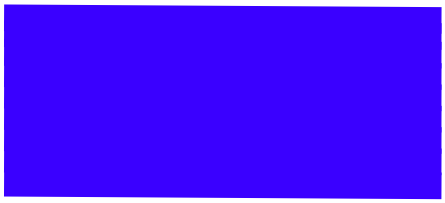
Francesco Caravenna Paolo Dai Pra

Introduzione alla Probabilità

Teoria con esempi ed esercizi

5 marzo 2012



Indice

Notazioni

Dato un insieme Ω e due suoi sottoinsiemi $A, B \subseteq \Omega$, useremo le notazioni standard

$$\begin{aligned} A \cup B &:= \{\omega \in \Omega : \omega \in A \text{ o } \omega \in B\}, \\ A \cap B &:= \{\omega \in \Omega : \omega \in A \text{ e } \omega \in B\}, \\ A^c &:= \{\omega \in \Omega : \omega \notin A\}, \\ A \setminus B &:= A \cap B^c, \\ A \triangle B &:= (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B), \end{aligned}$$

dove il simbolo ":=" indica una definizione. Le definizioni di unione e intersezione si estendono in modo naturale a una famiglia arbitraria $\{A_i\}_{i \in I}$ di sottoinsiemi di Ω:

$$\begin{aligned} \bigcup_{i \in I} A_i &:= \{\omega \in \Omega : \exists i \in I \text{ tale che } \omega \in A_i\}, \\ \bigcap_{i \in I} A_i &:= \{\omega \in \Omega : \forall i \in I \text{ si ha che } \omega \in A_i\}. \end{aligned}$$

Ricordiamo le leggi di De Morgan:

$$(A \cup B)^c = A^c \cap B^c, \qquad (A \cap B)^c = A^c \cup B^c,$$

e più in generale

$$\left(\bigcup_{i \in I} A_i\right)^c = \bigcap_{i \in I} A_i^c, \qquad \left(\bigcap_{i \in I} A_i\right)^c = \bigcup_{i \in I} A_i^c.$$

Indicheremo con $\mathbb{N} := \{1, 2, 3, \ldots\}$ i numeri naturali, zero escluso; quando vorremo includerlo, useremo la notazione $\mathbb{N}_0 := \{0, 1, 2, \ldots\}$. Adotteremo le notazioni standard per i numeri interi, razionali, reali e complessi, indicati rispettivamente con $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ e $\mathbb{C}$, e porremo $\mathbb{R}^+ := [0, \infty) = \{x \in \mathbb{R} : x \geq 0\}$ e $\mathbb{Q}^+ := \mathbb{Q} \cap [0, \infty)$.

Diremo che un numero $x \in \mathbb{R}$ è *positivo* se $x \geq 0$ e *strettamente positivo* se $x > 0$; analogamente, diremo che x è *negativo* se $x \leq 0$ e *strettamente negativo* se $x < 0$.

Si noti che con queste convenzioni 0 è sia positivo sia negativo. La parte positiva e negativa di un numero $x \in \mathbb{R}$ sono definite rispettivamente mediante $x^+ := \max\{x,0\}$ e $x^- := -\min\{x,0\} = \max\{-x,0\}$. Si noti che $x^+, x^- \geq 0$ e $x = x^+ - x^-$, mentre il valore assoluto di x è dato da $|x| = x^+ - x^-$.

Utilizzeremo gli aggettivi "crescente" e "decrescente" in senso debole: una funzione $f : \mathbb{R} \to \mathbb{R}$ sarà detta crescente (risp. decrescente) se per ogni $x > y$ si ha $f(x) \geq f(y)$ (risp. $f(x) \leq f(y)$). Una funzione costante è dunque sia crescente sia decrescente.

La cardinalità di un insieme A, cioè il numero dei suoi elementi, sarà denotata col simbolo $|A|$; per indicare che un insieme ha cardinalità finita scriveremo $|A| < \infty$. Un insieme A è detto numerabile se è in corrispondenza biunivoca con $\mathbb{N}$, cioè se esiste una applicazione $f : A \to \mathbb{N}$ iniettiva e suriettiva. Dati due insiemi A, B, il loro prodotto cartesiano $A \times B$ è definito come l'insieme di tutte le coppie (a,b) con $a \in A$ e $b \in B$. Queste definizioni verranno riprese e approfondite nel paragrafo 1.3.

Somme infinite

Data una successione di numeri reali $(x_n)_{n\in\mathbb{N}}$, è ben noto che la *somma della serie* corrispondente $\sum_{n=1}^{\infty} x_n$ è definita come il limite per $N \to \infty$ delle somme parziali $s_N := \sum_{n=1}^{N} x_n$, ammesso naturalmente che tale limite esista. Quando si ha a che fare con una famiglia di numeri reali $\{x_i\}_{i\in I}$ indicizzata da un insieme arbitrario I, per definire la *somma infinita* $\sum_{i\in I} x_i$ conviene procedere in modo differente.

Nel caso speciale di una famiglia a termini positivi ($x_i \geq 0$ per ogni $i \in I$) si pone

$$\sum_{i\in I} x_i \;:=\; \sup_{A\subseteq I,\,|A|<\infty} \sum_{j\in A} x_j \;\in\; [0,+\infty]\,,$$

dove $\sum_{j\in A} x_j$ è un'ordinaria somma finita, dal momento che $|A| < \infty$. Chiaramente si ha $\sum_{i\in I} x_i \in [0,+\infty]$ e inoltre $\sum_{i\in I} x_i = 0$ se e solo se $x_i = 0\ \forall i \in I$. Se $\sum_{i\in I} x_i < +\infty$, la famiglia $\{x_i\}_{i\in I}$ è detta *sommabile*: in questo caso i termini non nulli ($x_i > 0$) sono al più un'infinità numerabile (infatti $|\{i \in I : x_i > \frac{1}{n}\}| < \infty$ per ogni $n \in \mathbb{N}$).

Una famiglia $\{x_i\}_{i\in I}$ generica (non necessariamente a termini positivi) si dice *sommabile* se $\sum_{i\in I} |x_i| < \infty$, cioè se $\sum_{i\in I} x_i^+ < \infty$ e $\sum_{i\in I} x_i^- < \infty$. In questo caso si pone

$$\sum_{i\in I} x_i \;:=\; \sum_{i\in I} x_i^+ - \sum_{i\in I} x_i^-\,, \tag{0.1}$$

e si ha $\sum_{i\in I} x_i \in (-\infty,+\infty)$. Più in generale, se si fa l'ipotesi più debole che almeno una delle due somme $\sum_{i\in I} x_i^+$, $\sum_{i\in I} x_i^-$ sia finita, si può ancora definire $\sum_{i\in I} x_i$ mediante (0.1), tenendo presente che in questo caso $\sum_{i\in I} x_i \in [-\infty,+\infty]$.

Ricordiamo alcune proprietà delle somme infinite. Se $\{x_i\}_{i\in I}$ e $\{y_i\}_{i\in I}$ sono due famiglie sommabili, anche la famiglia $\{x_i + y_i\}_{i\in I}$ lo è e vale che

$$\sum_{i\in I} (x_i + y_i) \;=\; \sum_{i\in I} x_i + \sum_{i\in I} y_i\,.$$

Supponiamo ora di avere una famiglia $\{x_{i,j}\}_{(i,j)\in I\times J}$ i cui elementi sono indicizzati da uno spazio prodotto. Se la famiglia è a termini positivi ($x_{i,j} \geq 0$ per ogni $i \in I$, $j \in J$), oppure se è sommabile, vale la seguente versione del teorema di Fubini:

$$\sum_{(i,j)\in I\times J} x_{i,j} = \sum_{i\in I}\left(\sum_{j\in J} x_{i,j}\right) = \sum_{j\in J}\left(\sum_{i\in I} x_{i,j}\right).$$

La dimostrazione di queste proprietà non è difficile ma i dettagli sono piuttosto noiosi: il lettore interessato li può trovare nell'appendice A.

Capitolo 1
Spazi di probabilità discreti

Abstract In questo capitolo definiamo gli spazi di probabilità discreti, accompagnando la definizione con alcune rilevanti proprietà ed esempi significativi. Quindi introduciamo le nozioni di probabilità condizionata e di indipendenza.

1.1 Generalità

Nel corso di questo libro con la dicitura *esperimento aleatorio* indicheremo un'osservazione relativa ad un qualunque fenomeno (fisico, economico, sociale, ...) per il quale il risultato di tale osservazione non sia determinabile con certezza a priori. Il nostro obiettivo è di fornire una descrizione matematica di un esperimento aleatorio, definendo un *modello probabilistico*.

Il primo passo consiste nell'identificare un insieme Ω, detto *spazio campionario*, che contiene tutti gli esiti possibili dell'esperimento.

Esempio 1.1.
(i) Per il lancio di un dado ordinario a sei facce, lo spazio campionario naturale è $\Omega = \{1,2,3,4,5,6\}$
(ii) Per la rilevazione del numero di accessi giornalieri a un sito web, scelte possibili per lo spazio campionario sono $\Omega = \mathbb{N}_0$ oppure $\Omega = \{0,1,\ldots,10^{100}\}$.
(iii) Per la misurazione del tempo di attesa per l'accesso ad uno sportello di un ufficio postale, una scelta naturale di spazio campionario è data da $\Omega = [0,+\infty)$.

Il secondo ingrediente di un modello probabilistico è l'assegnazione di un "grado di fiducia", o *probabilità*, ai sottoinsiemi dello spazio campionario. Con riferimento all'Esempio 1.1, si vuol dare significato ad espressioni quali "probabilità che il numero ottenuto col dado sia pari", o "probabilità che il numero di accessi al sito web sia minore di 100", o "probabilità che il tempo di attesa sia compreso tra 3 e 10 minuti".

Vedremo più avanti in alcuni casi concreti come, sulla base di considerazioni sulla natura dell'esperimento aleatorio in esame, la scelta della probabilità risulti talvolta "naturale". Molto spesso, però, non è così. In ogni caso, la probabilità scelta va sottoposta a verifica sulla base di dati sperimentali, ottenuti da ripetizioni successive dell'esperimento. Tale problema di verifica è uno degli obbiettivi principali della *Statistica*, a cui accenneremo nel capitolo 7.

Comunque essa venga assegnata, una probabilità dovrà soddisfare ad alcune proprietà naturali. Tali proprietà risultano semplici da enunciare nel caso in cui lo spazio campionario Ω sia finito o numerabile. Rimuovendo tale ipotesi, la definizione di probabilità diviene più delicata. Per questa ragione, ci concentreremo innanzitutto sul caso di spazi di probabilità finiti o numerabili, rimandando l'analisi del caso generale al capitolo 4.

Definizione 1.1. Sia Ω un insieme finito o numerabile, e indichiamo con $\mathscr{P}(\Omega)$ la famiglia dei sottoinsiemi di Ω. Una funzione $P : \mathscr{P}(\Omega) \to [0,1]$ si dice *probabilità* se soddisfa alle seguenti proprietà:

(P1) $P(\Omega) = 1$.

(P2) (σ-*additività*) Per ogni successione $(A_n)_{n\in\mathbb{N}}$ di sottoinsiemi di Ω a due a due disgiunti, tale cioè che $A_n \cap A_m = \emptyset$ se $n \neq m$, si ha

$$P\left(\bigcup_{n=1}^{+\infty} A_n\right) = \sum_{n=1}^{+\infty} P(A_n)\,.$$

La coppia (Ω, P) è detta *spazio di probabilità discreto*, Ω è chiamato *spazio campionario* e i suoi sottoinsiemi sono detti *eventi*. Diremo che le proprietà (P1) e (P2) costituiscono il *sistema di assiomi* che definisce uno spazio di probabilità discreto.

L'interpretazione di uno spazio di probabilità discreto (Ω, P) è dunque la seguente: l'insieme Ω contiene tutti i possibili esiti di un esperimento aleatorio e, per ogni sottoinsieme $A \subseteq \Omega$, il numero $P(A) \in [0,1]$ esprime il "grado di fiducia" che si attribuisce all'eventualità che l'esito dell'esperimento sia un elemento di A. Un modo concreto di visualizzare questo "grado di fiducia" consiste nell'immaginare di ripetere l'esperimento aleatorio un numero elevato N di volte: indicando con M il numero di volte in cui l'esito ottenuto appartiene al sottoinsieme A, si ha che $P(A) \approx M/N$. Come vedremo nel capitolo 6, questa *interpretazione frequentista* della probabilità, introdotta in modo informale, riceve una giustificazione rigorosa a posteriori dalla *legge dei grandi numeri*.

Ritornando alla Definizione 1.1, la proprietà (P1) esprime il fatto che l'intero spazio campionario è un evento *certo*, ossia ha probabilità uno. La proprietà (P2) richiede una discussione più accurata. Iniziamo col dedurre due conseguenze degli assiomi (P1) e (P2).

Lemma 1.1. *Sia (Ω, P) uno spazio di probabilità discreto. Allora valgono le seguenti proprietà:*

(i) $P(\emptyset) = 0$.

(ii) *Se $A_1, A_2, \ldots, A_k$ sono eventi a due a due disgiunti (con $k \geq 2$), allora*

$$P(A_1 \cup A_2 \cup \cdots \cup A_k) \;=\; \sum_{j=1}^{k} P(A_j)\,. \tag{1.1}$$

Dimostrazione. (i) Sia $x = P(\emptyset) \in [0,1]$ e si definisca $A_n = \emptyset$ per ogni $n \in \mathbb{N}$. Chiaramente $(A_n)_{n\in\mathbb{N}}$ è una successione di sottoinsiemi disgiunti di Ω. Allora, per l'assioma (P2) e il fatto che $\bigcup_{n=1}^{+\infty} A_n = \emptyset$, si ha

$$x \;=\; P(\emptyset) \;=\; P\left(\bigcup_{n=1}^{+\infty} A_n\right) \;=\; \sum_{n=1}^{+\infty} P(A_n) \;=\; \sum_{n=1}^{+\infty} x\,.$$

Tale identità è possibile se e solo se $x = 0$.

(ii) Prolunghiamo la famiglia di eventi disgiunti $A_1, A_2, \ldots, A_k$ ad una successione infinita di eventi a due a due disgiunti, ponendo $A_n = \emptyset$ per $n > k$. Allora, per l'assioma (P2)

$$P(A_1 \cup A_2 \cup \cdots \cup A_k) \;=\; P\left(\bigcup_{j=1}^{+\infty} A_j\right) \;=\; \sum_{j=1}^{+\infty} P(A_j) \;=\; \sum_{j=1}^{k} P(A_j)\,.$$

Osservazione 1.1. Riscriviamo l'equazione (1.1) nel caso speciale $k = 2$:

$$P(A \cup B) \;=\; P(A) + P(B)\,, \qquad \forall A, B \subseteq \Omega, \quad A \cap B = \emptyset\,. \tag{1.2}$$

È interessante notare che la relazione (1.1) nel caso generale ($k \geq 2$) segue da (1.2) attraverso una semplice dimostrazione per induzione (esercizio!).

La proprietà (1.1) (o equivalentemente (1.2)), detta anche *additività finita*, è una condizione "naturale", che corrisponde ad un'idea intuitiva di probabilità (si pensi all'interpretazione frequentista). È pertanto significativo domandarsi se le coppie di assiomi $\{$(P1),(P2)$\}$ e $\{$(P1),(1.1)$\}$ siano equivalenti, cioè se da ciascuna coppia si possa dedurre l'altra. La risposta è affermativa nel caso in cui Ω sia un insieme finito, dal momento che non esistono successioni infinite di eventi disgiunti e non vuoti ($\mathscr{P}(\Omega)$ ha un numero finito di elementi). Se invece Ω è infinito, gli assiomi $\{$(P1),(P2)$\}$ sono strettamente più forti di $\{$(P1),(1.1)$\}$, cioè esistono funzioni $P : \Omega \to [0,1]$ che soddisfano (1.1) ma non (P1). Un esempio di una tale funzione per $\Omega = \mathbb{N}$, costruito usando l'assioma della scelta, è descritto nell'Appendice B (la cui lettura può essere omessa, essendo piuttosto sofisticati gli argomenti usati).

Dunque, la σ additività non è una conseguenza dell'additività finita. Benché la teoria della probabilità finitamente additiva sia sviluppata in una parte della letteratura matematica, motivata da diverse applicazioni, in questo testo considereremo soltanto la teoria della probabilità σ-additiva, che si adatta assai bene alla maggior parte delle applicazioni e che viene adottata nella grande maggioranza della letteratura. Le ragioni per cui l'assioma di σ-additività è rilevante rispetto al più debole (1.1) sono diverse, in parte non comprensibili in questa fase iniziale della

presentazione della teoria. Tuttavia, una implicazione rilevante della σ-additività è già descritta nel prossimo paragrafo.

1.1.1 Probabilità e densità

Data una probabilità P su uno spazio finito o numerabile Ω, possiamo associare a P una funzione p, detta *densità*, definita su Ω a valori in $[0,1]$, mediante la relazione

$$p(\omega) \;=\; P(\{\omega\})\,. \tag{1.3}$$

Si noti che p è definita su Ω, mentre P è definita su $\mathscr{P}(\Omega)$. Usando l'assioma (P2), è facile verificare che

$$P(A) \;=\; \sum_{\omega\in A} p(\omega)\,, \qquad \forall A \subseteq \Omega\,. \tag{1.4}$$

In particolare, prendendo $A = \Omega$ si ottiene

$$\sum_{\omega\in\Omega} p(\omega) \;=\; 1\,. \tag{1.5}$$

È anche possibile percorrere il cammino inverso. Supponiamo cioè che sia assegnata una funzione $p : \Omega \to [0,1]$ per cui valga la relazione (1.5). Possiamo allora *definire* una funzione $P : \mathscr{P}(\Omega) \to [0,1]$ mediante (1.4) ed è facile (anche se piuttosto noioso) mostrare che tale P è effettivamente una probabilità, cioè gli assiomi (P1) e (P2) sono soddisfatti. C'è dunque una corrispondenza biunivoca tra l'insieme delle probabilità P su Ω e l'insieme delle funzioni $p : \Omega \to [0,1]$ che soddisfano (1.5): tale corrispondenza è data dalle relazioni (1.3) e (1.4).

Questo argomento mostra che in uno spazio di probabilità discreto la probabilità è determinata dal suo valore sui *singoletti*, ossia sugli eventi costituiti da un solo elemento di Ω. Per gli spazi di probabilità più generali che vedremo più avanti, quest'ultima affermazione in generale è falsa.

1.1.2 Esempi

Concludiamo questo paragrafo con alcuni esempi di spazi di probabilità discreti.

Esempio 1.2. Sia Ω un insieme *finito*. Per $A \subseteq \Omega$, definiamo

$$P(A) \;:=\; \frac{|A|}{|\Omega|}\,,$$

ove $|\cdot|$ indica il numero di elementi di un insieme. Si vede facilmente che P è una probabilità, la cui densità, grazie alla definizione (1.4), è data da $p(\omega) = 1/|\Omega|$. Lo

spazio (Ω, P) così definito si dice *spazio di probabilità uniforme*. Esso è il modello probabilistico adeguato a descrivere gli esperimenti aleatori in cui tutti gli esiti si possono ritenere equiprobabili. Ad esempio: il lancio di un dado regolare, l'estrazione di un numero dalla ruota del lotto, la successione delle carte in un mazzo accuratamente mescolato...

Esempio 1.3. Sia Ω un insieme *finito*, e sia $H : \Omega \to \mathbb{R}$ una funzione arbitraria. Fissato un parametro reale $\beta \geq 0$, definiamo

$$p(\omega) := \frac{1}{Z(\beta)} e^{-\beta H(\omega)},$$

dove

$$Z(\beta) := \sum_{\omega \in \Omega} e^{-\beta H(\omega)}.$$

Si noti che la relazione (1.5) è verificata, di conseguenza la funzione p determina, mediante (1.4), una probabilità che, per evidenziare la dipendenza da β, indicheremo con P_β. Prendendo a prestito la terminologia della meccanica statistica, la probabilità P_β viene detta *misura di Gibbs* relativa alla *funzione Hamiltoniana* (o *energia*) H e alla *temperatura inversa* β. L'interpretazione è la seguente: gli elementi $\omega \in \Omega$ rappresentano gli stati di un sistema fisico, a cui è associata una energia $H(\omega)$; quando il sistema è in equilibrio termico alla temperatura assoluta T, ponendo $\beta = \frac{1}{k_B T}$ (dove k_B è la costante di Boltzmann) si ha che la probabilità di osservare il sistema in uno stato ω è data da $p(\omega) = P_\beta(\{\omega\})$.

Si noti che, nel caso $\beta = 0$ (temperatura infinita), $p(\cdot)$ non dipende da ω, pertanto P_0 non è altro che la probabilità uniforme su Ω. Consideriamo invece il limite $\beta \to +\infty$ di temperatura zero (assoluto). Indichiamo con $m := \min\{H(\omega) : \omega \in \Omega\}$ il minimo assoluto della Hamiltoniana, e introduciamo l'insieme (non vuoto)

$$A := \{\omega \in \Omega : H(\omega) = m\},$$

costituito dagli elementi di Ω con energia minima. Mostriamo ora che

$$\lim_{\beta \to +\infty} P_\beta(A) = 1. \tag{1.6}$$

In altre parole, nel limite $\beta \to +\infty$, P_β "si concentra" sugli elementi di minima energia. Per dimostrare (1.6) è sufficiente (perché?) mostrare che, per ogni $\omega \notin A$,

$$\lim_{\beta \to +\infty} P_\beta(\{\omega\}) = 0.$$

Si noti che

$$P_\beta(\{\omega\}) = \frac{1}{Z(\beta)} e^{-\beta H(\omega)},$$

e che

$$Z(\beta) \geq e^{-\beta m}.$$

Pertanto

$$P_\beta(\{\omega\}) \;\le\; \frac{e^{-\beta H(\omega)}}{e^{-\beta m}} \;=\; e^{-\beta[H(\omega)-m]}. \tag{1.7}$$

Essendo $\omega \notin A$, si ha $H(\omega) > m$, e (1.6) segue immediatamente da (1.7).

Esempio 1.4. Sia $\Omega = \mathbb{N}_0 = \{0,1,2,\ldots\}$ e poniamo

$$p(n) \;:=\; e^{-\lambda}\,\frac{\lambda^n}{n!},$$

dove $\lambda > 0$ è un parametro reale fissato. Ricordando che $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$, $\forall x \in \mathbb{R}$, la relazione (1.5) è verificata e dunque è possibile definire una probabilità P tramite (1.4). Come vedremo in seguito, tale probabilità è particolarmente utile nella descrizione delle *file di attesa*.

