriamo più il rapporto tra casi favorevoli e casi totali, ma quello tra casi favorevoli e casi contrari. Osserviamo che a partire dagli odd possiamo ricostruire la probabilità di H: dal momento che $P(H^c) = 1 - P(H)$, se gli odd a favore di H hanno un valore a, la probabilità di H sarà $P(H) = \frac{a}{1+a}$.
Andiamo a rivedere il Teorema di Bayes da questo nuovo punto di vista: dato un evento E di probabilità non nulla abbiamo

$$\frac{P(H \mid E)}{P(H^c \mid E)} = \frac{P(E \mid H)P(H)}{P(E)} \cdot \frac{P(E)}{P(E \mid H^c)P(H^c)} = \frac{P(E \mid H)}{P(E \mid H^c)} \cdot \frac{P(H)}{P(H^c)}.$$

In altre parole, gli odd a posteriori $\frac{P(H|E)}{P(H^c|E)}$ si ottengono dagli odd a priori $\frac{P(H)}{P(H^c)}$ moltiplicandoli per $\frac{P(E|H)}{P(E|H^c)}$. Questa quantità prende il nome di *rapporto di verosimiglianza* e misura quanto sia più probabile osservare il risultato E dell'esperimento a seconda che sia vera H o H^c.
La formula di Bayes per gli odd,

$$\frac{P(H \mid E)}{P(H^c \mid E)} = \frac{P(E \mid H)}{P(E \mid H^c)} \cdot \frac{P(H)}{P(H^c)}$$

contiene solo moltiplicazioni e la sua forma semplice ci fa ben sperare per il caso di esperimenti multipli. Se facciamo due esperimenti di esiti E_1 ed E_2, indipendenti tra loro condizionatamente a H e H^c, allora

$$\begin{aligned}\frac{P(H \mid E_1 \cap E_2)}{P(H^c \mid E_1 \cap E_2)} &= \frac{P(E_2 \mid H)}{P(E_2 \mid H^c)} \cdot \frac{P(H \mid E_1)}{P(H^c \mid E_1)} \\ &= \frac{P(E_2 \mid H)}{P(E_2 \mid H^c)} \cdot \frac{P(E_1 \mid H)}{P(E_1 \mid H^c)} \cdot \frac{P(H)}{P(H^c)}.\end{aligned}$$

Questa scrittura, che riguarda un doppio aggiornamento, è molto più semplice di quella ottenuta in precedenza. In essa possiamo leggere direttamente alcune proprietà, ad esempio che non è importante l'ordine in cui vediamo i risultati degli esperimenti. Inoltre possiamo ritrovare quanto già visto su esperimenti aventi esiti opposti: non si annullano a vicenda, perché il fattore di aggiornamento degli odd è il prodotto dei rapporti di verosimiglianza dei due esperimenti:

$$\frac{P(E_2 \mid H)}{P(E_2 \mid H^c)} \cdot \frac{P(E_1 \mid H)}{P(E_1 \mid H^c)} = \frac{1 - P(E_1 \mid H)}{1 - P(E_1 \mid H^c)} \cdot \frac{P(E_1 \mid H)}{P(E_1 \mid H^c)},$$

[10] La nozione di odd è in realtà più generale: non è necessario restringersi a un evento e al suo complementare, ma qui ci limiteremo a questo caso speciale. Per quanto riguarda il nome, adottiamo il più comune termine inglese, ma segnaliamo la traduzione *alea* usata da Li Calzi in *La matematica dell'incertezza* [14].

in cui in generale né il primo né il secondo fattore si semplificano, né lo fanno numeratore o denominatore.
Mettiamo ora assieme due osservazioni precedenti. Da un lato la formula di Bayes con gli odd è puramente moltiplicativa, cosa che la rende comoda in particolar modo in caso di esperimenti ripetuti. Dall'altro per il cervello umano la moltiplicazione non è così semplice come l'addizione. Se solo ci fosse un modo di passare da un prodotto a una somma, questo potrebbe aiutare la nostra intuizione.
Aver difficoltà con la moltiplicazione era un problema ancora più sentito dai matematici del passato, che non potevano appoggiarsi a strumenti di calcolo per fare (o controllare) i conti. Fu appunto per semplificare i conti e porre rimedio a questa situazione che Nepero[11] introdusse i logaritmi nel 1614.
Possiamo combinare questa idea con gli odd, prendendone i logaritmi, e ottenere i *log-odd*, ricavando immediatamente una terza versione della formula di Bayes,

$$\log\left(\frac{P(H \mid E)}{P(H^c \mid E)}\right) = \log\left(\frac{P(E \mid H)}{P(E \mid H^c)}\right) + \log\left(\frac{P(H)}{P(H^c)}\right),$$

che nel caso di due esperimenti diventa

$$\log\frac{P(H \mid E_1 \cap E_2)}{P(H^c \mid E_1 \cap E_2)} = \log\frac{P(E_2 \mid H)}{P(E_2 \mid H^c)} + \log\frac{P(E_1 \mid H)}{P(E_1 \mid H^c)} + \log\frac{P(H)}{P(H^c)}.$$

Non abbiamo specificato la base dei logaritmi, perché questo vale per ogni base ragionevole, anche se solitamente tre scelte la fanno da padrone: il logaritmo naturale in base e, il logaritmo in base 2 e quello in base 10. Dal punto di vista dei conti "spannometrici" a supporto dell'intuizione, il logaritmo in base 10 è particolarmente interessante, perché ci permette di lavorare direttamente con gli ordini di grandezza (gli esponenti).
Il fatto di prendere i log-odd ha un altro vantaggio, oltre a quello di rendere più semplici i calcoli. Per definizione le probabilità vivono nell'intervallo $[0, 1]$, ma abbiamo visto che gli estremi non si comportano bene con il Teorema di Bayes. Li abbiamo proibiti, in questo contesto, e dobbiamo ricordarci di evitarli. La rappresentazione coi log-odd ci aiuta parecchio, in questo senso.
Passando dalle probabilità agli odd, l'intervallo $[0, 1]$ viene trasformato nella semiretta $[0, +\infty)$. Col passo successivo ai log-odd la semiretta viene trasformata nell'intera retta reale: la probabilità 0 corrisponde a log-odd $-\infty$, la probabilità 1 a log-odd $+\infty$. Intuitivamente è più chiaro che non possiamo arrivare a quegli estremi. Possiamo anche notare che, al centro, un log-odd uguale a 0 per un evento E corrisponde in odd a $\frac{P(E)}{P(E^c)} = 1$. L'evento e il suo complementare hanno

[11] John Napier (1550 – 1617).

dunque la stessa probabilità, $\frac{1}{2}$: sono per noi indifferenti, non sappiamo scegliere, non sappiamo nulla.

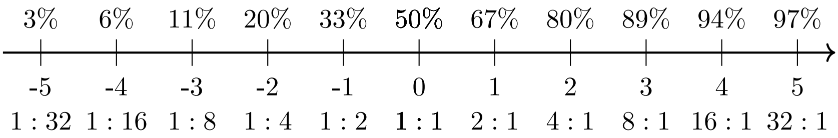

Figura 4.3: La nuova rappresentazione delle probabilità, usando i logaritmi in base 2. Sopra la retta abbiamo la probabilità di H, sotto la retta i $\log_2$-odd e, più sotto, gli odd.

Tornando alla formula di Bayes nella sua versione log-odd, possiamo notare che le informazioni che accumuliamo con ogni esperimento traslano la nostra convinzione lungo la retta. La traslazione è data dal logaritmo del rapporto di verosimiglianza, $\log\left(\frac{P(E|H)}{P(E|H^c)}\right)$, che avrà segno positivo quando $P(E \mid H) > P(E \mid H^c)$, cioè quando il risultato dell'esperimento è in favore dell'ipotesi H e viceversa segno negativo quando favorisce H^c. Ripetere un esperimento ottenendo più volte il medesimo risultato trasla più volte i log-odd della nostra convinzione, ogni volta della stessa quantità. La probabilità corrispondente, però, cambia tanto più lentamente quanto più lontani siamo da 0, cioè dal punto di equilibrio tra H e il suo complementare H^c. Questo perché, in modo molto razionale, la stessa quantità di evidenza ha effetti diversi a seconda di quanto siamo già convinti a priori.

Esempio 4.5.5. Se abbiamo che $\log_2\left(P(E \mid H)/P(E \mid H^c)\right) = 1$, questo cambierà la nostra prior del 17% se partiamo dal centro, cioè dal 50%, ma la cambierà solamente del 3% se siamo convinti al 94%.

È molto interessante notare che, a parità di modulo, l'evidenza ha più peso quando ci riporta verso il centro, in contrasto con la nostra convinzione, rispetto a quando ci allontana da esso, in sintonia con essa: un controesempio sperimentale ci dà più informazioni di un'ulteriore conferma. Possiamo aspettarcelo: se l'esperimento dà per l'n-esima volta una conferma di quello di cui già siamo convinti, questo ci stupisce sempre meno, viceversa qualcosa di opposto ci dà molto più da pensare e rimette in discussione quanto crediamo, o almeno dovrebbe farlo.
Infine, con questa rappresentazione è chiaro che per poter raggiungere la certezza avremmo bisogno di una quantità infinita di evidenza. Per quanto possiamo essere convinti di qualcosa, lo saremo solo a livello finito, per quanto grande. Quindi una quantità sufficiente di informazioni in contrasto con quello che crediamo ci porterà necessariamente a cambiare idea. Dobbiamo farlo, è dimostrato.

4.6 Esercizi

Problema 19
Abbiamo due contenitori: nel primo ci sono due monete d'oro, nel secondo una moneta d'oro e una d'argento. Scegliamo a caso uno dei due contenitori e ne estraiamo una moneta. Se la moneta che abbiamo estratto è d'oro, con che probabilità è d'oro anche la seconda moneta nel contenitore?

Problema 20 *(Kolmogorov Students' Contest in Probability Theory*[12]*, 2003)*
Su uno spazio di probabilità $(\Omega, \mathcal{F}, P)$ sono assegnati gli eventi F, G ed $(E_i)_{i=1}^n$ tali che $\bigcup_{i=1}^n E_i = \Omega$ e, per ogni $i = 1, \ldots, n$, $P(E_i) > 0$ e $P(F \mid E_i) \geqslant P(G \mid E_i)$. Possiamo dire che $P(F) \geqslant P(G)$?

Problema 21
Lanciamo contemporaneamente un dado a 20 facce e uno a 8 facce. Qual è la probabilità che facciamo più di 14 col primo e almeno 7 con il secondo?

Problema 22
Durante l'esplorazione di una caverna piena di rovine, l'avventuriero Tito ha trovato un sacchetto di pelle in cui ci sono 5 monete d'argento e 3 d'oro. Ciascuna moneta d'oro vale 10 monete d'argento. Prima di comunicare la sua scoperta ai compagni d'avventura, Tito pesca dal sacchetto due monete senza guardare. Qual è il valore medio (in monete d'argento) di quanto intasca in questo modo?

Problema 23
Giocando alla roulette francese, se si scommette un euro[13] su uno dei numeri tra 0 e 36, si ricevono 36 euro se esce il numero scelto e non si riceve nulla (perdendo quindi l'euro giocato) altrimenti. Supponendo che tutti i numeri abbiano la stessa probabilità di uscire, qual è il guadagno netto del giocatore sulla singola giocata?

---

[12] Competizione dedicata a studenti universitari che si svolge a Mosca dal 2001.

[13] Un'osservazione sulla roulette: dal punto di vista della probabilità è uno dei giochi più bilanciati dei casinò. Le case da gioco risolvono però il problema imponendo giocate minime ben superiori a un solo euro. Per saperne di più anche in questo caso è suggerito l'ottimo libro *Fate il nostro gioco* [4].

5. Altri esercizi

Questo capitolo contiene una raccolta di problemi in aggiunta a quelli visti al termine dei precedenti capitoli. La scelta dei problemi ha privilegiato la qualità rispetto alla quantità. Più ancora che per i problemi già incontrati, vale il consiglio di prendere per ciascuno il tempo necessario, dedicarci l'attenzione che merita e spendere un po' di energie prima di mollare e andare a sbirciare la soluzione. Avere percorso una parte della strada, anche se sbagliata, ci lascerà una certa familiarità con gli ingredienti del problema. Riflettere su cosa non ha funzionato nel nostro tentativo ci permetterà di capire e ricordare meglio la soluzione corretta.

Prima di passare ai problemi veri e propri, ricordiamo l'euristica consigliata da Pólya per risolvere i problemi, tratta dal già citato *Come risolvere i problemi di matematica* [19]. Come prima cosa, *leggere e capire* il problema, quali sono i dati e quali le incognite. Poi pensare a *possibili strategie* risolutive, considerando cosa ha funzionato (e cosa no) in problemi simili già incontrati, ma anche provando a formulare congetture più generali rispetto a quanto richiesto dal problema, oppure facendosi un'idea con qualche caso particolare. Una volta che abbiamo una strategia per risolvere il problema, dobbiamo *portarla a termine*. A questo punto se tutto va bene potremmo aver finito. Non bisogna però dimenticare l'ultimo passo: *rivedere* quello che si è fatto, controllando che il risultato abbia senso (se è una probabilità, è in $[0, 1]$?), ma anche osservando se la forma della soluzione non ci suggerisce altre possibili strade per arrivare al risultato. Quest'ultimo passaggio è spesso dimenticato, purtroppo, ma è uno dei più utili per scoprire cose nuove, molto più

che leggere soluzioni altrui. Se anche abbiamo risolto un problema enumerando tutti i possibili casi, riusciamo a posteriori a individuare qualche struttura che prima ci era sfuggita? Possiamo raccogliere in modo diverso i fattori in un prodotto, mettendo in evidenza una relazione cui non avevamo pensato? Se incontriamo nelle pagine seguenti o in qualunque altro allenamento un problema che abbiamo già visto (e di cui ci ricordiamo la soluzione), prendiamolo come una sfida: sapendo già dove dobbiamo arrivare, riusciamo a trovare una nuova strada?
Tra i problemi che seguono alcuni sono semplici, altri meno; alcuni hanno idee chiave nascoste, altri hanno una formulazione che cercherà di portarci fuori strada. Spero in ogni caso che almeno la buona parte, se non tutti, abbiano qualcosa da insegnare. Le soluzioni, come quelle dei problemi proposti negli altri capitoli, sono nel Capitolo 6.

Problema 24
Lanciamo due dadi a dodici facce. Qual è la probabilità che il prodotto dei numeri usciti sia multiplo di 5?

Problema 25
Dolores lancia un dado a 20 facce più volte, finché non esce un 20. Quante volte dovrà lanciarlo, in media?

Problema 26
Lanciamo una moneta equilibrata, finché non otteniamo sei teste consecutive. Quanti lanci dovremo fare in media?

Problema 27
Le finali del campionato di pallacanestro statunitense si giocano al meglio delle 7 partite. La prima delle due squadre finaliste, i Thunders, ha una probabilità $p > \frac{1}{2}$ di vincere ciascuna partita. Non sono possibili pareggi. Qual è la probabilità di vittoria dell'altra squadra finalista, i Cavaliers?

Problema 28
Nel mazzo di chiavi di Gianni ci sono 5 chiavi molto simili, di cui solo una apre la porta di casa. In un mese, Gianni rientra a casa a mezzogiorno metà dei giorni e a mezzanotte l'altra metà. Quando rientra a mezzogiorno prova le chiavi a caso una dopo l'altra, finché non riesce ad aprire la porta. Quando rientra a mezzanotte non riesce a vedere le chiavi, poiché la lampadina al suo pianerottolo non funziona ormai da anni, quindi prova le chiavi a caso, ma non riesce a tener traccia di quelle già provate, quindi succede che riprovi più volte la stessa chiave. Giovedì scorso Gianni ha aperto la porta di casa al terzo tentativo: con che probabilità era mezzanotte?

Problema 29
In un corso universitario tutti e 48 gli studenti iscritti hanno numeri di matricola consecutivi. Il professore, che ha molte paranoie sugli studenti che copiano, non vuole che a uno scritto ci siano, nella sua aula, studenti con numeri di matricola consecutivi. Se l'aula per l'esame ha 16 posti, in quanti modi può scegliere gli studenti in modo che soddisfino le sue condizioni? (Per fortuna degli studenti, le altre due professoresse in commissione sono molto più flessibili e dividono i rimanenti 32 studenti in due aule da 16 posti senza farsi troppi problemi.)

Problema 30
In un'urna[1] ci sono biglie di due colori, bianco e nero. Non sappiamo quante biglie di ciascun colore sono presenti nell'urna: sappiamo solo che la proporzione di biglie nere sul totale è p. Marguerite estrae dall'urna una biglia alla volta, ne guarda il colore e la reinserisce nell'urna. A questo punto mescola le biglie e ripete l'operazione finché non estrae una biglia nera. Se ripete più volte questo gioco, quale sarà in media la proporzione di biglie nere che avrà estratto?

Problema 31 *(Olimpiadi della Matematica, semifinale nazionale a squadre, 2014)*
Ehrenfesto, l'ingegnere degli dei, aveva degli automi che lo aiutavano a forgiare il bronzo nella sua officina. Uno di essi, ogni mattina, guardava quante once di bronzo c'erano nel crogiolo del dio; se questo numero era multiplo di 3 ne aggiungeva una, altrimenti tirava un dado a 8 facce e ne aggiungeva tante quante il numero uscito. Un secondo automa ogni sera guardava quante once di bronzo c'erano nel crogiolo, e se questo numero era multiplo di 3 usava parte del metallo per forgiare una spada del peso di 3 once. Sapendo che questa notte il crogiolo contiene 9 once di bronzo, qual è la probabilità che nei prossimi 20 giorni (20 mattine e 20 sere) vengano forgiate esattamente cinque spade?

Problema 32 *(American Mathematics Competition, 2010)*
Una rana fa 3 salti consecutivi, ciascuno di lunghezza esattamente 1 metro. La scelta della direzione per ogni salto è casuale e indipendente dai salti precedenti. Qual è la probabilità che dopo i 3 salti la rana non sia lontana più di un metro dal suo punto di partenza?

---

[1] Nei problemi di probabilità compaiono molto spesso le urne. È chiaro che non sono particolarmente interessanti, all'apparenza, dal punto di vista delle applicazioni nel mondo reale. Ma solo all'apparenza Questo tipo di formulazione permette di eliminare dal problema i dettagli irrilevanti, mantenendo solo le caratteristiche importanti e l'aspetto più intuitivo. Versioni differenti di questo stesso problema prevedono: genitori con figli suddivisi in base al genere, pesci in una vasca appartenenti a due specie diverse, premi di due tipi in un'attrazione al Luna Park e così via.

Problema 33 *(British Mathematical Olympiad, 1973)*
In una classe, l'insegnante propone alla classe un test con risposte vero/falso. La probabilità che un ragazzo risponda correttamente a una domanda è $p_{♂}$, la probabilità che una ragazza risponda correttamente è $p_{♀}$ e la probabilità che l'insegnante risponda correttamente è p. Sapendo che la probabilità che un alunno scelto a caso (ragazzo o ragazza) risponda allo stesso modo dell'insegnante è $\frac{1}{2}$, qual è il rapporto tra il numero dei ragazzi e quello delle ragazze nella classe?

Problema 34 *(Olimpiadi della Matematica, finale nazionale a squadre, 2014)*
Eolero, il dio dei 20, decide di fare un dono a Ellisseo: un otre contenente tutti i venti sfavorevoli che gli consentirà una serena navigazione verso Itôca a patto che rimanga sempre chiuso. L'otre è dotato di una combinazione a tre cifre decimali di cui Ellisseo sa che contiene due cifre 0 e una cifra 3, anche se non necessariamente in questo ordine. Spinti dalla curiosità, una notte i compagni di Ellisseo decidono di provare ad aprirlo. A turno lanciano una dracma: se esce testa sottraggono 1 a una delle tre cifre (scelta casualmente tra quelle maggiori di 0, ognuna con la stessa probabilità), mentre se esce croce sottraggono 1 a una delle tre cifre (scelta a caso tra quelle maggiori di 0, ognuna con la stessa probabilità) e contemporaneamente aggiungono 1 a una delle altre 2 (scelta a caso, ognuna con la stessa probabilità). Sapendo che la combinazione letta inizialmente sull'otre è 1-2-1, qual è la probabilità che riescano ad aprire l'otre esattamente dopo il secondo lancio di moneta?

Problema 35 *(Tratto da Ebert, 2003 [8])*
Tre prigionieri vengono convocati dal loro carceriere e fatti sedere, bendati, in una stanza. Il carceriere racconta loro quello che sta per succedere: "Tra non molto tornerò in questa stanza e metterò sulle vostre teste dei cappelli bianchi o neri, scegliendo indipendentemente per ciascuno di voi con probabilità $\frac{1}{2}$. A questo punto vi toglierò le bende e ognuno di voi potrà guardare il colore dei cappelli sulle teste degli altri due, ma non sulla propria. Non potrete comunicare tra voi e, al mio segnale, ciascuno di voi dovrà immediatamente provare a indovinare il colore del proprio cappello, oppure astenersi. Se almeno uno di voi indovinerà il proprio colore e nessuno sbaglierà il proprio colore, sarete rilasciati, altrimenti verrete riportati alle vostre celle."
Il carceriere esce per prendere i cappelli e i tre prigionieri hanno qualche minuto per accordarsi su una strategia. Possono trovarne una con cui hanno una probabilità maggiore di $\frac{1}{2}$ di essere rilasciati?