1.2 Proprietà fondamentali

Iniziamo con l'esporre alcune conseguenze quasi immediate degli assiomi (P1) e (P2). Qui e nel seguito, indichiamo con (Ω, P) un generico spazio di probabilità discreto.

Proposizione 1.1. *Siano $A, B \subseteq \Omega$. Allora valgono le seguenti proprietà:*

(i)
$$P(A^c) \;=\; 1 - P(A).$$

(ii) *Se $A \subseteq B$ allora*
$$P(B \setminus A) \;=\; P(B) - P(A).$$
In particolare
$$P(A) \;\le\; P(B).$$

(iii)
$$P(A \cup B) \;=\; P(A) + P(B) - P(A \cap B).$$
In particolare
$$P(A \cup B) \;\le\; P(A) + P(B).$$

Dimostrazione. (i) Per la proprietà di additività si ha

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

da cui la conclusione è immediata.

(ii) Basta osservare che, di nuovo per l'additività,

$$P(B) \;=\; P[A \cup (B \setminus A)] \;=\; P(A) + P(B \setminus A).$$

(iii) Scriviamo

$$A \cup B \;=\; [A \setminus (A \cap B)] \cup [B \setminus (A \cap B)] \cup (A \cap B).$$

I tre eventi nella precedente unione sono disgiunti. Dunque, usando l'additività e la relazione (ii), si ottiene

$$\begin{aligned} P(A \cup B) \;&=\; P[A \setminus (A \cap B)] + P[B \setminus (A \cap B)] + P(A \cap B) \\ &=\; P(A) - P(A \cap B) + P(B) - P(A \cap B) + P(A \cap B) \\ &=\; P(A) + P(B) - P(A \cap B)\,. \end{aligned}$$

L'identità della parte (iii) della Proposizione 1.1 può essere generalizzata all'unione di più di due eventi. Ad esempio, supponiamo di voler calcolare $P(A \cup B \cup C)$ per tre eventi A, B, C. Usando due volte l'identità appena citata

$$\begin{aligned} P(A \cup B \cup C) \;&=\; P((A \cup B) \cup C) \;=\; P(A \cup B) + P(C) - P((A \cup B) \cap C) \\ &=\; P(A) + P(B) - P(A \cap B) + P(C) - P((A \cap C) \cup (B \cap C)) \\ &=\; P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)\,. \end{aligned}$$

Non è difficile, a questo punto, "indovinare" la formula generale per l'unione di un numero finito arbitrario di eventi. Il seguente risultato è chiamato *formula di inclusione-esclusione*.

Proposizione 1.2. *Si considerino n eventi $A_1, A_2, \ldots, A_n$ di Ω. Allora*

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) \;=\; \sum_{k=1}^{n} \sum_{\substack{J \subseteq \{1,2,\ldots,n\} \\ \text{tali che } |J|=k}} (-1)^{k+1} P\left(\bigcap_{i \in J} A_i \right). \tag{1.8}$$

Dimostrazione. Dimostriamo per induzione su n che la relazione (1.8) è vera per ogni n-pla di eventi $A_1, A_2, \ldots, A_n$. Per $n = 1$ la formula (1.8) si riduce a $P(A_1) = P(A_1)$, e dunque non c'è nulla da dimostrare. Supponiamo allora che l'asserto sia vero per ogni $k \leq n$, e mostriamo che è vero per $n+1$. Siano $A_1, A_2, \ldots, A_n, A_{n+1}$ eventi. Usando il fatto che, per ipotesi induttiva, (1.8) vale per $n = 2$ otteniamo

$$\begin{aligned} &P(A_1 \cup A_2 \cup \cdots \cup A_n \cup A_{n+1}) \\ &\quad=\; P(A_1 \cup A_2 \cup \cdots \cup A_n) + P(A_{n+1}) - P((A_1 \cup A_2 \cup \cdots \cup A_n) \cap A_{n+1}) \\ &\quad=\; P(A_1 \cup A_2 \cup \cdots \cup A_n) + P(A_{n+1}) - P(B_1 \cup B_2 \cup \cdots \cup B_n)\,, \end{aligned} \tag{1.9}$$

dove abbiamo posto per comodità $B_i = A_i \cap A_{n+1}$, per $i = 1, 2, \ldots, n$. Usando nuovamente l'ipotesi induttiva, questa volta per n eventi, otteniamo

$$\begin{aligned} P(A_1 \cup A_2 \cup \cdots \cup A_n) &= \sum_{k=1}^{n} \sum_{\substack{J \subseteq \{1,2,\ldots,n\} \\ \text{tale che } |J|=k}} (-1)^{k+1} P\left(\bigcap_{i \in J} A_i\right) \\ &= \sum_{k=1}^{n} \sum_{\substack{J \subseteq \{1,2,\ldots,n+1\} \\ \text{tale che } |J|=k \text{ e } n+1 \notin J}} (-1)^{k+1} P\left(\bigcap_{i \in J} A_i\right), \end{aligned} \tag{1.10}$$

e analogamente

$$\begin{aligned} P(B_1 \cup B_2 \cup \cdots \cup B_n) &= \sum_{k=1}^{n} \sum_{\substack{J \subseteq \{1,2,\ldots,n\} \\ \text{tale che } |J|=k}} (-1)^{k+1} P\left(\bigcap_{i \in J} B_i\right) \\ &= \sum_{k=1}^{n} \sum_{\substack{J \subseteq \{1,2,\ldots,n\} \\ \text{tale che } |J|=k}} (-1)^{k+1} P\left(A_{n+1} \cap \left(\bigcap_{i \in J} A_i\right)\right) \\ &= -\sum_{k=2}^{n+1} \sum_{\substack{J \subseteq \{1,2,\ldots,n+1\} \\ \text{tale che } |J|=k \text{ e } n+1 \in J}} (-1)^{k+1} P\left(\bigcap_{i \in J} A_i\right). \end{aligned} \tag{1.11}$$

Sostituendo (1.10) e (1.11) nell'ultimo membro di (1.9), si ottiene

$$P(A_1 \cup A_2 \cup \cdots \cup A_n \cup A_{n+1}) = \sum_{k=1}^{n+1} \sum_{\substack{J \subseteq \{1,2,\ldots,n+1\} \\ \text{tale che } |J|=k}} (-1)^{k+1} P\left(\bigcap_{i \in J} A_i\right),$$

che è quanto si voleva dimostrare.

Va notato come le dimostrazioni dei risultati delle Proposizioni 1.1 e 1.2 usino solo l'additività finita e non la σ-additività, che invece gioca un ruolo fondamentale nella seguente.

Proposizione 1.3. *Sia $P : \mathscr{P}(\Omega) \to [0,1]$ una funzione che soddisfa* (P1) *e l'additività in* (1.1). *Allora le seguenti proprietà sono equivalenti:*

(a) *P è σ-additiva.*

(b) *Se $(A_n)_{n \in \mathbb{N}}$ è una successione crescente di eventi, cioè $A_n \subseteq A_{n+1}$ per ogni $n \in \mathbb{N}$, allora*
$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \to +\infty} P(A_n).$$

(c) *Se $(A_n)_{n \in \mathbb{N}}$ è una successione decrescente di eventi, cioè $A_{n+1} \subseteq A_n$ per ogni $n \in \mathbb{N}$, allora*
$$P\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n \to +\infty} P(A_n).$$

Dimostrazione. (a) $\Rightarrow$ (b). Per una data successione crescente $(A_n)_{n\in\mathbb{N}}$ di eventi, definiamo un'altra successione $(B_n)_{n\in\mathbb{N}}$ tramite $B_1 := A_1$, e $B_n := A_n \setminus A_{n-1}$ per $n \geq 2$. Per costruzione, gli eventi B_n sono a due a due disgiunti e, per ogni $n \in \mathbb{N}$,

$$\bigcup_{k=1}^{n} B_k = A_n\,.$$

Inoltre

$$\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n\,.$$

Allora, per la σ-additività,

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} A_n\right) &= P\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} P(B_n) = \lim_{n\to+\infty} \sum_{k=1}^{n} P(B_k) \\ &= \lim_{n\to+\infty} P\left(\bigcup_{k=1}^{n} B_k\right) = \lim_{n\to+\infty} P(A_n)\,. \end{aligned}$$

(b) $\Rightarrow$ (a). Sia (A_n) una successione di eventi a due a due disgiunti. Notando che la successione $(B_n)_{n\in\mathbb{N}}$ definita da $B_n := \bigcup_{k=1}^{n} A_k$ è crescente e usando l'additività finita e la (b), si ha:

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\bigcup_{n=1}^{\infty} B_n\right) = \lim_{n\to+\infty} P(B_n) = \lim_{n\to+\infty} \sum_{k=1}^{n} P(A_k) = \sum_{n=1}^{+\infty} P(A_n)\,.$$

(b) $\Rightarrow$ (c). Sia $(A_n)_{n\in\mathbb{N}}$ una successione decrescente di eventi. Posto $B_n := A_n^c$, $(B_n)_{n\in\mathbb{N}}$ è una successione crescente di eventi. Allora, usando (b), si ha

$$\begin{aligned} P\left(\bigcap_{n=1}^{\infty} A_n\right) &= P\left(\left(\bigcup_{n=1}^{\infty} B_n\right)^c\right) = 1 - P\left(\bigcup_{n=1}^{\infty} B_n\right) \\ &= 1 - \lim_{n\to+\infty} P(B_n) = \lim_{n\to+\infty} P(A_n)\,. \end{aligned}$$

(c) $\Rightarrow$ (b). Del tutto simile all'implicazione precedente. Si lasciano i dettagli al lettore.

Le proprietà in (b) e (c) nella Proposizione 1.3 vengono dette rispettivamente *continuità dal basso* e *continuità dall'alto* (della probabilità). Un utile corollario è il seguente.

Corollario 1.1 (Subadditività). *Sia $(A_n)_{n\in\mathbb{N}}$ una successione di eventi. Allora*

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n)\,.$$

Dimostrazione. Sia $B_n := \bigcup_{k=1}^{n} A_k$. Evidentemente $(B_n)_{n\in\mathbb{N}}$ è una successione crescente di eventi. Inoltre $\bigcup_{n\in\mathbb{N}} B_n = \bigcup_{n\in\mathbb{N}} A_n$. Per la parte (iii) della Proposizione 1.1, sappiamo che $P(A_1 \cup A_2) \leq P(A_1) + P(A_2)$. Con una facile dimostrazione per induzione, la precedente disuguaglianza si estende a:

$$P\left(\bigcup_{k=1}^{n} A_k\right) \leq \sum_{k=1}^{n} P(A_k)\,.$$

Ma allora, usando anche la Proposizione 1.3, si ha

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} A_n\right) &= P\left(\bigcup_{n=1}^{\infty} B_n\right) = \lim_{n\to+\infty} P(B_n) = \lim_{n\to+\infty} P\left(\bigcup_{k=1}^{n} A_k\right) \\ &\leq \lim_{n\to+\infty} \sum_{k=1}^{n} P(A_k) = \sum_{n=1}^{+\infty} P(A_n)\,. \end{aligned}$$

1.3 Il calcolo combinatorio

Ricordiamo dall'Esempio 1.2 che uno spazio di probabilità discreto (Ω, P) si dice *uniforme* se Ω è un insieme finito e si ha $P(A) = \frac{|A|}{|\Omega|}$, per ogni $A \subseteq \Omega$. Pertanto, il calcolo della probabilità di un evento in uno spazio uniforme si riduce a contarne il numero di elementi. I problemi di conteggio, anche in insiemi abbastanza semplici, sono tipicamente non banali e vanno affrontati con attenzione. Lo strumento matematico fondamentale in questo contesto è il *calcolo combinatorio*, che ora descriviamo.

1.3.1 Principi basilari

Dati due insiemi A, B, si dice che A è *in corrispondenza biunivoca* con B se esiste un'applicazione $f : A \to B$ biunivoca, cioè iniettiva e suriettiva. Chiaramente A è in corrispondenza biunivoca con B se e soltanto se B è in corrispondenza biunivoca con A: si scrive talvolta "A e B sono in corrispondenza biunivoca", che rende palese la simmetria della relazione (si tratta in effetti di una relazione di equivalenza). Dato $n \in \mathbb{N}$, si dice che un insieme A *ha cardinalità* n e si scrive $|A| = n$ se A è in corrispondenza biunivoca con l'insieme $\{1, 2, \ldots, n\}$. Si noti che la proprietà "A ha cardinalità n" è la formalizzazione matematica dell'affermazione intuitiva "A ha n elementi". *In questo paragrafo considereremo solo insiemi finiti*, cioè insiemi che hanno cardinalità n per un opportuno $n \in \mathbb{N}$.

Per determinare la cardinalità di un insieme, la strategia tipica consiste nel ricondurre il calcolo all'applicazione combinata (talvolta non banale) di alcuni principi o osservazioni basilari. Una prima osservazione, elementare ma molto utile,

è che se un insieme A è in corrispondenza biunivoca con un insieme B, allora $|A| = |B|$. Un'altra osservazione, anch'essa molto intuitiva, è la seguente: se A, B sono due sottoinsiemi (di uno stesso spazio) *disgiunti*, cioè tali che $A \cap B = \emptyset$, allora $|A \cup B| = |A| + |B|$. Più in generale, se $A_1, \ldots, A_k$ sono sottoinsiemi a due a due disgiunti, tali cioè che $A_i \cap A_j = \emptyset$ per $i \neq j$, allora $|\bigcup_{i=1}^{k} A_i| = \sum_{i=1}^{k} |A_i|$. La dimostrazione di queste osservazioni è semplice ed è lasciata per esercizio.

Un principio leggermente meno elementare riguarda la cardinalità degli *insiemi prodotto*. Ricordiamo che, dati due insiemi A, B, il loro prodotto cartesiano $A \times B$ è definito come l'insieme delle coppie ordinate (a,b), con $a \in A$ e $b \in B$. Vale allora la relazione $|A \times B| = |A||B|$. Il modo più semplice per convincersi della validità di questa formula consiste nel disporre gli elementi di $A \times B$ in una tabella rettangolare, dopo aver numerato gli elementi dei due insiemi. Più precisamente, se $A = \{a_1, a_2, \ldots, a_m\}$, $B = \{b_1, b_2, \ldots, b_k\}$, possiamo elencare gli elementi dell'insieme $A \times B$ nel modo seguente:

$$\begin{array}{cccccc} (a_1,b_1) & (a_1,b_2) & \cdots & (a_1,b_{k-1}) & (a_1,b_k) \\ (a_2,b_1) & (a_2,b_2) & \cdots & (a_2,b_{k-1}) & (a_2,b_k) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ (a_m,b_1) & (a_m,b_2) & \cdots & (a_m,b_{k-1}) & (a_m,b_k) \end{array},$$

da cui è chiaro che $|A \times B| = m \cdot k = |A| \cdot |B|$.

Diamo ora una dimostrazione più formale. Per $x \in A$ indichiamo con $\{x\} \times B$ il sottoinsieme di $A \times B$ costituito dagli elementi che hanno x come prima componente, cioè $\{x\} \times B := \{(x,b) : b \in B\}$. Possiamo quindi scrivere $A \times B = \cup_{x \in A}(\{x\} \times B)$, e si noti che questa unione è disgiunta, poiché $(\{x_1\} \times B) \cap (\{x_2\} \times B) = \emptyset$ se $x_1 \neq x_2$. Per l'osservazione enunciata sopra si ha dunque $|A \times B| = \sum_{x \in A} |\{x\} \times B|$. Si noti ora che l'insieme $\{x\} \times B$ è in corrispondenza biunivoca con B, per qualunque $x \in A$: la corrispondenza è data semplicemente da $(x,b) \mapsto b$. Di conseguenza $|\{x\} \times B| = |B|$ e si ottiene la formula $|A \times B| = \sum_{x \in A} |B| = |A|\,|B|$.

Per induzione si estende facilmente la formula al caso di più di due fattori: più precisamente, se A_1, …, A_k sono insiemi finiti, l'insieme prodotto $A_1 \times \cdots \times A_k$, definito come l'insieme delle k-uple $(a_1, \ldots, a_k)$, con $a_i \in A_i$, ha cardinalità data dalla formula $|A_1 \times \cdots \times A_k| = |A_1| \cdots |A_k| = \prod_{i=1}^{k} |A_i|$. Un'ulteriore estensione di questa formula, elementare ma non banale, conduce a quello che è noto come il principio fondamentale del calcolo combinatorio. Prima di vedere di che cosa si tratta, discutiamo qualche applicazione delle formule appena viste.

1.3.2 Disposizioni con ripetizione

Dato un insieme $A = \{a_1, \ldots, a_n\}$ di cardinalità $n \in \mathbb{N}$ e dato $k \in \mathbb{N}$, le funzioni definite su $\{1, \ldots, k\}$ a valori in A sono dette *disposizioni con ripetizione di k elementi estratti da A*. È facile vedere che le disposizioni con ripetizione sono in corrispondenza biunivoca naturale con gli elementi dell'insieme $A^k := A \times \cdots \times A$

(k volte): la corrispondenza è quella che a $(x_1,\dots,x_k) \in A^k$ associa la funzione $f : \{1,\dots,k\} \to A$ definita da $f(i) := x_i$. La formula sulla cardinalità degli insiemi prodotto dà $|A^k| = |A|^k = n^k$: ci sono dunque n^k possibili disposizioni con ripetizione di k elementi estratti da un insieme di n elementi.

Una disposizione con ripetizione può dunque essere vista come una sequenza *ordinata* $(x_1,\dots,x_k)$ di elementi $x_i \in A$, *non necessariamente distinti*: si può cioè avere $x_i = x_j$ per $i \neq j$. Sottolineiamo che l'ordine in cui compaiono gli elementi è importante: per esempio, (a_1,a_2) e (a_2,a_1) sono due disposizioni differenti.

Esempio 1.5. (1) I compleanni di un gruppo ordinato di 4 persone costituiscono una disposizione con ripetizione di 4 elementi estratti dall'insieme dei giorni dell'anno, che ha cardinalità 366 (contando il 29 febbraio). Sono dunque possibili $366^4 \approx 1.8 \cdot 10^{10}$ sequenze distinte di compleanni.

(2) Per compilare una colonna di una schedina del Totocalcio occorre scegliere, per ciascuna delle 13 partite in esame, tra la vittoria della squadra di casa (1), il pareggio (x) o la vittoria della squadra in trasferta (2). Una colonna compilata è dunque una disposizione con ripetizione di 13 elementi estratti dall'insieme $\{1,x,2\}$ e di conseguenza ci sono $3^{13} \approx 1.6 \cdot 10^6$ modi possibili di compilare una colonna.

(3) Le possibili "parole" (anche prive di significato) costituite da 10 lettere dell'alfabeto inglese coincidono con le disposizioni con ripetizione di 10 elementi estratti da un insieme che ne continene 26: il loro numero è dunque pari a $26^{10} \approx 1.4 \cdot 10^{14}$. Le parole che effettivamente hanno un significato (per esempio nella lingua inglese) sono naturalmente molte meno: anche includendo i termini tecnici, il numero totale di parole di qualunque lunghezza della lingua inglese non supera il milione. Di conseguenza, la probabilità che digitando una sequenza di dieci lettere a caso si ottenga una parola di senso compiuto è certamente minore di $10^6/(1.4 \cdot 10^{14}) < 10^{-8}$.

Osservazione 1.2. Se $B = \{b_1,\dots,b_k\}$ è un insieme di cardinalità $k \in \mathbb{N}$, si indica con A^B l'insieme di tutte le funzioni da B in A. L'insieme A^B è in corrispondenza biunivoca con A^k: una corrispondenza è per esempio quella che a $(x_1,\dots,x_k) \in A^k$ associa la funzione $f \in A^B$ definita da $f(b_i) := x_i$. Come conseguenza della formula sulla cardinalità degli insiemi prodotto, otteniamo dunque che $|A^B| = |A|^k = n^k$, cioè $|A^B| = |A|^{|B|}$.

1.3.3 Il principio fondamentale del calcolo combinatorio

Un esempio molto ricorrente nelle applicazioni è quello in cui gli elementi di un insieme possano essere determinati attraverso scelte successive. Per esempio, sia E l'insieme delle funzioni iniettive da $\{1,\dots,k\}$ in A (si noti che necessariamente $k \leq n$). Possiamo determinare ogni funzione $f \in E$ scegliendo innanzitutto la prima componente $f(1)$ come un elemento qualunque di A, quindi scegliendo la seconda

componente $f(2)$ come un elemento qualunque di $A \setminus \{f(1)\}$, e così via. Abbiamo n esiti possibili per la scelta di $f(1)$, $(n-1)$ per la scelta di $f(2)$, ..., $(n-k+1)$ per la scelta di $f(k)$. Per analogia con gli insiemi prodotto, dovrebbe essere intuitivamente chiaro che $|E| = n \cdot (n-1) \cdots (n-k+1)$. Si noti che l'*insieme dei valori* ammissibili per $f(i)$ dipende dagli esiti delle scelte precedenti, tuttavia *il numero di valori* ammissibili è sempre lo stesso, pari a $n-i+1$. Generalizzando questo esempio, giungiamo al *principio fondamentale del calcolo combinatorio*, che possiamo formulare come segue.

Teorema 1.1 (Principio fondamentale del calcolo combinatorio). *Supponiamo che gli elementi di un insieme E possano essere determinati mediante k scelte successive, in cui ogni scelta abbia un numero fissato di esiti possibili: la prima scelta ha n_1 esiti possibili, la seconda scelta ne ha n_2, ..., la k-esima scelta ne ha n_k, dove $n_1, \ldots, n_k \in \mathbb{N}$. Supponiamo inoltre che sequenze distinte di esiti determinino elementi distinti di E. Allora $|E| = n_1 \cdot n_2 \cdots n_k$.*

Così enunciato, questo principio può apparire un po' vago (per esempio, il concetto di "scelta" non è stato definito precisamente). Una riformulazione matematicamente precisa del Teorema 1.1, con la relativa dimostrazione, è data dal Teorema 1.2 più in basso, che comporta tuttavia notazioni abbastanza pesanti e risulta di poco aiuto per l'applicazione del principio a casi concreti. Nella pratica, si fa tipicamente riferimento all'enunciato del Teorema 1.1. L'idea cruciale è che gli elementi dell'insieme E possono essere messi *in corrispondenza biunivoca* con le sequenze di esiti delle scelte, che hanno una struttura di spazio prodotto, da cui segue la formula per la cardinalità. La condizione che sequenze distinte di esiti determinino elementi distinti di E serve proprio a garantire che la corrispondenza sia biunivoca: la mancata verifica di questa condizione è la principale fonte di errori nell'applicazione del principio. Qualche esempio chiarirà la situazione.

Esempio 1.6. (a) Un mazzo di carte da poker è costituito da 52 carte, identificate dal *seme* (cuori ♡, quadri ♢, fiori ♣, picche ♠) e dal *tipo* (un numero da 1 a 10 oppure J, Q, K). Indichiamo con E l'insieme delle carte di numero pari (figure escluse) e di colore rosso (cioè di cuori o di quadri). Ogni elemento di E può essere determinato attraverso due scelte successive: la scelta del seme, che ha 2 esiti possibili (cuori e quadri), e la scelta del tipo, che ne ha 5 (cioè $2,4,6,8,10$). Segue dunque che $|E| = 2 \cdot 5 = 10$.

(b) Dato un mazzo di carte da poker, si chiama *full* un sottoinsieme di 5 carte costituito dall'unione di un *tris* (un sottoinsieme di 3 carte dello stesso tipo) e di una *coppia* (un sottoinsieme di 2 carte dello stesso tipo). Indichiamo con E l'insieme dei possibili full. Sottolineiamo che gli elementi di E sono *sottoinsiemi* di 5 carte, non disposizioni: in particolare, le carte non sono ordinate.

Gli elementi di E possono essere determinati univocamente attraverso 4 scelte successive: 1) il tipo del tris; 2) il tipo della coppia; 3) i semi delle carte che compaiono nel tris; 4) i semi delle carte che compaiono nella coppia. Per la prima scelta ci sono 13 esiti possibili, per la seconda scelta, qualunque sia

l'esito della prima scelta, ci sono 12 esiti possibili (chiaramente i due tipi devono essere differenti, perché non esistono cinque carte dello stesso tipo). Per la terza scelta, occorre scegliere tre semi nell'insieme {cuori, quadri, fiori, picche}: per enumerazione diretta, è facile vedere che ci sono 4 esiti possibili; analogamente, per la quarta scelta occorre scegliere due semi e per questo ci sono 6 esiti possibili (ritorneremo nell'Esempio 1.8 sul modo di contare i sottoinsiemi). Applicando il Teorema 1.1 si ottiene dunque che $|E| = 13 \cdot 12 \cdot 4 \cdot 6 = 3\,744$.

(c) Dato un mazzo di carte da poker, indichiamo con E l'insieme delle *doppie coppie*, cioè i sottoinsiemi di 5 carte costituiti dall'unione di due *coppie* di tipi diversi, più una quinta carta di tipo diverso dai tipi delle due coppie.
Per determinare $|E|$ si potrebbe essere tentati di procedere analogamente al caso dei full, attraverso sei scelte successive: 1) il tipo della prima coppia; 2) il tipo della seconda coppia; 3) il tipo della "quinta carta"; 4) i semi delle carte che compaiono nella prima coppia; 5) i semi delle carte che compaiono nella seconda coppia; 6) il seme della "quinta carta". Ci sono 13 esiti possibili per la prima scelta, 12 per la seconda scelta, 11 per la terza, 6 per la quarta, 6 per la quinta, 4 per la sesta: si otterrebbe dunque $|E| = 13 \cdot 12 \cdot 11 \cdot 6^2 \cdot 4 = 247\,104$. Tuttavia questo risultato è errato.
La ragione è che le scelte 1) e 2) sono ambigue, dal momento che non esiste una "prima" e una "seconda" coppia. In effetti, sequenze distinte di esiti delle sei scelte sopra elencate non conducono a elementi distinti di E: ciascun elemento di E, cioè ciascuna doppia coppia, viene infatti selezionata *esattamente due volte*. Per esempio, la doppia coppia $\{5\heartsuit, 5\diamondsuit, 6\heartsuit, 6\clubsuit, 7\spadesuit\}$ viene determinata sia con l'esito "5" della scelta 1) e l'esito "6" della scelta 2), sia viceversa. Per tale ragione, il risultato corretto è $|E| = 247\,104/2 = 123\,552$, cioè la metà di quanto ottenuto in precedenza.
Un modo alternativo di ottenere il risultato corretto è di riunire le scelte 1) e 2) nell'unica scelta 1bis) "i tipi delle due coppie", che ha $13 \cdot 12/2 = 78$ esiti possibili (anche su questo torneremo nell'Esempio 1.8). Le scelte 1bis), 3), 4), 5) e 6) permettono di applicare correttamente il Teorema 1.1, ottenendo $|E| = 78 \cdot 11 \cdot 6^2 \cdot 4 = 123\,552$.

Passiamo ora a riformulare il principio fondamentale del calcolo combinatorio in modo matematicamente più preciso. Occorre innanzitutto esprimere astrattamente il concetto di "scelta", in modo che possa essere applicato a insiemi arbitrari, che non abbiano necessariamente una struttura di spazio prodotto. Dato un insieme E, definiamo una *scelta su* E come una *partizione* di E, vale a dire una famiglia di sottoinsiemi $\{E_1, \dots, E_m\}$ tali che $E = \bigcup_{i=1}^m E_i$ e $E_i \cap E_j = \emptyset$ per $i \neq j$. Il "numero di esiti della scelta" è per definizione il numero m di elementi della partizione. Intuitivamente, l'indice i numera gli "esiti" della scelta mentre l'insieme E_i corrisponde agli elementi di E compatibili con l'esito i della scelta. Per esempio, riconsideriamo l'insieme E delle funzioni iniettive da $\{1, \dots, k\}$ in $A = \{a_1, \dots, a_n\}$: la scelta della prima componente corrisponde alla partizione $\{E_1, \dots, E_n\}$ definita da $E_i := \{f \in E : f(1) = a_i\}$ e ha dunque n esiti possibili.

Estendiamo ora la definizione: fissato un insieme E, *due scelte successive su* E sono il dato di una partizione $\{E_1, \dots, E_m\}$ di E (che rappresenta la prima scelta) e, per ogni elemento E_i di tale partizione, una partizione $\{E_{i,1}, \dots, E_{i,k_i}\}$ di E_i (che rappresenta la seconda scelta). Si noti che la seconda scelta non è "una scelta su E", come definita sopra, ma piuttosto una scelta su E_i, per ogni $i = 1, \dots, m$. In particolare, il numero k_i di esiti della seconda scelta può in generale dipendere

dall'esito i della prima scelta. Nel caso in cui ciò accada, cioè se $k_i = k$ per ogni $i = 1, \dots, m$, diremo che *la seconda scelta ha un numero fissato k di esiti possibili*. Ritornando all'esempio delle funzioni iniettive da $\{1, \dots, k\}$ in $A = \{a_1, \dots, a_n\}$, la scelta delle prime due componenti è un esempio di due scelte successive su E: infatti, per ogni elemento $E_i = \{f \in E : f(1) = a_i\}$ della prima scelta, la seconda scelta è data dalla partizione $\{E_{i,j}\}_{j \in \{1,\dots,n\} \setminus \{i\}}$ definita da $E_{i,j} := \{f \in E : f(1) = a_i, f(2) = a_j\}$. In particolare, la seconda scelta ha un numero fissato $n-1$ di esiti possibili.[1]

Il passaggio da due a k scelte successive è solo notazionalmente più complicato. Per definizione, dato un insieme E, *k scelte successive su E* sono il dato di una famiglia di partizioni, definite nel modo seguente:

- la prima scelta è una partizione $\{E_1, \dots, E_{n_1}\}$ di E;
- per ogni $2 \le j \le k$ e per ogni elemento $E_{i_1,\dots,i_{j-1}}$ della $(j-1)$-esima scelta, la j-esima scelta è una partizione $\{E_{i_1,\dots,i_{j-1},\ell}\}_{\ell \in \{1,\dots,n_j^*\}}$ di $E_{i_1,\dots,i_{j-1}}$, dove il numero n_j^* di elementi della partizione (cioè il numero di esiti della j-esima scelta) può in generale dipendere dagli esiti delle scelte precedenti: $n_j^* = n_j^*(i_1, \dots, i_{j-1})$.

Nel caso in cui $n_j^*(i_1, \dots, i_{j-1}) = n_j$ non dipenda da $i_1, \dots, i_{j-1}$, diremo che la j-esima scelta ha un numero fissato n_j di esiti possibili.

Alla luce delle notazioni appena introdotte, possiamo riformulare il Teorema 1.1 nel modo seguente.

Teorema 1.2. *Siano definite k scelte successive su un insieme E, tali che ogni scelta abbia un numero fissato di esiti possibili: la prima scelta ha n_1 esiti possibili, la seconda scelta ne ha n_2, ..., la k-esima scelta ne ha n_k, dove $n_1, \dots, n_k \in \mathbb{N}$. Supponiamo che gli elementi di E siano determinati univocamente dalle k scelte, cioè $|E_{i_1,\dots,i_k}| = 1$ per ogni scelta di $i_1, \dots, i_k$. Allora la cardinalità di E è pari a $n_1 \cdot n_2 \cdots n_k$.*

Dimostrazione. Per definizione di scelte successive, $E = \bigcup_{i=1}^{n_1} E_i$; a sua volta $E_i = \bigcup_{j=1}^{n_2} E_{i,j}$, eccetera: di conseguenza vale la relazione

$$E = \bigcup_{i_1=1}^{n_1} \dots \bigcup_{i_k=1}^{n_k} E_{i_1,\dots,i_k}. \tag{1.12}$$

Mostriamo che questa unione è disgiunta, cioè $E_{i_1,\dots,i_k} \cap E_{i'_1,\dots,i'_k} = \emptyset$ se $(i_1, \dots, i_k) \neq (i'_1, \dots, i'_k)$. Se $(i_1, \dots, i_k) \neq (i'_1, \dots, i'_k)$ significa che $i_j \neq i'_j$ per qualche $1 \le j \le k$: prendendo il più piccolo di tali valori di j, possiamo supporre che $i_1 = i'_1$, ..., $i_{j-1} = i'_{j-1}$ mentre $i_j \neq i'_j$. Per definizione di scelte successive, $E_{i_1,\dots,i_k} \subseteq E_{i_1,\dots,i_{j-1},i_j}$ e analogamente $E_{i'_1,\dots,i'_k} \subseteq E_{i'_1,\dots,i'_{j-1},i'_j} = E_{i_1,\dots,i_{j-1},i'_j}$, per cui basta mostrare che $E_{i_1,\dots,i_{j-1},i_j} \cap E_{i_1,\dots,i_{j-1},i'_j} = \emptyset$. Ma per ipotesi gli insiemi $\{E_{i_1,\dots,i_{j-1},\ell}\}_{\ell \in \{1,\dots,n_j\}}$ formano una partizione di $E_{i_1,\dots,i_{j-1}}$, in particolare sono a due a due disgiunti: quindi $E_{i_1,\dots,i_{j-1},i_j} \cap E_{i_1,\dots,i_{j-1},i'_j} = \emptyset$ poiché $i_j \neq i'_j$. Essendo l'unione in (1.12) disgiunta e ricordando che per ipotesi $|E_{i_1,\dots,i_k}| = 1$, si ottiene

$$|E| = \sum_{i_1=1}^{n_1} \dots \sum_{i_k=1}^{n_k} |E_{i_1,\dots,i_k}| = n_1 \cdot n_2 \cdots n_k,$$

che è la relazione voluta.

Ritornando per un'ultima volta all'insieme E delle funzioni iniettive da $\{1, \dots, k\}$ in $A = \{a_1, \dots, a_n\}$, è facile verificare che valgono le ipotesi del Teorema 1.2: le scelte di $f(1)$, $f(2)$, ..., $f(k)$ costituiscono k scelte successive su E e gli elementi di E sono univocamente determinati da queste scelte. Dato che la scelta di $f(i)$ ha $n-i+1$ esiti possibili, segue dal Teorema 1.1 che vale la formula $|E| = n(n-1)\cdots(n-k+1)$ ottenuta in precedenza.

[1] Si noti che è risultato conveniente indicizzare gli esiti della seconda scelta usando l'insieme $\{1, \dots, n\} \setminus \{i\}$ piuttosto che $\{1, \dots, n-1\}$.

1.3.4 Disposizioni semplici e permutazioni

Dato un insieme $A = \{a_1, \ldots, a_n\}$ di cardinalità $n \in \mathbb{N}$ e dato $k \in \mathbb{N}$, abbiamo visto che le funzioni da $\{1, \ldots, k\}$ a valori in A sono dette disposizioni con ripetizione di k elementi estratti da A, e hanno cardinalità n^k. Se $k \leq n$, le funzioni *iniettive* da $\{1, \ldots, k\}$ in A sono dette *disposizioni semplici (o senza ripetizione) di k elementi estratti da A*. Abbiamo già incontrato questo insieme di funzioni nello scorso paragrafo e abbiamo visto che la sua cardinalità è data dalla formula $n(n-1)\cdots(n-k+1)$.

Nel caso speciale in cui $k = n$, le disposizioni semplici di n elementi estratti da A sono dette *permutazioni di A*. Si osservi che una permutazione $f : A \to A$ può essere vista come una elencazione ordinata di tutti gli elementi di A, cioè $(f(1), f(2), \ldots, f(n))$. È interessante notare che l'insieme delle permutazioni di un insieme fissato A costituisce un *gruppo* rispetto alla composizione di applicazioni, che è non commutativo per $n \geq 3$. A meno di corrispondenze biunivoche, non costa niente considerare il caso "speciale" $A = \{1, \ldots, n\}$: in questo caso, il gruppo delle permutazioni è indicato con S_n. Munito della probabilità uniforme, lo spazio S_n ha proprietà interessanti e per certi versi sorprendenti, alcune delle quali verranno discusse nel paragrafo 2.2.

Introduciamo il simbolo $n!$, detto "n fattoriale", definito da

$$n! := n(n-1)\cdots 1 = \prod_{i=1}^{n} i \quad \text{per } n \in \mathbb{N}, \qquad 0! := 1 \,. \tag{1.13}$$

Per quanto abbiamo visto, la cardinalità del gruppo delle permutazioni di un insieme di n elementi è pari a $n!$ (in particolare $|S_n| = n!$). Analogamente, possiamo riscrivere la cardinalità delle disposizioni semplici di k elementi estratti da un insieme che ne contiene n come $n!/(n-k)!$.

Esempio 1.7. Supponiamo di mischiare un mazzo di carte da poker. La sequenza ordinata delle carte che ne risulta è una permutazione delle carte del mazzo. Il numero delle possibili sequenze ottenute in questo modo è dunque pari $52! \approx 8 \cdot 10^{67}$.

1.3.5 Combinazioni

Sia $A = \{a_1, \ldots, a_n\}$ un insieme di cardinalità $n \in \mathbb{N}$ e sia $k \in \mathbb{N}_0$ con $0 \leq k \leq n$. I sottoinsiemi di A di cardinalità k sono detti *combinazioni di k elementi estratti da A*. Se una disposizione semplice corrisponde a una sequenza ordinata, una combinazione può essere vista come una collezione *non ordinata* di elementi. Indichiamo con $\mathscr{C}_{n,k}$ l'insieme delle combinazioni e con $\mathscr{D}_{n,k}$ l'insieme delle disposizioni semplici di k elementi estratti da A.

Per $k = 0$ si ha $\mathscr{C}_{n,0} = \{\emptyset\}$ e dunque $|\mathscr{C}_{n,0}| = 1$. Per determinare $|\mathscr{C}_{n,k}|$ per $k \in \{1, \ldots, n\}$, ricordiamo che ci sono $|\mathscr{D}_{n,k}| = n!/(n-k)!$ disposizioni (cioè sequenze

ordinate) di k elementi distinti estratti da A. Dato che nelle combinazioni l'ordine degli elementi non conta, dobbiamo identificare le disposizioni che danno origine alla stessa combinazione, cioè che selezionano lo stesso sottoinsieme di A: dato che ci sono $k!$ riordinamenti possibili (cioè permutazioni) di k elementi fissati, si ottiene $|\mathscr{C}_{n,k}| = |\mathscr{D}_{n,k}|/k! = \binom{n}{k}$, dove abbiamo introdotto il *coefficiente binomiale*, definito da

$$\binom{n}{k} := \frac{n!}{k!\,(n-k)!}\,, \qquad \text{per } n \in \mathbb{N}_0\,,\ k \in \{0,\dots,n\}\,.$$

Si noti che la formula $|\mathscr{C}_{n,k}| = \binom{n}{k}$ vale anche per $k = 0$.

Questo argomento può essere formalizzato in modo più preciso (e più tecnico). Cominciamo con un piccolo risultato preparatorio. Siano D, E due insiemi finiti e sia $g : D \to E$ un'applicazione suriettiva. Per ogni $y \in E$, introduciamo il sottoinsieme $g^{-1}(y) := \{x \in D :\ g(x) = y\}$ costituito dagli elementi di D che vengono mandati da g in y. Supponiamo che valga la seguente proprietà: esiste $k \in \mathbb{N}$ tale che, per ogni $y \in E$, si ha $|g^{-1}(y)| = k$ (cioè per ogni $y \in E$ esistono esattamente k elementi $x \in D$ che vengono mandati da g in y). Segue allora che $|D| = k\,|E|$. La dimostrazione è semplice: possiamo sempre scrivere $E = \bigcup_{y\in D} g^{-1}(y)$ e inoltre l'unione è disgiunta (esercizio). Quindi $|E| = \sum_{y\in D} |g^{-1}(y)| = \sum_{y\in D} k = k\,|D|$.

Fissiamo ora $k \in \{1,\dots,n\}$ e definiamo una applicazione $g : \mathscr{D}_{n,k} \to \mathscr{C}_{n,k}$ nel modo seguente: data $f \in \mathscr{D}_{n,k}$, definiamo $g(f) := \mathrm{Im}(f)$, dove $\mathrm{Im}(f)$ indica l'immagine di f (ricordiamo che f è una funzione iniettiva da $\{1,\dots,k\}$ in A). È immediato verificare che g è ben definita, cioè effettivamente $g(f) \in \mathscr{C}_{n,k}$ per ogni $f \in \mathscr{D}_{n,k}$, e che g è suriettiva. Se mostriamo che $|g^{-1}(B)| = k!$, per ogni $B \in \mathscr{C}_{n,k}$, si ottiene $|\mathscr{D}_{n,k}| = k!\,|\mathscr{C}_{n,k}|$ e quindi la formula $|\mathscr{C}_{n,k}| = \binom{n}{k}$ è dimostrata.

Indichiamo con S_B l'insieme delle permutazioni di B, cioè le applicazioni $\pi : B \to B$ biunivoche, e fissiamo un elemento arbitrario $f_0 \in g^{-1}(B)$. È molto facile convincersi che, per ogni $\pi \in S_B$, si ha $\pi \circ f_0 \in g^{-1}(B)$: infatti l'applicazione $\pi \circ f_0$ è iniettiva, perché lo sono sia f_0 sia π, e $\mathrm{Im}(\pi \circ f_0) = \mathrm{Im}(f_0) = B$, perché π è una permutazione di B. Risulta dunque ben posta l'applicazione $H : S_B \to g^{-1}(B)$ definita da $H(\pi) := \pi \circ f_0$. Supponiamo che $H(\pi_1) = H(\pi_2)$: per ogni $b \in B$, se $i \in \{1,\dots,k\}$ è tale che $f_0(i) = b$ (tale i esiste perché $\mathrm{Im}(f_0) = B$), otteniamo $(\pi_1 \circ f_0)(i) = (\pi_2 \circ f_0)(i)$, cioè $\pi_1(b) = \pi_2(b)$; dato che $b \in B$ è arbitrario, segue che $\pi_1 = \pi_2$, dunque l'applicazione H è iniettiva. Se ora consideriamo un arbitrario $f \in g^{-1}(B)$, è facile costruire $\pi \in S_B$ tale che $\pi \circ f_0 = f$, cioè $H(\pi) = f$, quindi l'applicazione H è suriettiva. Avendo mostrato che H è biunivoca, segue che gli insiemi S_B e $g^{-1}(B)$ sono in corrispondenza biunivoca e dunque $|g^{-1}(B)| = |S_B| = k!$ se $B \in \mathscr{C}_{n,k}$, che è quanto restava da dimostrare.

Elenchiamo alcune semplici proprietà dei coefficienti binomiali:

$$\binom{n}{0} = \binom{n}{n} = 1\,, \qquad \binom{n}{k} = \binom{n}{n-k}\,, \qquad \forall n \in \mathbb{N}_0\,,\ \forall k \in \{0,\dots,n\}\,.$$

Vale inoltre la relazione

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}\,, \qquad \forall n \in \mathbb{N}\,,\ \forall k \in \{0,\dots,n\}\,, \tag{1.14}$$

come si verifica facilmente.[2] Ricordiamo infine la formula nota come *binomio di*

[2] Esiste anche una dimostrazione combinatoria della relazione (1.14). Le combinazioni di k elementi estratte da $\{1,\dots,n\}$, che sono $\binom{n}{k}$, possono essere infatti divise in due sottoinsiemi: quelle che contengono 1 e quelle che non lo contengono; le prime sono in corrispondenza biunivoca con le combinazioni di $(k-1)$ elementi estratti da $\{2,\dots,n\}$, e sono dunque $\binom{n-1}{k-1}$, mentre le seconde sono in corrispondenza biunivoca con le combinazioni di k elementi estratti da $\{2,\dots,n\}$, che sono

Newton:

$$(a+b)^n \;=\; \sum_{k=0}^{n} \binom{n}{k} a^k\, b^{n-k}, \qquad \forall n \in \mathbb{N}_0\,,\ \forall a,b \in \mathbb{R}\,, \tag{1.15}$$

che si dimostra per induzione usando (1.14).

Esempio 1.8. Ritornando brevemente all'Esempio 1.6, il numero di modi di scegliere 3 "semi" tra i quattro possibili $\{\heartsuit, \diamondsuit, \clubsuit, \spadesuit\}$ è pari al numero di combinazioni di 3 elementi estratti da un insieme che ne contiene 4 ed è dunque dato da $\binom{4}{3} = 4$ (come avevamo concluso per enumerazione diretta). Analogamente, il numero di modi di scegliere 2 semi è pari a $\binom{4}{2} = 6$ e il numero di modi di scegliere due "tipi" tra i 13 possibili è pari a $\binom{13}{2} = 78$.

Una "mano" a Poker è un sottoinsieme di 5 carte distinte estratte da un mazzo che ne contiene 52. Il numero di possibili mani è dato dunque da $\binom{52}{5} = 2\,598\,960$. Ricordando l'Esempio 1.6, le probabilità di fare *full* oppure *doppia coppia* valgono rispettivamente $3\,744/2\,598\,960 \approx 0.14\%$ e $123\,552/2\,598\,960 \approx 4,8\%$.