Problema 36 *(Kolmogorov Students' Contest in Probability Theory, 2004)*
Vogliamo truccare due dadi a sei facce in modo che ciascuno degli 11 possibili valori della somma delle loro facce abbia la stessa probabilità di uscire. In quanti modi possiamo farlo?

Problema 37 *(Marcus Moore)*

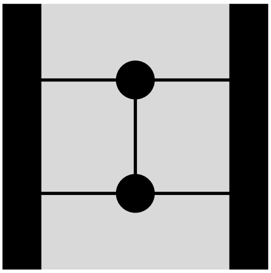

Figura 5.1: La cartina del viaggiatore

Un viaggiatore si avvicina a un fiume che, secondo la sua cartina (in Figura 5.1) ha due isole e 5 ponti. Nei giorni precedenti c'è stata una piena del fiume e ciascuno dei ponti è stato distrutto, indipendentemente dagli altri, con probabilità $\frac{1}{2}$. Con che probabilità il viaggiatore potrà attraversare il fiume?

Problema 38
Due squadre si sfidano in una gara di matematica, avendo a disposizione una scorta infinita di problemi. Ogni problema vale 10 punti. Se la differenza tra i punteggi delle due squadre è minore di 20 punti, diciamo che le squadre sono in *testa a testa*. Quando una squadra ha 20 o 30 punti in più dell'altra si dice che è in *vantaggio*, quando ne ha 40 in più ha *vinto*. Per rendere le cose più divertenti, nel momento in cui una squadra è in vantaggio deve rinunciare a uno dei suoi giocatori. A questo punto o la squadra vince e la gara finisce, oppure il divario si riduce, le squadre tornano testa a testa e il giocatore può rientrare. Le squadre sono abbastanza equilibrate, quindi la probabilità di andare in vantaggio è la stessa per ciascuna squadra. Se invece una squadra è in vantaggio, l'assenza di un giocatore si fa sentire, quindi la probabilità di vincere è un terzo della probabilità di tornare testa a testa. Qual è la probabilità che la squadra vincitrice sia stata in svantaggio[2] durante la partita?

Problema 39
Franco e Francesco, per passare il tempo, hanno inventato un gioco. Comincia Franco, lanciando una moneta: se esce croce passa il turno (e la moneta) a Francesco, mentre se esce testa continua a lanciare. Se Francesco riceve la moneta, procede nello stesso modo. Vince il primo giocatore che ottiene tre teste di fila. Se esce testa con probabilità p, qual è la probabilità che vinca Francesco?

[2] Ovviamente una squadra è in *svantaggio* nel momento in cui l'altra squadra è in vantaggio.

Problema 40 *(Olimpiadi della Matematica, gara distrettuale, 2016)*
Alberto e Barbara giocano a biliardino. Prima di iniziare, decidono che la partita finirà non appena uno dei due avrà fatto 3 gol più dell'altro. Sapendo che, per ogni pallina giocata, sia Alberto che Barbara hanno il 50% di probabilità di segnare, qual è la probabilità che la partita non termini prima del ventunesimo gol?

Problema 41
Nel famoso triello (duello a tre) del film *Il buono, il brutto e il cattivo* si affrontano il Biondo, Tuco e Sentenza. Nella versione che consideriamo qui, i tre sparano in ordine, ciascuno a un bersaglio di sua scelta. Ogni colpo che va a segno è mortale. Tuco, che spara per primo, colpisce il proprio bersaglio con probabilità 40%, il Biondo, secondo a sparare, colpisce sempre il proprio bersaglio, mentre Sentenza, il terzo a sparare, colpisce con probabilità 50%. Il triello continua finché non rimane in piedi solo uno dei tre. Tutti conoscono le abilità proprie e altrui. Qual è la massima probabilità di sopravvivere che ha Tuco e con che strategia può ottenerla?

Problema 42
A Pisa piove spesso, ma in modo poco prevedibile, e spesso soffia il vento, che distrugge gli ombrelli. La probabilità che piova in un determinato giorno è $\frac{3}{5}$, indipendentemente da quanto è successo nei giorni precedenti. Di conseguenza Giorgia la mattina scende in strada e, se sta piovendo, apre l'ombrello che porta sempre con sé. Se però soffia il vento, cosa che avviene con probabilità $\frac{1}{2}$ indipendentemente dalla pioggia, preferisce affrontare il maltempo e inzupparsi d'acqua. Qual è la probabilità che Giorgia non si bagni per quattro giorni consecutivi?

Problema 43
Lázló propone a sua figlia Judit la seguente sfida: Judit dovrà giocare tre partite a scacchi e vincerne due consecutive contro Lázló stesso e sua sorella Zsuzsanna. Judit può scegliere se giocare la prima e la terza contro Lázló e la seconda contro Zsuzsanna, oppure viceversa la prima e la terza contro la sorella e la seconda contro il padre. Se la probabilità che Judit ha di vincere con la sorella è $\frac{3}{5}$ e quella di vincere con il padre è $\frac{7}{9}$, quale delle due opzioni le conviene scegliere?

Problema 44 *(Kolmogorov Students' Contest in Probability Theory, 2004)*
Per un volo low-cost sono stati venduti biglietti numerati per tutti e 121 i posti. I primi 120 passeggeri salgono a bordo e si siedono dove capita. L'ultimo passeggero a salire, invece, vuole assolutamente sedersi al proprio posto e, se lo trova occupato, chiede a chi è seduto di lasciarglielo. Se qualcuno si alza per lasciare il posto a chi ne ha diritto, una volta in piedi si dirigerà verso il posto segnato sul proprio biglietto, eventualmente chiedendo a chi ci lo occupa di alzarsi e così via. In media, quanti passeggeri si devono alzare e cambiare posto?

Problema 45 *(Olimpiadi della Matematica, gara distrettuale, 2017)*
Un'urna contiene 8 palline, sulle quali sono scritti i numeri da 1 a 8. Federica pesca due palline di seguito, cancella il numero scritto sulla prima e lo sostituisce con il suo doppio, cancella il numero sulla seconda pallina e lo sostituisce con il quadruplo di esso. Reinserisce quindi le due palline nell'urna (ad esempio, se Federica ha pescato le palline 3 e 7 in quest'ordine, reinserirà nell'urna due palline con i numeri 6 e 28). Infine, estrae nuovamente una pallina: qual è la probabilità che la pallina estratta abbia il numero 8?

Problema 46 *(Olimpiadi della Matematica, semifinale nazionale a squadre, 2016)*
Dopo aver superato il primo ufficio Bogon, Jacob deve visitarne altri trenta, numerati da 2 a 31. Per ogni k, l'ufficio recante il numero k ha una probabilità $\frac{1}{k^2}$, indipendentemente dagli altri, di rifiutare l'ordine presidenziale. Qual è la probabilità che nessuno dei trenta uffici rimanenti rifiuti il modulo?

Problema 47 *(Tratto da Justicz, Scheinerman e Winkler, 1990 [12])*
Daniele sceglie due punti a caso nell'intervallo $[0, 1]$, individuando così un segmento. Prende poi altri due punti e individua un secondo intervallo. Continua in questo modo per 2020 volte. Qual è la probabilità che tra gli intervalli costruiti da Daniele ce ne sia uno che interseca tutti gli altri?

Problema 48 *(Olimpiadi della Matematica, gara a squadre locale, 2004)*
Per riprendersi dalle fatiche di questa gara, i 7 componenti di una squadra hanno organizzato una spaghettata aglio, olio e peperoncino. Le dosi di pasta che vengono servite sono uguali per tutti, tranne che per il capitano, che ha diritto a una razione doppia, e il consegnatore, che avendo corso tanto ha diritto a una razione tripla. Sapendo che nel sugo sono stati messi 4 peperoncini interi e che i piatti sono stati fatti a caso dopo aver mescolato bene la pasta con il sugo, determinare la probabilità che almeno un commensale si ritrovi nel piatto più di un peperoncino.

Problema 49 *(Olimpiadi della Matematica, gara distrettuale, 2014)*
Cinque amici devono scendere da una seggiovia a cinque posti e possono farlo andando in tre direzioni differenti: a sinistra, dritto oppure a destra. Scendendo da una seggiovia è facile scontrarsi con i propri compagni di risalita. Per esempio: se io decido di andare dritto e qualcuno alla mia sinistra di andare a destra, ci scontriamo; lo stesso accade se io decido di andare a destra e qualcuno alla mia destra va dritto (o a sinistra); se invece qualcuno va nella mia stessa direzione non ci scontriamo; e così via. Se ciascuno dei cinque amici sceglie a caso dove andare, con probabilità $\frac{1}{3}$ per ciascuna direzione, qual è la probabilità che non ci siano scontri?

Problema 50 *(Olimpiadi della Matematica, finale nazionale gara a squadre, 2016)* Luke Randomwalker sta cercando un posto sicuro dove nascondersi, e per farlo viaggia in incognito a bordo di navi mercantili. Le navi scelte da Luke seguono rotte che collegano tra loro n pianeti. Tra di essi vi sono: Coruscantor, dove Luke si trova all'inizio; Banahch-Torsk, un ameno pianeta doppio dove Luke si ferma immediatamente (se ci passa); Taodana, sede del covo di Maz Karamata, dove ci sono così tante spie del Prim'Ordine che è certo che qualcuno lo riconosca e lo uccida. Da ogni pianeta (esclusi Taodana e Banahch-Torsk) partono rotte unidirezionali verso esattamente altri due pianeti, e da al massimo uno di questi due esiste una successione di rotte che consente di tornare al pianeta appena lasciato. Ogni volta che lascia un pianeta, Luke sceglie a caso tra le due rotte possibili (con uguale probabilità) e si ferma solamente se arriva su Banahch-Torsk oppure se viene ucciso su Taodana. Sapendo che la probabilità che arrivi sano e salvo su Banahch-Torsk è $\frac{1}{2016}$, quanto vale n come minimo?

6. Soluzioni

In questo capitolo troviamo le soluzioni di tutti i problemi proposti nel testo, raccolti in sezioni corrispondenti ai capitoli nei quali sono stati enunciati.

6.1 Combinatoria – Riscaldamento

Soluzione al Problema 1

Per la soluzione possiamo usare la medesima idea vista per l'Esempio 1.3.3. Quello che cambia (oltre ai numeri e al fatto che l'ordine dei colori sia fissato) è che nella piazza Anfiteatro di Lucca non è univocamente identificato il primo edificio, quindi dobbiamo contare anche i diversi modi in cui possiamo iniziare a contare. La risposta è quindi

$$26 \cdot \binom{25}{7} = \frac{26!}{18! \cdot 7!} = 12\,498\,200.$$

Possiamo farlo anche in un modo diverso, che ci permette poi di lasciare facilmente cadere l'ipotesi dell'ordine fissato per i colori. Possiamo infatti contare quanti sono i modi diversi di scegliere le prime case di ogni colore, $\binom{26}{8}$, e poi scegliere da quale di queste 8 case partire, per un totale di

$$\binom{26}{8} \cdot 8 = \frac{26! \cdot 8}{18! \cdot 8!} = \frac{26!}{18! \cdot 7!}.$$

Se ora lasciamo cadere l'ipotesi che i colori siano ordinati, abbiamo ancora $\binom{26}{8}$ modi di scegliere le prime case di un certo colore, che però ora moltiplichiamo per gli 8! modi di ordinare gli 8 colori. □

Soluzione al Problema 2

È un altro problema imparentato con l'Esempio 1.3.3. Pensiamo a 21 come a una fila di 21 noci di cocco. Scrivere 21 come somma di 5 interi positivi significa dividere la fila in 5 parti non vuote, cioè mettere dei bastoncini tra le noci di cocco. Abbiamo 4 bastoncini da piazzare e 20 posti in cui metterli, quindi la risposta è $\binom{20}{4}$. □

Soluzione al Problema 3

Qui la difficoltà aggiuntiva (o se vogliamo la differenza) rispetto agli altri problemi simili che abbiamo visto sta nel fatto che ammettiamo addendi nulli (o scatole vuote, se lo vogliamo vedere figurativamente). Come facciamo a ricondurci al caso che conosciamo?

Scriviamo un numero n come somma di 5 interi positivi, cosa che possiamo fare in $\binom{n-1}{4}$ modi, poi togliamo uno da ciascuno degli addendi. In questo modo abbiamo scritto $n-5$ come somma di 5 numeri naturali (eventualmente nulli). A noi interessa scrivere 21 in questo modo, quindi $n - 5 = 21$ e i modi sono $\binom{25}{4}$. □

Soluzione al Problema 4

Vediamole in ordine.

i. In questo caso ci basta scrivere la definizione di coefficiente binomiale e sfruttare la commutatività del prodotto,
$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!} = \frac{n!}{(n-k)! \cdot [n-(n-k)]!} = \binom{n}{n-k}.$$
Oltre a svolgere il conto, avremmo anche potuto osservare che selezionare k elementi tra n è equivalente a scegliere $n - k$ elementi da scartare tra n.

ii. Questo è il caso limite del precedente, ma merita qualche parola in più: in quanti modi possiamo scegliere n oggetti tra n disponibili? Solamente in 1 modo. Viceversa, potremmo discutere sul fatto che ci sia solo un modo di scegliere 0 oggetti tra n (non scegliere alcun oggetto), ma è quello che esce sostituendo 0 a k nella definizione data, poiché $0! = 1$, ed è anche consistente con la proprietà vista sopra.

iii. Questa proprietà è molto interessante: osservando il triangolo di Tartaglia, notiamo che la somma sulla riga n-esima è uguale a 2^n, cioè che sommando tutti i coefficienti binomiali che hanno n nella posizione superiore otteniamo 2^n. Per capire come mai, ci appoggiamo al binomio di Newton, ossia allo sviluppo di un binomio elevato a potenza n. Sappiamo che
$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} \cdot a^k \cdot b^{(n-k)}.$$

Possiamo vederlo in modo combinatorio come segue: quando andiamo a fare l'elevamento a potenza, stiamo svolgendo il prodotto tra n fattori, ciascuno dei quali è una copia della somma $a + b$. Da ogni copia possiamo prendere una a o una b. In quanti modi possiamo avere k fattori a e $n - k$ fattori b?

Dobbiamo solamente scegliere in quali k delle n copie di $a + b$ peschiamo le a, cosa che possiamo fare in $\binom{n}{k}$ modi. Tornando al quesito iniziale, possiamo a questo punto prendere $a = b = 1$ e abbiamo nel secondo membro la somma cercata, con il primo membro che diventa uguale a 2^n.

iv. Questa è la proprietà alla base della costruzione del triangolo di Tartaglia. Immaginiamo di avere $n + 1$ oggetti in fila e di sceglierne $k + 1$ tra di essi.

Consideriamo due casi possibili (disgiunti): tutti quelli in cui prendiamo l'ultimo oggetto e tutti quelli in cui non lo prendiamo. Essi sono, rispettivamente, $\binom{n}{k}$, perché dobbiamo scegliere altri k oggetti tra gli n rimanenti, e $\binom{n}{k+1}$, perché avendo escluso l'ultimo oggetto, dobbiamo scegliere tutti i $k + 1$ oggetti tra i rimanenti n.

Se questa dimostrazione non ci piace, possiamo sempre usare le definizioni:

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

□

Soluzione al Problema 5

I rapinatori si muovono solamente lungo le strade e devono necessariamente fare 6 isolati verso Nord e 12 isolati verso Ovest. Vogliono anche fare meno strada possibile, quindi non si muoveranno mai verso Est o verso Sud.

Vogliamo allora contare i percorsi costituiti esattamente da 6 spostamenti verso Nord e da 12 verso Ovest. Possiamo vederli come parole di $6 + 12 = 18$ lettere, di cui 6 "N" e 12 "O".

Ci siamo ricondotti quindi al conteggio degli anagrammi di

"NNNNNNOOOOOOOOOOOO".

In tutto i percorsi di lunghezza minima sono $\binom{18}{6} = 18\,564$. □

Soluzione al Problema 6

I casi totali sono tanti quanti i percorsi di lunghezza minima che vanno al covo, già contati nel Problema 5: $\binom{18}{6}$. Quanti sono quelli che portano alla cattura? Tutti quelli che passano dal primo posto di blocco, più tutti quelli che passano dal secondo posto di blocco meno quelli che passano da entrambi[1], usando il principio di inclusione-esclusione. Possiamo contare il numero dei possibili percorsi esattamente come prima, eventualmente scomponendo il percorso in sotto-percorsi.

Contiamo allora quanti sono i percorsi che passano dal primo posto di blocco: $\binom{4}{1} \cdot 1 \cdot \binom{13}{5}$ perché dobbiamo raggiungere il nodo $(4, 2)$, passare dal posto di blocco e poi andare dal nodo $(5, 2)$ al covo in $(13, 7)$. In modo simile contiamo quanti passano dal secondo, $\binom{13}{4} \cdot 1 \cdot \binom{4}{1}$, e quanti da entrambi, $\binom{4}{1} \cdot 1 \cdot \binom{8}{3} \cdot 1 \cdot \binom{4}{1}$. In tutto i percorsi di lunghezza minima che portano alla cattura sono

$$4 \cdot \left[\binom{13}{4} + \binom{13}{5} \right] - 4 \cdot \left[\binom{8}{3} \cdot 4 \right] = 4 \cdot \left[\binom{14}{5} - 4 \cdot \binom{8}{3} \right] = 7112.$$

La probabilità che vengano catturati è allora $\frac{7112}{18\,564} = \frac{254}{663} \approx 38{,}3\%$, facendo il rapporto con il numero di percorsi totali di lunghezza minima. ☐

Soluzione al Problema 7

Indichiamo con B_n il numero di partizioni distinte di un insieme di cardinalità n. Sappiamo che $B_0 = 1$, perché l'insieme vuoto ha una sola partizione possibile. Se passiamo al caso $n = 1$, abbiamo nuovamente $B_1 = 1$. Possiamo provare a continuare ancora per qualche passo, ma è meglio provare a caratterizzare i B_n ricorsivamente.

Supponiamo di avere un insieme E di n elementi numerati da 1 a n e di andare a prendere, in una sua partizione, l'insieme $S \subseteq E$ a cui appartiene l'elemento n. A questo punto ci rimangono un certo numero di insiemi nella partizione che, tutti assieme, hanno un numero k di elementi (di E), con $0 \leqslant k \leqslant n - 1$. Per ciascuno di questi valori di k, possiamo scegliere questi k elementi in $\binom{n-1}{k}$ modi, dal momento che abbiamo già usato l'elemento n, e per ciascuna scelta abbiamo B_k partizioni possibili dei k elementi rimasti.

Abbiamo allora che

$$B_n = \sum_{k=0}^{n-1} \binom{n-1}{k} B_k, \qquad n \geqslant 1,$$

i cui primi valori sono $B_0 = B_1 = 1$, $B_2 = 2$, $B_3 = 5$, $B_4 = 15$, $B_5 = 52$, $B_6 = 203$, $B_7 = 877$, $B_8 = 4140$, $B_9 = 21\,147$, $B_{10} = 115\,975$. ☐

---

[1] Osserviamo che questo dipende dalla particolare disposizione scelta dei posti di blocco: ci sono posizionamenti della polizia tali per cui non esistono percorsi di lunghezza minima che passano da entrambi.