Esempio 1.9. Si consideri un'urna contenente N palline, di cui m rosse e $N-m$ verdi (con $m \leq N$). Supponiamo di eseguire n estrazioni successive, secondo uno dei seguenti due schemi di estrazione:

- *Estrazioni con reimmissione*. Dopo ogni estrazione, la pallina estratta viene reinserita nell'urna.
- *Estrazioni senza reimmissione*. Le palline estratte non vengono reinserite. In questo caso dev'essere $n \leq N$.

Calcoliamo, per ciascuno dei due schemi, la probabilità che esattamente k delle n palline stratte siano rosse.

Caso di estrazioni con reimmissione. Supponiamo di numerare le palline da 1 a N e, per fissare le idee, assumiamo che le palline rosse siano quelle numerate da 1 a m. L'esito di n estrazioni successive può essere interpretato come una disposizione con ripetizione di n elementi presi dall'insieme $\{1,2,\ldots,N\}$. Sia dunque Ω l'insieme di tali disposizioni, e P la probabilità uniforme su Ω. Sappiamo che $|\Omega| = N^n$. Denotiamo infine con A l'insieme delle disposizioni contenenti esattamente k palline rosse. Si tratta di calcolare

$$P(A) \;=\; \frac{|A|}{|\Omega|}\,.$$

Per determinare $|A|$, utilizziamo il principio fondamentale. Un elemento di A è determinato dalle seguenti scelte successive.

- Si scelgono le k posizioni (su n possibili) in cui mettere le palline rosse: per questa scelta ci sono $\binom{n}{k}$ esiti possibili.
- Si dispongono k palline rosse (prese dalle m presenti nell'urna) nelle posizioni prescelte: ci sono m^k tali disposizioni.
- Si dispongono $(n-k)$ palline verdi (prese dalle $(N-m)$ presenti nell'urna) nelle rimanenti posizioni: ci sono $(N-m)^{n-k}$ tali disposizioni.

$\binom{n-1}{k}$.

Si ottiene pertanto

$$|A| \;=\; \binom{n}{k} m^k (N-m)^{n-k},$$

da cui segue facilmente che

$$P(A) \;=\; \binom{n}{k} \left(\frac{m}{N}\right)^k \left(1-\frac{m}{N}\right)^{n-k}.$$

Questo risultato verrà reinterpretato più avanti, nell'Esempio 1.17.

Caso di estrazioni senza reimmissione. Enumeriamo le palline come nel caso precedente. Un naturale spazio campionario, in cui la probabilità uniforme esprime la casualità dell'estrazione, è quello delle *disposizioni senza ripetizione*. Poichè, tuttavia, l'evento "il numero di palline rosse estratte è k" non dipende dall'ordine di estrazione, è forse ancora più naturale scegliere come spazio campionario l'insieme delle *combinazioni*. Sia dunque Ω l'insieme dei sottoinsiemi di n elementi dell'insieme $\{1,2,\ldots,N\}$, e P la probabilità uniforme su di esso. L'evento di cui vogliamo calcolare la probabilità è

$$A \;=\; \{\omega \in \Omega : |\,\omega \cap \{1,2,\ldots,m\}| = k\} \;=\; \{\omega \in \Omega : |\,\omega \cap \{m+1,\ldots,N\}| = n-k\}.$$

Chiaramente $A = \emptyset$ se $k > m$ oppure se $(n-k) > (N-m)$. Se invece $k \le m$ e $(n-k) \le (N-m)$, ogni elemento di A è determinato da due scelte successive: occorre scegliere k elementi da $\{1,2,\ldots,m\}$ e $(n-k)$ da $\{m+1,\ldots,N\}$. Di conseguenza possiamo scrivere

$$|A| \;=\; \binom{m}{k}\binom{N-m}{n-k},$$

dove usiamo la convenzione secondo cui $\binom{i}{j} = 0$ se $j < 0$ o $j > i$. Ricordando che $|\Omega| = \binom{N}{n}$, possiamo dunque concludere che

$$P(A) \;=\; \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}.$$

*1.3.6 Coefficienti multinomiali**

Dato un insieme A, ricordiamo che una partizione di A è una famiglia $\{A_1,\ldots,A_r\}$ di sottoinsiemi di A tali che $A_i \cap A_j = \emptyset$, per ogni $i \neq j$, e $\bigcup_{i=1}^r A_i = A$. Sottolineiamo che l'ordine dei sottoinsiemi che compaiono nella partizione è importante: $\{A_1,A_2,A_3\}$ e $\{A_2,A_1,A_3\}$ sono due partizioni distinte. Supponiamo ora che $|A| = n \in \mathbb{N}_0$ e siano assegnati $r \in \mathbb{N}$ e $k_1,\ldots,k_r \in \mathbb{N}_0$ tali che $k_1 + \ldots + k_r = n$: ci proponiamo di calcolare il numero di partizioni $\{A_1,\ldots,A_r\}$ di A tali che $|A_1| = k_1, \ldots, |A_r| = k_r$.

Il caso $r = 1$ è banale: esiste infatti un'unica partizione di A in un solo sottoinsieme, che è naturalmente A stesso. Il caso $r = 2$ è strettamente connesso con le com-

binazioni: infatti una partizione in due sottoinsiemi $\{A_1, A_2\}$ è identificata da A_1, poiché $A_2 = A_1^c$; dovendo essere $|A_1| = k_1$, il numero di tali partizioni coincide con il numero di combinazioni di k_1 elementi estratti da A, pari a $\binom{n}{k_1} = \frac{n!}{k_1!(n-k_1)!} = \frac{n!}{k_1!k_2!}$. A questo punto non è difficile indovinare la formula nel caso generale: il numero di partizioni $\{A_1, \ldots, A_r\}$ di A tali che $|A_1| = k_1, \ldots, |A_r| = k_r$ (dove $k_1 + \ldots + k_r = n$) è dato da

$$\frac{n!}{k_1!\, k_2! \cdots k_r!}\,. \tag{1.16}$$

La dimostrazione è una applicazione del principio fondamentale del calcolo combinatorio. Una partizione $\{A_1, \ldots, A_r\}$ può essere infatti determinata mediante le seguenti scelte successive:

- si sceglie A_1 come sottoinsieme di A di cardinalità k_1, in $\binom{n}{k_1}$ modi possibili;
- si sceglie A_2 come sottoinsieme di $A \setminus A_1$ di cardinalità k_2, in $\binom{n-k_1}{k_2}$ modi possibili;
- si procede analogamente fino alla scelta di A_{r-1}, per cui ci sono $\binom{n-(k_1+k_2+\ldots+k_{r-2})}{k_{r-1}}$ possibilità;
- una volta scelto A_{r-1}, l'ultimo elemento della partizione è automaticamente determinato dalla relazione $A_r = A \setminus (A_1 \cup \ldots \cup A_{r-1})$.

Applicando il Teorema 1.1, il numero di partizioni cercato è dato dunque da

$$\binom{n}{k_1} \cdot \binom{n-k_1}{k_2} \cdot \binom{n-(k_1+k_2)}{k_3} \cdots \binom{n-(k_1+k_2+\ldots+k_{r-2})}{k_{r-1}}\,,$$

che non è altro che una riscrittura di (1.16), come si verifica dopo qualche semplificazione.

La quantità in (1.16), per $r \in \mathbb{N}$ e $n, k_1, \ldots, k_r \in \mathbb{N}_0$ con $k_1 + \ldots + k_r = n$, è detta *coefficiente multinomiale*: può essere infatti vista come un'estensione del coefficiente binomiale, a cui si riduce per $r = 2$. I coefficienti multinomiali ricorrono frequentemente nei cosiddetti problemi di occupazione. Un caso molto importante è descritto nell'esempio seguente.

Esempio 1.10. Supponiamo di disporre casualmente n palline numerate in r urne, dove $n \in \mathbb{N}_0$, $r \in \mathbb{N}$ (ogni urna può contenere un numero qualunque di palline). Chiediamoci qual è la probabilità che la prima urna contenga k_1 palline, la seconda ne contenga k_2, …, la r-esima ne contenga k_r, dove naturalmente $k_1, \ldots, k_r \in \mathbb{N}_0$ sono tali che $k_1 + \ldots + k_r = n$.

Lo spazio di probabilità naturale per questo esperimento aleatorio è l'insieme Ω delle disposizioni con ripetizione di n elementi presi dall'insieme $\{1, \ldots, r\}$ (che rappresenta le urne), vale a dire

$$\Omega_{n,r} := \big\{ f : \{1, \ldots, n\} \to \{1, \ldots, r\} \big\}\,, \tag{1.17}$$

munito della probabilità P uniforme, visto che la disposizione delle palline nelle urne avviene in modo completamente casuale. L'evento che ci interessa (che la prima

urna contenga k_1 palline, la seconda ne contenga k_2, ..., la r-esima ne contenga k_r) è rappresentato da

$$\mathscr{O}_{k_1,\ldots,k_r} := \{f \in \Omega_{n,r} : |f^{-1}(1)| = k_1, \ldots, |f^{-1}(r)| = k_r\}.$$

Essendo lo spazio di probabilità uniforme, dobbiamo calcolare la cardinalità di $\mathscr{O}_{k_1,\ldots,k_r}$, cioè il numero di disposizioni di palline nelle urne tali che la prima urna ne contenga k_1, la seconda ne contenga k_2, ..., la r-esima ne contenga k_r. Un momento di riflessione mostra che la risposta è data proprio dal coefficiente multinomiale (1.16).[3] Di conseguenza, per ogni scelta di $k_1,\ldots,k_r \in \mathbb{N}_0$ con $k_1+\ldots+k_r = n$, si ha

$$P(\mathscr{O}_{k_1,\ldots,k_r}) = \frac{|\mathscr{O}_{k_1,\ldots,k_r}|}{|\Omega_{n,r}|} = \frac{1}{r^n}\frac{n!}{k_1!\cdots k_r!}, \tag{1.18}$$

poiché sappiamo che $|\Omega_{n,r}| = r^n$.

1.4 Probabilità condizionata

Nello studio di un modello probabilistico, risulta interessante studiare l'influenza che l'occorrere di un dato evento B ha sulla probabilità di occorrenza di un altro evento A.

Esempio 1.11. Nelle estrazioni per una ruota del Lotto, vengono estratte "a caso" 5 palline da un'urna contenente palline numerate da 1 a 90. Supponiamo di giocare due numeri su quella ruota, e precisamente l'1 e il 3. Una persona presente all'estrazione, mi avvisa che dei 5 numeri estratti 3 sono dispari. Qual è la probabilità di fare "ambo" sulla base di questa informazione? E qual è la probabilità in assenza di tale informazione?

È chiaro che la soluzione di tale problema richiede che si definisca il significato di calcolare una probabilità *sulla base* di una data informazione. Prima di proporre una definizione formale, cerchiamo una soluzione "ragionevole". Lo spazio campionario in questione è Ω = "insieme di tutte le cinquine di numeri tra 1 e 90". Assumendo l'equità dell'estrazione, scegliamo come probabilità P quella uniforme. Due eventi compaiono nell'enunciato del problema: A = "i cinque numeri estratti contengono l'1 e il 3", B = "dei cinque numeri estratti 3 sono dispari". In assenza dell'informazione sull'occorrenza di B, scriveremmo semplicemente

$$P(A) = \frac{|A|}{|\Omega|}.$$

Poichè la scelta di un elemento di A corrisponde alla scelta di tre numeri diversi da 1 e 3, si ha che $|A| = \binom{88}{3}$, e quindi

[3] Infatti, indicando con $A := \{1,\ldots,n\}$ l'insieme delle palline e con $A_i \subseteq A$ le palline contenute nell'urna i-esima, le disposizioni cercate sono in corrispondenza biunivoca con le partizioni $\{A_1,\ldots,A_r\}$ di A tali che $|A_1| = k_1, \ldots, |A_r| = k_r$.

$$P(A) = \frac{\binom{88}{3}}{\binom{90}{5}} = \frac{20}{8010} \simeq 0.0025.$$

Assumere l'occorrenza di B significa escludere la possibilità che la cinquina estratta non sia in B. Inoltre, anche sapendo che la cinquina estratta è in B, non vi è alcun motivo per rimuovere l'ipotesi di equiprobabilità degli elementi di B. Dunque, la procedura "naturale" consiste nel rimpiazzare lo spazio campionario Ω con B, e calcolare le probabilità dei sottoinsiemi di B secondo la probabilità uniforme su B. Poichè A non è un sottoinsieme di B, si tratterà di calcolare la probabilità di $A \cap B$ secondo la probabilità uniforme su B. Concludiamo allora che l'oggetto più ragionevole per esprimere la probabilità di A *condizionata* all'occorrenza di B è

$$\frac{|A \cap B|}{|B|}.$$

Come esercizio, calcoliamo tale probabilità. Gli elementi di $A \cap B$ sono costituiti dalle cinquine contenenti $1,3$, un altro numero dispari diversi da 1 e 3, e due numeri pari. Dunque

$$|A \cap B| = 43\binom{45}{2}.$$

Inoltre, poichè gli elementi di B contengono 3 numeri dispari e 2 pari,

$$|B| = \binom{45}{3}\binom{45}{2}.$$

Infine

$$\frac{|A \cap B|}{|B|} = \frac{43}{\binom{45}{3}} = \frac{6}{1980} \simeq 0.003,$$

che è maggiore della probabilità in assenza di informazioni.

L'esempio appena trattato assieme all'osservazione che, se P è la probabilità uniforme

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

motiva la definizione che segue.

Definizione 1.2. Sia (Ω, P) uno spazio di probabilità discreto, A e B due eventi per cui $P(B) > 0$. La *probabilità di A condizionata a B* si denota con $P(A|B)$ ed è definita da

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

Alcune proprietà formali della probabilità condizionata sono sintetizzate nella seguente proposizione.

Proposizione 1.4. *Sia B un evento fissato, con $P(B) > 0$, e consideriamo la funzione*

$$\begin{array}{ccc} \mathscr{P}(\Omega) & \longrightarrow & [0,1] \\ A & \to & P(A|B). \end{array}$$

Tale funzione è una probabilità su Ω.

La dimostrazione, che consiste nella verifica della validità degli assiomi (P1) e (P2), è lasciata per esercizio. Vale la pena sottolineare che, fissato un evento A, la funzione $B \mapsto P(A|B)$ *non* è una probabilità.

La seguente proposizione fornisce una caratterizzazione della probabilità condizionata che ne motiva ulteriormente la definizione.

Proposizione 1.5. *Sia B un evento fissato, con $P(B) > 0$. Allora $P(\cdot\,|B)$ è l'unica probabilità Q su Ω con le seguenti proprietà:*

(1) $Q(B) = 1$*;*

(2) *per ogni coppia di eventi E, F con $P(F) > 0$ si ha $\frac{Q(E)}{Q(F)} = \frac{P(E)}{P(F)}$.*

Dimostrazione. È immediato verificare che $P(\cdot\,|B)$ soddisfa le proprietà elencate. Viceversa, sia Q una probabilità che soddisfa le proprietà 1, 2 e sia A un evento arbitrario. Applicando la proprietà 2 agli eventi $E = A \cap B$ e $F = B$, visto che $Q(B) = 1$ per la proprietà 1 si ottiene la relazione

$$Q(A \cap B) \;=\; \frac{Q(A \cap B)}{Q(B)} \;=\; \frac{P(A \cap B)}{P(B)} \;=:\; P(A|B)\,. \tag{1.19}$$

Osserviamo ora che possiamo scrivere

$$A \;=\; A \cap (B \cup B^c) \;=\; (A \cap B) \cup (A \cap B^c)\,.$$

Dato che $A \cap B \subseteq B$ e $A \cap B^c \subseteq B^c$, gli eventi $A \cap B$ e $A \cap B^c$ sono disgiunti, quindi $Q(A) = Q(A \cap B) + Q(A \cap B^c)$. Per la proprietà 1 si ha $Q(B) = 1$, quindi $Q(A \cap B^c) \leq Q(B^c) = 0$ e di conseguenza $Q(A) = Q(A \cap B)$. Ricordando l'equazione (1.19), abbiamo dimostrato che $Q(A) = P(A|B)$, cioè Q coincide con la probabilità condizionata a B.

In molte situazioni, la nozione di probabilità condizionata è utile nella costruzione stessa di un modello probabilistico: talvolta è "naturale" assegnare il valore di alcune probabilità condizionate, e da esse dedurre il valore di probabilità non condizionate.

Esempio 1.12. Due urne contengono, rispettivamente, 3 palline rosse e 1 verde e 1 pallina rossa e 1 verde. Si sceglie, con ugual probabilità, una delle due urne e poi, dall'urna scelta, si estrae una pallina. Qual è la probabilità di estrarre una pallina rossa?

Denotiamo con a e b le due urne. Come spazio campionario, si può scegliere l'insieme costituito dalle coppie $(a,r), (a,v), (b,r), (b,v)$, dove la prima componente indica l'urna scelta e la seconda il colore della pallina estratta. L'evento

$A = \{(a,r),(a,v)\}$ corrisponde a "l'urna scelta è la a", l'evento $R = \{(a,r),(b,r)\}$ corrisponde a "la pallina estratta è rossa". Dev'essere senz'altro $P(A) = 1/2$, visto che le urne vengono scelte con uguale probabilità. Inoltre, *supponendo* di aver scelto l'urna a, la probabilità di estrarre una pallina rossa è $3/4$. Perciò porremo $P(R|A) = 3/4$. Analogamente $P(R|A^c) = 1/2$. Il procedimento per dedurre $P(R)$ dai dati a disposizione è indicato dal risultato che segue.

Proposizione 1.6. *Sia $(B_n)_{n=1}^N$ una sequenza di eventi finita ($N < +\infty$) o infinita ($N = +\infty$) tali che*

(a) *Per ogni n*
$$P(B_n) > 0.$$

(b) *Gli eventi sono a due a due disgiunti, cioè*
$$B_n \cap B_m = \emptyset$$
se $n \neq m$.

(c)
$$\bigcup_{n=1}^{N} B_n = \Omega.$$

Allora, per ogni evento A,
$$P(A) = \sum_{n=1}^{N} P(A|B_n)P(B_n).$$

Tale identità prende il nome di formula delle probabilità totali.

Dimostrazione. Si osservi che
$$A = \bigcup_{n=1}^{N} (A \cap B_n),$$
e gli eventi di quest'ultima unione sono disgiunti. Usando l'additività di P e la definizione di probabilità condizionata, si ha
$$P(A) = \sum_{n=1}^{N} P(A \cap B_n) = \sum_{n=1}^{N} P(A|B_n)P(B_n).$$

Supponiamo ora che A e B siano due eventi tali che $P(A) > 0$, $P(B) > 0$, sicché entrambe le probabilità condizionate $P(A|B)$ e $P(B|A)$ sono definite. È pressoché immediato verificare la seguente relazione.

Teorema 1.3. *(Formula di Bayes) Se $P(A) > 0$ e $P(B) > 0$, allora*
$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}. \tag{1.20}$$

Dimostrazione. La formula di Bayes (1.20) è equivalente a

$$P(B|A)P(A) = P(A|B)P(B),$$

che è vera in quanto entrambi i membri sono uguali a $P(A \cap B)$.

Nell'ipotesi che $0 < P(B) < 1$, usando la formula delle probabilità totali, la formula di Bayes può essere riscritta nella forma

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}. \tag{1.21}$$

Analogamente, se $(B_n)_{n=1}^N$ è una sequenza di eventi soddisfacenti alle ipotesi della Proposizione 1.6, si ha

$$P(B_n|A) = \frac{P(A|B_n)P(B_n)}{P(A)} = \frac{P(A|B_n)P(B_n)}{\sum_{k=1}^N P(A|B_k)P(B_k)}. \tag{1.22}$$

Le versioni (1.21) e (1.22) della formula di Bayes sono quelle che più spesso capita di usare negli esercizi.

La formula di Bayes, a dispetto della sua semplicità, è una delle formule fondamentali della Probabilità, ed è all'origine di un'intera area della Statistica, la *Statistica Bayesiana*. La rilevanza della formula di Bayes nelle applicazioni, si può già apprezzare in applicazioni semplici, come quella che segue.

Esempio 1.13. Per determinare la presenza di un certo virus viene elaborato un test clinico avente la seguente efficacia: se il virus è presente allora il test risulta positivo il 99% dei casi; se il virus è assente il test risulta positivo il 2% dei casi. E' noto che 2 persone su 10.000 hanno il virus. Supponiamo che un individuo scelto a caso risulti positivo al test. Con quale sicurezza possiamo affermare che sia malato?

Come accade sovente negli esercizi in cui si applica la formula di Bayes, non è rilevante descrivere nel dettaglio lo spazio campionario. Si considerino gli eventi, così descritti in modo informale: $A =$ "l'individuo è malato"; $B =$ "il test è risultato positivo". I dati del problema sono:

$$\begin{aligned} P(A) &= 0.0002 \\ P(B|A) &= 0.99 \\ P(B|A^c) &= 0.02. \end{aligned} \tag{1.23}$$

Calcoliamo $P(A|B)$. Utilizzando la formula di Bayes e la formula delle probabilità totali, si ha

$$P(A|B) = P(B|A)\frac{P(A)}{P(B)} = P(B|A)\frac{P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)} \simeq 0.01$$

che è estremamente bassa. Quindi, anche se un individuo risulta positivo al test, è molto improbabile che sia malato. Questo test dunque darà una grande percentuale di falsi positivi.

E se avessimo voluto specificare per bene lo spazio campionario? Si sarebbe potuto procedere così. Definiamo

$$\Omega = \{(m,p),(m,n),(s,p),(s,n)\} = \{m,s\} \times \{p,n\}$$

dove m e s indicano la presenza (m) o l'assenza del virus, p e n il risultato del test: p = positivo, n = negativo. Qual è la probabilità P su Ω? Per individuare P dobbiamo usare i dati del problema. Si noti che gli eventi A e B definiti sopra, corrispondono ai seguenti sottoinsiemi di Ω:

$$A = \{(m,p),(m,n)\}$$
$$B = \{(m,p),(s,p)\}.$$

Usando i dati in (1.23), è effettivamente possibile calcolare la probabilità di tutti i sottoinsiemi di Ω (provarci!), da cui si può dedurre il valore di $P(A|B)$. Tuttavia, per rispondere al quesito posto, questi dettagli sono poco rilevanti.

1.5 Indipendenza di eventi

Si è visto come la probabilità condizionata $P(A|B)$ rappresenti la probabilità dell'evento A sotto la condizione del verificarsi dell'evento B. E' possibile che tale condizione non modifichi la probabilità di A, ossia

$$P(A|B) = P(A). \tag{1.24}$$

Usando la definizione di probabilità condizionata, si vede che l'identità (1.24) equivale a:

$$P(A \cap B) = P(A)P(B). \tag{1.25}$$

L'identità in (1.25), rispetto a quella in (1.24) ha il vantaggio di essere esplicitamente simmetrica in A e B, e di essere definita (e banalmente vera) anche quando $P(B) = 0$. Essa viene dunque scelta per caratterizzare la nozione di *indipendenza*.

Definizione 1.3. In uno spazio di probabilità discreto (Ω, P), due eventi A e B si dicono *indipendenti* se

$$P(A \cap B) = P(A)P(B).$$

Esempio 1.14. Da due mazzi di carte da Poker si estraggono due carte, una per mazzo. Lo spazio campionario naturale è l'insieme delle coppie (i, j) nel prodotto cartesiano $\Omega = X \times X$, dove X è l'insieme delle carte di un mazzo. Possiamo assumere che la scelta sia "casuale", cioè descritta dalla probabilità P uniforme su Ω. Consideriamo due eventi A e B, di cui l'evento A dipende solo dall'estrazione dal primo mazzo, l'evento B solo dall'estrazione dal secondo mazzo. In altre parole, se $F, G \subseteq X$, A e B sono della forma:

$$A = \{(i,j) \in \Omega : i \in F\}$$
$$B = \{(i,j) \in \Omega : j \in G\}.$$

Si noti che $|\Omega| = 52^2$, $|A| = 52|F|$, $|B| = 52|G|$, $|A \cap B| = |F||G|$. Ne segue facilmente che

$$P(A \cap B) = \frac{|A \cap B|}{|\Omega|} = \frac{|F||G|}{52^2} = \frac{|A|}{|\Omega|}\frac{|B|}{|\Omega|} = P(A)P(B).$$

Dunque A e B sono indipendenti. Notare che gli eventi A e B si riferiscono a due *ripetizioni* dello stesso esperimento aleatorio. L'indipendenza esprime il fatto che l'esito di un esperimento non "influenza" l'esito dell'altro esperimento. Questo contesto di *prove indipendenti ripetute*, rilevante in molti aspetti della Probabilità e della Statistica, è quello in cui la nozione di indipendenza appare in modo naturale.

L'esempio 1.14 si può facilmente generalizzare al caso di 3 o più mazzi di carte. In questo caso, P è la probabilità uniforme si $\Omega = X^n$, $F_1, F_2, \ldots, F_n \subseteq X$ e, per $i = 1,2,\ldots,n$, $A_i = \{\omega = (\omega_1, \omega_2, \ldots, \omega_n) \in X^n : \omega_i \in F_i\}$. Una semplice generalizzazione dell'argomento visto sopra, mostra che

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = \prod_{i=1}^{n} P(A_i). \tag{1.26}$$

Naturalmente se avessimo considerato solo alcuni degli A_i, per esempio la coppia A_i, A_j con $i \neq j$, lo stesso facile calcolo sulle cardinalità avrebbe mostrato che

$$P(A_i \cap A_j) = P(A_i)P(A_j).$$

Tutto ciò suggerisce che, nell'estendere la nozione di indipendenza a tre o più eventi, dobbiamo richiedere la proprietà "moltiplicativa" (1.26), ma anche il fatto che se una famiglia di eventi è costituita da eventi indipendenti, anche ogni sua sottofamiglia è costituita da eventi indipendenti. Queste due richieste, non sono implicate l'una dall'altra, come mostrano i seguenti esempi.

Esempio 1.15. Sia $\Omega = \{1,2,3,4\}$, P = Probabilità uniforme, $A = \{1,2\}$, $B = \{2,3\}$, $C = \{1,3\}$. Si vede immediatamente che le coppie (A,B), (B,C) e (A,C) sono formate da eventi indipendenti. Tuttavia:

$$P(A \cap B \cap C) = P(\emptyset) = 0 \neq P(A)P(B)P(C) = 1/8.$$

Esempio 1.16. Sia $\Omega = \{1,2,\ldots,6\}^2$, con la probabilità uniforme. Siano

$$A = \{(i,j) : j = 1,2, \text{ o } 5\}$$
$$B = \{(i,j) : j = 4,5, \text{ o } 6\}$$
$$C = \{(i,j) : i + j = 9\}.$$

Si ha

$$P(A\cap B)=\frac{1}{6}\neq\frac{1}{4}=P(A)P(B)$$
$$P(A\cap C)=\frac{1}{36}\neq\frac{1}{18}=P(A)P(C)$$
$$P(B\cap C)=\frac{1}{12}\neq\frac{1}{18}=P(B)P(C)$$

ma

$$P(A\cap B\cap C)=\frac{1}{36}=P(A)P(B)P(C).$$

Definizione 1.4. Sia I un qualunque insieme di indici, e sia $\{A_i : i\in I\}$ una famiglia di eventi in uno spazio di probabilità discreto (Ω,P). Diremo che tali eventi sono *indipendenti* se per ogni sottoinsieme finito J di I, si ha

$$P\left(\bigcap_{j\in J}A_j\right)=\prod_{j\in J}P(A_j).$$

La proposizione che segue afferma che se in una famiglia di eventi indipendenti si rimpiazzano alcuni eventi con i loro complementari, si ottiene ancora una famiglia di eventi indipendenti.

Proposizione 1.7. *Sia $\{A_i : i\in I\}$una famiglia di eventi indipendenti, $I'\subseteq I$, e definiamo*

$$B_i=\begin{cases}A_i^c \text{ se } i\in I'\\ A_i \text{ se } i\in I\setminus I'.\end{cases}$$

Allora $\{B_i : i\in I\}$ è una famiglia di eventi indipendenti.

Dimostrazione. Sia $J\subset I$ finito, e sia $J'=J\cap I'$. Dobbiamo mostrare che

$$P\left(\bigcap_{j\in J}B_j\right)=\prod_{j\in J}P(B_j). \tag{1.27}$$

Possiamo supporre che

$$J=\{j_1,j_2,\dots,j_m\},\ J'=\{j_1,\dots,j_k\},$$

dove $k\leq m$. Se $k=0$, la (1.27) segue immediatamente dall'indipendenza di $\{A_i : i\in I\}$. Supponiamo $k=1$. Usando l'indipendenza di $\{A_i : i\in I\}$, si ha

$$\begin{aligned}P\left(B_{j_1}\cap B_{j_2}\cap\cdots\cap B_{j_k}\right)&=P\left(A_{j_1}^c\cap A_{j_2}\cap\cdots\cap A_{j_k}\right)\\&=P\left(\left[A_{j_2}\cap\cdots\cap A_{j_k}\right]\setminus\left[A_{j_1}\cap A_{j_2}\cap\cdots\cap A_{j_k}\right]\right)\\&=P(A_{j_2})\cdots P(A_{j_k})-P(A_{j_1})P(A_{j_2})\cdots P(A_{j_k})\\&=[1-P(A_{j_1})]P(A_{j_2})\cdots P(A_{j_k})\\&=P(A_{j_1}^c)P(A_{j_2})\cdots P(A_{j_k})\\&=P(B_{j_1})P(B_{j_2})\cdots P(B_{j_k}).\end{aligned}$$

A questo punto si procede per induzione su k, per trattare tutti i casi $0 \leq k \leq m$. Si lasciano i semplici dettagli al lettore.

Nel prossimo esempio vedremo un calcolo classico basato sull'indipendenza.

Esempio 1.17. Si eseguono N prove ripetute di un gioco in cui la probabilità di vincere è $p \in [0,1]$. Si assuma che i risultati di prove distinte siano indipendenti.

a. Qual è la probabilità di vincere $n \leq N$ volte sugli N tentativi?

b. Qual è la probabilità di vincere per la prima volta all'n-esimo tentativo?

Come spazio campionario possiamo scegliere

$$\Omega = \{x = (x_1, x_2, \ldots, x_N) : x_i \in \{0,1\}\} = \{0,1\}^N,$$

dove $x_i = 1$ significa che l'i-esima prova del gioco è stata vinta. Definiamo per $t \in \{0,1\}$ l'evento

$$A_i(t) = \{x \in \Omega : x_i = t\}.$$

L'indipendenza delle prove effettuate si traduce nel fatto che gli eventi $A_1(t_1)$, $A_2(t_2)$, $\ldots$, $A_N(t_N)$ sono indipendenti, per ogni scelta di $t_1, \ldots, t_N$. Inoltre è chiaro che per ogni $y \in \Omega$

$$\{y\} = \bigcap_{i=1}^{N} A_i(y_i).$$

Per l'indipendenza, allora,

$$P(\{y\}) = \prod_{i=1}^{N} P(A_i(y_i)).$$

Inoltre, poichè per ogni tentativo la probabilità di vincere è p,

$$P(A_i(t)) = \begin{cases} p & \text{se } t = 1 \\ 1-p & \text{se } t = 0 \end{cases}$$

o, equivalentemente, $P(A_i(t)) = p^t(1-p)^{1-t}$. Ne segue allora che

$$P(\{y\}) = \prod_{i=1}^{N} p^{y_i}(1-p)^{1-y_i} = p^{\sum_{i=1}^N y_i}(1-p)^{N-\sum_{i=1}^N y_i}.$$

Gli eventi di cui vogliamo calcolare la probabilità sono

$$A = \{y \in \Omega : \sum_{i=1}^{N} y_i = n\}$$

e

$$B = A_1(0) \cap \cdots \cap A_{n-1}(0) \cap A_n(1).$$

Da quanto visto sopra,

$$\begin{aligned} P(A) &= \sum_{y\in A} P(\{y\}) \\ &= \sum_{y\in A} p^n(1-p)^{N-n} \\ &= |A|p^n(1-p)^{N-n} \\ &= \binom{N}{n} p^n(1-p)^{N-n}, \end{aligned}$$

dove abbiamo usato il fatto che scegliere un elemento di A equivale a scegliere gli n tentativi vincenti sugli N a disposizione. Si noti che l'espressione ottenuta per la probabilità di A è analoga a quella trovata nell'esempio 1.9 per lo schema di estrazioni con reimmissione, con $p := \frac{m}{N}$. Infatti lo schema di estrazioni con reimmissione è un esempio di prove ripetute e indipendenti.

Più facile è calcolare la probabilità di B:

$$P(B) = P(A_1(0))\cdots P(A_{n-1}(0))P(A_n(1)) = (1-p)^{n-1}p.$$

I due esempi che seguono, liberamente tratti da casi giudiziari reali, mostrano quanto la non comprensione della nozione di condizionamento possa condurre a conclusioni errate.

Esempio 1.18. Le indagini relative ad un omicidio hanno condotto alle seguenti conclusioni:

- il colpevole possiede un determinato set di caratteristiche (individuate da un testimone, per es. capelli rossi, zoppicante, automobile verde, ecc.) che lo rendono piuttosto raro: si stima che una frazione $p \ll 1$ di popolazione possegga tali caratteristiche;
- il colpevole risiede in una città di n abitanti; inoltre $np \simeq 0.05$.

Una ricerca su un database di individui schedati ha identificato un (unico) individuo che possiede i due requisiti sopra citati. Per affermarne la colpevolezza, l'accusa argomenta come segue.

1. La probabilità che nella città vi siano almeno due individui con le caratteristiche individuate dal testimone è circa $(np)^2 = 0.0025$.
2. Pertanto con probabilità $1-(np)^2 = 0.9975$ quello trovato è l'unico individuo con tali caratteristiche: la sua colpevolezza è dunque accertata con probabilità 0.9975.

Questo argomento è sbagliato, in particolare la conclusione. Vediamo perché. Una ragionevole assunzione è che ogni individuo possegga la caratteristica individuata con probabilità p, indipendentemente dagli altri. Pertanto, per quanto visto nell'Esempio 1.17, se A_k denota l'evento "in città ci sono k individui con quella caratteristica, si ha

$$P(A_k) = \binom{n}{k} p^k(1-p)^{n-k}.$$

In particolare

$$P(A_0) = (1-p)^n \qquad\qquad P(A_1) = np(1-p)^{n-1}.$$

Quindi la probabilità che vi siano almeno due individui con i requisiti richiesti è

$$P\left[(A_0 \cup A_1)^c\right] = 1-(1-p)^n - np(1-p)^{n-1} \simeq \frac{n(n-1)}{2}p^2 \simeq \frac{1}{2}(np)^2,$$

dove la prima approssimazione si ottiene con uno sviluppo di Taylor al secondo ordine della funzione $f(p) = 1-(1-p)^n - np(1-p)^{n-1}$ attorno a $p = 0$. Il punto 1. dell'argomento dell'accusa va dunque corretto di un fattore $1/2$. Ma non è certo questo l'errore più rilevante, anche perché la correzione rende l'argomento ancor più stringente. L'accusa infatti non ha tenuto conto del fatto che un individuo con le caratteristiche individuate *è già stato trovato*. La quantità probante non ''e dunque la "probabilità che vi siano almeno due individui con i requisiti richiesti", bensì la "probabilità che ve ne siano almeno due *condizionata* alla conoscenza che la ricerca ha identificato un individuo che li possiede. Consideriamo quindi gli eventi $B =$ "vi sono almeno due individui con i requisiti richiesti" e $C =$ " la ricerca ha identificato esattamente un individuo che li possiede". Se k è il numero di residenti della città inseriti nel database, e assumiamo che la frequenza delle caratteristiche in questione tra gli schedati sia la stessa che nel resto della popolazione. Allora

$$P(C) = kp(1-p)^{k-1}.$$

Inoltre, $B \cap C = B' \cap C$ dove $B' =$ "tra gli individui *non* schedati almeno uno ha le caratteristiche richieste". Inoltre gli eventi B' e C, riferendosi a gruppi distinti di individui, sono indipendenti. Pertanto

$$P(B \cap C) = P(B' \cap C) = P(B')P(C) = \left[1-(1-p)^{n-k}\right] P(C),$$

da cui

$$P(B|C) = 1-(1-p)^{n-k} \simeq (n-k)p \simeq np = 0.05$$

almeno nel caso, verosimile, che $k \ll n$. Dunque, sulla base delle conoscenze acquisite, la probabilità che l'individuo trovato sia l'unico con le caratteristiche date, e quindi che sia colpevole, è

$$P(B^c|C) = 0.95,$$

probabilmente non sufficiente a fugare "ogni ragionevole dubbio".

Per apprezzare la sottigliezza della questione, consideriamo il seguente quesito. Immaginiamo di avere l'informazione del testimone, ma di non aver condotto alcuna ricerca su database. Qual è la probabilità che il colpevole sia l'unico individuo con le caratteristiche richieste? In questo caso, l'unica informazione disponibile è che "esiste almeno un individuo con le caratteristiche richieste", che corrisponde all'evento A_0^c. Osservando che $B = (A_0 \cup A_1)^c$,

$$P(C|A_0^c) = P\left[(A_0 \cup A_1)^c | A_0^c\right] = \frac{P\left[(A_0 \cup A_1)^c\right]}{P\left(A_0^c\right)}.$$

Usando la precedente stima per $P\left[(A_0 \cup A_1)^c\right]$ e l'analoga stima

$$P\left(A_0^c\right) = 1 - (1-p)^n \simeq np,$$

otteniamo

$$P\left[(A_0 \cup A_1)^c | A_0^c\right] \simeq \frac{1}{2} np$$

che differisce di un fattore $\frac{1}{2}$ dal risultato trovato prima!

Esempio 1.19. Una donna venne assassinata, il marito era il principale sospettato. Nel corso delle indagini si scoprì che il marito aveva più volte picchiato la moglie. L'accusa affermò che questo rappresentasse un importante indizio per la colpevolezza. La difesa ribatté che, secondo i dati forniti dalla Polizia di Stato, tra gli uomini che picchiano le loro mogli, solo 1 su 10000 finisce poi per assassinarla. Pertanto tale dato contribuisce solo in modo molto marginale alla tesi di colpevolezza. In primo grado il giudice accolse la tesi della difesa.

In secondo grado l'accusa (ci piace pensare con l'aiuto di un matematico) rilevò il seguente errore nell'argomento della difesa. La frazione $\frac{1}{10000}$ fornisce una stima della probabilità che una donna venga ammazzata dal marito *condizionata* al fatto che il marito la picchiasse. Ma noi non solo sappiamo che il marito la picchiava, ma anche che la donna è stata effettivamente assassinata. Quindi la probabilità che utilizza l'intera informazione disponibile è: probabilità che "una donna venga ammazzata dal marito" *condizionata* a "la donna veniva picchiata dal marito ed è stata ammazzata".

Per formalizzare il problema, consideriamo una popolazione numerosa e sufficientemente omogenea di donne sposate, e consideriamo i seguenti eventi, relativi ad un individuo casualmente scelto in questa popolazione:

$A =$ "la donna viene assassinata"
$B =$ "la donna è stata picchiata dal marito"
$C =$ "la donna viene assassinata dal marito".

Possiamo assumere che $P(C|B) = \frac{1}{10000}$. Inoltre i dati dell Polizia indicano che, nella totalità della popolazione, circa una donna su 100000 viene assassinata, cioè possiamo assumere $P(A) = \frac{1}{100000}$. La probabilità che desideriamo calcolare è $P(C|A \cap B)$. Usando il fatto che $C \subseteq A$ otteniamo

$$P(C|A \cap B) = \frac{P(A \cap B \cap C)}{P(A \cap B)} = \frac{P(B \cap C)}{P(A \cap B)} = \frac{P(C|B)}{P(A|B)}.$$

I dati a disposizione non ci consentono di calcolare $P(A|B)$. Si noti che la conoscenza che una donna viene picchiata dal marito, rende assai più probabile il suo assassinio da parte del marito. È quanto meno plausibile assumere che tutte le donne, vengano o meno picchiate dal marito, hanno grosso modo la stessa probabilità

di essere assassinate da *una persona diversa dal marito*, cioè

$$P(A \setminus C|B) = P(A \setminus C);$$

in altre parole, $A \setminus C$ e B sono indipendenti. Abbiamo allora

$$\begin{aligned} P(C|A \cap B) = \frac{P(C|B)}{P(A|B)} = \frac{P(C|B)}{P(C|B) + P(A \setminus C|B)} = \frac{P(C|B)}{P(C|B) + P(A \setminus C)} \\ \geq \frac{P(C|B)}{P(C|B) + P(A)} = \frac{\frac{1}{10000}}{\frac{1}{10000} + \frac{1}{100000}} = \frac{10}{11}. \end{aligned}$$

Quindi, la sola informazione che avesse picchiato la moglie, rende la probabilità che l'imputato sia colpevole almeno $\frac{10}{11}$! L'imputato è stato poi condannato...

Capitolo 2
Esempi rilevanti di probabilità discrete

Abstract In questo capitolo vedremo all'opera le nozioni di probabilità viste nel capitolo 1, applicate ad alcuni esempi rilevanti e non banali. Con l'eccezione di un'osservazione non essenziale nel sottoparagrafo 2.2.1, la nozione di indipendenza (paragrafo 1.5) non è necessaria per i paragrafi 2.2, 2.3 e 2.5, mentre viene usata nel paragrafo 2.6. Infine, il contenuto del paragrafo 2.7, incluso in questo capitolo per omogeneità di argomento, è di carattere più avanzato e richiede il concetto di variabile casuale, sviluppato nel capitolo 3.

2.1 Permutazioni aleatorie

In questo capitolo vedremo all'opera le nozioni di probabilità viste nel capitolo 1, applicate ad alcuni esempi rilevanti e non banali. Con l'eccezione di un'osservazione non essenziale nel sottoparagrafo 2.2.1, la nozione di indipendenza (paragrafo 1.5) non è necessaria per i paragrafi 2.2, 2.3 e 2.5, mentre viene usata nel paragrafo 2.6. Infine, il contenuto del paragrafo 2.7, incluso in questo capitolo per omogeneità di argomento, è di carattere più avanzato e richiede il concetto di variabile casuale, sviluppato nel capitolo 3.

2.2 Permutazioni aleatorie

Come già visto nel capitolo 1, denotiamo con S_n l'insieme delle funzioni biiettive dall'insieme $\{1,2,\ldots,n\}$ in sè. S_n è un gruppo non commutativo (per $n \geq 3$) se dotato dell'operazione di composizione. In tutti gli esempi che vedremo, assumeremo che la probabiltà P su S_n sia quella uniforme. Naturalmente anche P dipende da n, ma non c'è ragione di appesantire le notazioni chiamandola, ad esempio, P_n. Lo spazio di probabilità (S_n, P) è un buon modello per l'esperimento aleatorio che consiste nel *mescolare* accuratamente n oggetti, e quindi osservare l'ordinamento ottenuto.

In questo paragrafo esaminiamo alcune proprietà interessanti dello spazio (S_n, P), prendendo spunto da alcuni problemi.

2.2.1 Cicli

Problema 2.1. Un gruppo di n amici affitta una casa per una vacanza. Dopo alcuni giorni tutti convengono che sia il caso di fare delle pulizie, ma si stenta a trovare dei volontari. Laura, che è volonterosa e bizzarra, avanza la seguente proposta. Ognuno scrive il proprio nome su una carta. Quindi le n carte vengono accuratamente mescolate e distribuite. Laura allora leggerà ad alta voce il nome sulla sua carta. Quindi la persona il cui nome è stato letto leggerà a sua volta il nome sulla sua carta; si prosegue così finchè non viene letto il nome di Laura. A questo punto, le persone il cui nome è stato chiamato formeranno la squadra per le pulizie.

(i) Qual è la probabilità che Laura si trovi a dover fare le pulizie da sola?

(ii) Qual è la probabilità che tutti debbano fare le pulizie?

(iii) Più in generale, qual è la probabilità che la squadra delle pulizie sia composta da m persone?

Soluzione 2.1. È conveniente riformulare il problema con un linguaggio più formale. Etichettiamo gli n amici con i numeri $1,2,\ldots,n$, assumendo che il numero 1 corrisponda a Laura. L'esito del mescolamento delle n carte può essere descritto da un elemento $\sigma \in S_n$: la carta in mano alla persona i ha il nome della persona $\sigma(i)$.