Soluzione al Problema 8

Innanzitutto se vogliamo che esista una funzione iniettiva da A a B è necessario che B abbia almeno tanti elementi quanti A, cioè $\#A \leqslant \#B$. A questo punto osserviamo che possiamo scegliere l'immagine del primo elemento di A in $\#B$ modi, quella del secondo in $\#B-1$ modi e così via, fino all'immagine dell'ultimo elemento di A, che potrà essere scelta in $\#B-(\#A-1)$ modi. Quindi abbiamo

$$\#B \cdot (\#B-1) \cdot \dots \cdot (\#B-(\#A-1)) = \frac{(\#B)!}{(\#B-\#A)!}.$$

Possiamo vedere questo risultato anche in un altro modo: ci interessano i riordinamenti degli elementi di B, trascurando però tutti gli elementi oltre quello in posizione $\#A$. □

Soluzione al Problema 9

Siccome stiamo chiedendo che le funzioni siano suriettive, devono esserci in A almeno tanti elementi quanti ce ne sono in B, cioè $\#A \geqslant \#B$. Sappiamo[2] che le funzioni da A a B sono $\#B^{\#A}$, quindi questo impone un limite superiore al numero di funzioni suriettive: tra tutte le funzioni, infatti, ci sono ad esempio quelle che hanno nell'immagine tutto B tranne un unico elemento. Vogliamo quindi toglierle dal conteggio di tutte le funzioni. Quante sono le funzioni che escludono un solo elemento? Possiamo scegliere l'elemento escluso in $\#B$ modi e, per ciascuna scelta, ci sono $(\#B-1)^{\#A}$ funzioni, quindi dobbiamo sottrarre $\#B \cdot (\#B-1)^{\#A}$ a $\#B^{\#A}$. A questo punto dobbiamo considerare le funzioni che escludono due elementi di B dall'immagine, infatti le abbiamo sottratte due volte, una volta per ciascuno dei due elementi. Insomma, dobbiamo continuare con il principio di inclusione-esclusione. Vediamo esplicitamente quante ne dobbiamo aggiungere: dobbiamo contare in quanti modi possiamo scegliere due elementi tra $\#B$, cioè $\binom{\#B}{2}$ e, per ciascuno di questi modi, quante sono le funzioni da A all'insieme B tranne questi due elementi, cioè $(\#B-2)^{\#A}$. Ricordiamo anche che queste vanno aggiunte, perché sottratte due volte, poi sarà il turno di quelle con tre elementi esclusi, che sono state sottratte tre volte, ma anche ri-aggiunte tre volte e devono quindi essere sottratte di nuovo. Mettendo tutto assieme abbiamo l'uguaglianza

$$\#B^{\#A} - \#B \cdot (\#B-1)^{\#A} + \binom{\#B}{2} \cdot (\#B-2)^{\#A} - \dots = \\ = \sum_{i=0}^{\#B} (-1)^i \cdot \binom{\#B}{i} \cdot (\#B-i)^{\#A}. \quad \square$$

[2]Per approfondimenti consultare l'Appendice.

6.2 La strada di Kolmogorov

Soluzione al Problema 10

I multipli di 3 che compaiono sul dado sono 3, 6, 9 e 12, mentre di multipli di 7 c'è solo il 7. Non ci sono numeri che siano multipli di 3 e 7, quindi non ne contiamo alcuno due volte. Di questi numeri, 3 sono dispari e 2 sono pari, quindi la probabilità cercata è

$$3 \cdot \frac{1}{9} + 2 \cdot \frac{1}{18} = \frac{4}{9}. \qquad \square$$

Soluzione al Problema 11

Abbiamo una bella collezione di eventi disgiunti, in numero infinito, ma numerabile: tutti i singoletti dei numeri dispari e i numeri pari. La probabilità che esca un numero dispari è la somma della serie geometrica[3], $\sum_{k=1}^{+\infty} 3^{-k} = \frac{1}{2}$. La probabilità che esca un numero pari è allora $1 - \frac{1}{2} = \frac{1}{2}$. Chiamiamo E l'insieme dei numeri pari positivi, allora $\frac{1}{2} = P(\{0\}) + P(E) = 2 \cdot P(E)$, da cui $P(\{0\}) = P(E) = \frac{1}{4}$. $\square$

Soluzione al Problema 12

Possiamo risolvere questo esercizio andando a considerare tutti i casi favorevoli, ossia 3 biglie bianche e 1 nera, 2 bianche e 2 nere, oppure 3 nere e 1 bianca. Tuttavia dovremmo tenere conto dei diversi modi di ottenere le varie combinazioni, oltre alle rispettive probabilità. Se invece calcoliamo la probabilità che le cose non vadano come vogliamo, abbiamo meno casi (e più semplici) da considerare.

Le biglie possono essere tutte del medesimo colore se sono tutte bianche, oppure tutte nere. La probabilità che siano tutte bianche è $\frac{16}{27} \cdot \frac{15}{26} \cdot \frac{14}{25} \cdot \frac{13}{24}$, che siano tutte nere è $\frac{11}{27} \cdot \frac{10}{26} \cdot \frac{9}{25} \cdot \frac{8}{24}$. Sommando le probabilità di questi due casi disgiunti abbiamo $\frac{51\,600}{421\,200} = \frac{43}{351}$. Dal momento che a noi interessa la probabilità dell'evento complementare, per avere la risposta non ci resta che sottrarre questo numero da 1:

$$1 - \frac{43}{351} = \frac{308}{351}. \qquad \square$$

Soluzione al Problema 13

Numeriamo le invitate da 1 a n e mettiamoci nei panni dell'invitata i. La probabilità che costei riceva il proprio pacchetto è $\frac{1}{n}$. Possiamo chiamare R_i l'evento "invitata i-esima riceve il proprio regalo". Quello che vogliamo calcolare è allora la probabilità dell'unione di questi R_i al variare di i tra 1 e n: $P\left(\bigcup_{i=1}^{n} R_i\right)$. Si tratta però di eventi non disgiunti. Infatti più invitate potrebbero ricevere il proprio pacchetto. Dobbiamo quindi calcolare le probabilità delle intersezioni e usare il principio di inclusione–esclusione.

[3] Un breve richiamo sulla serie geometrica è in Appendice.

Se consideriamo due invitate i e j, la probabilità che entrambe ricevano il proprio regalo è $\frac{1}{n} \cdot \frac{1}{n-1}$, corrispondente all'evento $R_i \cap R_j$. Osserviamo che questa probabilità dipende, come già $P(R_i)$, solamente da quante invitate sono coinvolte e non dalle loro identità. In generale, se guardiamo l'evento in cui k invitate ricevono il proprio regalo, esso avrà probabilità $\frac{1}{n} \cdot \frac{1}{n-1} \cdot \dots \cdot \frac{1}{n-(k-1)} = \frac{(n-k)!}{n!}$. Per ogni k, però, abbiamo $\binom{n}{k}$ modi di scegliere k invitate tra le partecipanti. Allora la probabilità cercata, ossia la probabilità dell'unione, sarà data da

$$P\left(\bigcup_{i=1}^{n} R_i\right) = \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} \frac{(n-k)!}{n!} = \sum_{k=1}^{n} (-1)^{k+1} \frac{1}{k!}.$$

Il problema può dirsi concluso: per valori assegnati di n basta andare a calcolarsi questa somma (finita) a segni alterni. Tuttavia, chi avesse già incontrato la funzione esponenziale espressa come serie potrebbe riconoscere l'ultimo membro come l'approssimazione al grado n-esimo di $1 - e^{-1}$ (e saper dare quindi più facilmente un'approssimazione del valore cercato).
Abbiamo anche la risposta al problema complementare: "Qual è la probabilità che nessuna abbia davanti a sé il proprio pacchetto?". Al crescere del numero delle invitate n, questa quantità tende a $e^{-1} \approx 37\%$. □

Soluzione al Problema 14
Questo fatto non è vero in generale. Consideriamo ad esempio $\Omega = \{0, 1, 2\}$ con le due algebre $\mathcal{A} = \{\varnothing, \{0\}, \{1, 2\}, \Omega\}$ e $\mathcal{B} = \{\varnothing, \{0, 1\}, \{2\}, \Omega\}$. L'unione $\mathcal{A} \cup \mathcal{B} = \{\varnothing, \{0\}, \{0, 1\}, \{1, 2\}, \{2\}, \Omega\}$ ha cardinalità 6 (quindi sappiamo che ha qualche problema) e in particolare non contiene $\{0, 2\}$, unione di $\{0\}$ e $\{2\}$ e il suo complementare $\{1\}$. □

Soluzione al Problema 15
In generale non è vero: ci basta considerare Ω infinito e osserviamo immediatamente che $\mathcal{A}$ non può contenere Ω stesso né essere chiusa rispetto al complementare. Una possibilità per costruire un'algebra che contenga tutti gli eventi finiti è quella di prendere la famiglia di tutti gli eventi finiti o a complementare finito. Se vogliamo è in un certo senso l'algebra generata dagli eventi finiti. □

6.3 Costruire spazi di probabilità

Soluzione al Problema 16
La prima osservazione da fare è che siccome il poligono ha un numero dispari di vertici, allora non è possibile che il triangolo ottenuto sia rettangolo: o è acutangolo, o è ottusangolo. La seconda è che il circocentro del poligono coincide con quello

del triangolo. La terza è che il circocentro di un triangolo è interno al triangolo solamente se il triangolo è acutangolo.
Come possiamo caratterizzare un triangolo acutangolo o, equivalentemente, ottusangolo? Il discrimine tra i due è dato dal triangolo rettangolo, che non possiamo avere, ma che ha la caratteristica che due dei suoi vertici giacciono alle estremità del medesimo diametro e che tutti e tre i vertici appartengono dunque alla medesima semicirconferenza (attenzione, non basta che siano nella medesima semicirconferenza, come vedremo tra un attimo, ma occorre che due siano sul diametro che la delimita).
Cosa succede se lasciamo cadere l'ipotesi che due punti giacciano alle estremità di un diametro, ma conserviamo quella che i tre punti siano nella medesima semicirconferenza? Guardiamo il più lungo arco di circonferenza delimitato da due di questi punti e che non contenga il terzo: esso è strettamente più lungo di una semicirconferenza, quindi ogni angolo alla circonferenza che insiste su questo arco ha ampiezza maggiore di un angolo retto. In particolare l'angolo individuato dai due estremi di quest'arco e avente vertice nel terzo punto è ottuso. Abbiamo allora una caratterizzazione dei triangoli ottusangoli.

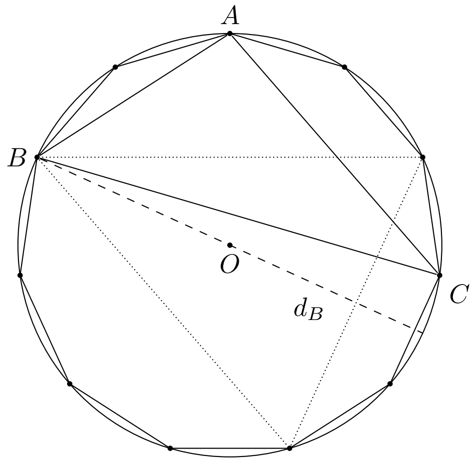

Figura 6.1: Un triangolo ottusangolo e uno acutangolo in un 11-gono

Per concludere il problema dobbiamo contare in quanti modi possiamo prendere tre vertici tra $2n+1$, cioè in quanti modi possiamo costruire un triangolo, e in quanti modi possiamo prenderli in modo che il triangolo sia ottusangolo. La risposta alla prima domanda è $\binom{2n+1}{3}$. Per rispondere alla seconda, invece, usiamo la caratterizzazione appena vista, con una piccola accortezza. Chiamiamo A il vertice in cui abbiamo l'angolo ottuso, B il vertice successivo, in senso antiorario, e C il terzo vertice. Cominciamo scegliendo il vertice B, cosa che possiamo fare in $2n+1$ modi. Tracciamo il diametro d_B passante per B: tutti e tre i punti devono stare necessariamente dallo stesso lato (quello sinistro, percorrendo il diametro partendo da B) rispetto a d_B. Osserviamo che metà dei punti rimanenti giacciono in questa semicirconferenza, cioè n, e tra questi dobbiamo sceglierne 2, in $\binom{n}{2}$ modi possibili. Riassumendo, la probabilità di ottenere un triangolo ottusangolo è

$$\frac{(2n+1)\binom{n}{2}}{\binom{2n+1}{3}} = \frac{(2n+1)n(n-1)}{2} \cdot \frac{3!}{(2n-1)2n(2n+1)} = \frac{3n-3}{2(2n-1)},$$

quindi la probabilità cercata è il suo complementare,

$$1 - \frac{3n-3}{2(2n-1)} = \frac{n+1}{4n-2}.$$

Se avessimo voluto considerare anche il caso con $2n+2$ vertici, con $n \geqslant 1$, ossia un numero di vertici pari e maggiore di 3, l'argomento sarebbe stato molto simile. Avremmo però dovuto tener traccia anche dei triangoli rettangoli, per cui il circocentro cade sull'ipotenusa (e quindi, in senso stretto, non all'interno del triangolo). $\square$

Soluzione al Problema 17

Questo non è un esercizio di probabilità, ma di teoria degli insiemi, ma questo non lo rende meno interessante. Quanto visto sulla generazione degli intervalli aperti a partire da quelli chiusi ci suggerisce di usare unioni con estremi "perturbati" da un termine $\frac{1}{n}$. Limitiamoci al caso $[0,1]$ (il caso $\mathbb{R}$ è del tutto analogo, a patto di prendere il complementare rispetto all'insieme precedente).

Procediamo a ritroso: partiamo da $[a,b]$, che è il nostro obiettivo. Come prima cosa passiamo al complementare, $[a,b]^c = [0,a) \cup (b,1]$. Di questi due, il secondo è già della forma che ci interessa, quindi ci aspettiamo di non doverlo modificare ulteriormente. Per quanto riguarda $[0,a)$, abbiamo già visto che possiamo scriverlo come unione numerabile nel modo seguente: $[0,a) = \bigcup_{n\in\mathbb{N}} \left[0, a-\frac{1}{n}\right]$. Questo apparentemente non ci va bene, perché è un'unione di intervalli chiusi, ma il complementare di ciascuno di questi intervalli è del tipo ammesso, quindi lo saranno anche loro stessi e la loro unione. Non solo, possiamo in realtà usare l'associatività dell'unione per individuare una soluzione:

$$\begin{aligned}[a,b] = ([0,a) \cup (b,1])^c &= \left(\bigcup_{n\in\mathbb{N}} \left[0, a-\tfrac{1}{n}\right] \cup (b,1] \right)^c \\ &= \left(\bigcup_{n\in\mathbb{N}} \left(\left[0, a-\tfrac{1}{n}\right] \cup (b,1] \right) \right)^c = \left(\bigcup_{n\in\mathbb{N}} \left(a-\tfrac{1}{n}, b\right]^c \right)^c. \end{aligned}$$ $\square$

Soluzione al Problema 18

Risolviamo il problema più in generale, per n naturale positivo e k naturale minore o uguale di n: se Francesco lancia n monete, con che probabilità ottiene k teste (e quindi $n-k$ croci)? Il nostro spazio di probabilità Ω è lo spazio delle n-uple di 0 e 1 (o di T e C). La tribù è quella generata dai rettangoli e la probabilità è, componente per componente, quella vista nell'Esempio 3.3.4. Qual è l'evento del quale vogliamo calcolare la probabilità? È quello che contiene tutti gli esiti (cioè le stringhe di lunghezza n) in cui esattamente k componenti sono T e i restanti $n-k$ sono C. Queste stringhe sono $\binom{n}{k}$. Per ciascuna stringa dobbiamo calcolare la probabilità, per poi andarle a sommare, dal momento che i singoletti degli esiti sono tutti eventi disgiunti. Per fortuna, possiamo osservare che ciascuna di queste stringhe ha la medesima probabilità (poiché il prodotto è commutativo): ogni T contribuisce con un fattore p, mentre ogni C dà un fattore $1-p$, per una probabilità di ciascuna stringa uguale a $p^k(1-p)^{n-k}$. La probabilità di avere esattamente k teste con n monete è allora $\binom{n}{k}p^k(1-p)^{n-k}$. Questo tipo di esperimento prende il nome di modello *binomiale*. Nell'esempio specifico, ci basta sostituire $n = 8$ e $k = 8 - 5 = 3$, ottenendo così $56p^3(1-p)^5$. □

6.4 Sapendo che...

Soluzione al Problema 19

La risposta è $\frac{2}{3}$. Un modo di vederlo è il seguente: abbiamo tre situazioni possibili, o abbiamo pescato la sola moneta d'oro nel secondo contenitore, oppure abbiamo pescato una delle due monete d'oro dal primo. I casi sono tre perché quest'ultimo va contato doppio, dal momento che possiamo prendere l'una o l'altra moneta. A questo punto i casi favorevoli sono 2 e abbiamo concluso.

Le prime volte che si vedono problemi di questo tipo, però, è normale (ed è bene!) non fidarsi troppo della propria intuizione e, se il tempo non è un problema, mettersi a scrivere per bene il tutto, come faremo ora. Abbiamo i seguenti eventi: C, "il contenitore scelto ha due monete d'oro", O, "la moneta estratta è d'oro", assieme ai loro complementari. Il problema ci chiede $P(C \mid O)$, che possiamo calcolare col Teorema di Bayes:

$$P(C \mid O) = \frac{P(O \mid C) \cdot P(C)}{P(O)} = \frac{P(O \mid C) \cdot P(C)}{P(O \mid C) \cdot P(C) + P(O \mid C^c) \cdot P(C^c)}.$$

Per concludere non ci resta che sostituire i valori: $P(C) = \frac{1}{2}$, $P(O \mid C) = 1$ e $P(O \mid C^c) = \frac{1}{2}$, ottenendo

$$P(C \mid O) = \frac{1 \cdot \frac{1}{2}}{\frac{1}{2} + \frac{1}{4}} = \frac{2}{3}.$$ □

Soluzione al Problema 20

La cosa ci può apparire ragionevole: dalla formula di fattorizzazione abbiamo

$$P(F) = \sum_{i=1}^{n} P(F \mid E_i)P(E_i) \geq \sum_{i=1}^{n} P(G \mid E_i)P(E_i) = P(G).$$

Nella prima somma abbiamo dei prodotti in cui il secondo fattore compare identico nell'addendo corrispondente nella seconda somma, mentre il primo fattore è sempre maggiore o uguale del termine corrispondente nella seconda somma. Ma la formula di fattorizzazione richiede non solo che l'unione degli eventi E_i ricopra Ω, ma anche che sia un'unione disgiunta, ipotesi che non è data in questo problema. Vediamo se questo può essere il punto di svolta per un controesempio.

Prendiamo $\Omega = \{1, 2, 3\}$ con la probabilità P definita uguale a $\frac{1}{3}$ sui singoletti. Definiamo, con $n = 2$, $E_1 = \{1, 2\}$ ed $E_2 = \{1, 3\}$, e anche $F = \{1\}$ e $G = \{2, 3\}$. A questo punto abbiamo:

$$P(F \mid E_1) = P(\{1\}) = \frac{1}{3} = P(\{2\}) = P(G \mid E_1)$$
$$P(F \mid E_2) = P(\{1\}) = \frac{1}{3} = P(\{3\}) = P(G \mid E_2),$$

quindi le condizioni sono soddisfatte. Tuttavia, allo stesso tempo,

$$P(F) = P(\{1\}) = \frac{1}{3} < \frac{2}{3} = P(\{2, 3\}) = P(G),$$

che contraddice $P(F) \geqslant P(G)$. □

Soluzione al Problema 21

Chiamiamo S il successo con il primo dado (quindi almeno 15) e T il successo con il secondo (quindi almeno 7). Vogliamo calcolare la probabilità dell'evento $P(S \cap T)$, cosa che possiamo fare sfruttando l'indipendenza tra i due dadi. Otteniamo

$$P(S \cap T) = P(S) \cdot P(T) = \frac{6}{20} \cdot \frac{2}{8} = \frac{3}{40}.$$

Avremmo anche potuto prendere una strada diversa, osservando che abbiamo $P(S \cap T) = 1 - P(S^c \cup T^c)$. In questo caso vogliamo calcolare $P(S^c \cup T^c)$, con qualche attenzione poiché l'unione non è disgiunta: abbiamo un'intersezione da non dimenticare. Possiamo calcolarne la probabilità sfruttando l'indipendenza tra i due dadi:

$$P(S^c \cup T^c) = P(S^c) + P(T^c) - P(S^c \cap T^c) = \frac{14}{20} + \frac{6}{8} - \frac{21}{40} = \frac{37}{40}.$$

Questo ci porta, per fortuna, allo stesso risultato. □

Soluzione al Problema 22

Dobbiamo considerare 3 casi: entrambe le monete sono d'argento, entrambe le monete sono d'oro, oppure le monete sono una d'argento e una d'oro. Nel primo caso il bottino è 2, nel secondo caso è 20, nel terzo è 11. Per ciascun caso calcoliamo la probabilità:

$$P(\{A, A\}) = \frac{\binom{5}{2}}{\binom{8}{2}} = \frac{10}{28}, \quad P(\{O, O\}) = \frac{\binom{3}{2}}{\binom{8}{2}} = \frac{3}{28}, \quad P(\{A, O\}) = \frac{\binom{5}{1}\binom{3}{1}}{\binom{8}{2}} = \frac{15}{28}.$$

Ora dobbiamo pesare ciascun possibile bottino per la corrispondente probabilità e sommare tutti i contributi:

$$2 \cdot \frac{5}{14} + 20 \cdot \frac{3}{28} + 11 \cdot \frac{15}{28} = \frac{245}{28} = \frac{35}{4}. \qquad \square$$

Soluzione al Problema 23

Se esce il numero su cui ha scommesso, cosa che succede con probabilità $\frac{1}{37}$, il giocatore vince 36 euro; a questi dobbiamo però sottrarre l'euro giocato. Nei rimanenti casi, invece, perde l'euro giocato, con probabilità $\frac{36}{37}$. Facendo la media pesata dei guadagni abbiamo allora $(36 - 1) \cdot \frac{1}{37} + (-1) \cdot \frac{36}{37} = -\frac{1}{37} \approx -0{,}027$. In sostanza, per ogni euro giocato il giocatore perde in media quasi 3 centesimi. $\square$

6.5 Altri esercizi

Soluzione al Problema 24

Possiamo risolvere questo problema andando a vedere quali sono le possibili combinazioni dei risultati dei due dadi che ci restituiscono un risultato utile. Oppure possiamo chiederci quand'è che le cose vanno male. Per avere un multiplo di 5, infatti, abbiamo bisogno che almeno uno dei due dadi dia come risultato 5 o 10. La probabilità che non esca alcuno dei due su un dado è $\frac{10}{12} = \frac{5}{6}$ e che ciò accada per entrambi i dadi è $\frac{25}{36}$. Pertanto, la probabilità che il prodotto sia multiplo di 5 è il complementare: $\frac{11}{36}$. $\square$

Soluzione al Problema 25

La probabilità che 20 esca al primo lancio è $\frac{1}{20}$, che esca (per la prima volta) al secondo lancio è $\frac{19}{20} \cdot \frac{1}{20}$, al terzo $(\frac{19}{20})^2 \cdot \frac{1}{20}$ e così via; al k-esimo lancio è $(\frac{19}{20})^{k-1} \cdot \frac{1}{20}$. Quindi il numero medio di lanci è

$$L = \frac{1}{20} + 2 \cdot \frac{19}{20} \cdot \frac{1}{20} + 3 \cdot \left(\frac{19}{20}\right)^2 \cdot \frac{1}{20} + \cdots = \sum_{k=1}^{+\infty} k \cdot \left(\frac{19}{20}\right)^{k-1} \cdot \frac{1}{20}.$$

Somiglia a una serie geometrica come quella che conosciamo, ma non è proprio lei, purtroppo, a causa di quel fattore k. Dobbiamo trovare un modo per liberarcene.