La squadra per le pulizie si ottiene applicando *ripetutamente* σ a 1:

$$\sigma(1),\ \sigma\circ\sigma(1) =: \sigma^2(1),\ \ldots,\ \sigma^{k-1}(1),\ 1,$$

dove $k \geq 1$ è il più piccolo numero intero tale che $\sigma^k(1) = 1$. La sequenza

$$(1, \sigma(1), \sigma^2(1), \ldots, \sigma^{k-1}(1))$$

viene detta *ciclo* di *lunghezza k*. La costruzione fatta a partire dall'elemento 1 può essere ripetuta a partire da un elemento arbitrario. È chiaro che ogni elemento $i \in \{1,2,\ldots,n\}$ appartiene ad uno ed un solo ciclo: in altre parole una permutazione individua una partizione in cicli di $\{1,2,\ldots,n\}$. Il quesito (iii) del problema in esame, che contiene gli altri due come casi particolari, può essere pertanto riformulato come segue: qual è la probabilità che il ciclo contenente 1 abbia lunghezza m?

Definiamo

$$C_m := \{\sigma \in S_n :\ \text{il ciclo contenente 1 ha lunghezza } m\}.$$

Si noti che $C_1 = \{\sigma \in S_n : \sigma(1) = 1\}$, e che c'è una naturale corrispondenza biunivoca tra C_1 e l'insieme delle permutazioni di $\{2,3,\ldots,n\}$, da cui si deduce che

$$|C_1| = (n-1)! \;\Rightarrow\; P(C_1) = \frac{1}{n}.$$

In altre parole, la probabilità che Laura si trovi da sola a fare le pulizie è pari a $\frac{1}{n}$. Questo risponde alla domanda (i).

Consideriamo ora la domanda (ii), cioè calcoliamo $P(C_n)$. Per contare gli elementi di C_n osserviamo che se $\sigma \in C_n$ essa ha un unico ciclo, che si può rappresentare nella forma

$$(1, \sigma(1), \sigma^2(1), \ldots, \sigma^{n-1}(1)).$$

Ma la scrittura precedente si può *interpretare* come una permutazione di $\{1,2,\ldots,n\}$ con 1 al primo posto. Abbiamo appena notato che ci sono $(n-1)!$ tali permutazioni, per cui $|C_n| = (n-1)!$. Segue in particolare che $P(C_n) = \frac{1}{n}$, che risponde alla domanda (ii).

In modo più rigoroso, mostriamo che l'applicazione φ definita da

$$(\varphi(\sigma))(k) \;:=\; \sigma^{k-1}(1)\,,$$

dove si intende che σ^0 sia la funzione identica, è una biiezione tra C_n e C_1. È evidente che $\varphi(\sigma) \in C_1$ per ogni $\sigma \in C_n$ e che l'applicazione φ è iniettiva. Per mostrare che φ è suriettiva, e dunque biiettiva, basta mostrare che ammette inversa destra, cioè che esiste un'applicazione $\psi : C_1 \to C_n$ tale che $\varphi \circ \psi =$ identità su C_1. Mostriamo che tale ψ è data da

$$(\psi(\tau))(k) \;:=\; \tau(\tau^{-1}(k)+1)\,, \qquad \text{dove } n+1 := 1\,,$$

cioè si ha $\varphi(\psi(\tau)) = \tau$ per ogni $\tau \in C_1$. Per verificare quest'ultimo fatto, per definizione

$$\varphi(\psi(\tau))(1) \;=\; (\psi(\tau))^0(1) \;=\; 1 \;=\; \tau(1)\,,$$

dato che $\tau \in C_1$. Inoltre, assumendo che per $m \le k$ si abbia $\varphi(\psi(\tau))(m) = \tau(m)$, si ha

$$\begin{aligned}\varphi(\psi(\tau))(k+1) &= (\psi(\tau))^k(1) = (\psi(\tau))[(\psi(\tau))^{k-1}(1)] \\ &= (\psi(\tau))[\varphi(\psi(\tau))(k)] = (\psi(\tau))(\tau(k)) = \tau(\tau^{-1}(\tau(k))+1) = \tau(k+1).\end{aligned}$$

Abbiamo dunque mostrato per induzione che $\varphi(\psi(\tau))(k) = \tau(k)$ per ogni $k \in \{1,2,\ldots,n\}$.

A questo punto abbiamo gli strumenti per calcolare $|C_m|$ per ogni valore di m, cioè per rispondere alla domanda (iii). Infatti, gli elementi $\sigma \in C_m$ possono essere determinati dalle seguenti *tre scelte successive*:

- si scelgono gli m elementi del ciclo contenente 1, per cui ci sono $\binom{n-1}{m-1}$ esiti possibili (uno degli m elementi dev'essere 1);
- si sceglie uno dei cicli formati da questi m elementi: come abbiamo appena visto nella risposta alla domanda (ii), ci sono $(m-1)!$ tali cicli;
- si scelgono i valori di σ sui rimanenti $n-m$ elementi: dato che σ permuta in modo arbitrario tali elementi, per questa scelta ci sono $(n-m)!$ esiti possibili.

Per il principio fondamentale del calcolo combinatorio, si ottiene

$$|C_m| = \binom{n-1}{m-1}(m-1)!(n-m)! = (n-1)! \;\Rightarrow\; P(C_m) = \frac{1}{n}.$$

Concludendo, la probabilità che la squadra per le pulizie sia composta da m elementi è $\frac{1}{n}$, in particolare non dipende da m.

Avendo acquistato un po' di familiarità con le permutazioni, consideriamo il problema seguente.

Problema 2.2. Lo stesso gruppo di n amici decide di usare il metodo proposto da Laura per suddividersi in sottogruppi, corrispondenti alla partizione in cicli determinata della permutazione.

- Qual è la probabilità che si formi un sottogruppo, necessariamente unico, con più di $n/2$ persone?

Soluzione 2.2. Sia $m > \frac{n}{2}$, e sia

$$D_m := \{\sigma \in S_n : \sigma \text{ ha un ciclo di lunghezza } m\}.$$

In realtà D_m è ben definito anche per $m \leq n/2$. Tuttavia in questo caso un ciclo di lunghezza m non è necessariamente unico, e l'argomento che vedremo ora per determinare il numero dei suoi elementi non si può applicare. Assumiamo perciò $m > n/2$. Gli elementi di D_m possono essere determinati attraverso le seguenti scelte successive:

- si scelgono gli m elementi che compaiono nel ciclo "grande", per cui ci sono $\binom{n}{m}$ esiti possibili;
- si sceglie il ciclo sugli m elementi fissati: per questa scelta ci sono $(m-1)!$ esiti possibili;
- si permutano in modo arbitrario i rimanenti $n-m$ elementi, per cui ci sono $(n-m)!$ esiti possibili.

Pertanto

$$|D_m| = \binom{n}{m}(m-1)!(n-m)! = \frac{n!}{m} \;\Rightarrow\; P(D_m) = \frac{1}{m}.$$

Si osservi che se fosse $m \leq \frac{n}{2}$, la possibile non unicità dei cicli di lunghezza m conduce a "contare più di una volta" la stessa permutazione, e quindi il precedente conteggio risulta non corretto: questo si evince anche dal fatto che $\sum_{m=1}^{n} \frac{1}{m} > 1$ per ogni $n \in \mathbb{N}$.

Per rispondere al quesito del problema, dobbiamo calcolare

$$P\left(\bigcup_{\frac{n}{2}<m\leq n} D_m\right) = \sum_{\frac{n}{2}<m\leq n} P(D_m) = \sum_{\frac{n}{2}<m\leq n} \frac{1}{m}.$$

Denotiamo con p_n quest'ultima probabilità, e sia $\lfloor \frac{n}{2} \rfloor$ la parte intera di $\frac{n}{2}$. Intuitivamente, sostituendo la somma con un integrale si ottiene $p_n \approx \int_{\lfloor n/2 \rfloor}^{n} \frac{1}{x}\,dx \approx \log n - \log(n/2) = \log 2$. Precisiamo ora questa relazione in modo rigoroso. Usando le disuguaglianze (da verificare per esercizio!), valide per ogni $x > 0$

$$0 \le x - \log(1+x) \le x^2,$$

abbiamo che, se $m > \lfloor \frac{n}{2} \rfloor$

$$0 \le \frac{1}{m} - \log\left(\frac{m+1}{m}\right) \le \frac{1}{m^2} \le \frac{4}{n^2}.$$

Perciò

$$0 \le \sum_{\frac{n}{2} < m \le n} \frac{1}{m} - \sum_{\frac{n}{2} < m \le n} \log\left(\frac{m+1}{m}\right) \le \frac{2}{n},$$

da cui, essendo

$$\sum_{\frac{n}{2} < m \le n} \log\left(\frac{m+1}{m}\right) = \log\left(\frac{n+1}{\lfloor \frac{n}{2} \rfloor + 1}\right),$$

si deduce che

$$\log\left(\frac{n+1}{\lfloor \frac{n}{2} \rfloor + 1}\right) \le p_n \le \log\left(\frac{n+1}{\lfloor \frac{n}{2} \rfloor + 1}\right) + \frac{2}{n}.$$

In particolare

$$\lim_{n \to +\infty} p_n = \log 2.$$

In altre parole, per grandi valori di n, la probabilità che si formi un sottogruppo con più di $n/2$ persone è approssimativamente $\log 2 \simeq 0,693147181$ e dunque (approssimativamente) *non dipende da n*, un risultato non evidente a priori. Per $n > 50$, $|p_n - \log 2| \le 0.01$.

Il risultato appena ottenuto permette di trovare una soluzione al seguente difficile problema.

Problema 2.3. Il docente di un corso di Probabilità frequentato da 100 studenti propone ai suoi allievi quanto segue. Si preparano 100 buste, numerate da 1 a 100, e 100 carte, su ciascuna delle quali è scritto il nome di uno studente del corso (senza ripetizioni, si escludano omonimie). Quindi le carte vengono inserite, casualmente, una per ogni busta. Le buste, chiuse ma non sigillate, vengono quindi disposte sulla cattedra di un aula. Gli studenti entrano nell'aula uno per volta. Ogni studente apre a suo piacimento 50 buste, e comunica al docente se, tra le buste aperte, c'è quella con il proprio nome. Quindi le richiude ed esce dall'aula, e da quel momento non può più comunicare con i colleghi che ancora devono entrare in aula. Il docente alzerà il voto dell'esame di tre punti a tutti, solo nel caso in cui *ciascuno studente* trovi la busta contenente la carta con il proprio nome. Gli studenti non possono comunicare dopo l'inizio delle aperture, ma possono concordare una strategia *a priori*. Si determini una strategia che conduca al successo (cioè all'aumento di tre punti per tutti) con probabilità almeno 0.3.

Soluzione 2.3. È assolutamente non ovvio che questo problema abbia soluzione. Le strategie "banali" falliscono miseramente. Supponiamo, ad esempio, che gli studenti

non si accordino per nulla, ad esempio che ognuno di essi scelga a caso, indipendentemente dagli altri, le 50 buste da aprire. In questo caso è facile mostrare che ognuno avrebbe probabilità $\frac{1}{2}$ di trovare il proprio nome e, vista l'indipendenza delle scelte, la probabilità che *tutti* trovino il proprio nome sarebbe $\frac{1}{2^{100}}$: irrisoria! Si può fare naturalmente di peggio: se fossero così sciocchi da accordarsi di aprire tutti le *stesse* 50 buste, la probabilità di successo sarebbe nulla. Quello che non è ovvio è se sia possibile fare meglio.

Per semplificare le notazioni, poniamo $n := 100$ ed etichettiamo i cento nomi con i numeri $1, 2, \ldots, n$. Denotiamo inoltre con $\sigma(k)$ il numero (nome) all'interno della busta numero k. Tale σ è evidentemente un elemento di S_n, e la probabilità uniforme su S_n corrisponde al fatto che i nomi nelle buste vengono inseriti a caso. Lo scopo di ogni studente k è di aprire la busta numero j con $\sigma(j) = k$. Supponiamo che gli studenti si accordino per seguire la seguente strategia. Ogni studente k apre per prima la busta k, e ne legge il contenuto $\sigma(k)$. Quindi apre la busta $\sigma(k)$ leggendone il contenuto $\sigma^2(k)$, e così via. Se, nella permutazione σ, l'elemento k appartiene ad un ciclo di lunghezza $m \leq \frac{n}{2}$, la m-sima busta aperta è la busta $\sigma^{m-1}(k)$, il cui contenuto è $\sigma^m(k) = k$: questo significa che lo studente trova la carta col proprio nome! Segue pertanto che se non ci sono in σ cicli di lunghezza maggiore di $n/2$, *ogni studente troverà sicuramente la busta contenente il proprio nome*. Perciò

$$\text{probabilità di successo della strategia } \geq 1 - p_n,$$

dove p_n è la probabilità calcolata nel problema precedente. Avendo visto che $p_n \simeq \log 2 \simeq 0.69$, abbiamo ottenuto quanto richiesto (per scrupolo, per $n = 100$, si calcola $p_n \simeq 0.688$). Sottolineiamo che il limite inferiore ottenuto alla probabilità di successo è approssimativamente indipendente da n, se n è abbastanza grande.

Per capire meglio la strategia, definiamo per $m = 1, \ldots, n$ l'evento $B_m := \{$lo studente numero m trova la carta col proprio nome$\}$, e poniamo $B := \{$tutti gli studenti trovano la carta col proprio nome$\} = \bigcap_{m=1}^{n} B_m$. Non è difficile convincersi del fatto che $P(B_m) = 0.5$ per ogni $m = 1, \ldots, n$, *qualunque sia la strategia seguita!* Di conseguenza $P(B) \leq 0.5$. Con la strategia proposta abbiamo mostrato che $P(B) = P(\bigcap_{m=1}^{n} B_m) \geq 0.3$. In particolare, gli eventi $\{B_m\}_{1 \leq m \leq n}$ sono "molto sovrapposti" (quindi tutt'altro che indipendenti). Il cuore della soluzione consiste proprio nel determinare una strategia tale che, se si verifica il primo evento A_1, con grande probabilità si verificano tutti gli altri eventi A_m con $m \geq 1$.

2.2.2 Punti fissi

Problema 2.4. Una comitiva di n turisti si sta imbarcando per un viaggio aereo. La loro guida ha tutte le carte d'imbarco (nominative), che deve distribuire ai turisti prima dell'imbarco. Per la fretta e la confusione le distribuisce a caso. Qual è la probabilità che qualcuno dei turisti riceva effettivamente la propria carta d'imbarco? Qual è la probabiltà che esattamente m turisti ricevano la propria carta d'imbarco?

Soluzione 2.4. Al solito, etichettiamo con $\{1,2,\dots,n\}$ gli n turisti e sia $\sigma(i)$ il numero (nome) sulla carta d'imbarco ricevuta dal turista i. Chiaramente $\sigma \in S_n$. L'i-esimo turista riceve la propria carta d'imbarco se $\sigma(i) = i$, cioè se i è un *punto fisso* della permutazione σ. Dunque, i quesiti del problema si possono riformulare come segue: qual è la probabilità che una permutazione abbia almeno un punto fisso? E qual è la probabilità che abbia esattamente m punti fissi?

Per $m = 0,1,\dots,n$ e $i = 1,2,\dots,n$, introduciamo gli eventi

$$A_m := \{\sigma \in S_n : \ \sigma \text{ ha esattamente } m \text{ punti fissi}\}, \qquad C_i := \{\sigma \in S_n : \sigma(i) = i\}.$$

Chiaramente

$$A_0^c = C_1 \cup C_2 \cup \cdots \cup C_n.$$

Per la formula di inclusione-esclusione (Proposizione 1.2)

$$P(A_0^c) = \sum_{k=1}^{n} \sum_{\substack{J \subseteq \{1,2,\dots,n\} \\ \text{tale che } |J|=k}} (-1)^{k+1} P\left(\bigcap_{i \in J} C_i\right). \tag{2.1}$$

Fissiamo dunque $k \in \{1,2,\dots,n\}$, e sia $J \subseteq \{1,2,\dots,n\}$ tale che $|J| = k$. Si ha

$$\bigcap_{i \in J} C_i = \{\sigma \in S_n : \sigma(i) = i \text{ per ogni } i \in J\}.$$

Le permutazioni che lasciano fissi gli elementi di J sono in naturale corrispondenza biunivoca con le permutazioni di $\{1,2,\dots,n\} \setminus J$, e quindi

$$\left|\bigcap_{i \in J} C_i\right| = (n-k)! \ \Rightarrow \ P\left(\bigcap_{i \in J} C_i\right) = \frac{(n-k)!}{n!}.$$

Poichè i sottoinsiemi J di $\{1,2,\dots,n\}$ con k elementi sono $\binom{n}{k}$, si ha

$$\sum_{\substack{J \subseteq \{1,2,\dots,n\} \\ \text{tale che } |J|=k}} P\left(\bigcap_{i \in J} C_i\right) = \binom{n}{k} \frac{(n-k)!}{n!} = \frac{1}{k!}.$$

Inserendo quest'ultima uguaglianza in (2.1) otteniamo

$$P(A_0^c) = \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k!} = 1 - \sum_{k=0}^{n} \frac{(-1)^k}{k!}.$$

È inoltre ben noto che

$$\left| \sum_{k=0}^{n} \frac{(-1)^k}{k!} - e^{-1} \right| \leq \frac{1}{(n+1)!}.$$

Quindi l'approssimazione

$$P(A_0^c) \simeq 1 - e^{-1} \simeq 0.632$$

è eccellente per valori non troppo piccoli di n (già per $n = 6$ i due numeri hanno le prime tre cifre decimali uguali). Dunque, la probabilità che almeno un passeggero riceva la sua carta di imbarco è "quasi" indipendente dal numero di passeggeri!

Resta da determinare $P(A_m)$ per $m \geq 1$. Nel seguito usiamo la notazione

$$q_n := \sum_{k=0}^{n} \frac{(-1)^k}{k!},$$

che, come appena visto, è la probabilità che dell'insieme delle permutazioni di un insieme di n elementi che non hanno alcun punto fisso, cioè $P(A_0) = q_n$.
Per $J \subseteq \{1,2,\ldots,n\}$, $|J| = m$, sia

$$B_J := \{\sigma \in S_n : \ \sigma(j) = j \text{ per ogni } j \in J, \ \sigma(i) \neq i \text{ per ogni } i \notin J\}.$$

Ogni elemento di B_J può essere identificato con una permutazione degli elementi di J^c che non ha alcun punto fisso. Per quanto appena visto, ci sono $q_{n-m}(n-m)!$ tali permutazioni. Pertanto

$$|B_J| = q_{n-m}(n-m)! \ \Rightarrow \ P(B_J) = q_{n-m}\frac{(n-m)!}{n!}.$$

Infine essendo

$$A_m = \bigcup_{J:|J|=m} B_J,$$

ed essendo la precedente l'unione di insiemi disgiunti,

$$P(A_m) = \binom{n}{m} q_{n-m}\frac{(n-m)!}{n!} = \frac{q_{n-m}}{m!}.$$

Se m non è troppo vicino a n

$$P(A_m) \simeq \frac{e^{-1}}{m!}.$$

2.3 La passeggiata aleatoria semplice

Consideriamo un moto *discreto* sull'insieme dei numeri interi $\mathbb{Z}$ che avvenga con le seguenti regole:

- la posizione iniziale, cioè all'istante $k = 0$, è 0;
- se $x \in \mathbb{Z}$ è la posizione all'istante k, allora le posizioni possibili all'istante $k+1$ sono $x+1$ e $x-1$;

- fissato l'istante finale n, tutti i cammini *possibili* che terminano all'istante n sono equiprobabili.

Un cammino possibile è dunque identificabile con un vettore $(s_0, s_1, \ldots, s_n)$, dove $s_0 = 0$ e, per $k = 0, 1, \ldots, n-1$, $|s_{k+1} - s_k| = 1$. È facile vedere che l'applicazione $(s_0, s_1, \ldots, s_n) \mapsto (x_1, x_2, \ldots, x_n)$ data da $x_k := s_k - s_{k-1}$ fornisce un biiezione tra l'insieme dei cammini possibili e l'insieme $\{-1, 1\}^n$, la cui inversa è data, per $k \geq 1$, da $s_k = \sum_{j=1}^{k} x_k$. In particolare, ci sono 2^n cammini possibili che terminano all'istante n. Le variabili x_k rappresentano gli *incrementi* del cammino, mentre le variabili s_k rappresentano le *posizioni* del cammino nei diversi istanti.

Data la biiezione appena citata, è indifferente lavorare con le posizioni o con gli incrementi: per semplicità, scegliamo come spazio campionario Ω_n del nostro modello lo spazio degli incrementi, cioè $\Omega_n := \{-1, 1\}^n$. Come suggerito dalle regole citate sopra, muniamo Ω_n della probabilità uniforme, che indichiamo con P. Lo spazio di probabilità (Ω_n, P) viene chiamato *passeggiata aleatoria semplice e simmetrica* (di n passi). L'aggettivo *semplice* sta ad indicare che gli incrementi possono assumere solo i valori ± 1, mentre *simmetrica* indica che le sequenze di incrementi sono tutte equiprobabili. Si tratta del più semplice modello per un moto aleatorio, tuttavia di rilevanza teorica e applicativa fondamentale. La seguente è una lista di domande "classiche" relative al nostro modello.

- Con quale probabilità e con quale frequenza la passeggiata raggiunge un determinato valore?
- Quanto tempo trascorre la passeggiata aleatoria al di sopra di un determinato valore?
- Quali sono i valori "tipici" di s_n, per n grande? (Il senso concreto dell'aggettivo "tipici" sarà chiarito in seguito.)

Prima di analizzare in dettaglio alcune di queste domande, c'è una questione che merita di essere approfondita. Consideriamo un evento che dipende solo dai primi n incrementi della passeggiata aleatoria, come ad esempio "la posizione s_n al tempo n è uguale a 0". Per descrivere questo evento, è naturale considerare il sottoinsieme di Ω_n dato da

$$A_n := \{(x_1, \ldots, x_n) \in \Omega_n : x_1 + x_2 + \cdots + x_n = 0\}. \tag{2.2}$$

La scelta di Ω_n non è tuttavia obbligata: è altrettanto legittimo adottare come spazio campionario Ω_N, per un qualunque valore di $N \geq n$, e definire l'analogo sottoinsieme di Ω_N in termini delle prime n variabili $x_1, \ldots, x_n$:

$$A_N := \{(x_1, \ldots, x_N) \in \Omega_N : x_1 + x_2 + \cdots + x_n = 0\}.$$

Questa "ambiguità" non crea problemi, perché le probabilità degli eventi A_n e A_N è la stessa. Più precisamente, indicando per chiarezza con P_n e P_N le probabilità su Ω_n e Ω_N rispettivamente, si ha $P_n(A_n) = P_N(A_N)$ per ogni $n \geq N$. La dimostrazione è semplice: dato che $A_N = A_n \times \{-1, 1\}^{N-n}$, si può scrivere

$$P_N(A_N) \;=\; \frac{|A_N|}{|\Omega_N|} \;=\; \frac{|A_n|\,2^{N-n}}{2^N} \;=\; \frac{|A_n|}{2^n} \;=\; P_n(A_n)\,.$$

Si noti che non si è usata in alcun modo la forma esplicita dell'evento A_n, data dall'equazione (2.2), ma solo il fatto che $A_n \subseteq \Omega_n$ e che $A_N = A_n \times \{-1,1\}^{N-n}$. Abbiamo ottenuto un'importante conclusione: per calcolare la probabilità di un evento che dipende solo dai primi n incrementi (equivalentemente, dalle prime n posizioni) della passeggiata aleatoria, si può scegliere come spazio campionario Ω_N, per un qualunque valore di $N \geq n$.

Data questa arbitrarietà nella scelta dello spazio Ω_N, risulta *naturale* (almeno per un matematico...) considerare lo spazio campionario dato dai cammini di lunghezza infinita $\Omega_\infty := \{-1,1\}^{\mathbb{N}}$, che contiene in modo canonico Ω_N per ogni $N \in \mathbb{N}$. Il problema è di definire la "giusta" probabilità su Ω_∞, che estenda (in un senso da precisare) la probabilità uniforme su Ω_n. Infatti lo spazio Ω_∞ è infinito (dunque la probabilità uniforme non ha senso) e non è neppure numerabile, dunque la nozione stessa di probabilità che abbiamo introdotto nel capitolo 1 non si può applicare. Questo problema ammette una soluzione positiva, che però richiede una nozione più generale di spazio di probabilità, e sarà affrontato nel capitolo 4.

2.3.1 Considerazioni preliminari

Indichiamo con $\mathbf{x} := (x_1, x_2, \ldots, x_n)$ l'elemento generico di Ω_n. Per $n \in \mathbb{N}$ e $m \in \mathbb{Z}$, introduciamo l'evento

$$\{s_n = m\} \;:=\; \{\mathbf{x} \in \Omega_n \,:\, x_1 + x_2 + \cdots + x_n = m\}\,. \tag{2.3}$$

Sottolineiamo che "$\{s_n = m\}$" è semplicemente una *notazione* (che riprenderemo e generalizzeremo nel capitolo 3) per indicare l'evento definito dal membro destro in (2.3). Per la probabilità dell'evento $\{s_n = m\}$ scriveremo semplicemente $P(s_n = m)$. È facile vedere che se n è pari, necessariamente s_n è pari e, analogamente, se n è dispari, necessariamente s_n è dispari, cioè $P(s_n = m) = 0$ se n e m non hanno la stessa parità. Pertanto è sufficiente considerare probabilità del tipo

$$P(s_{2n} = 2m)\,, \qquad\qquad P(s_{2n+1} = 2m+1),$$

dove $n \in \mathbb{N}$ e $m \in \mathbb{Z}$. Inoltre, essendo gli incrementi di modulo 1, $P(s_{2n} = 2m) > 0$ se e solo se $|m| \leq n$, mentre $P(s_{2n+1} = 2m+1) > 0$ se e solo se $-n-1 \leq m \leq n$.

Consideriamo ora un elemento $\mathbf{x} = (x_1, x_2, \ldots, x_{2n}) \in \Omega_{2n}$. Sia k il numero di incrementi positivi di $\mathbf{x}$, cioè $k := |\{i : x_i = 1\}|$. Essendo $2n-k$ gli incrementi negativi, segue che la posizione finale del cammino corrispondente è $s_{2n} = x_1 + \cdots + x_{2n} = k - (2n-k) = 2(k-n)$. Pertanto $\mathbf{x} \in \Omega_{2n}$ è un elemento dell'evento $\{s_{2n} = 2m\}$ se e solo se il numero di incrementi positivi di $\mathbf{x}$ è $k = n+m$. Da questo segue facilmente che

$$|\{s_{2n} = 2m\}| \;=\; \binom{2n}{n+m},$$

e quindi

$$P(s_{2n} = 2m) \;=\; \frac{|\{s_{2n} = 2m\}|}{|\Omega_{2n}|} \;=\; \frac{1}{2^{2n}}\binom{2n}{n+m}. \tag{2.4}$$

Lasciamo al lettore verificare, in modo del tutto analogo, che se si considera lo spazio Ω_{2n+1} dei cammini di lunghezza dispari $2n+1$, si ha

$$P(s_{2n+1} = 2m+1) \;=\; \frac{1}{2^{2n+1}}\binom{2n+1}{n+m+1}. \tag{2.5}$$

2.3.2 Il problema della ricorrenza

Ci interessiamo ora alla probabilità che la passeggiata aleatoria ritorni al punto di partenza: per $n \in \mathbb{N}$ poniamo

$$R_n \;:=\; P(s_k = 0 \;\text{ per qualche } k = 1,2,\ldots,n). \tag{2.6}$$

L'obiettivo è di capire il comportamento di R_n per $n \to \infty$.

Ricordando che $s_k = 0$ è possibile solo per k pari, definiamo

$$\begin{aligned} u_{2n} &:= P(s_{2n} = 0), \\ f_{2n} &:= P(s_2 \neq 0, \ldots, s_{2(n-1)} \neq 0, s_{2n} = 0). \end{aligned}$$

Si noti che u_{2n} è la probabilità che la passeggiata aleatoria valga 0 al passo $2n$, mentre f_{2n} è la probabilità che la passeggiata aleatoria ritorni a 0 *per la prima volta* al passo $2n$. Dire che la passeggiata aleatoria visita zero in un qualche passo $k \leq 2n$ è equivalente a dire che il primo ritorno a zero avviene prima di $2n$ passi: vale cioè l'uguaglianza di eventi

$$\{s_k = 0 \text{ per qualche } k = 1,\ldots,2n\} \;=\; \bigcup_{m=1}^{n} \{s_2 \neq 0, \ldots, s_{2(m-1)} \neq 0, s_{2m} = 0\},$$

e inoltre gli eventi che appaiono nell'unione sono a due a due disgiunti. Si ha pertanto l'uguaglianza

$$R_{2n} \;=\; \sum_{k=1}^{n} f_{2k}. \tag{2.7}$$

Si noti che, grazie a (2.4), si ha

$$u_{2n} \;=\; \frac{1}{2^{2n}}\binom{2n}{n}.$$

Nel prossimo paragrafo determineremo una formula esplicita per f_{2n}:

$$f_{2n} \;=\; \frac{1}{2n}u_{2n-2} \;=\; \frac{1}{2n-1}u_{2n}\,,$$

che tuttavia non è particolarmente utile per determinare il comportamento asintotico, per n grande, delle somme parziali $\sum_{k=1}^{n} f_{2k}$ (e dunque di R_{2n}, grazie a (2.7). Sarà invece fondamentale il seguente Lemma.

Lemma 2.1. *Per ogni $n > 0$*

$$u_{2n} \;=\; \sum_{k=1}^{n} f_{2k}u_{2(n-k)}\,.$$

Dimostrazione. Sia $A := \{s_{2n} = 0\}$, per cui $u_{2n} = P(A)$. L'evento A si può scrivere come unione dei seguenti n eventi disgiunti:

$$A_k \;:=\; \{s_2 \neq 0, \ldots, s_{2k-2} \neq 0, s_{2k} = 0, s_{2n} = 0\}\,,$$
$$A \;=\; \bigcup_{k=1}^{n} A_k\,.$$

Contiamo i cammini in A_k. La cardinalità di A_k è uguale al numero di cammini di lunghezza $2k$ che ritornano a 0 la prima volta dopo $2k$ passi ($2^{2k} f_{2k}$) moltiplicato il numero di cammini di lunghezza $2n-2k$ che terminano in 0 ($2^{2(n-k)} u_{2(n-k)}$). Pertanto

$$P(A_k) \;=\; \frac{1}{2^{2n}}|A_k| \;=\; \frac{1}{2^{2n}}\,2^{2k} f_{2k}\,2^{2(n-k)}\,u_{2(n-k)} \;=\; f_{2k}\,u_{2(n-k)}\,.$$

Essendo $P(A) = \sum_{k=1}^{n} P(A_k)$, la conclusione segue facilmente.

Lemma 2.2. *Siano $(a_n)_{n\geq 0}$ e $(b_n)_{n\geq 1}$ due successioni di numeri reali positivi tali che $a_0 > 0$ e, per ogni $n \geq 1$,*

$$a_n \;=\; \sum_{k=1}^{n} b_k a_{n-k}\,.$$

Supponiamo che $a_n \leq 1$ per ogni $n \geq 1$. Allora $\sum_{n=1}^{+\infty} b_n \leq 1$ e

$$\sum_{n=1}^{+\infty} b_n = 1 \quad\Longleftrightarrow\quad \sum_{n=1}^{+\infty} a_n = +\infty.$$

Dimostrazione. Usando il fatto che, per somme infinite ad addendi positivi è lecito permutare l'ordine degli addendi, abbiamo, posto $s := \sum_{n=1}^{+\infty} a_n$,

$$s \;=\; \sum_{n=1}^{+\infty} a_n \;=\; \sum_{n=1}^{+\infty}\sum_{k=1}^{n} b_k a_{n-k} \;=\; \sum_{k=1}^{\infty} b_k \sum_{n=k}^{+\infty} a_{n-k} \;=\; \sum_{k=1}^{+\infty} b_k \sum_{m=0}^{+\infty} a_m \;=\; (a_0 + s)\sum_{k=1}^{+\infty} b_k\,.$$

Ciò implica immediatamente che, se $s < +\infty$, allora

$$\sum_{k=1}^{+\infty} b_k = \frac{s}{a_0+s} < 1.$$

Resta da dimostrare che, se $s = +\infty$, si ha $\sum_{k=1}^{+\infty} b_k = 1$. Possiamo scrivere

$$\sum_{n=1}^{N} a_n \;=\; \sum_{n=1}^{N}\sum_{k=1}^{n} b_k a_{n-k} \;=\; \sum_{k=1}^{N} b_k \sum_{n=k}^{N} a_{n-k} \;=\; \sum_{k=1}^{N} b_k \sum_{n=0}^{N-k} a_n\,. \tag{2.8}$$

Trattandosi di somme a termini positivi, valgono le seguenti disuguaglianze:

$$\sum_{k=1}^{N} b_k \sum_{n=0}^{N-k} a_n \;\le\; \sum_{k=1}^{N} b_k \sum_{n=0}^{N} a_n = \sum_{k=1}^{N} b_k \left(a_0 + \sum_{n=1}^{N} a_n\right) \tag{2.9}$$

e, per $m \le N$,

$$\sum_{k=1}^{N} b_k \sum_{n=0}^{N-k} a_n \ge \sum_{k=1}^{m} b_k \sum_{n=0}^{N-k} a_n \ge \sum_{k=1}^{m} b_k \sum_{n=0}^{N-m} a_n \ge \sum_{k=1}^{m} b_k \left(a_0 + \sum_{n=1}^{N} a_n - m\right), \tag{2.10}$$

dove, nell'ultimo passaggio, abbiamo usato il fatto che $a_n \le 1$ per ogni $n \ge 1$. Supponiamo ora $\sum_{n=0}^{+\infty} a_n = +\infty$. Da (2.8) e (2.9) segue che

$$\sum_{k=1}^{N} b_k \;\ge\; \frac{\sum_{n=1}^{N} a_n}{a_0 + \sum_{n=1}^{N} a_n}\,,$$

da cui, passando al limite per $N \to +\infty$, si ottiene

$$\sum_{k=1}^{+\infty} b_k \;\ge\; 1\,. \tag{2.11}$$

Usando (2.8) e (2.10) otteniamo

$$\sum_{k=1}^{m} b_k \;\le\; \frac{\sum_{n=1}^{N} a_n}{a_0 + \sum_{n=1}^{N} a_n - m}\,,$$

da cui, passando prima al limite per $N \to +\infty$ e poi per $m \to +\infty$, otteniamo

$$\sum_{k=1}^{+\infty} b_k \;\le\; 1\,, \tag{2.12}$$

che, assieme a (2.11) conclude la dimostrazione.

Usando il Lemma 2.2 con $a_n = u_{2n}$ e $b_n = f_{2n}$, segue che

$$\lim_{n\to+\infty} R_{2n} = \sum_{n=1}^{+\infty} f_{2n} = 1 \iff \sum_{n} u_{2n} = +\infty.$$

L'affermazione $\lim_{n\to+\infty} R_{2n} = 1$ si può esprimere dicendo che la probabilità che la passeggiata aleatoria torni al punto di partenza entro i primi n passi tende a 1 per $n \to +\infty$, e viene chiamata *proprietà di ricorrenza*. Per stabilire se tale affermazione sia vera o falsa, dobbiamo dunque studiare la convergenza della serie $\sum_n u_{2n}$. A tale scopo usiamo la seguente celebre formula di approssimazione.

Proposizione 2.1. *(Formula di Stirling) Per ogni* $n \geq 1$

$$n! \;=\; \left(\frac{n}{e}\right)^n \sqrt{2\pi n}\, e^{\frac{\theta(n)}{12n}},$$

dove $0 \leq \theta(n) \leq 1$.

Dimostrazione. Diamo ora una dimostrazione, piuttosto elementare, dell'esistenza di una costante $C > 0$ tale che

$$n! \;=\; C\left(\frac{n}{e}\right)^n \sqrt{n}\, e^{\frac{\theta(n)}{12n}}. \tag{2.13}$$

Questo sarà sufficiente per i risultati di questo paragrafo. Nel prossimo paragrafo dimostreremo che in realtà $C = \sqrt{2\pi}$.

Notiamo anzitutto che (2.13) è equivalente a dimostrare che esiste una costante c $(= \log C)$ tale che per ogni $n \geq 1$

$$0 \;\leq\; \log n! - \left(n+\frac{1}{2}\right)\log n + n - c \;\leq\; \frac{1}{12n}. \tag{2.14}$$

Poniamo $d_n := \log n! - \left(n+\frac{1}{2}\right)\log n + n$. È semplice (esercizio!) mostrare che (2.14) segue dai seguenti due fatti, con $c = \lim_{n\to\infty} d_n$:

(i) La successione d_n è decrescente.

(ii) La successione $d_n - \frac{1}{12n}$ 'e crescente.

Quindi, non ci rimane che dimostrare le affermazioni (i) e (ii).

Con semplici calcoli si ottiene:

$$d_n - d_{n+1} \;=\; \left(n+\frac{1}{2}\right)\log\frac{n+1}{n} - 1 \;=\; \frac{2n+1}{2}\log\frac{1+\frac{1}{2n+1}}{1-\frac{1}{2n+1}} - 1.$$

A questo punto si usa la serie di Taylor $\log(1+t) \;=\; \sum_{k=1}^{\infty}(-1)^{k+1}\frac{t^k}{k}$, convergente per $|t| < 1$, per ottenere

$$\log\frac{1+t}{1-t} \;=\; \log(1+t) - \log(1-t) \;=\; \sum_{k=1}^{\infty}(-1)^{k+1}\frac{t^k}{k} + \sum_{k=1}^{\infty}\frac{t^k}{k} \;=\; 2\sum_{k=0}^{+\infty}\frac{t^{2k+1}}{2k+1},$$

che converge anch'essa per $|t| < 1$. Usando tale serie per $t = \frac{1}{2n+1}$ si trova

$$d_n - d_{n+1} \;=\; (2n+1)\sum_{k=0}^{+\infty} \frac{1}{2k+1}\frac{1}{(2n+1)^{2k+1}} - 1 \;=\; \sum_{k=1}^{+\infty} \frac{1}{2k+1}\frac{1}{(2n+1)^{2k}} \;\geq\; 0\,, \tag{2.15}$$

essendo quest'ultima una serie a termini positivi. Ciò dimostra (i).

Usando di nuovo (2.15) e il fatto che $2k+1 \geq 3$ per $k \geq 1$, si ottiene

$$\begin{aligned} d_n - d_{n+1} &= \sum_{k=1}^{+\infty} \frac{1}{2k+1}\frac{1}{(2n+1)^{2k}} \;\leq\; \frac{1}{3}\sum_{k=1}^{+\infty} \frac{1}{(2n+1)^{2k}} \;=\; \frac{1}{3}\frac{(2n+1)^{-2}}{1-(2n+1)^{-2}} \\ &= \frac{1}{3[(2n+1)^2-1]} \;=\; \frac{1}{12n(n+1)} \;=\; \frac{1}{12n} - \frac{1}{12(n+1)}\,, \end{aligned}$$

dove abbiamo usato la somma della serie geometrica

$$\sum_{n=1}^{+\infty} x^n \;=\; \frac{1}{1-x} \qquad \text{per } |x| < 1\,.$$

In altri termini, abbiamo dimostrato che

$$d_n - \frac{1}{12n} \leq d_{n+1} - \frac{1}{12(n+1)},$$

cioè la relazione (ii).

Dimostriamo finalmente la ricorrenza della passeggiata aleatoria semplice e simmetrica su $\mathbb{Z}$.

Teorema 2.1.

$$\lim_{n\to+\infty} R_{2n} \;=\; 1\,.$$

Dimostrazione. Come osservato prima, la tesi è equivalente a

$$\sum_n u_{2n} \;=\; +\infty\,. \tag{2.16}$$

Usando la formula di Stirling, abbiamo

$$u_{2n} = \frac{1}{2^{2n}}\binom{2n}{n} = \frac{1}{2^{2n}}\frac{(2n)!}{(n!)^2} = \frac{1}{2^{2n}}\frac{C\left(\frac{2n}{e}\right)^{2n}\sqrt{2n}e^{\frac{\theta(2n)}{24n}}}{C^2\left(\frac{n}{e}\right)^{2n} n e^{\frac{\theta(n)}{6n}}} = \frac{\sqrt{2}}{C\sqrt{n}}\exp\left[\frac{\theta(2n)}{24n} - \frac{\theta(n)}{6n}\right],$$

da cui segue

$$\lim_{n\to+\infty} \frac{u_{2n}}{1/\sqrt{n}} \;=\; \frac{\sqrt{2}}{C}\,.$$

Da ciò, per il criterio del confronto asintotico tra serie, si ricava (2.16), e questo conclude la dimostrazione.

Concludiamo questo paragrafo con la nozione di passeggiata aleatoria semplice e simmetrica multi-dimensionale. Per $d \geq 1$, consideriamo i cammini di lunghezza

n "generati" dallo spazio di incrementi

$$\Omega_n^d = \{\mathbf{x} = (x_1, x_2, \ldots, x_n) : x_i \in \{-1,1\}^d \text{ per } i = 1,2,\ldots,n\}.$$

Ciò significa che consideriamo cammini uscenti dall'origine di $\mathbb{Z}^d$ e la cui posizione al tempo k è $s_k := x_1 + x_2 + \cdots + x_k \in \mathbb{Z}^d$, dove $x_i \in \{-1,1\}^d$. Tutti questi cammini si assumono equiprobabili, cioè la probabilità P su Ω_n^d è quella uniforme. Si noti che se $A \subseteq \Omega_n = \{-1,1\}^n$ allora $A^d = A \times A \times \cdots \times A \subseteq \Omega_n^d$ e vale la formula

$$\left|A^d\right| \;=\; |A|^d\,.$$

Consideriamo allora l'evento $S_{2n}^{(d)} := \{s_{2n} = 0\}$ (in Ω_N^d, con $N \geq 2n$), dove 0 denota l'origine di $\mathbb{Z}^d$, e denotiamo con $u_{2n}^{(d)}$ la sua probabilità. In particolare, $u_{2n}^{(1)} = u_{2n}$. Poichè $s_{2n} = 0$ se e solo se tutte le d componenti sono uguali a zero, si ha

$$S_{2n}^{(d)} \;=\; \left(S_{2n}^{(1)}\right)^d.$$

Pertanto

$$u_{2n}^{(d)} \;=\; P\big(S_{2n}^{(d)}\big) \;=\; \frac{\left|S_{2n}^{(d)}\right|}{\left|\Omega_N^d\right|} \;=\; \frac{\left|S_{2n}^{(1)}\right|^d}{|\Omega_N|^d} \;=\; (u_{2n})^d\,. \tag{2.17}$$

In modo analogo al caso unidimensionale, possiamo definire

$$f_{2k}^{(d)} \;:=\; P(s_2 \neq 0, \ldots, s_{2(k-1)} \neq 0, s_{2k} = 0)\,.$$

Allora la quantità $R_{2n}^{(d)}$, definita da

$$R_{2n}^{(d)} \;:=\; \sum_{k=1}^{n} f_{2k}^{(d)}\,,$$

è la probabilità che la passeggiata aleatoria torni all'origine entro $2n$ passi. La relazione

$$u_{2n}^{(d)} \;=\; \sum_{k=1}^{n} f_{2k}^{(d)} u_{2(n-k)}^{(d)}$$

si dimostra esattamente come nel caso $d = 1$. Dunque, applicando il Lemma 2.2, abbiamo che *la passeggiata aleatoria è ricorrente*, cioè $\lim_{n\to\infty} R_{2n}^{(d)} = 1$, *se e solo se* $\sum_n u_{2n}^{(d)} = +\infty$. D'altra parte, essendo $u_{2n}^{(d)} = (u_{2n})^d$ e avendo visto che u_{2n} è asintoticamente equivalente a $n^{-1/2}$, deduciamo che $u_{2n}^{(d)}$ è asintoticamente equivalente a $n^{-d/2}$. Poichè la serie di termine generale $n^{-\alpha}$ converge se e solo se $\alpha > 1$, possiamo concludere quanto segue.

Teorema 2.2. *La passeggiata aleatoria semplice e simmetrica in dimensione d è ricorrente per $d = 1,2$ e non è ricorrente per $d \geq 3$.*

Un modo suggestivo per esprimere la non ricorrenza della passeggiata aleatoria in dimensione ≥ 3, consiste nell'affermare che "la passeggiata aleatoria in dimen-

sione ≥ 3 ha una probabilità strettamente positiva di non ritornare mai all'origine". Per dare un significato a questa affermazione sarebbe però necessario introdurre lo spazio campionario delle passeggiate aleatorie di lunghezza infinita. Quindi, per il momento, la descrizione rigorosa della non ricorrenza è: "se $d \geq 3$, esiste una costante $\varepsilon > 0$ tale che per ogni $n \in \mathbb{N}$ la probabilità che la passeggiata in dimensione ≥ 3 non sia mai tornata all'origine *nei primi n passi* è maggiore di ε".