Scriviamo il problema nella sua forma generale, sostituendo a $\frac{1}{20}$ una generica probabilità di successo p. Allora il numero medio di tentativi sarà

$$L = \sum_{k=1}^{+\infty} k \cdot (1-p)^{k-1} \cdot p = p + 2(1-p)p + 3(1-p)^2 p + \cdots$$

e per avere una serie geometrica dovremmo sottrarre $(1-p)p$ dal secondo addendo, $2(1-p)^2 p$ dal terzo e così via, sottraendo $(k-1)(1-p)^{k-1}p$ dal k-esimo addendo. Con un po' di occhio, possiamo accorgerci che si tratta degli stessi termini presenti nella somma per L, moltiplicati per $1-p$. Quindi possiamo scrivere

$$L - (1-p)L = \sum_{k=0}^{+\infty}(1-p)^k p = p\sum_{k=0}^{+\infty}(1-p)^k = \frac{p}{1-(1-p)},$$

da cui ricaviamo $L = \frac{1}{p}$.
C'è un'altra via possibile: riscriviamo L in un modo diverso. Cosa succede al primo lancio? Con probabilità p abbiamo ottenuto il successo che cercavamo, mentre con probabilità $1-p$ siamo al punto di partenza, ma con un lancio già fatto. Quindi in questo secondo caso faremo in media $L+1$ lanci. In altre parole $L = p + (1-p)(L+1)$, che risolvendo in L dà $L = \frac{1}{p}$.
Nel caso di Dolores $p = \frac{1}{20}$ quindi, qualunque strada prendiamo, serviranno in media 20 tentativi[4] per ottenere un 20. $\square$

Soluzione al Problema 26

Possiamo vedere alcune somiglianze con il Problema 25. Stiamo in effetti generalizzando quanto visto in quel caso: invece di un solo successo, vogliamo averne 6 consecutivi. Con questo parallelo in mente, proviamo a rispondere a una domanda più generale: quanti lanci servono in media per ottenere n teste consecutive.
Quello cui puntiamo è una formula ricorsiva, di cui già sappiamo il valore nel caso $n = 1$. Infatti, se indichiamo con N_1 il numero di lanci che ci occorrono in media per avere una testa, siccome la probabilità di successo è $p = \frac{1}{2}$, abbiamo $N_1 = \frac{1}{p} = 2$.
Ora procediamo per induzione: supponiamo di conoscere il numero N_{n-1} e di voler ottenere N_n. Per ottenere n teste consecutive, dobbiamo innanzitutto ottenerne $n-1$, cosa che richiederà in media N_{n-1} lanci. A questo punto con probabilità $\frac{1}{2}$ al lancio successivo uscirà testa, portandoci al nostro obiettivo di n teste consecutive. Con probabilità $\frac{1}{2}$, però, uscirà croce e saremo punto e a capo, con $N_{n-1}+1$ lanci già fatti.

---

[4] Se il caso generale vi ricorda qualcosa è perché abbiamo già parlato di questo tipo di processo, detto geometrico, nell'Esempio 3.3.5, di cui ora sappiamo che ha valor medio $\frac{1}{p}$.

Mettendo insieme le due possibilità, abbiamo quindi

$$N_n = \frac{1}{2}(N_{n-1} + 1) + \frac{1}{2}(N_{n-1} + 1 + N_n),$$

che risolvendo in N_n dà $N_n = 2(N_{n-1} + 1)$. Possiamo ora risolvere la relazione di ricorrenza[5] così ottenuta, ricordando che $N_1 = 2$. Ricaviamo $N_n = 2^{n+1} - 2$, che per $n = 6$ dà 126.

Anche per questo problema vediamo una seconda strategia risolutiva, più esplicita. Chiamiamo in questo caso N il numero di lanci che ci aspettiamo di fare in media, quello che prima avevamo chiamato N_6. Vogliamo riscrivere N usando un'unione di casi a intersezione nulla. Per trovare questi casi, guardiamo cosa succede ai primi lanci.

Cominciamo dal primo lancio. Se esce croce (cosa che accade con probabilità $\frac{1}{2}$), dopo il primo lancio siamo al punto di partenza: il numero di lanci che ci aspettiamo di dover fare è ancora N, cui però dobbiamo sommare quello già fatto. In sostanza questo primo caso dà un contributo $\frac{N+1}{2}$ a N, cioè $N = \frac{N+1}{2} + R$ dove R è un contributo che viene da casi in cui il primo lancio è testa. Guardiamo allora come è fatto questo R.

Se il primo lancio è testa, spezziamo di nuovo in due i casi possibili, a seconda del risultato del secondo lancio. Anche in questo caso, se otteniamo una croce siamo di nuovo punto a capo, e ripartiamo con due lanci già fatti. Questo succede con probabilità $\frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2}$, perché è richiesto che il primo lancio sia testa e il secondo croce. In questo caso, il contributo a R e quindi a N sarà $\frac{N+2}{4}$.

Se invece entrambi i primi lanci sono testa, passiamo a guardare il terzo. Con probabilità $\frac{1}{8}$ la sequenza dei primi 3 lanci sarà di 2 teste e una croce, quindi ripartiamo da capo con 3 lanci già fatti, per un contributo a N di $\frac{N+3}{8}$. In alternativa, se i risultati dei primi 3 lanci sono 3 teste, passiamo al quarto, in cui nuovamente le cose vanno male se il risultato è croce. Questa sequenza di 3 teste e una croce ha probabilità $\frac{1}{16}$ che ci rimanda al via con 4 lanci già fatti, per un contributo a N di $\frac{N+4}{16}$.

Ancora una volta, se nei primi 4 lanci abbiamo 4 teste, separiamo il caso in cui la quinta sia croce, che ci fa ripartire da capo con un contributo di $\frac{N+5}{32}$, da quello in qui anche la quinta sia testa. In questo caso siamo arrivati con 5 teste consecutive al senso lancio e, finalmente, la musica è un po' diversa. Da un lato se la sesta è una croce ripartiamo da capo, con 6 lanci in più, per un contributo a N di

[5] Se non sappiamo risolvere in generale una relazione di ricorrenza, in questo caso possiamo fare tutte le sostituzioni, visto che ci sono solo 6 termini.

$\frac{N+6}{64}$. Dall'altro, se la sesta faccia uscita è anch'essa testa, abbiamo finito, per un contributo uguale a $\frac{6}{64}$.
Ora mettiamo assieme tutti questi contributi e abbiamo:

$$N = \frac{N+1}{2} + \frac{N+2}{4} + \frac{N+3}{8} + \frac{N+4}{16} + \frac{N+5}{32} + \frac{N+6}{64} + \frac{6}{64} = \frac{63N+126}{64},$$

da cui $N = 126$. □

Soluzione al Problema 27

Non c'è possibilità di pareggio né nelle singole partite, né nelle finali complessivamente, quindi è equivalente calcolare la probabilità di vittoria dei Cavaliers o dell'evento complementare, la vittoria dei Thunders (a patto, nel secondo caso, di sottrarre la probabilità ottenuta da 1 per avere la risposta). Possiamo anche vedere la sequenza di partite come 7 lanci di una moneta sbilanciata, che dà testa con probabilità p e croce con probabilità $1-p$. I Cavaliers vinceranno le finali se nei lanci della moneta ci saranno esattamente 4, 5, 6 o 7 croci o, equivalentemente, esattamente 3, 2, 1 o 0 teste.
Abbiamo già visto, nel Problema 18, come calcolare la probabilità che ci siano esattamente k teste nei 7 lanci: dobbiamo contare quante sono le possibili sequenze con questo numero di monete, cioè $\binom{7}{k}$, e sommare, per ciascuna di esse, la sua probabilità, che è costantemente uguale a $p^k(1-p)^{7-k}$, per un contributo di $\binom{7}{k}p^k(1-p)^{7-k}$. Dobbiamo ora sommare queste proprietà per gli eventi, disgiunti, corrispondenti a $k = 0, \ldots, 3$. Abbiamo allora

$$P(\text{Cavaliers}) = \sum_{k=0}^{3} \binom{7}{k} p^k (1-p)^{7-k} = 1 - \sum_{k=4}^{7} \binom{7}{k} p^k (1-p)^{7-k}.$$

Possiamo osservare che non abbiamo usato l'ipotesi $p > \frac{1}{2}$: essa non ci occorre per calcolare la probabilità richiesta, ma da essa possiamo ricavare che $P(\text{Cavaliers}) < \frac{1}{2}$.
Osserviamo infatti che avremmo anche potuto scrivere

$$P(\text{Cavaliers}) = \sum_{k=0}^{3} \binom{7}{k} p^k (1-p)^{7-k}$$
$$P(\text{Thunders}) = \sum_{k=4}^{7} \binom{7}{k} p^k (1-p)^{7-k}.$$

Entrambe le somme hanno 4 addendi, che possiamo confrontare a due a due, sfruttando le proprietà del binomiale. Prendiamo ad esempio $k = 2$ nella prima somma e $k = 5$ nella seconda e facciamo la differenza:

$$\binom{7}{2} p^2 (1-p)^5 - \binom{7}{5} p^5 (1-p)^2 = \binom{7}{5} p^2 (1-p)^2 [(1-p)^3 - p^3].$$

Il segno dipende solo dall'ultima parentesi. Concentrandoci su quella otteniamo

$$(1-p)^3 - p^3 = -2p^3 + 3p^2 - 3p + 1 = -(2p-1)(p^2 - p + 1),$$

che è una quantità negativa per $p > \frac{1}{2}$. Questo ci dice che, dei due termini considerati, quello che contribuisce a $P(\text{Cavaliers})$ è più piccolo. Possiamo ottenere disuguaglianze simili per gli altri termini nelle due somme. Concludiamo grazie a queste disuguaglianze e al fatto che i due eventi sono complementari. □

Soluzione al Problema 28

Cominciamo col calcolare le probabilità che Gianni apra la porta al terzo tentativo, nei due casi. Se è mezzogiorno, evento che indichiamo con il simbolo ☼, la probabilità che apra la porta al terzo tentativo è: $\frac{4}{5} \cdot \frac{3}{4} \cdot \frac{1}{3} = \frac{1}{5}$. Se invece è rientrato a mezzanotte, indicata con ☾, la probabilità è: $\frac{4}{5} \cdot \frac{4}{5} \cdot \frac{1}{5} = \frac{16}{125}$. Come possiamo usare queste probabilità?

Stiamo cercando la probabilità che Gianni sia rientrato a mezzanotte, sapendo che ha aperto la porta al terzo tentativo. Abbiamo quindi bisogno di usare il Teorema di Bayes, assieme alla formula di fattorizzazione, poiché sappiamo che i due orari di rientro sono equiprobabili e complementari. Allora

$$P(☾ \mid 3) = \frac{P(3 \mid ☾) \cdot P(☾)}{P(3 \mid ☾) \cdot P(☾) + P(3 \mid ☼) \cdot P(☼)} = \frac{8}{125} \cdot \frac{250}{41} = \frac{16}{41}. \qquad □$$

Soluzione al Problema 29

Il problema (di combinatoria) ci suggerisce di usare il coefficiente binomiale, ma come spesso accade non in modo diretto. Pensiamo a un problema leggermente diverso: supponiamo di avere un certo numero n di studenti e di volerne scegliere 16, per ora senza vincoli. Sappiamo come farlo, è il solito coefficiente binomiale: $\binom{n}{16}$.

Ora vogliamo inserire la condizione sui numeri di matricola non consecutivi: supponiamo, senza perdere generalità, che i numeri di matricola siano semplicemente i numeri $1, \ldots, n$ e di aver messo in ordine crescente di matricola i 16 scelti. Come possiamo essere sicuri di non averne di consecutivi? Il primo estratto non ci dà problemi. Il secondo, invece, potrebbe essere consecutivo al primo: come possiamo evitarlo? Aggiungiamo 1 al suo numero di matricola. In questo modo siamo certi che non sia il successivo. Passiamo ora al terzo: in questo caso aggiungere 1 non basta. Infatti se il suo numero fosse stato consecutivo a quello del secondo prima della traslazione, lo sarebbe ancora. Quindi, per essere sicuri, dobbiamo aggiungere 2 al suo numero di matricola. Allo stesso modo aggiungeremo 3 al numero di matricola del quarto e così via, fino ad aggiungere 15 al numero di matricola del sedicesimo. Questo ci dice che ora i numeri, invece che andare da 1 a n, possono arrivare fino a $n + 15$.

Ora vogliamo tornare al nostro problema: abbiamo un modo per contare in quanti modi è possibile scegliere 16 numeri non consecutivi tra $n+15$: sono $\binom{n}{16}$. Allo stesso tempo, gli studenti sono in tutto 48, quindi dobbiamo ricavare n, sapendo che $n+15=48$: abbiamo $n=33$. La risposta è $\binom{33}{16}$. □

Soluzione al Problema 30

Cominciamo con una risoluzione un po' particolare. Condizioniamo al numero dell'estrazione: tra le biglie estratte per prime, la proporzione delle nere è p. Tra le seconde biglie estratte, la proporzione delle biglie nere è ancora p, e così via. Osserviamo che grazie al condizionamento ogni volta lasciamo fuori i giochi conclusi dall'estrazione di una biglia nera.

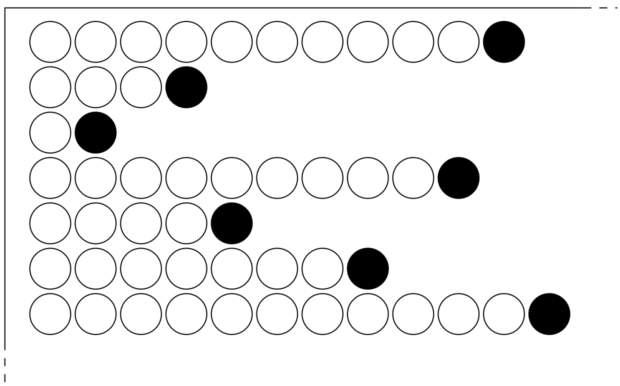

Figura 6.2: Alcune delle estrazioni di Marguerite

Graficamente quello che stiamo facendo è leggere la tabella delle estrazioni (ad esempio quella rappresentata in Figura 6.2) in due modi diversi. Se ogni riga rappresenta una partita, conterrà una sequenza (possibilmente vuota) di palline bianche, seguite da una nera, ma le sequenze avranno lunghezze diverse. Guardando la stessa tabella in verticale, abbiamo sulla colonna n-esima le palline estratte come n-esime nei vari giochi (alcune righe saranno vuote, ma le ignoriamo). In ciascuna colonna la proporzione di biglie nere sul totale di quelle effettivamente pescate sarà uguale a p. Siccome la proporzione di biglie nere tra quelle estratte per prime è p, ed è la stessa tra quelle estratte per seconde e così via per ogni estrazione, la proporzione di biglie nere estratte è p.

Questa soluzione ha richiesto un'idea un po' particolare. Vediamo allora un altro modo di risolvere il medesimo problema. Se una partita dura n estrazioni significa che Marguerite ha estratto $n-1$ biglie bianche seguite da una biglia nera. Questo avviene con probabilità $(1-p)^{n-1}\cdot p$. Possiamo allora considerare il numero di biglie estratte in media da Marguerite in ogni partita. Sapendo che solo una di queste è nera, ci basterà prendere il reciproco del numero medio di biglie estratte e avremo la proporzione di biglie nere estratte. Il numero medio di biglie estratte è

$$\sum_{n=1}^{+\infty} n \cdot (1-p)^{n-1} \cdot p = p \sum_{n=1}^{+\infty} n \cdot (1-p)^{n-1}.$$

Per calcolare questa somma possiamo ricondurci alla serie geometrica: sappiamo che $0 < 1-p < 1$, quindi la serie geometrica $\sum_{n=0}^{+\infty}(1-p)^n = [1-(1-p)]^{-1} = p^{-1}$. Se calcoliamo la sua derivata in p, abbiamo

$$\sum_{n=1}^{+\infty} n \cdot (1-p)^{n-1} = \sum_{n=0}^{+\infty} n \cdot (1-p)^{n-1} = \frac{1}{[1-(1-p)]^2} = \frac{1}{p^2},$$

da cui otteniamo che in media Marguerite estrarrà p^{-1} biglie, di cui una sola nera, ossia che la proporzione di biglie nere sul totale è p. □

Soluzione al Problema 31

Il comportamento del primo automa ci suggerisce di guardare alle once di bronzo nel crogiolo in modulo 3[6]: abbiamo 3 classi possibili, 0, 1 e 2. Se siamo in 0, il primo automa ci porta in 1 con probabilità 1. Se invece siamo in 1 o 2, dobbiamo guardare al risultato del lancio del dado: in 2 casi su 8 (se il dado dà 3 o 6) l'automa ci lascia dove siamo, in 3 su 8 ci sposta in 0. Sempre in 3 casi su 8 ci sposta in 2 se siamo in 1 o in 1 se siamo in 2.

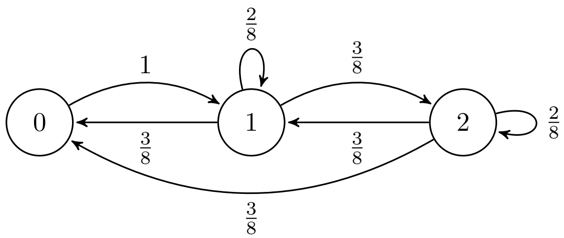

Figura 6.3: Come ci manda in giro il primo automa

Non ci dobbiamo però dimenticare del secondo automa. Questo agisce solamente se siamo in 0 e, sottraendo 3 once, non ci fa cambiare classe. Potremmo però avere un problema: se siamo in 0 con 0 once, il secondo automa non potrebbe forgiare alcuna spada: ma dal nostro schema non abbiamo la possibilità di accorgercene. Per fortuna sappiamo che sere e mattine si alternano, che partiamo con più di 3 once e che, lasciato 0, per tornarci dobbiamo accumulare almeno altre 3 once. Possiamo dunque stare tranquilli: non finiremo il bronzo. In altre parole, lo schema in Figura 6.3 descrive completamente il problema.

[6]Per approfondire l'argomento è possibile consultare *Aritmetica modulare* [7], edito in questa stessa collana.

Possiamo allora ripensare l'esercizio come una passeggiata casuale su questo diagramma: ci spostiamo da una classe all'altra con le probabilità scritte sulle frecce. Se teniamo traccia del percorso fatto, cioè se scriviamo una dopo l'altra tutte le classi per cui passiamo, la probabilità cercata è quella di avere una stringa di 20 numeri che inizia con "1" e che contiene cinque "0". Infatti la prima mattina, vedendo nel crogiolo 9 once di bronzo, il primo automa ne aggiunge una. Questo succederà ogni volta che compare uno "0" nella stringa: a meno che non si tratti dell'ultimo carattere, il successivo sarà un "1".
Per quello che dobbiamo fare non ci interessa davvero distinguere tra le volte in cui scriviamo "1" e quelle in cui scriviamo "2": possiamo considerarli come un unico carattere, che indicheremo con ✠. Dobbiamo però aggiornare il nostro schema, che ora sarà a due soli stati, "0" e "✠", con probabilità 1 di transizione da "0" a "✠", $\frac{3}{8}$ di transizione da "✠" a "0" e $\frac{5}{8}$ di rimanere in "✠", come in Figura 6.4.

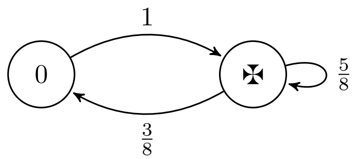

Figura 6.4: Lo schema semplificato

Ricapitoliamo: la stringa inizia con "✠" e la lettera che segue uno "0" è sempre un "✠", a meno che "0" non sia in ultima posizione. Inoltre un contributo "✠" arriva con probabilità $\frac{5}{8}$, indipendentemente dalla lettera precedente, mentre un contributo "0✠" (o solo "0" se è l'ultima lettera) compare con probabilità $\frac{3}{8}$, indipendentemente dalla lettera precedente. Non ci resta che considerare due casi.

1. La stringa inizia e finisce con "✠". Dobbiamo allora mettere nella stringa 5 coppie "0✠" (per un totale di 10 lettere) e 9 "✠". Possiamo farlo in $\binom{14}{5}$ modi, ciascuno dei quali ha probabilità $\left(\frac{3}{8}\right)^5 \cdot \left(\frac{5}{8}\right)^9$. Il primo caso contribuisce complessivamente con $\binom{14}{5} \cdot \left(\frac{3}{8}\right)^5 \cdot \left(\frac{5}{8}\right)^9$.

2. La stringa inizia con "✠" e finisce con "0". Dobbiamo quindi mettere nella stringa 4 coppie "0✠" e 10 "✠". Possiamo farlo in $\binom{14}{4}$ modi, ciascuno dei quali ha probabilità $\left(\frac{3}{8}\right)^4 \cdot \left(\frac{5}{8}\right)^{10} \cdot \left(\frac{3}{8}\right)$, dove l'ultimo termine indica la probabilità di arrivare dalla penultima lettera (necessariamente una "✠") allo "0" finale. Il contributo complessivo di questo secondo caso è $\binom{14}{4} \cdot \left(\frac{3}{8}\right)^5 \cdot \left(\frac{5}{8}\right)^{10}$.

I due casi sono disgiunti, quindi la probabilità cercata è

$$\frac{3^6 \cdot 5^9 \cdot 7^2 \cdot 11 \cdot 13}{8^{15}} = \frac{9\,976\,763\,671\,875}{35\,184\,372\,088\,832} \approx 0{,}28. \qquad \square$$

Soluzione al Problema 32

Vediamo qualche possibile attacco. Possiamo pensare ai tre salti come a tre sbarrette metalliche rigide, unite agli estremi con dei perni sui quali possono ruotare. Fissato il punto di partenza, i punti più lontani che possiamo raggiungere sono su una circonferenza di raggio 3. È facile convincerci che siamo in grado di raggiungere ogni punto all'interno di questo cerchio, modificando opportunamente gli angoli tra le sbarrette. Potremmo allora pensare che per trovare la probabilità cercata basti considerare il rapporto tra l'area del cerchio di raggio 3 e la superficie del cerchio di raggio 1, che contiene tutti i punti che non distano più di 1 dall'origine. Il rapporto tra queste due aree è $\frac{1}{9}$, ma non è la risposta giusta.

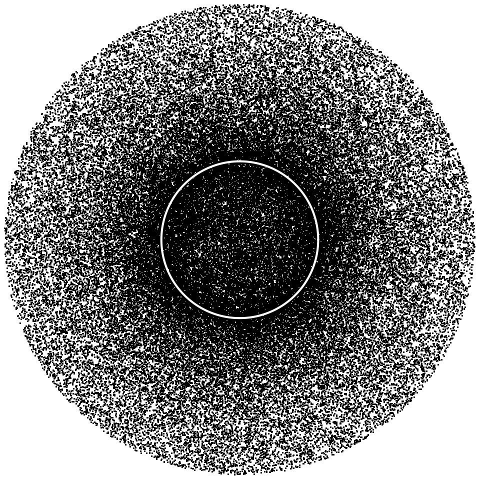

Figura 6.5: Simulazione dei punti d'arrivo della rana. Cerchiamo la probabilità che i punti siano all'interno della circonferenza di raggio 1, in bianco nell'immagine.

Infatti non è vero che la probabilità di raggiungere una certa zona è proporzionale alla sua area, come vediamo in Figura 6.5. Anche intuitivamente ci aspettiamo che la probabilità di arrivare vicino alla circonferenza di raggio 3 sia poco probabile: solo poche combinazioni di angoli ci portano nei pressi del bordo della figura.

Proviamo allora qualcosa di diverso: il problema presenta una chiara simmetria rotazionale attorno al punto di partenza, cioè zone sovrapponibili con una rotazione attorno al centro hanno necessariamente la stessa probabilità. Possiamo allora supporre fissata la prima sbarretta e far ruotare solo le altre due. Il secondo estremo della seconda sbarretta traccia una circonferenza passante per l'origine e, centrata in ciascun punto di questa circonferenza, abbiamo un'ulteriore circonferenza spazzata dalla terza sbarretta. In questo caso abbiamo che l'insieme di tutti i punti raggiungibili, a prima sbarretta fissata, è un cerchio di raggio 2 centrato nel secondo estremo della prima sbarretta. Di questi punti quelli che vanno bene sono

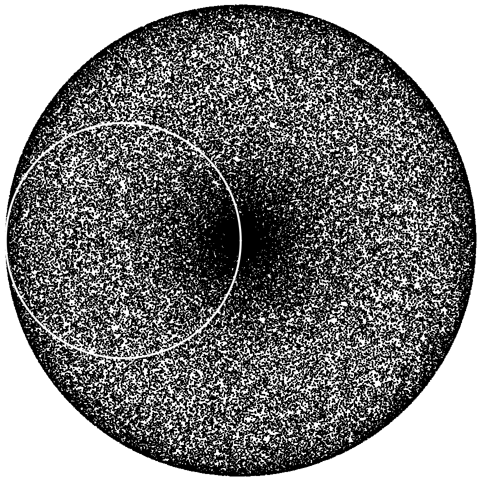

Figura 6.6: Simulazione dei punti d'arrivo della rana nel caso in cui il primo salto sia fissato con arrivo in $(1, 0)$. L'area che ci interessa è quella delimitata dalla circonferenza bianca, visibile sulla sinistra.

all'interno di un cerchio di raggio 1. Il rapporto tra le due aree è $\frac{1}{4}$ che è la risposta giusta, come vedremo, ma ottenuta con un metodo sbagliato. Infatti abbiamo ancora una volta lo stesso problema di prima: i punti di arrivo della rana non sono distribuiti uniformemente nemmeno in questo caso, come vediamo in Figura 6.6.

Cerchiamo ancora un'altra strategia, perché ormai abbiamo capito che procedere con le aree non ci porterà a una soluzione. L'idea di fissare una sbarretta non è cattiva, ma invece di fissare la prima, fissiamo la seconda: da un punto di vista fisico stiamo considerando il sistema di riferimento cartesiano in cui la seconda sbarretta ha la prima estremità in $(-0,5, 0)$ e la seconda in $(0,5, 0)$.

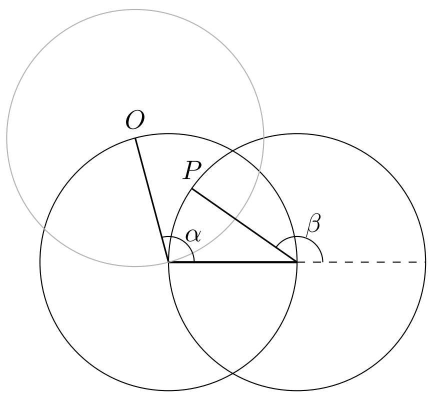

Figura 6.7: Gli angoli della rana. In grigio chiaro la circonferenza di raggio 1 centrata nel punto di partenza.

Ci concentriamo sugli angoli: chiamiamo α quello tra le prime due sbarrette e β quello tra la seconda e la terza, come nella Figura 6.7. Ci piacciono gli angoli perché variano uniformemente (a differenza del secondo estremo della terza sbarretta nei tentativi precedenti).

Vogliamo capire quando succede (al variare di α e β) che la seconda estremità P della terza sbarretta sia all'interno del cerchio di raggio 1 centrato in O, prima estremità della prima sbarretta. Misuriamo entrambi gli angoli rispetto all'asse orizzontale su cui giace la seconda sbarretta.

Cominciamo con i casi limite: per qualunque valore di α tra 0 e 360 gradi (o tra 0 e 2π se usiamo i radianti), se $\beta = \alpha$, allora le tre sbarrette sono tre lati di un parallelogramma. Quindi OP sarà lungo 1 come la seconda sbarretta, caso che ci va bene.

In generale, poi, per $0 \leqslant \alpha \leqslant 180°$, vanno bene tutti i valori di β tali che $\alpha \leqslant \beta \leqslant 180°$. Per $180° \leqslant \alpha \leqslant 360°$, vanno bene le ampiezze $\alpha \leqslant \beta \leqslant 360°$.

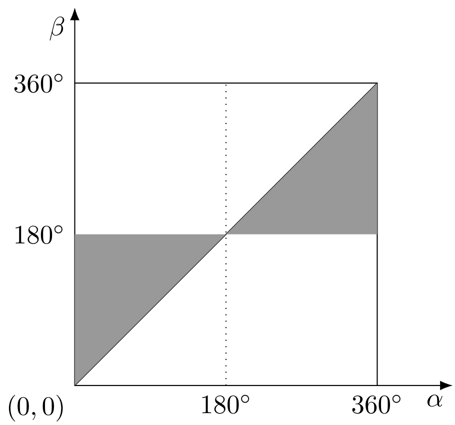

Figura 6.8: Quali coppie di angoli vanno bene alla rana.

Se rappresentiamo nel piano α, β le coppie di ampiezze che vanno bene, otteniamo la Figura 6.8, in cui ora possiamo leggere la probabilità come rapporto di aree, dal momento che la probabilità su ciascun asse è proporzionale alla lunghezza dei segmenti. La probabilità cercata è quindi $\frac{1}{4}$. $\square$

Soluzione al Problema 33

Indichiamo con m il numero di ragazzi nella classe e con f il numero di ragazze. Vogliamo scrivere la probabilità che un alunno a caso sia d'accordo con l'insegnante (probabilità che sappiamo essere uguale a $\frac{1}{2}$) spezzando i due casi possibili, cioè che l'alunno sia un ragazzo o una ragazza, con la formula di fattorizzazione.

Abbiamo allora

$$\begin{aligned}\frac{1}{2} &= P\,(\text{accordo} \mid \male) \cdot P(\male) + P\,(\text{accordo} \mid \female) \cdot P(\female) \\ &= [p \cdot p_{\male} + (1-p) \cdot (1-p_{\male})] \cdot \frac{m}{m+f} \\ &\quad + [p \cdot p_{\female} + (1-p) \cdot (1-p_{\female})] \cdot \frac{f}{m+f},\end{aligned} \tag{6.1}$$

dove abbiamo usato il fatto che un alunno e l'insegnante possono essere d'accordo o se rispondono entrambi correttamente o se sbagliano entrambi.
Questa equazione sembra avere due incognite. Tuttavia stiamo cercando il rapporto $\frac{m}{f}$, quindi non è detto che sia necessaria un'altra equazione. Moltiplichiamo primo e secondo membro per $\frac{m+f}{f}$,

$$\frac{m+f}{f} \cdot \frac{1}{2} = [p \cdot p_{\male} + (1-p) \cdot (1-p_{\male})] \cdot \frac{m}{f} + p \cdot p_{\female} + (1-p) \cdot (1-p_{\female})$$

per poi raccogliere i termini con $\frac{m}{f}$ a primo membro

$$\frac{m}{f}\left\{\frac{1}{2} - [p \cdot p_{\male} + (1-p) \cdot (1-p_{\male})]\right\} = p \cdot p_{\female} + (1-p) \cdot (1-p_{\female}) - \frac{1}{2}$$

e risolvere, assumendo che $p \cdot p_{\male} + (1-p) \cdot (1-p_{\male}) \neq \frac{1}{2}$

$$\frac{m}{f} = \frac{p \cdot p_{\female} + (1-p) \cdot (1-p_{\female}) - \frac{1}{2}}{\frac{1}{2} - (p \cdot p_{\male} + (1-p) \cdot (1-p_{\male}))}.$$

La risposta trovata ha un aspetto abbastanza strano: cosa succede in particolare quando $p_{\male} = p_{\female}$? Avremmo in tal caso $\frac{m}{f} = \frac{2a-1}{1-2a} = -1$, risposta che non sembra avere molto senso. In realtà, con questa condizione, dalla (6.1) otteniamo che $p \cdot p_{\male} + (1-p) \cdot (1-p_{\male}) = \frac{1}{2}$, quindi l'ultima uguaglianza non sarebbe alla nostra portata. In effetti in questo caso non ci possiamo nemmeno aspettare una soluzione, perché non abbiamo più nulla che ci permetta di distinguere ragazzi e ragazze.
Il risultato ottenuto è problematico anche quando $p = \frac{1}{2}$, perché, sostituendo nella (6.1) avremmo $1 = \frac{1}{2}$. Non solo, esaminando ancora la (6.1), possiamo notare che non tutti i valori di $p_{\male}$ e $p_{\female}$ in $(0,1)$ sono accettabili: scoprire quali sono i vincoli che questi devono soddisfare può essere un esercizio nell'esercizio. $\square$

Soluzione al Problema 34

In precedenza abbiamo parlato di invarianti. Questo problema non riguarda propriamente un invariante. Tuttavia la somma delle tre cifre varia in modo controllato, a seconda del risultato del lancio della moneta: se esce croce la somma delle tre cifre rimane invariata, se invece esce testa, la somma diminuisce di uno. La somma iniziale è uguale a 4, mentre la somma delle cifre nella configurazione finale è 3, quindi per avere successo dopo due lanci esatti, dobbiamo necessariamente avere una testa e una croce. Ma non abbiamo ancora finito: in questo problema l'ordine dei lanci è significativo e, comunque, dobbiamo tenere traccia di quali sono le cifre che vengono scelte a caso per essere cambiate.

Affinché i compagni di Ellisseo abbiano qualche speranza di successo, devono comparire un 3 e due 0. Avendo solo due mosse a disposizione l'unica possibilità è ottenere $(0, 3, 0)$, cioè bisogna sottrarre 1 a ciascuna delle cifre agli estremi e aggiungere 1 alla cifra centrale. Questa è solo una delle $\binom{3}{2} = 3$ possibili combinazioni che aprono l'otre, quindi anche se i compagni di Ellisseo la ottengono, apriranno l'otre con probabilità $\frac{1}{3}$.

Mettiamoci nei panni dei compagni di Ellisseo: in quali[7] modi possiamo ottenere $(0, 3, 0)$ e quali sono le probabilità corrispondenti?

Abbiamo detto che devono uscire una testa e una croce: consideriamo quindi le due permutazioni possibili dei lanci. Se escono, nell'ordine, testa e croce (con probabilità $\frac{1}{4}$), al primo lancio dobbiamo sottrarre 1 a una delle tre cifre. Con probabilità $\frac{1}{3}$ andiamo a diminuire la cifra centrale e abbiamo sicuramente perso. Con probabilità $\frac{2}{3}$, invece, sottraiamo 1 dalla prima o dall'ultima cifra, portandola a 0. Questa cifra non può più diminuire!

In questo caso all'uscita della croce sulla seconda moneta andiamo a sottrarre 1 dal 2 centrale con probabilità $\frac{1}{2}$ e perdiamo, oppure con probabilità sempre $\frac{1}{2}$ portiamo a zero anche la rimanente cifra 1 e abbiamo ancora qualche probabilità di vincere. Quanta? Dobbiamo incrementare di 1 la cifra 2 al centro, con probabilità $\frac{1}{2}$, altrimenti perdiamo. La probabilità che vinciamo se escono una testa e una croce in quest'ordine è $\frac{2}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{6}$. Quindi la probabilità che escano una testa e una croce in quest'ordine e che vinciamo è $\frac{1}{4} \cdot \frac{1}{6} = \frac{1}{24}$, usando l'identità per la probabilità dell'intersezione (4.1).

Resta da considerare il caso, disgiunto, che escano nell'ordine croce e testa (con probabilità $\frac{1}{4}$). Come prima cosa dobbiamo far diminuire una cifra e le cose sono del tutto analoghe a quanto visto nel caso precedente: con probabilità $\frac{1}{3}$ andiamo a toccare la cifra centrale e abbiamo perso, con probabilità $\frac{2}{3}$ portiamo una cifra 1 a 0 e continuiamo. Dobbiamo aggiungere 1 a una cifra, ma siccome siamo ancora al lancio della prima moneta (la croce) di sicuro non andremo ad aggiungerla allo 0 che è appena comparso (dobbiamo far aumentare un'altra cifra), quindi la probabilità di non perdere subito è $\frac{1}{2}$, quella di far crescere il 2 centrale a 3.

In questo caso consideriamo l'ultima modifica, una diminuzione di 1 che non può toccare lo 0, ma solo il rimanente 1 o il 3, ciascuno con uguale probabilità. La

[7] È importante determinare quali sono e non solo quanti: i casi non sono equiprobabili.

probabilità di ottenere $(0,3,0)$ se le monete sono croce e testa, nell'ordine, è quindi di nuovo $\frac{2}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{6}$ e la probabilità che escano croce e testa e sblocchiamo il lucchetto è ancora $\frac{1}{4} \cdot \frac{1}{6} = \frac{1}{24}$.

Sommiamo questi due contributi, dal momento che sono eventi disgiunti, e abbiamo una probabilità di successo uguale a $\frac{1}{12}$. In alternativa, possiamo anche scrivere il risultato con la formula di fattorizzazione in cui due addendi (quelli corrispondenti a due teste o due croci) sono identicamente nulli:

$$\begin{aligned} P((0,3,0)) &= P((0,3,0) \mid CC) \cdot P(CC) + P((0,3,0) \mid CT) \cdot P(CT) \\ &\quad + P((0,3,0) \mid TC) \cdot P(TC) + P((0,3,0) \mid TT) \cdot P(TT) \\ &= 0 \cdot \frac{1}{4} + \frac{1}{6} \cdot \frac{1}{4} + \frac{1}{6} \cdot \frac{1}{4} + 0 \cdot \frac{1}{4} \\ &= \frac{1}{12}. \end{aligned}$$

La probabilità che i compagni di Ellisseo liberino i 20 intrappolati è allora uguale alla probabilità che ottengano la sequenza $(0,3,0)$ e che essa sia la combinazione corretta, cioè $\frac{1}{12} \cdot \frac{1}{3} = \frac{1}{36}$, come un doppio 6 lanciando due dadi. $\square$

Soluzione al Problema 35

Cominciamo a vedere alcune possibili strategie. Se tutti i prigionieri provano a indovinare il colore del proprio cappello, la probabilità che vengano rilasciati è $(\frac{1}{2})^3 = \frac{1}{8}$. Possono sicuramente fare di meglio con una strategia molto semplice: se decidono un portavoce tra loro prima di iniziare e solo il portavoce cerca di indovinare il proprio colore, mentre gli altri due passano, la probabilità di essere rilasciati sale a $\frac{1}{2}$.

In entrambe queste soluzioni, però, non stanno usando le informazioni che possono avere dal guardare i cappelli altrui. In altre parole stanno considerando solo i singoli cappelli (i propri) e non le terne di cappelli. Se pensiamo invece ai tre cappelli tutti insieme, la probabilità che siano tutti dello stesso colore è $2 \cdot \frac{1}{8} = \frac{1}{4}$, mentre la probabilità che siano due di un colore e uno di un altro è $2 \cdot \binom{3}{2} \cdot \frac{1}{8} = \frac{3}{4}$ (abbiamo infatti due modi di scegliere il colore predominante e $\binom{3}{2}$ modi di scegliere quali due dei tre cappelli saranno dello stesso colore).

I tre prigionieri adottano quindi la seguente strategia: chi vede (sulle teste degli altri due) cappelli di due colori diversi passa, chi vede cappelli del medesimo colore dichiara per il proprio cappello l'altro colore. Questa strategia fallisce solamente quando i cappelli sono tutti dello stesso colore. Questo accade, come abbiamo visto, in 1 caso su 4.

Con questa strategia, la probabilità che i tre siano rilasciati è allora $\frac{3}{4} > \frac{1}{2}$. $\square$

Soluzione al Problema 36

Come prima cosa ci chiediamo cosa significhi truccare i dadi: vuol dire assegnare a ogni coppia dado-faccia (d, f) una probabilità $p_{d,f}$ in modo che siano soddisfatte le seguenti condizioni,

$$\sum_{f=1}^{6} p_{d,f} = 1, \quad d = 1, 2$$

$$P(S = s) = \sum_{f=1}^{s-1} p_{1,f} \cdot p_{2,s-f} = \frac{1}{11}, \quad s = 2, \ldots, 12,$$

in cui abbiamo indicato con $S = s$ l'evento “la somma dei due dadi è s” e poniamo $p_{d,f} = 0$ se $f > 6$.

Partiamo dalle somme più semplici, 2 e 12, ciascuna delle quali può essere ottenuta in un solo modo:

$$P(S = 2) = p_{1,1} \cdot p_{2,1} = \frac{1}{11}, \quad P(S = 12) = p_{1,6} \cdot p_{2,6} = \frac{1}{11}.$$

Abbiamo questi prodotti, ma non abbiamo ottenuto informazioni sui valori delle $p_{d,f}$ coinvolte. Se proviamo a continuare con le somme successive, abbiamo

$$P(S = 3) = p_{1,2} \cdot p_{2,1} + p_{1,1} \cdot p_{2,2} = \frac{1}{11}, \quad P(S = 11) = p_{1,6} \cdot p_{2,5} + p_{1,5} \cdot p_{2,6} = \frac{1}{11},$$

che coinvolgono le probabilità incontrate prima, ciascuna delle quali moltiplicata con qualcosa di nuovo. Non sembra molto promettente. Proviamo allora a vedere cosa succede nel caso più complicato, la somma 7:

$$\begin{aligned} P(S = 7) &= \sum_{f=1}^{6} p_{1,f} \cdot p_{2,7-f} \\ &= p_{1,1} \cdot p_{2,6} + p_{1,2} \cdot p_{2,5} + p_{1,3} \cdot p_{2,4} + p_{1,4} \cdot p_{2,3} + p_{1,5} \cdot p_{2,2} + p_{1,6} \cdot p_{2,2}. \end{aligned}$$

Tra i numerosi addendi possiamo osservare che il primo e l'ultimo sono particolari: coinvolgono le stesse facce che compaiono nei casi più semplici visti prima. Abbiamo però un piccolo problema: siccome i prodotti sono scambiati, come possiamo usare in questo caso i risultati appena ottenuti? Possiamo usare la disuguaglianza tra le medie[8],

$$P(S = 7) \geqslant p_{1,1} \cdot p_{2,6} + p_{1,6} \cdot p_{2,2} \geqslant 2 \cdot \sqrt{p_{1,1} \cdot p_{2,6} \cdot p_{1,6} \cdot p_{2,2}} = \frac{2}{11},$$

[8] Per questo problema tutto quello che ci serve sapere è che la media aritmetica di un insieme di numeri non negativi è maggiore o uguale della media geometrica del medesimo insieme (la nota disuguaglianza AM-GM). Per saperne di più il rimando d'obbligo è alle *Schede Olimpiche* [10].

che è maggiore del valore richiesto. Questo ci mostra che la richiesta non può essere soddisfatta in alcun modo[9]. □

Soluzione al Problema 37

D'istinto potremmo provare a considerare tutti i possibili percorsi per calcolare la probabilità che ciascuno di essi sia percorribile. Tuttavia dobbiamo stare attenti, perché i percorsi non sono indipendenti tra loro (ad esempio, se condividono un ponte).

Proviamo allora a cambiare il punto di vista: facciamo entrare in gioco una barca a vela, sul fiume, che vorrebbe navigare oltre queste isole, ma non riesce a passare sotto i ponti. Il percorso della barca e quello del viaggiatore si incrociano: se uno dei due riesce a passare, l'altro è necessariamente bloccato. Non solo: le due situazioni sono perfettamente simmetriche, come possiamo vedere in figura.

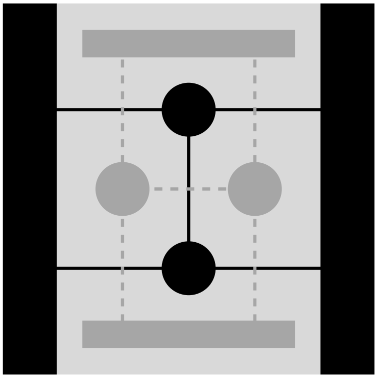

Figura 6.9: Ponti e barche a vela

A questo punto la probabilità non può che essere $\frac{1}{2}$.

È possibile generalizzare questo problema ad altre configurazioni di isole e ponti, ma non in tutti i casi è possibile usare la simmetria per la soluzione. Possiamo farlo, ad esempio, con 6 o 12 isole. □

Soluzione al Problema 38

Per comodità chiamiamo le due squadre Pari e Dispari. Se sono testa a testa saremo in 0, se i Dispari sono in vantaggio saremo in 1, se i Dispari vincono arriviamo in 3 (e ci fermiamo). Viceversa se i Pari sono in vantaggio saremo in 2 e se vincono

[9]Questo in realtà un problema di algebra olimpica, mascherato da problema di probabilità, con una domanda da problema di combinatoria. Per risolverlo è necessario accorgersi che l'idea fondamentale è la disuguaglianza tra le medie. Per arrivarci dobbiamo pensare alle probabilità in modo astratto, sfruttando l'indipendenza dei dadi.

arriviamo in 4. Dal testo ricaviamo che le probabilità di transizione sono quelle in Figura 6.10.

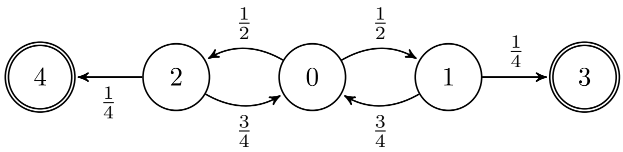

Figura 6.10: Lo schema della sfida

Stiamo cercando la probabilità che la squadra vincitrice sia stata in svantaggio. Chiamiamo p questa probabilità. Per la simmetria dello schema, abbiamo che $\frac{p}{2}$ è la probabilità che i Dispari abbiano vinto e siano stati in svantaggio e similmente $\frac{p}{2}$ è la probabilità che i Pari abbiano vinto e siano stati in svantaggio. Allo stesso tempo $1-p$ è la probabilità che la squadra vincente non sia mai stata in svantaggio, dal momento che la gara termina necessariamente con la vittoria di una delle due squadre. Anche $1-p$ è equamente suddiviso tra le due squadre.
Concentriamoci ora sulla vittoria della squadra Dispari senza che essa sia mai andata in svantaggio. Una possibilità è lo schema di gara sia stato $0-1-3$, oppure potrebbe essere stato $0-1-0-1-3$. In generale non possono comparire 2 e a maggior ragione 4, quindi ci sarà un certo numero (maggiore o uguale di 1) di 01, seguito da un 3 conclusivo. Il primo 01 ha probabilità $\frac{1}{2}$, il 3 finale ha probabilità $\frac{1}{4}$, ogni altro 01 nel mezzo ha probabilità $\frac{3}{4}\cdot\frac{1}{2}=\frac{3}{8}$. Tutte queste vie alla vittoria sono disgiunte, quindi ne possiamo sommare le probabilità

$$\frac{1-p}{2}=\frac{1}{2}\cdot\frac{1}{4}+\frac{1}{2}\cdot\frac{3}{8}\cdot\frac{1}{4}+\frac{1}{2}\cdot\left(\frac{3}{8}\right)^2\cdot\frac{1}{4}+\cdots$$

Possiamo scriverla come somma di una serie geometrica:

$$\frac{1-p}{2}=\frac{1}{8}\cdot\sum_{k=0}^{+\infty}\left(\frac{3}{8}\right)^k=\frac{1}{8}\cdot\frac{8}{5}=\frac{1}{5}.$$

Quindi $1-p=\frac{2}{5}$ e di conseguenza $p=\frac{3}{5}$. □

Soluzione al Problema 39

Possiamo calcolare indifferentemente la probabilità di vittoria dell'uno o dell'altro giocatore. Concentriamoci quindi su Franco, il primo a lanciare. Può vincere ottenendo subito tre teste, oppure passando il turno un certo numero di volte, purché Francesco non vinca e gli ripassi il turno ogni volta. È abbastanza comodo calcolare la probabilità che un giocatore passi il turno, cioè che non lanci tre teste di fila: questa probabilità è $1-p^3$. Ricordiamoci ora che Franco vince se c'è un numero pari di passaggi di turno, prima della sequenza di tre teste, quindi

$$P(\text{Franco}) = p^3 \sum_{k=0}^{+\infty} (1-p^3)^{2k} = p^3 \sum_{k=0}^{+\infty} (1-2p^3+p^6)^k = \frac{1}{2-p^3},$$

dove ci siamo ricondotti alla forma nota della serie geometrica. Di conseguenza la probabilità di vittoria di Francesco è $\frac{1-p^3}{2-p^3}$. Osserviamo che per p prossimo a 1, la probabilità di vittoria di Francesco è molto bassa, perché è molto probabile che Franco possa sfruttare la sua posizione di vantaggio. Viceversa, per p piccola, tendente a 0, la probabilità di vittoria di entrambi tende a $\frac{1}{2}$. Addirittura per $p = 0$ la probabilità di vittoria di ciascuno è proprio $\frac{1}{2}$. Ma cosa vuol dire che $p = 0$? Non uscirà mai testa, quindi in realtà il gioco andrà avanti all'infinito, senza fine. Ha quindi un certo significato filosofico dire che una partita che nessuno dei due può vincere è vinta con probabilità $\frac{1}{2}$ da ciascuno. □

Soluzione al Problema 40

Cominciamo dai primi passi: la partita non può finire al primo o al secondo gol. Non solo, la differenza reti tra i due giocatori ha la stessa parità del numero di gol segnati fino a quel momento, poiché $a - b$ e $a + b$ hanno la stessa parità.

La partita potrebbe finire al terzo gol. La probabilità che questo accada è uguale alla probabilità che chiunque abbia segnato il primo gol segni i due successivi, cioè è uguale a $\frac{1}{4}$.

Proseguendo, la partita non può concludersi col quarto gol: se non si è conclusa al terzo, significa che uno dei due ha fatto 2 gol e l'altro 1 solo, quindi al quarto gol il distacco potrà essere uguale a 0 o a 2. In generale, la partita non potrà mai finire a un turno pari, perché sarebbe finita al turno precedente. Ultima osservazione: per finire al quinto gol la partita non deve essere finita al terzo gol e i due gol successivi (il quarto e il quinto) devono essere stati segnati da chi era già in vantaggio (di un gol) al terzo.

Questi primi ragionamenti ci mostrano due cose: che la partita non può terminare in un turno pari e che possiamo scrivere la probabilità che la partita termini (in un turno dispari) a partire dalla probabilità che non sia terminata nel turno dispari precedente. Questo ci suggerisce di costruire una ricorsione per calcolare la probabilità che la partita non termini a un certo turno dispari.

Chiamiamo quindi $p(2k+1)$ la probabilità che la partita non termini col $(2k+1)$-esimo gol (e che non sia finita prima). Per $k > 1$ questa probabilità è uguale alla probabilità $p(2k-1)$ che la partita non sia finita al turno $2k-1$, moltiplicata per la probabilità che il giocatore in vantaggio al turno $2k-1$ non abbia segnato entrambi i gol successivi, cioè $1 - \frac{1}{4} = \frac{3}{4}$. Quindi $p(2k+1) = \frac{3}{4} \cdot p(2k-1)$.

Abbiamo trovato la ricorsione che cercavamo, di cui conosciamo il termine iniziale (per $k = 0$) $p(1) = 1$. Possiamo allora risolverla, trovando che $p(2k+1) = (\frac{3}{4})^k$. Il problema chiede la probabilità che la partita non termini prima del ventunesimo gol, quindi che non sia ancora terminata quando viene segnato il diciannovesimo gol, cioè $p(19) = (\frac{3}{4})^9 \approx 7{,}5\%$. □

Soluzione al Problema 41

Se Tuco spara a Sentenza e lo colpisce (con probabilità $\frac{2}{5}$), verrà sicuramente colpito dal Biondo allo sparo successivo. Se Tuco manca Sentenza, il Biondo sparerà a quest'ultimo, uccidendolo, in modo da minimizzare la probabilità di essere colpito al prossimo round, visto che Tuco ha una mira peggiore di quella di Sentenza. Tocca nuovamente a Tuco, che avrà ora la possibilità di sparare al Biondo, colpendolo con probabilità $\frac{2}{5}$, per una probabilità complessiva uguale a $\frac{3}{5} \cdot \frac{2}{5} = \frac{6}{25}$. Se manca anche questa volta, verrà colpito. Se decide di sparare a Sentenza, allora, la sua probabilità di sopravvivere è $\frac{6}{25} = 24\%$.
Se Tuco spara al Biondo e lo manca (con probabilità $\frac{3}{5}$), il Biondo sparerà a Sentenza e Tuco avrà una seconda chance di sopravvivere, sparando di nuovo al Biondo e colpendolo (in questa sequenza) con probabilità $\frac{6}{25}$. Se invece lo colpisce, con probabilità $\frac{2}{5}$, comincerà un duello con Sentenza in cui la probabilità che Tuco sopravviva è

$$\frac{1}{2} \cdot \frac{2}{5} + \frac{1}{2} \cdot \frac{3}{5} \cdot \frac{1}{2} \cdot \frac{2}{5} + \cdots = \frac{1}{5} \cdot \sum_{k=0}^{+\infty} \left(\frac{3}{10}\right)^k = \frac{2}{7}.$$

Sparando al Biondo la sua probabilità di sopravvivenza è $\frac{6}{25} + \frac{2}{5} \cdot \frac{2}{7} = \frac{62}{175} \approx 35\%$.
Tuco sta per sparare al Biondo, pensando con rammarico che ha meno chance di sopravvivere così che non in un duello standard con il Biondo, quando si rende conto di aver trovato la strategia vincente e spara in aria, garantendosi una probabilità di successo del 40% al prossimo giro, quando avrà la sua unica possibilità di colpire il Biondo. □

Soluzione al Problema 42

Il meteo nei quattro giorni è indipendente, quindi possiamo calcolare la probabilità che Giorgia non si bagni in un giorno ed elevarla alla quarta. Per calcolare la probabilità di non bagnarsi in un giorno, consideriamo il suo complementare: Giorgia si bagna solamente se piove e c'è vento, cosa che accade con probabilità $\frac{3}{5} \cdot \frac{1}{2} = \frac{3}{10}$, usando l'ipotesi di indipendenza tra pioggia e vento.
La probabilità di non bagnarsi in un singolo giorno è allora $\frac{7}{10}$ e quella di non bagnarsi per quattro giorni consecutivi è $\frac{2401}{10\,000} \approx 24\%$. □

Soluzione al Problema 43

Se Judit volesse massimizzare il numero di partite vinte, le converrebbe giocare due volte contro il padre e una sola volta contro la sorella, dal momento che $\frac{3}{5} < \frac{7}{9}$. La richiesta del problema, però, è un'altra: deve vincerne due di seguito.

Per poterlo fare, Judit deve assolutamente vincere la partita centrale, quindi le conviene affrontare in quella Lázló e nelle altre due Zsuzsanna. Se gioca in questo modo la sua probabilità di vittoria è

$$\frac{3}{5} \cdot \frac{7}{9} \cdot \frac{3}{5} + \frac{3}{5} \cdot \frac{7}{9} \cdot \frac{2}{5} + \frac{2}{5} \cdot \frac{7}{9} \cdot \frac{3}{5} = \frac{49}{75} \approx 65\%,$$

mentre con l'altra sequenza è

$$\frac{7}{9} \cdot \frac{3}{5} \cdot \frac{7}{9} + \frac{7}{9} \cdot \frac{3}{5} \cdot \frac{2}{9} + \frac{2}{9} \cdot \frac{3}{5} \cdot \frac{7}{9} = \frac{77}{135} \approx 57\%. \qquad \square$$

Soluzione al Problema 44

Consideriamo un problema più generale: un aereo con p posti e p passeggeri, di cui i primi $p-1$ si siedono a caso e il passeggero p pretende di avere il proprio posto. Cominciamo a calcolare la probabilità, per p generico, che nessuno dei primi $p-1$ passeggeri sia seduto al posto numero p. I primi $p-1$ passeggeri possono scegliere i loro posti in $p!$ modi. Quelli tra loro che non occupano il posto numero p sono $(p-1)!$; quindi la probabilità che il posto p sia libero è $\frac{1}{p}$.

Cerchiamo una formula ricorsiva per il numero medio N_p di passeggeri che si alzano su un aereo con p posti. Chiaramente per $p = 1$ non c'è bisogno di fare alzare nessuno, quindi $N_1 = 0$. Se $p = 2$ ci sono due posti e due passeggeri. Il primo passeggero occuperà il proprio posto con probabilità $\frac{1}{2}$: in questo caso nessuno dovrà alzarsi. Sempre con probabilità $\frac{1}{2}$ il primo passeggero avrà sbagliato sedile e dovrà alzarsi. In media si alza $\frac{1}{2}$ persona. Vediamo anche il terzo caso. Con probabilità $\frac{1}{3}$ nessuno si deve alzare, mentre con probabilità $\frac{2}{3}$ ci ritroviamo nel caso precedente, con un passeggero che si è già alzato, quindi $N_3 = \frac{2}{3}(1 + N_2)$. Ora abbiamo capito la nostra formula ricorsiva: per $p > 1$ generico,

$$N_p = \frac{p-1}{p} \cdot (1 + N_{p-1}),$$

da cui

$$\begin{aligned} N_p &= \frac{p-1}{p} \left[1 + \frac{p-2}{p-1}(1 + N_{p-2}) \right] \\ &= \frac{p-1}{p} + \frac{p-2}{p} + \frac{p-2}{p} \cdot N_{p-2} \\ &= \cdots \\ &= \frac{1}{p} \cdot \sum_{k=1}^{p-1} k = \frac{p-1}{2}, \end{aligned}$$

in cui abbiamo usato il fatto che $N_1 = 1$. Quindi nel caso $p = 121$ in media si alzeranno 60 persone. □

Soluzione al Problema 45

A seconda di quali sono le prime due palline pescate, possiamo ritrovarci con 0, 1, 2 o 3 palline con il numero 8. Proviamo a calcolare quali sono le probabilità di queste configurazioni, contando quante delle 56 possibili estrazioni ordinate delle prime due palline danno luogo alle varie situazioni. Cominciamo col caso più facile: abbiamo un solo modo per avere 3 palline col numero 8: dobbiamo estrarre, nell'ordine, 4 e 2. Per avere 2 palline col numero 8 abbiamo due possibili vie: peschiamo il 4 come prima e non peschiamo né il 2 né l'8 (né il 4, ovviamente) come seconda (possiamo farlo in 5 modi distinti) oppure peschiamo una biglia che non sia 2, 4 o 8 come prima e la 2 come seconda (anche questo possiamo farlo in 5 modi distinti). Vediamo in quanti modi possiamo ritrovarci senza alcun 8: dobbiamo pescare l'8 come prima e non pescare il 2 come seconda (in 6 modi), oppure non pescare né 4 né 8 come prima e l'8 come seconda (anche questo in 6 modi). Non rimangono che i casi in cui c'è una sola pallina 8, che possiamo contare sottraendo quelli ottenuti finora da 56: $56 - 1 - 5 - 5 - 6 - 6 = 33$.
A questo punto possiamo sommare le probabilità nei vari casi (distinti) in cui abbiamo 3, 2 e 1 pallina 8,

$$\frac{1}{56} \cdot \frac{3}{8} + \frac{10}{56} \cdot \frac{2}{8} + \frac{33}{56} \cdot \frac{1}{8} = \frac{3 + 20 + 33}{56} \cdot \frac{1}{8} = \frac{1}{8}.$$

Seguendo questa strada è facile incappare in errori di calcolo. Proviamo a iniziare dalla fine: se abbiamo pescato una pallina 8, questa può essere quella originale, oppure un 2 o un 4 trasformato. Nel primo caso, necessariamente la pallina non è stata estratta tra le prime due, quindi la probabilità è $\frac{6}{8} \cdot \frac{1}{8}$, prodotto delle probabilità di non averla pescata all'inizio e averla pescata all'ultima estrazione.
Se l'8 era in precedenza un 2, vuol dire che è stato pescato alla seconda estrazione e poi ancora alla terza, con una probabilità $(\frac{1}{8})^2$. Se prima era un 4, allora è stato pescato alla prima e alla terza estrazione, anche in questo caso con probabilità $(\frac{1}{8})^2$.
Mettendo assieme i contributi abbiamo $\frac{1}{8} \cdot (\frac{6}{8} + \frac{1}{8} + \frac{1}{8}) = \frac{1}{8}$. □

Soluzione al Problema 46

L'unica sequenza di eventi che ci va bene è quella in cui nessun ufficio rifiuta il modulo. Il k-esimo ufficio non rifiuta il modulo con probabilità $(1 - \frac{1}{k^2})$, indipendentemente dagli altri, quindi la probabilità che nessun ufficio rifiuti è il prodotto di questi contributi al variare di k tra 2 e 31, ossia $\prod_{k=2}^{31}(1 - \frac{1}{k^2})$. Non resta che fare i conti; poiché non abbiamo molta voglia di farli, cercheremo di limitarli il più possibile. I fattori sono della forma

$$1 - \frac{1}{k^2} = \frac{k^2 - 1}{k^2} = \frac{(k-1) \cdot (k+1)}{k^2}$$

e se scriviamo tre fattori consecutivi abbiamo

$$\begin{aligned}\left(1 - \frac{1}{(k-1)^2}\right)\left(1 - \frac{1}{k^2}\right)\left(1 - \frac{1}{(k+1)^2}\right) &= \frac{(k-2)k}{(k-1)^2} \cdot \frac{(k-1)(k+1)}{k^2} \cdot \frac{k(k+2)}{(k+1)^2} \\ &= \frac{k-2}{k-1} \cdot 1 \cdot \frac{k+2}{k+1}.\end{aligned}$$

I fattori si semplificano tra loro, quindi di tutto il prodotto rimarranno solo alcuni termini dal primo e dall'ultimo fattore, cioè

$$\prod_{k=2}^{31}\left(1 - \frac{1}{k^2}\right) = \frac{1}{2} \cdot \frac{32}{31} = \frac{16}{31} \approx 51{,}6\%. \qquad \square$$

Soluzione al Problema 47

Il numero 2020 sembra assolutamente arbitrario, quindi proviamo a vedere alcuni casi più semplici. Partiamo con il primo caso non banale, quello in cui abbiamo due coppie di punti nell'intervallo unitario. Proviamo anche a riformulare il problema: siccome ogni punto viene preso indipendentemente dagli altri, possiamo pensare che Daniele prenda 4 punti uniformemente a caso nell'intervallo $[0, 1]$ appaiandoli poi a caso. Tutti i modi di appaiare i punti hanno la stessa probabilità, quindi stiamo trasformando il problema in un problema di combinatoria, nel quale andremo a fare il rapporto tra il numero di configurazioni favorevoli e quello di tutte le configurazioni possibili. In particolare, possiamo notare che usando questo punto di vista non ci interessano le lunghezze dei vari segmenti, né il fatto che i punti siano nell'intervallo unitario: possiamo pensarli come 4 punti equidistanziati. Inoltre non serve distinguere l'ordine nelle coppie di estremi: ci interessano le coppie non ordinate.

Con 4 punti, che chiamiamo 1, 2, 3 e 4, abbiamo tre possibilità di appaiare tra loro i punti: $\{\{1, 2\}, \{3, 4\}\}$, $\{\{1, 3\}, \{2, 4\}\}$ e $\{\{1, 4\}, \{2, 3\}\}$. Di queste tre, solo la prima non ha alcun intervallo che interseca tutti gli altri (cioè l'altro). Quindi la probabilità cercata in questo caso è $\frac{2}{3}$.

Passiamo al caso successivo, in cui abbiamo 6 punti, che chiamiamo $1, 2, \ldots, 6$. Anche qui elenchiamo tutti i modi possibili di appaiare i punti: sono 15, perché abbiamo 6! modi di ordinare i punti, da dividere per $3 \cdot 2!$ perché per ogni coppia non importa l'ordine e ancora per 3! perché non ci interessa l'ordine delle tre coppie. Segniamo con una spunta (✓) quelli che soddisfano la condizione di intersezione e con una croce (✗) quelli che non la soddisfano:

$$\{1,2\}\begin{cases}\{3,4\},\{5,6\} & ✗\\ \{3,5\},\{4,6\} & ✗\\ \{3,6\},\{4,5\} & ✗\end{cases}\quad \{1,3\}\begin{cases}\{2,4\},\{5,6\} & ✗\\ \{2,5\},\{4,6\} & ✓\\ \{2,6\},\{4,5\} & ✓\end{cases}\quad \{1,4\}\begin{cases}\{2,3\},\{5,6\} & ✗\\ \{2,5\},\{3,6\} & ✓\\ \{2,6\},\{3,5\} & ✓\end{cases}$$

$$\{1,5\}\begin{cases}\{2,3\},\{4,6\} & ✓\\ \{2,4\},\{3,6\} & ✓\\ \{2,6\},\{3,4\} & ✓\end{cases}\quad \{1,6\}\begin{cases}\{2,3\},\{4,5\} & ✓\\ \{2,4\},\{3,5\} & ✓\\ \{2,5\},\{3,4\} & ✓\end{cases}$$

In questo elenco possiamo leggere due cose: da un lato anche per sei punti la probabilità di avere un intervallo che interseca tutti gli altri è $\frac{2}{3}$. La cosa ci fa pensare che si tratti di un valore costante indipendente dal numero di punti. Dall'altro, il modo in cui i casi favorevoli e quelli non favorevoli sono distribuiti nell'elenco non ci piace molto: proviamo a scriverli in modo diverso, partendo (quasi) dal centro, cioè dal punto 3:

$$\{3,1\}\begin{cases}\{2,4\},\{5,6\} & ✗\\ \{2,5\},\{4,6\} & ✓\\ \{2,6\},\{4,5\} & ✓\end{cases}\quad \{3,2\}\begin{cases}\{1,4\},\{5,6\} & ✗\\ \{1,5\},\{4,6\} & ✓\\ \{1,6\},\{4,5\} & ✓\end{cases}\quad \{3,4\}\begin{cases}\{1,2\},\{5,6\} & ✗\\ \{1,5\},\{2,6\} & ✓\\ \{1,6\},\{2,5\} & ✓\end{cases}$$

$$\{3,5\}\begin{cases}\{1,2\},\{4,6\} & ✗\\ \{1,4\},\{2,6\} & ✓\\ \{1,6\},\{2,4\} & ✓\end{cases}\quad \{3,6\}\begin{cases}\{1,2\},\{4,5\} & ✗\\ \{1,4\},\{2,5\} & ✓\\ \{1,5\},\{2,4\} & ✓\end{cases}$$

In questo modo, qualunque sia il compagno di 3, abbiamo sempre un appaiamento dei rimanenti quattro punti che non soddisfa la condizione di intersezione e due che la soddisfano. Ora, per dimostrare che qualunque sia il numero n di intervalli la probabilità cercata è uguale a $\frac{2}{3}$, lasciamoci ispirare da quest'ultimo elenco dei casi per $n = 3$ e proviamo a procedere in modo analogo, partendo dal centro, qualunque sia n.

Se Daniele ha preso n intervalli, abbiamo $2n$ punti da appaiare. Chiamiamo *Inizio* l'insieme dei primi n punti, $\{1, \ldots, n\}$ e *Fine* l'insieme degli ultimi n punti, ossia $\{n+1, \ldots, 2n\}$. Anche se gli appaiamenti sono casuali, vogliamo mostrare un algoritmo per navigare tra i vari intervalli in modo deterministico: in sostanza avremo una regola deterministica per prendere uno dei due estremi dell'intervallo tra i punti disponibili, mentre il secondo estremo sarà casuale. Vediamo come fare. Cominciamo prendendo come primo punto $a_1 = n$: a_1 sarà appaiato, cioè formerà un intervallo, con un certo altro punto b_1. Questo era il nostro caso base. Ora proseguiamo in modo induttivo come segue: supponiamo di avere determinato i punti $a_1, \ldots, a_k$ e i loro compagni $b_1, \ldots, b_k$ e di voler scegliere a_{k+1}. Vogliamo rimanere il più possibile vicino al centro, quindi se l'ultimo punto che abbiamo preso, cioè b_k, era all'Inizio, prendiamo come a_{k+1} il più piccolo tra i punti ancora liberi (cioè non appaiati) della Fine; viceversa, se b_k era alla Fine, prendiamo come a_{k+1} il più grande dei punti liberi dell'Inizio. A questo punto appaiamo a_{k+1} con l'altro estremo dell'intervallo di cui è lui stesso estremo e chiamiamo questo punto b_{k+1}.

Prima di continuare facciamo un paio di osservazioni. Per come stiamo prendendo gli a_k, non ci saranno ad alcun istante spazi vuoti tra il più piccolo e il più grande degli a incontrati fino a quel punto: non è detto che gli a siano tutti in fila (potrebbero esserci dei b in mezzo), ma sicuramente non ci sono numeri non appaiati in mezzo a loro. Inoltre, chiamiamo *crescenti* quegli intervalli in cui $a_k < b_k$ e *decrescenti* quelli in cui $a_k > b_k$. Per come li stiamo enumerando, il k-esimo intervallo è crescente se e solo se b_k è alla Fine. Se a un certo passo del nostro algoritmo i punti liberi all'Inizio sono tanti quanti i punti liberi alla Fine, questa situazione bilanciata non cambierà fintanto che continuiamo a incontrare intervalli crescenti. Nel momento in cui troviamo un intervallo decrescente, l'Inizio avrà due punti liberi in meno rispetto alla Fine. Questa situazione rimane invariata se continuiamo a incontrare intervalli decrescenti, mentre torna all'equilibrio non appena troviamo un intervallo crescente.

Continuiamo con il nostro algoritmo fino a che non restano solo 4 punti liberi, cosa che accade dopo $n-2$ iterazioni. Chiamiamo questi punti $\alpha_1 < \alpha_2 < \alpha_3 < \alpha_4$. Ci sono due sole possibilità: se l'ultimo intervallo incontrato (l'intervallo $n-2$) era crescente, α_1 e α_2 sono all'Inizio e α_3 e α_4 sono alla Fine; viceversa se l'intervallo $n-2$ era decrescente, solo α_1 è all'Inizio e tutti gli altri tre punti liberi sono alla Fine. Non ci restano che due appaiamenti da fare, con questi quattro punti liberi; possiamo farli in tre modi: $\{[\alpha_1, \alpha_2], [\alpha_3, \alpha_4]\}$, $\{[\alpha_1, \alpha_3], [\alpha_2, \alpha_4]\}$ e $\{[\alpha_1, \alpha_4], [\alpha_2, \alpha_3]\}$.

Nel secondo e nel terzo caso abbiamo certamente un intervallo che interseca tutti gli altri. Infatti $\alpha_1 < a_k < \alpha_3$ per ogni $k = 1, \ldots, n-2$, dal momento che non ci sono buchi tra gli a_k e che α_1 è sicuramente all'Inizio e α_3 è sicuramente alla Fine e quindi $[\alpha_1, \alpha_3]$ e $[\alpha_1, \alpha_4]$ intersecano tutti gli altri intervalli. Per dimostrare che la probabilità cercata è proprio $\frac{2}{3}$ non ci resta che far vedere che nel primo caso non ci può essere alcun intervallo che interseca tutti gli altri.

Se ci fosse un tale intervallo, sicuramente non è $[\alpha_1, \alpha_2]$ o $[\alpha_3, \alpha_4]$, poiché questi non si intersecano tra loro. Dovrebbe essere allora uno degli intervalli incontrati in precedenza. Ma se l'intervallo $n-2$ è crescente, questo non può essere: dal momento che tutti gli a_k con $k = 1, \ldots, n-2$ sono compresi tra α_2 e α_3, ciascuno

dei primi $n-2$ intervalli può intersecare al più uno tra i due intervalli finali. Non resta che considerare il caso in cui l'intervallo $n-2$ è decrescente. Supponiamo che l'intervallo che interseca tutti gli altri sia quello incontrato come j-esimo: siccome esso deve intersecare tutti gli intervalli, deve in particolare intersecare $[\alpha_3, \alpha_4]$, quindi deve essere crescente, perché il suo b_j deve stare alla Fine. Tra j ed $n-2$ quindi ci deve essere almeno un istante l con $j \leqslant l \leqslant n-2$ tale che l'intervallo $l-1$ è crescente, mentre l'intervallo l è crescente. Se $l-1$ è un intervallo crescente, a_l sta all'Inizio e in particolare $a_l < n$, poiché $l \neq 1$, ma anche $a_l < a_j$, poiché $l > j$. Ma l'intervallo l è decrescente e l'intervallo j è crescente, quindi $b_l < a_l < a_j < b_j$ e i due intervalli non si intersecano, in contraddizione coll'ipotesi che j fosse l'intervallo che intersecava tutti gli altri.
La probabilità che ci sia un intervallo che interseca tutti gli altri, allora, è indipendente da $n > 1$ ed è uguale a $\frac{2}{3}$. □

Soluzione al Problema 48

Affrontiamo il problema usando come punto di partenza le porzioni: ce ne sono 7 (una per ogni giocatore) più quella aggiuntiva per il capitano e le due aggiuntive per il consegnatore, per un totale di 10 porzioni. Consideriamo i peperoncini come una quadrupla ordinata (p_1, p_2, p_3, p_4): possono essere distribuiti nelle porzioni in 10^4 modi possibili, 10 per ogni peperoncino. Questi saranno i nostri casi totali: poiché abbiamo usato le porzioni e "ordinato" i peperoncini, i casi sono tutti equiprobabili tra loro. Questo non sarebbe accaduto se ci fossimo concentrati sui piatti e non sulle porzioni.
Nei casi favorevoli potremmo avere più di un peperoncino per porzione, cosa che complica il conto delle possibili combinazioni. Consideriamo quindi i casi in cui le cose vanno male, così da avere al più un peperoncino per porzione. Questo non elimina completamente le difficoltà di conteggio, perché dobbiamo comunque tenere traccia dei casi speciali del capitano e del consegnatore, visto che il problema chiede di rispondere in termini di piatti e non di porzioni.
Siccome i peperoncini sono in porzioni diverse e non vogliamo che capitano o consegnatore ne abbiano più di uno, abbiamo quattro situazioni possibili, disgiunte tra loro, che consideriamo separatamente.

1. Capitano e consegnatore hanno un peperoncino ciascuno e i 2 rimanenti sono in 2 delle altre 5 porzioni. In quanti modi può succedere? Possiamo scegliere il peperoncino del consegnatore in 4 modi: questo può finire in una qualsiasi delle sue porzioni quindi possiamo scegliere la porzione "piccante" in 3 modi. Indipendentemente (quindi moltiplicando) possiamo scegliere il peperoncino del capitano in 3 modi e la corrispondente porzione in 2 modi.

Per i due peperoncini rimanenti abbiamo rispettivamente 5 e 4 possibilità, per un totale di $4 \cdot 3 \cdot 3 \cdot 2 \cdot 5 \cdot 4 = 1440$ casi. Osserviamo che scegliere prima il peperoncino del capitano e poi quello del consegnatore sarebbe del tutto equivalente: cambierebbe solo l'ordine dei primi fattori.

2. Il consegnatore ha un peperoncino e tutti gli altri sono in porzioni normali distinte. Il consegnatore contribuisce con $4 \cdot 3$ casi, mentre il resto della squadra (capitano escluso) dà un fattore $5 \cdot 4 \cdot 3$. Complessivamente abbiamo allora 720 casi.
3. Il capitano ha un peperoncino e tutti gli altri sono in porzioni normali distinte. Abbiamo ormai capito come funziona: i casi sono $4 \cdot 2 \cdot 5 \cdot 4 \cdot 3 = 480$.
4. Consegnatore e capitano restano all'asciutto e tutti i peperoncini sono in porzioni normali distinte. In questa situazione abbiamo $5 \cdot 4 \cdot 3 \cdot 2 = 120$ casi possibili.

Per concludere non ci resta che sommare i contributi delle quattro situazioni e sottrarli da 10^4 per trovare il numero di casi favorevoli. La somma è 7240, che divideremo per 10 000 (i casi totali) per ricavare la probabilità: $\frac{181}{250} = 72{,}4\%$. □

Soluzione al Problema 49

Come prima cosa riscriviamo il problema in una forma più facile da rappresentare: possiamo vedere le configurazioni di uscita dei 5 amici dalla seggiovia come stringhe di 5 caratteri pescati nell'alfabeto $\{-1, 0, 1\}$ a seconda che decidano di andare, rispettivamente, a sinistra, dritto o a destra. Poiché ciascun amico sceglie uniformemente e indipendentemente dagli altri, tutte le stringhe hanno la medesima probabilità, uguale a $\frac{1}{3^5}$. Ogni stringa è un esito, quindi possiamo considerare l'insieme di tutte le stringhe senza collisioni e moltiplicare la sua cardinalità per la probabilità appena trovata.

Il problema è diventato dunque un esercizio di combinatoria: dobbiamo solo capire quali e quante siano le stringhe che vanno bene. Per non avere scontri tra gli sciatori non ci devono essere intersezioni fra le traiettorie: tutti i -1 devono venire prima degli 0 che a loro volta devono venire prima degli 1.

Una possibile strategia per risolverlo è la seguente: cerchiamo di riformulare quest'ultima richiesta in un modo diverso, per ricondurci a un problema noto. Possiamo vederla nel modo seguente: come possiamo scrivere 5 come somma di tre numeri naturali, cioè il numero di -1, 0 e 1? Come già visto nel Problema 3, possiamo fare questa somma in $\binom{5+3-1}{3-1} = \binom{7}{2} = 21$ modi.

La probabilità cercata è allora $21 \cdot \frac{1}{243} = \frac{7}{81}$. □

Soluzione al Problema 50

Questo è proprio un problema olimpico: la probabilità coinvolta non è tantissima, ma per risolverlo serve avere un po' di occhio. Quindi è sicuramente un buon esercizio di allenamento. Cominciamo estraendo dal testo qualche informazione. Come prima cosa, notiamo che la probabilità di arrivare su Banahch-Torsk è esattamente $\frac{1}{2016}$, ma che allo stesso tempo ogni percorso che possiamo fare tra i pianeti ha una probabilità della forma $\frac{1}{2^k}$ per qualche k naturale. Dovremo quindi trovare un modo di scrivere $\frac{1}{2016}$ come somma (eventualmente infinita) di potenze di $\frac{1}{2}$. Parcheggiamo per un attimo questo pensiero e continuiamo a leggere il testo. Il viaggio di Luke può finire solo o su Taodana o su Banahch-Torsk, quindi la probabilità che finisca su Taodana è $\frac{2015}{2016}$. Questo confronto ci dice, a livello molto informale, che le strade che portano a Taodana sono molto più numerose di quelle che portano a Banahch-Torsk.

Inoltre la richiesta che per ogni pianeta ci sia al più una strada che riporta al punto di partenza ci dice che una volta che entriamo in un ciclo (cioè una strada che ci riporta al punto dove siamo), ogni pianeta del ciclo avrà una rotta in uscita per il successivo pianeta del ciclo e una seconda che non si può riconnettere al ciclo stesso e che quindi finirà prima o poi in Banahch-Torsk o Taodana. Siccome la probabilità di finire in Taodana è molto più alta, come detto, ci aspettiamo che la maggior parte di questi percorsi laterali porti verso Taodana. Non solo, visto che vogliamo minimizzare il numero n di pianeti, vorremo che quasi tutti questi percorsi laterali finiscano direttamente in Taodana.

Attenzione, non stiamo ancora scrivendo la soluzione: stiamo cercando di capire quale possa essere la struttura del problema e quindi della soluzione. In questa fase di esplorazione possiamo permetterci di essere imprecisi, a patto di sistemare bene tutti questi dettagli nel momento in cui scriviamo la soluzione vera e propria. Per capire se ci sono dei cicli o no, dobbiamo andare a scrivere $\frac{1}{2016}$ come somma di potenze di $\frac{1}{2}$: se ci occorre una somma infinita avremo dei cicli, altrimenti possiamo farne a meno. Consideriamo con attenzione il numero 2016: non dista moltissimo da una potenza di 2, visto che $2016 = 2048 - 32$. Raccogliendo possiamo scrivere come $2016 = 2^{11}(1 - 2^{-6})$; quindi, tornando alla probabilità,

$$\frac{1}{2016} = \frac{1}{2^{11}(1 - 2^{-6})} = \frac{1}{2^{11}} \cdot \frac{1}{1 - \frac{1}{2^6}}.$$

Nell'ultimo termine, il secondo fattore ci ricorda la somma di una serie geometrica di ragione $\frac{1}{2^6}$. In altre parole, stiamo dicendo che possiamo scrivere $\frac{1}{2016}$ come

$$\frac{1}{2016} = \frac{1}{2^{11}} + \frac{1}{2^{11}} \cdot \frac{1}{2^6} + \frac{1}{2^{11}} \cdot \left(\frac{1}{2^6}\right)^2 + \cdots .$$

Come leggiamo questo risultato? Innanzitutto osserviamo che Banahch-Torsk non può essere raggiunto prima dell'undicesimo passo: infatti abbiamo 11 scelte da fare prima di poterci arrivare. Nel passo precedente, se non arriviamo a Banahch-Torsk, dobbiamo finire in un altro pianeta, dal quale possiamo ritornare (per questioni di economia nel numero di pianeti) a questo bivio. Senza bisogno di aggiungere altri pianeti, possiamo ipotizzare che il cammino ci riporti sul quinto pianeta che abbiamo toccato dopo Coruscantor, dando origine a un ciclo che ci può riportare a Banahch-Torsk dopo altri 6 passi o altri 12 e così via.
Al momento abbiamo quindi 3 pianeti con nome e altri 10 che stanno sulla rotta tra Coruscantor e Banahch-Torsk. Escludendo l'ultimo pianeta toccato prima di arrivare a Banahch-Torsk, ogni rotta in uscita da un pianeta deve portarci in Taodana o in altri pianeti al di fuori di quelli già elencati. La scelta più economica è quindi che queste rotte portino su Taodana, lasciando il totale dei pianeti coinvolti a 13.
Nella Figura 6.11 mostriamo che questa configurazione è possibile e soddisfa tutte le nostre condizioni. □

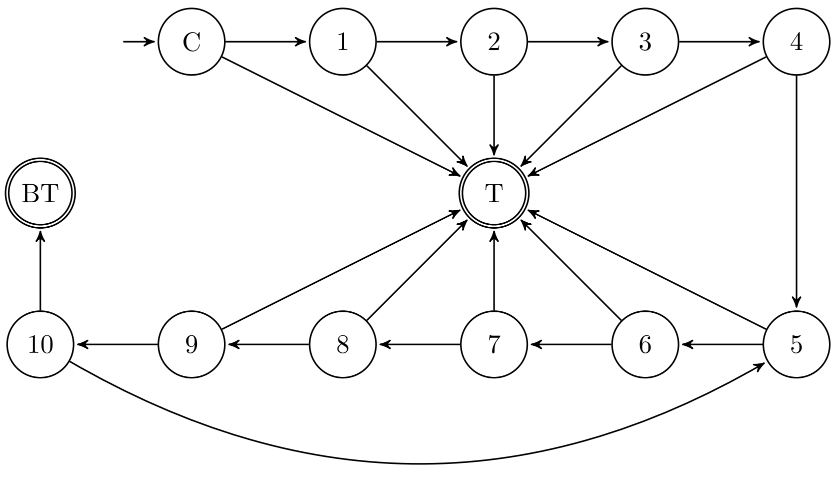

Figura 6.11: Viaggi tra i pianeti (identificati con le loro iniziali): tutte le frecce hanno probabilità $\frac{1}{2}$

A. Richiami vari

A.1 Teoria elementare degli insiemi

Un insieme può essere caratterizzato per estensione, andando a elencarne tutti gli elementi. È il modo forse più naturale, ma è possibile solamente se l'insieme è finito ed è pratico solo se l'insieme ha pochi elementi. In alternativa, possiamo caratterizzare un insieme mediante le proprietà soddisfatte da tutti e soli i suoi elementi. In questo caso parliamo di definizione intensiva.

Se però dobbiamo lavorare con più di un insieme, ci piacerebbe avere un modo per confrontarli e identificarli. Diciamo che due insiemi A e B sono uguali e scriviamo $A = B$ se ciascuno è sottoinsieme dell'altro, $A \subseteq B$ e $B \subseteq A$, ossia se tutti gli elementi di A sono anche elementi di B e viceversa.

D'altra parte ci sono, soprattutto in combinatoria, occasioni in cui non ci interessa sapere quali sono gli elementi di un insieme, ma solamente quanti sono. Di conseguenza vogliamo identificare due insiemi che abbiano lo stesso numero di elementi (anche se gli elementi non sono gli stessi). Da questo punto di vista un insieme con sei foglie non è diverso da un insieme con sei palline o con sei punti. Questo è molto "matematico": estraiamo e astraiamo dagli oggetti solo quelle proprietà che ci interessano, ignorando tutte le altre.

Chiamiamo il numero di elementi di un insieme A *cardinalità* di A e lo indichiamo con la notazione $\#A$. Finché abbiamo a che fare con insiemi finiti, non ci sono troppi problemi; ma nel momento in cui passiamo a insiemi infiniti, abbiamo bisogno

di un po' più di precisione. Diciamo allora che due insiemi A e B hanno la stessa cardinalità, cioè sono *equipotenti*, se esiste una funzione biettiva $f : A \to B$. In particolare un insieme equipotente all'insieme $\mathbb{N}$ dei numeri naturali ha cardinalità (infinita) numerabile e questa quantità è denotata con $\aleph_0$, il primo dei numeri cardinali (cioè usati per indicare le cardinalità) infiniti[1].
Se, dati due insiemi A e B, esiste una funzione iniettiva $f : A \to B$, allora $\#A \leqslant \#B$. Se inoltre non esiste una funzione biettiva tra i due, possiamo dire che $\#A < \#B$. Attenzione: nel momento in cui iniziamo a maneggiare gli infiniti dobbiamo procedere con estrema cautela. Infatti non è necessariamente vero (all'interno della teoria degli insiemi) che valga la proprietà di tricotomia, ossia che dati due insiemi debba essere vera una delle seguenti: $\#A < \#B$, $\#A = \#B$ o $\#B < \#A$. Il problema si ha con i cardinali infiniti e, in particolare, la tricotomia equivale all'assioma della scelta.

Esempio A.1.1. Consideriamo l'insieme $2\mathbb{N}$ dei numeri naturali pari. Esso è un sottoinsieme proprio dell'insieme $\mathbb{N}$ dei numeri naturali. Tuttavia $2\mathbb{N}$ e $\mathbb{N}$ hanno la stessa cardinalità. Infatti possiamo prendere $f : \mathbb{N} \to 2\mathbb{N}$ tale che $f(n) = 2n$. La funzione f è biettiva: la sua inversa è $f^{-1} : 2\mathbb{N} \to \mathbb{N}$ con $f^{-1}(2m) = m$. Allora i due insiemi $2\mathbb{N}$ e $\mathbb{N}$ sono equipotenti, ossia ci sono tanti numeri naturali pari quanti numeri naturali.

Dobbiamo quindi procedere con cautela, come mostrato anche dal seguente risultato.

Teorema A.1.1 (Cantor[2]-Bernstein[3])**.** *Dati due insiemi A e B, se esistono due funzioni iniettive $f : A \to B$ e $g : B \to A$, allora esiste almeno una funzione biettiva tra i due insiemi. In altri termini, se $\#A \leqslant \#B$ e $\#B \leqslant \#A$, allora $\#A = \#B$.*

Questa affermazione sulla cardinalità di due insiemi sembra assolutamente ovvia. Ma, come abbiamo visto nell'Esempio A.1.1, l'uso degli infiniti può trarre in inganno. Perciò il teorema è necessario; la sua dimostrazione, comunque, non è per niente banale.

Esempio A.1.2. Dato un insieme A, le funzioni da A a $\{0, 1\}$ formano un insieme di cardinalità $2^{\#A}$. Infatti una funzione da A a $\{0, 1\}$ associa a ogni elemento di A uno tra 0 e 1 e la scelta per un elemento di A non influenza quella per gli

[1] Il fatto che $\aleph_0$ sia il primo cardinale infinito suggerisce che ce ne siano degli altri, più grandi. Così è, in effetti, e vedremo un esempio nelle prossime pagine.

[2] Georg Cantor (1845 – 1918).

[3] Felix Bernstein (1878 – 1956).

altri. Quindi abbiamo 2 scelte per ciascun elemento e i fattori 2 devono essere moltiplicati tra loro. Gli elementi di A sono $\#A$, da cui il risultato.

In generale possiamo dire qualcosa di più: le funzioni da un insieme A a un insieme B formano un insieme di cardinalità $\#B^{\#A}$. Per questo motivo l'insieme delle funzioni da A a B si scrive B^A.

Proposizione A.1.2. *Dato un insieme A di cardinalità eventualmente infinita (anche più che numerabile) l'insieme delle parti di A (o insieme potenza di A) $\mathcal{P}(A)$ ha cardinalità $\#\mathcal{P}(A) = 2^{\#A}$.*

Dimostrazione. Come detto in precedenza, per mostrare che un insieme ha una certa cardinalità, quello che possiamo fare è costruire una relazione biunivoca (o una codifica) dal nostro insieme a un insieme che sappiamo avere la cardinalità cercata. Sappiamo anche che un insieme che ha proprio $2^{\#A}$ elementi è l'insieme delle funzioni da A in $\{0,1\}$. Quello che ci resta da fare, dunque, è far vedere che i sottoinsiemi di A sono tanti quanti le funzioni da A in $\{0,1\}$. Per ogni sottoinsieme $S \subseteq A$ definiamo la funzione $f_S : A \to \{0,1\}$ come segue: $f_S(a) = \mathbb{1}_S(a)$. In sostanza, codifichiamo con un 1 la presenza dell'elemento nel sottoinsieme, con 0 la sua assenza. Viceversa per ogni funzione $f : A \to \{0,1\}$ possiamo definire $S_f = f^{-1}(1)$, cioè il sottoinsieme di Ω contenente tutti gli a per cui $f(a) = 1$. Si verifica facilmente che le relazioni $S \to f_S$ e $f \to S_f$ sono entrambe iniettive[4], quindi $\#\mathcal{P}(A) \leqslant 2^{\#A} \leqslant \#\mathcal{P}(A)$ (Teorema di Cantor-Bernstein) e abbiamo l'uguaglianza cercata. $\square$

Teorema A.1.3 (Cantor)**.** *Non esiste alcuna funzione suriettiva da un insieme A al suo insieme delle parti $\mathcal{P}(A)$. In particolare, quindi, $\#A < \#\mathcal{P}(A)$.*

Dimostrazione. Cominciamo osservando che $\mathcal{P}(A)$ contiene una copia di A: la funzione $i : A \to \mathcal{P}(A)$ che manda ogni elemento di A nel suo singoletto è iniettiva, quindi $\#A \leqslant \#\mathcal{P}(A)$.

Procediamo ora per assurdo e supponiamo di avere una funzione $f : A \to \mathcal{P}(A)$ suriettiva. Consideriamo l'insieme $N = \{a \in A \,:\, a \notin f(a)\}$ degli elementi di A che non appartengono alla propria immagine mediante f. Dal momento che f è suriettiva su $\mathcal{P}(A)$, esiste un elemento $\alpha \in A$ tale che $f(\alpha) = N$. A questo punto abbiamo una contraddizione: se $\alpha \in N$, allora dalla definizione di N segue $\alpha \notin f(\alpha) = N$. Se invece $\alpha \notin N$, allora $\alpha \in f(\alpha) = N$. Dunque non può esistere una funzione suriettiva da A a $\mathcal{P}(A)$.

[4]Sono anche suriettive e in particolare l'una è l'inversa dell'altra.

Se non possono esistere funzioni suriettive, non possono in particolare esistere funzioni biiettive: pertanto i due insiemi hanno cardinalità diversa e vale la disuguaglianza stretta. □

Proposizione A.1.4. *L'insieme $\mathcal{P}(\mathbb{N})$ ha cardinalità uguale a quella dei numeri reali.*

Dimostrazione. L'idea di Cantor che sta alla base di questa dimostrazione è far vedere che possiamo identificare i sottoinsiemi dei numeri naturali con i numeri reali nell'intervallo $[0, 1]$. Questo intervallo, a sua volta, ha tanti elementi quanti tutti i numeri reali.
Cominciamo con la prima parte. Per prima cosa, forti di quanto visto sopra, identifichiamo $\mathcal{P}(\mathbb{N})$ con l'insieme delle successioni binarie. Vogliamo interpretarle come rappresentazioni binarie dei numeri reali in $[0, 1]$, considerandole come se fossero le cifre dopo la virgola. In questo modo abbiamo tutti e soli[5] i numeri reali in $[0, 1]$.

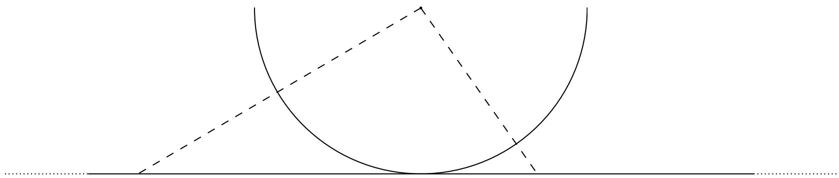

Figura A.1: Una biiezione tra $(0, 1)$ e $\mathbb{R}$

Ora vogliamo identificare $[0, 1]$ con l'intera retta reale. In realtà identifichiamo l'intervallo aperto $(0, 1)$ con la retta reale. Per fare ciò deformiamo il segmento senza estremi $(0, 1)$ in una semicirconferenza, anch'essa senza estremi. Prendiamo la retta reale e disegniamola in modo che sia tangente al punto medio della semicirconferenza. A questo punto possiamo tracciare le semirette uscenti dal centro della semicirconferenza che intersecano la semicirconferenza stessa, come rappresentato in Figura A.1. Ciascuna di esse incontra la retta reale in uno e un solo punto. Abbiamo così stabilito una relazione biunivoca tra ogni punto della

[5] In realtà stiamo un po' imbrogliando: come nel caso delle rappresentazioni decimali abbiamo il problema dei numeri che finiscono con un 9 periodico o con uno 0 periodico. Questi numeri vengono "contati" due volte, quindi abbiamo un problema simile con i numeri che finiscono con 1 periodico o con 0 periodico, che possiamo caratterizzare come i numeri con un numero finito di cifre. Tuttavia con un po' di accortezza siamo in grado di aggirare questo ostacolo, tenendo conto che questi numeri sono in quantità numerabile.

semicirconferenza (e quindi ogni numero nell'intervallo $(0,1)$) e ogni punto della retta reale (cioè ogni numero reale). $\square$

I risultati precedenti danno due informazioni interessanti riguardo ad alcuni insiemi che abbiamo appena visto.

Corollario A.1.5. *L'insieme $\mathbb{R}$ dei numeri reali ha cardinalità $2^{\aleph_0}$ strettamente maggiore della cardinalità $\aleph_0$ dell'insieme dei numeri naturali.*

Corollario A.1.6. *L'insieme $\mathcal{P}(\mathbb{R})$ ha cardinalità $2^{(2^{\aleph_0})}$, che in particolare è più grande di quella di $\mathbb{R}$.*

Dimostrazione. La prima parte segue dalla Proposizione A.1.2. La seconda parte dal Teorema A.1.3. $\square$

Due ultime curiosità, prima di andare oltre la cardinalità. Ci sono infinite cardinalità infinite, di cui $\aleph_0$ non è che la prima. Osserviamo però che non abbiamo detto che la cardinalità $2^{\aleph_0}$ dei reali (detta anche *continuo* e indicata con $\mathfrak{c}$) sia il secondo numero cardinale infinito (cioè $\aleph_1$). Non lo abbiamo fatto perché non è (necessariamente) vero: è la famosa *ipotesi del continuo.*

Teorema A.1.7 (Leggi di De Morgan[6])**.** *Se A e B sono due insiemi, valgono le seguenti identità:*

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

Vediamo ora alcuni modi di scrivere la differenza simmetrica tra due insiemi $A \triangle B$, definita come $A \triangle B = (A \setminus B) \cup (B \setminus A)$. La dimostrazione del prossimo risultato è un esercizio di teoria degli insiemi, che richiede le leggi di De Morgan.

Proposizione A.1.8. *Dati due insiemi A e B, i seguenti insiemi sono uguali: $A \triangle B$, $(A \cup B) \setminus (A \cap B)$, $(A \cup B) \cap (A \cap B)^c$, $(A \cup B) \cap (A^c \cup B^c)$, $(A \cap B^c) \cup (A^c \cap B)$, $A^c \triangle B^c$.*

La differenza simmetrica non è il solo insieme che potrà essere utile scrivere in modi diversi. C'è, per ogni insieme A, una particolare collezione di sue rappresentazioni diverse: la collezione delle partizioni di A. Dato un insieme A, una sua *partizione* è una famiglia $\mathcal{S}$ di sottoinsiemi di A tali che ogni elemento di A appartiene a uno e uno solo degli insiemi in $\mathcal{S}$. In altre parole, $\bigcup_{S \in \mathcal{S}} S = A$ ed è un'unione

[6] Augustus De Morgan (1806 – 1871).

disgiunta: due insiemi non coincidenti $S, T \in \mathcal{S}$ hanno intersezione vuota. Vale la pena osservare che nella definizione di partizione non è richiesto che A sia non vuoto. L'insieme vuoto ha un'unica partizione, l'insieme vuoto stesso[7].
Per tornare verso la combinatoria e la probabilità, possiamo chiederci quante siano, per un insieme finito, le partizioni possibili. Se abbiamo un insieme finito di cardinalità n, il numero delle sue partizioni è B_n, l'n-esimo numero di Bell[8] (come mostrato nel Problema 7).

A.2 La serie geometrica

Una successione, cioè una funzione dai numeri naturali a un qualche altro insieme, solitamente quello dei numeri reali, si dice *geometrica* se è costante il rapporto tra due elementi consecutivi della successione (detto *ragione*). In altre parole la serie è geometrica se $\frac{a_{n+1}}{a_n} = r$, indipendentemente da n.
Data una successione geometrica, possiamo considerare le sue somme parziali cioè, fissato m, la quantità

$$\sum_{i=0}^{m} a_i = a_0 \sum_{i=0}^{m} r^i = a_0 \left(\frac{1 - r^{m+1}}{1 - r} \right),$$

per $r \neq 1$. Se $r = 1$ la somma parziale dei primi $m + 1$ termini è banalmente $a_0(m + 1)$. Se il valore assoluto di r è minore di 1, la successione delle somme parziali converge e possiamo quindi parlare di *somma della serie geometrica*:

$$\sum_{i=0}^{+\infty} a_i = a_0 \sum_{i=0}^{+\infty} r^i = a_0 \left(\frac{1}{1 - r} \right).$$

A.3 Teorema di Carathéodory

Il Teorema di Carathéodory è lo strumento che permette di definire la probabilità su Ω dandone il valore su una piccola parte degli eventi e non su tutti quelli che stanno nella tribù. Per poter fare ciò, però, non possiamo prendere una famiglia qualunque di eventi, ma dobbiamo prenderne una abbastanza ricca. Una possibilità è quella di prendere un'algebra: se abbiamo una funzione su quest'algebra che si

[7] Si potrebbe fare un'osservazione filosofica che la partizione dell'insieme vuoto non è l'insieme vuoto stesso, ma è semplicemente a esso isomorfa: infatti nel secondo caso gli elementi che non sono nell'insieme vuoto sono loro stessi insiemi (i sottoinsiemi dell'insieme vuoto). Un bel grattacapo, che possiamo lasciare tranquillamente ai logici e ai teorici degli insiemi.

[8] Eric Temple Bell (1883 – 1960).

comporta come una probabilità, allora la possiamo estendere a una probabilità vera e propria definita sulla tribù generata da $\mathcal{A}$.

Teorema A.3.1 (Carathéodory)**.** *Dati un insieme* Ω*, un'algebra* $\mathcal{A}$ *di sottoinsiemi di* Ω *e una funzione* $P_0 : \mathcal{A} \to [0, 1]$ *che ha le proprietà di una probabilità, allora esiste un'unica probabilità* P *su* $(\Omega, \sigma(\mathcal{A}))$ *che coincide con* P_0 *su* $\mathcal{A}$*.*

Bibliografia

[1] Apollodoro, Guidorizzi G., *Biblioteca*, Adelphi, Milano 1995.

[2] Barsanti M., Conti F., De Lellis C., Franzoni T., *Le Olimpiadi della Matematica* (seconda edizione), Zanichelli, Bologna 2002.

[3] Callegari E., *Combinatoria per problemi*, Scienza Express, Trieste 2021.

[4] Canova P., Rizzuto D., *Fate il nostro gioco*, add editore, Torino 2016.

[5] Càssola C., *Geometria piana per le gare di matematica*, Scienza Express, Trieste 2018.

[6] Conti F., Barsanti M., Franzoni T., *Le Olimpiadi della Matematica* (prima edizione), Zanichelli, Bologna 1994.

[7] Damantino S., Campeotto E., *Aritmetica modulare*, Scienza Express, Trieste 2020.

[8] Ebert T., Merkle W., Vollmer H., *On the Autoreducibility of Random Sequences*, SIAM Journal on Computing, 32(6):1542–1569, 2003.

[9] Enzensberger H. M., Berner R. S., *Il mago dei numeri*, Einaudi, Torino 2017.

[10] Gobbino M., *Schede olimpiche*, Unione Matematica Italiana, Bologna 2012.

[11] van der Hoeven J., *The Jolly Writer: Your Guide to GNU TeXmacs*, Scypress, Chevreuse 2020.

[12] Justicz J., Scheinerman E. R., Winkler P. M., *Random Intervals*, The American Mathematical Monthly, 97(10):881–889, 1990.

[13] Kahneman D., *Pensieri lenti e veloci*, Mondadori, Milano 2013.

[14] Li Calzi M., *La matematica dell'incertezza*, il Mulino, Bologna 2016.

[15] Mlodinow L., *La passeggiata dell'ubriaco*, Rizzoli, Milano 2010.

[16] Mosteller F., *Fifty Challenging Problems in Probability with Solutions*, Dover Publications, New York 1987.

[17] Orwell G., *La fattoria degli animali*, Mondadori, Milano 2016.

[18] Paolini G., *La matematica delle Olimpiadi*, La Scuola, Brescia 2012.

[19] Pólya G., *Come risolvere i problemi di matematica*, UTET Università, Torino 2016.

[20] Schneps L., Colmez C., *Math on Trial*, Basic Books, New York 2013.

[21] Shan Z., Wang S., *Probability and Expectation*, East China Normal University Press; World Scientific, Shanghai; Singapore 2016.

[22] Trombetta M., *Calcolo combinatorio*, Scienza Express, Trieste 2018.

Indice

Finito di stampare nel gennaio 2021 da Tipografia Monteserra
via Torino, 12 - 56010 Vicopisano (PI)
tel. 050 799477 fax 050 796931
www.tipografiamonteserra.it
per conto di Scienza Express edizioni