2.4 Le statistiche di Maxwell-Boltzmann, Bose-Einstein e Fermi-Dirac

Nell'esempio 1.10 abbiamo parlato di n palline disposte in r urne, ma lo stesso schema si può applicare a contesti completamente diversi, come ad esempio:

1. gli esiti del lancio di n dadi regolari (o di n lanci di uno stesso dado regolare) a r facce;
2. la disposizione di n passeggeri su un treno composto da r vagoni;
3. la distribuzione di n incidenti in un certo periodo di r giorni;
4. le configurazioni di un sistema di n particelle, ciascuna delle quali può assumere r stati distinti.

In tutti questi casi, spesso non si è interessati a conoscere *quali*, ma soltanto *quante* "palline" ci siano in ciascuna "urna". Se indichiamo con k_i il numero di "palline" nell'"urna" i-esima, ci si concentra cioè sui numeri $k_1, \ldots, k_r \in \mathbb{N}_0$, che sono detti *numeri di occupazione*. Per descrivere esperimenti aleatori che coinvolgono tali numeri, risulta dunque naturale introdurre lo spazio campionario

$$\widehat{\Omega}_{n,r} := \big\{(k_1, \ldots, k_r) \in (\mathbb{N}_0)^r : k_1 + \ldots + k_r = n\big\}. \tag{2.18}$$

Per quanto visto nell'esempio 1.10, una probabilità P_{MB} "naturale" sullo spazio $\widehat{\Omega}_{n,r}$ è quella corrispondente alla densità p_{MB} definita da

$$p_{MB}\big((k_1, \ldots, k_r)\big) = P_{MB}\big(\{(k_1, \ldots, k_r)\}\big) := \frac{1}{r^n} \frac{n!}{k_1! \cdots k_r!}, \tag{2.19}$$

in accordo con (1.18). Per ragioni storiche legate alla meccanica statistica, la probabilità P_{MB} è detta *statistica di Maxwell-Boltzmann* ed è efficacemente utilizzata in svariate situazioni, quali ad esempio quelle descritte nei punti (1), (2), (3) in alto. L'applicazione della probabilità P_{MB} alla situazione (4) è invece delicata e merita una discussione più approfondita.

Supponiamo che un sistema fisico sia composto da n particelle, dove ogni particella può assumere r stati distinti. Facciamo inoltre l'ipotesi che, qualunque siano gli stati delle n particelle, l'energia totale del sistema sia sempre la stessa. Allora, quando il sistema è in equilibrio termico, secondo i principi della meccanica statistica *tutte le configurazioni del sistema sono equiprobabili*.

Il punto fondamentale è decidere che cosa si debba intendere con *configurazione del sistema*. A priori sembra naturale identificare la configurazione del sistema con la n-upla degli stati assunti dalle particelle che lo compongono. In questa ottica, affermare che tutte le configurazioni sono equiprobabili significa considerare la probabilità uniforme sullo spazio $\Omega_{n,r}$ definito in (1.17), cioè sulle disposizioni con ripetizione di n elementi estratti dall'insieme $\{1,\dots,r\}$ (che rappresenta i possibili stati). Di conseguenza, se siamo interessati ai numeri di occupazione, cioè a quante (e non quali) particelle assumano un certo stato, la probabilità "giusta" sullo spazio $\widehat{\Omega}_{n,r}$ dei numeri di occupazione, definito in (2.18), sembrerebbe proprio essere la statistica di Maxwell-Boltzmann P_{MB}, introdotta in (2.19).

Tuttavia, se le n particelle che compongono il sistema sono *indistinguibili*, la statistica di Maxwell-Boltzmann conduce a predizioni errate. La ragione è che, in base ai principi della meccanica quantistica, la scelta dello spazio $\Omega_{n,r}$ definito in (1.17) per descrivere le configurazioni del sistema è fondamentalmente sbagliata: infatti, essendo le particelle indistinguibili, non esiste alcun esperimento che permetta di distinguere due elementi di $\Omega_{n,r}$ con gli stessi numeri di occupazione. Di conseguenza, le configurazioni del sistema devono essere *identificate* con le r-uple dei numeri di occupazione, cioè con gli elementi di $\widehat{\Omega}_{n,r}$: in altri termini, non ha fisicamente senso chiedersi *quali*, ma soltanto *quante* particelle assumano un certo stato. L'affermazione che tutte le configurazioni sono equiprobabili conduce dunque a munire $\widehat{\Omega}_{n,r}$ della *probabilità uniforme*. Tale probabilità, che indicheremo con P_{BE}, è detta *statistica di Bose-Einstein* ed è dunque definita dalla densità

$$p_{BE}\big((k_1,\dots,k_r)\big) \;=\; P_{BE}\big(\{(k_1,\dots,k_r)\}\big) \;:=\; \frac{1}{|\widehat{\Omega}_{n,r}|}\,. \tag{2.20}$$

Gli esperimenti hanno effettivamente mostrato che la statistica di Bose-Einstein si applica con successo a sistemi composti da particelle (o atomi) indistinguibili a spin intero, che vengono dette *bosoni* (un esempio è costituito dai fotoni). Per sistemi composti da particelle a spin semi-intero, dette *fermioni* (ne sono un esempio protoni, neutroni, elettroni, neutrini, ...), il principio di esclusione di Pauli impone l'ulteriore restrizione che *due particelle non possono assumere lo stesso stato*. Questo significa che i numeri di occupazione possono assumere soltanto i valori 0 e 1: in particolare, si deve avere $n \le r$. Di conseguenza, le configurazioni del sistema sono descritte dall'insieme ristretto

$$\widehat{\Omega}'_{n,r} \;:=\; \big\{(k_1,\dots,k_r) \in \{0,1\}^r : k_1+\ldots+k_r = n\big\}\,. \tag{2.21}$$

Dovendo essere tutte le configurazioni equiprobabili, la probabilità "corretta" su $\widehat{\Omega}'_{n,r}$ è quella uniforme, detta *statistica di Fermi-Dirac* e indicata con P_{FD}:

$$p_{FD}\big((k_1,\dots,k_r)\big) \;=\; P_{FD}\big(\{(k_1,\dots,k_r)\}\big) \;:=\; \frac{1}{|\widehat{\Omega}'_{n,r}|}\,. \tag{2.22}$$

Anche in questo caso, gli esperimenti hanno mostrato che la probabilità P_{FD} descrive correttamente il comportamento di sistemi composti da fermioni.

Osserviamo che i valori di $|\widehat{\Omega}_{n,r}|$ e $|\widehat{\Omega}'_{n,r}|$, che compaiono nelle definizioni (2.20) e (2.22), possono essere esplicitati:

$$|\widehat{\Omega}_{n,r}| = \binom{n+r-1}{n}, \qquad |\widehat{\Omega}'_{n,r}| = \binom{r}{n}. \tag{2.23}$$

La seconda relazione segue immediatamente dal fatto che $\widehat{\Omega}'_{n,r}$ è in corrispondenza biunivoca naturale con l'insieme delle combinazioni di n elementi estratti da $\{1,\dots,r\}$: la corrispondenza è quella che a $(k_1,\dots,k_r) \in \widehat{\Omega}'_{n,r}$ associa il sottoinsieme degli indici $i \in \{1,\dots,r\}$ per cui $k_i = 1$ (si ricordi che $k_i \in \{0,1\}$). Per dimostrare la prima relazione in (2.23), procediamo in modo analogo, mostrando che $\widehat{\Omega}_{n,r}$ è in corrispondenza biunivoca con l'insieme delle combinazioni di $(r-1)$ elementi estratti da $\{1,\dots,n+r-1\}$, che sappiamo avere cardinalità pari a $\binom{n+r-1}{r-1} = \binom{n+r-1}{n}$. Un sottoinsieme di $(r-1)$ elementi di $\{1,\dots,n+r-1\}$ può essere indicato con $\{t_1,\dots,t_{r-1}\}$, con $1 \le t_1 < t_2 < \dots < t_{r-1} \le n+r-1$. Se ora definiamo

$$k_1 := t_1 - 1\,, \qquad k_i := t_i - t_{i-1} - 1\,, \ \text{per } 1 < i < r\,, \qquad k_r := (n+r-1) - t_{r-1}\,,$$

si ha per costruzione $k_i \in \mathbb{N}_0$ e $k_1 + \dots + k_r = n$, cioè $(k_1,\dots,k_r) \in \widehat{\Omega}_{n,r}$. È facile vedere che la corrispondenza ora descritta è biunivoca: la corrispondenza inversa è quella che a $(k_1,\dots,k_r)$ associa il sottoinsieme $\{t_1,\dots,t_{r-1}\}$ di $\{1,\dots,n+r-1\}$ definito da

$$t_1 := k_1 + 1\,, \qquad t_i := t_{i-1} + k_i + 1\,, \ \text{per } 2 \le i \le r-1\,.$$

Una rappresentazione grafica di questa corrispondenza si ottiene disegnando i numeri di occupazione mediante palline separate da sbarrette: per esempio, per $n = 7$ e $r = 4$, all'elemento $(1,4,0,2) \in \widehat{\Omega}_{7,4}$ corrisponde la stringa "$\circ \,|\, \circ\circ\circ\circ \,|\, |\, \circ\circ$" composta da $n+r-1 = 10$ simboli. Le sbarrette identificano allora un sottoinsieme di $\{1,\dots,10\}$ di cardinalità $r-1 = 4$.

Le statistiche di Bose-Einstein e Fermi-Dirac, definite dalle relazioni (2.20) e (2.22), sono dette *statistiche quantistiche*, mentre la statistica di Maxwell-Boltzmann, definita in (2.19), è talvolta indicata come *statistica classica*.

Esempio 2.1. Si consideri un sistema di n particelle non interagenti, ciascuna delle quali può assumere r stati differenti, tutti con la stessa energia. Qual è la probabilità $p_{n,r}$ che gli stati assunti dalle particelle siano tutti diversi? Per ipotesi l'energia totale del sistema è indipendente dagli stati delle singole particelle, per cui si può applicare l'analisi svolta in questo paragrafo. Osserviamo inoltre che se $n > r$ si ha ovviamente $p_{n,r} = 0$, per cui supporremo d'ora in avanti che $n \le r$.

In termini di numeri di occupazione $(k_1,\dots,k_r)$, dire che gli stati assunti dalle particelle siano tutti diversi significa imporre $k_i \in \{0,1\}$ per ogni $1 \le i \le r$. In altri termini, $p_{n,r}$ è la probabilità dell'insieme $\widehat{\Omega}'_{n,r}$ definito in (2.21). Il valore di $p_{n,r}$ dipende naturalmente dal tipo di particelle considerate. Se le particelle sono fermioni indistinguibili, sappiamo che $\widehat{\Omega}'_{n,r}$ è l'intero spazio di configurazioni ammissibili, per cui ovviamente $p^{FD}_{n,r} = 1$, in accordo col principio di esclusione di

Pauli. Se invece le n particelle sono bosoni indistinguibili, occorre usare la statistica di Bose-Einstein definita in (2.20), da cui si ottiene

$$p_{n,r}^{BE} = P_{BE}(\widehat{\Omega}'_{n,r}) = \frac{|\widehat{\Omega}'_{n,r}|}{|\widehat{\Omega}_{n,r}|} = \frac{\binom{r}{n}}{\binom{n+r-1}{n}} = \prod_{i=0}^{n-1} \frac{r-i}{r+i} = \prod_{i=0}^{n-1} \frac{1-\frac{i}{r}}{1+\frac{i}{r}}, \tag{2.24}$$

avendo usato le relazioni in (2.23). Infine, se le particelle sono distinguibili, usando la statistica di Maxwell-Boltzmann (2.19) si ottiene

$$p_{n,r}^{MB} = P_{MB}(\widehat{\Omega}'_{n,r}) = \sum_{(k_1,\dots,k_r)\in\widehat{\Omega}'_{n,r}} p_{MB}\big((k_1,\dots,k_r)\big) = \frac{n!}{r^n}|\widehat{\Omega}'_{n,r}| = \prod_{i=0}^{n-1}\left(1-\frac{i}{r}\right), \tag{2.25}$$

dove si è usata la seconda relazione in (2.23) e il fatto che $p_{MB}\big((k_1,\dots,k_r)\big) = n!/r^n$ per ogni $(k_1,\dots,k_r)\in\widehat{\Omega}'_{n,r}$, poiché $k_i\in\{0,1\}$ per ogni $1\le i\le r$. Dalle relazioni (2.24) e (2.25) segue in particolare che $p_{n,r}^{BE} < p_{n,r}^{MB}$ per ogni $n\ge 1$.

Per avere un'idea più precisa dei risultati ottenuti, supponiamo che r sia grande e che $n = o(r^{2/3})$. Più precisamente, fissiamo un'arbitraria successione positiva $(\varepsilon_r)_{r\in\mathbb{N}}$ tale che $\lim_{r\to\infty}\varepsilon_r = 0$ e studiamo il comportamento asintotico di $p_{n,r}^{MB}$ e $p_{n,r}^{BE}$ nel limite $r\to\infty$, assumendo che $n\le \varepsilon_r r^{2/3}$. Dato che $(1+x) = \exp(x+O(x^2))$ per $x\to 0$, da (2.25) si ottiene

$$p_{n,r}^{MB} = \exp\left(-\sum_{i=0}^{n-1}\frac{i}{r} + \sum_{i=0}^{n-1} O\left(\frac{i^2}{r^2}\right)\right) = \exp\left(-\frac{1}{2}\,\frac{n^2}{r} + o(1)\right), \tag{2.26}$$

dove abbiamo usato il fatto che $\sum_{i=0}^{n-1} i = \frac{1}{2}n(n-1)$ e $\sum_{i=0}^{n-1} i^2 = O(n^3)$. Analogamente, da (2.24) si ottiene:

$$p_{n,r}^{BE} = \exp\left(-\frac{n^2}{r} + o(1)\right). \tag{2.27}$$

Le due relazioni (2.26) e (2.27) da un lato ci dicono che, sia per la statistica di Maxwell-Boltzmann sia per quella di Bose-Einstein, il regime "critico" per la probabilità $p_{n,r}$ è $n\approx\sqrt{r}$: in altri termini, $p_{n,r}\approx 1$ se $n\ll\sqrt{r}$ mentre $p_{n,r}\approx 0$ se $n\gg\sqrt{r}$. Dall'altro lato, se già sapevamo che $p_{n,r}^{BE} < p_{n,r}^{MB}$ per ogni $n\ge 1$, le relazioni (2.26) e (2.27) mostrano che nel regime $n\sim c\sqrt{r}$, con $c\in(0,\infty)$, la differenza tra $p_{n,r}^{MB}$ e $p_{n,r}^{BE}$ resta positiva anche nel limite di r grande: si ha in effetti $p_{n,r}^{BE}\to e^{-c}$ e $p_{n,r}^{MB}\to e^{-c/2}$.

Ricordiamo infine che la statistica di Maxwell-Boltzmann, adottata per descrivere un sistema di "particelle distinguibili", si può applicare anche a sistemi "macroscopici" come quelli descritti nei punti (1), (2) e (3) all'inizio del paragrafo. In particolare, $p_{n,r}^{MB}$ coincide con la probabilità che, se n passeggeri salgono su un treno composto da r vagoni e si dispongono a caso, nessun vagone contenga più di un passeggero.

2.4.1 La condensazione di Bose-Einstein

L'analisi che ha portato all'introduzione delle statistiche di Bose-Einstein (2.20) e Fermi-Dirac (2.22) è basata sull'ipotesi che l'energia totale del sistema sia indipendente dagli stati delle particelle che lo compongono. Nel caso più realistico in cui l'energia del sistema non sia costante, si procede in modo del tutto analogo a quanto visto nell'Esempio 1.3: se Ω è lo spazio delle configurazioni del sistema, $H : \Omega \to \mathbb{R}$ è la funzione energia e $\beta \in [0,\infty)$ è l'inverso della temperatura assoluta, quando il sistema è in equilibrio termico alla temperatura β la probabilità di osservare il sistema in una configurazione $\omega \in \Omega$ è data da

$$P_\beta\big(\{\omega\}\big) := \frac{e^{-\beta H(\omega)}}{Z(\beta)}, \qquad \text{dove} \qquad Z(\beta) := \sum_{\omega\in\Omega} e^{-\beta H(\omega)}.$$

Il punto cruciale qui è la scelta dello spazio delle configurazioni Ω. Se il sistema è composto da n particelle *indistinguibili*, ciascuna delle quali può assumere r stati, si dovrà scegliere $\Omega = \widehat{\Omega}_{n,r}$ come in (2.18) se le particelle sono bosoni e $\Omega = \widehat{\Omega}'_{n,r}$ come in (2.21) se le particelle sono fermioni. Se invece le particelle sono distinguibili (per esempio se sono diverse tra loro) si sceglierà $\Omega = \Omega_{n,r}$ come in (1.17).

Come abbiamo già osservato nell'Esempio 1.3, per $\beta = 0$ (limite di temperatura infinita) la probabilità $P_\beta(\{\omega\})$ non dipende più da ω, dunque P_0 non è altro che la probabilità uniforme su Ω. Ritroviamo in questo modo le statistiche di Bose-Einstein ($P_0 = P_{BE}$ se $\Omega = \widehat{\Omega}_{n,r}$, cf. (2.20) e di Fermi-Dirac ($P_0 = P_{FD}$ se $\Omega = \widehat{\Omega}'_{n,r}$, cf. (2.22). Sempre dall'Esempio 1.3 sappiamo che per $\beta \to \infty$ (limite di temperatura nulla) la probabilità P_β si concentra sull'insieme dei minimi assoluti della funzione H. In questo regime un sistema di bosoni indistinguibili dà origine a un fenomeno molto interessante, noto come *condensazione di Bose-Einstein*, in cui gli effetti quantistici diventano visibili su grande scala.

Per studiare un modello quantitativo di tale fenomeno, consideriamo un sistema di n particelle indistinguibili, ciascuna delle quali può assumere r stati, che indicheremo con $\{1,2,\dots,r\}$. Allo stato i corrisponde un'energia ε_i della singola particella, e supporremo che $\varepsilon_1 < \varepsilon_2 < \ldots < \varepsilon_r$. Facciamo ora l'ipotesi che le particelle siano debolmente interagenti, in modo che l'energia totale del sistema sia data approssimativamente dalla somma delle energie delle singole particelle. In altri termini, considerando lo spazio $\Omega = \widehat{\Omega}_{n,r}$ dei numeri di occupazione, cf. (2.18), a ogni configurazione $\omega = (k_1,\dots,k_r) \in \widehat{\Omega}_{n,r}$ del nostro sistema di n particelle corrisponde un'energia $H(\omega) = k_1\varepsilon_1 + \ldots + k_r\varepsilon_r$. La misura di Gibbs $P_\beta = P^{BE}_{n,r,\beta}$ associata a tale energia è la probabilità su $\widehat{\Omega}_{n,r}$ definita per ogni $\omega = (k_1,\dots,k_r) \in \widehat{\Omega}_{n,r}$ da

$$P^{BE}_{n,r,\beta}\big(\{(k_1,\dots,k_r)\}\big) := \frac{e^{-\beta(k_1\varepsilon_1+\ldots+k_r\varepsilon_r)}}{Z^{BE}_{n,r,\beta}},$$

dove

$$Z^{BE}_{n,r,\beta} := \sum_{(k_1,\dots,k_r)\in(\mathbb{N}_0)^r : k_1+\ldots+k_r=n} e^{-\beta(k_1\varepsilon_1+\ldots+k_r\varepsilon_r)}. \tag{2.28}$$

È possibile derivare un'espressione chiusa per la serie di potenze con coefficienti $Z^{BE}_{n,r,\beta}$:

$$\begin{aligned} \Gamma_{r,\beta}(x) &:= \sum_{n\in\mathbb{N}_0} x^n Z^{BE}_{n,r,\beta} = \sum_{n\in\mathbb{N}_0} \sum_{\substack{(k_1,\dots,k_r)\in(\mathbb{N}_0)^r \\ k_1+\dots+k_r=n}} x^{k_1+\dots+k_r} e^{-\beta(k_1\varepsilon_1+\dots+k_r\varepsilon_r)} \\ &= \sum_{(k_1,\dots,k_r)\in(\mathbb{N}_0)^r} (xe^{-\beta\varepsilon_1})^{k_1}\cdots(xe^{-\beta\varepsilon_r})^{k_r} = \frac{1}{(1-xe^{-\beta\varepsilon_1})\cdots(1-xe^{-\beta\varepsilon_r})}\,, \end{aligned} \tag{2.29}$$

dove l'ultima uguaglianza è valida se $|xe^{-\beta\varepsilon_i}|<1$ per ogni $i=1,\dots,r$, ossia per ogni $x\in\mathbb{C}$ con $|x|<e^{\beta\varepsilon_1}$ (ricordiamo che $\varepsilon_1<\varepsilon_2<\dots<\varepsilon_r$).

Consideriamo ora l'evento A_ℓ che ℓ o più tra le n particelle *non* siano nello stato fondamentale (quello di energia minima), ossia:

$$A_\ell := \{(k_1,\dots,k_r)\in\widehat{\Omega}_{n,r} : k_1\le n-\ell\}\,. \tag{2.30}$$

Possiamo esprimere la probabilità di A_ℓ nel modo seguente:

$$\begin{aligned} P^{BE}_{n,r,\beta}(A_\ell) &= \sum_{(k_1,\dots,k_r)\in A_\ell} P^{BE}_{n,r,\beta}(\{(k_1,\dots,k_r)\}) \\ &= \frac{1}{Z^{BE}_{n,r,\beta}} \sum_{k_1=0}^{n-\ell} e^{-\beta k_1\varepsilon_1} \sum_{\substack{(k_2,\dots,k_r)\in(\mathbb{N}_0)^{r-1} \\ k_2+\dots+k_r=n-k_1}} e^{-\beta(k_2\varepsilon_2+\dots+k_r\varepsilon_r)} \\ &= \frac{1}{Z^{BE}_{n,r,\beta}} \sum_{k_1=0}^{n-\ell} e^{-\beta k_1\varepsilon_1} \tilde{Z}^{BE}_{n-k_1,r-1,\beta} = \frac{1}{Z^{BE}_{n,r,\beta}} \sum_{m=\ell}^{n} e^{-\beta(n-m)\varepsilon_1} \tilde{Z}^{BE}_{m,r-1,\beta}\,, \end{aligned} \tag{2.31}$$

dove $\tilde{Z}^{BE}_{n-k_1,r-1,\beta}$ indica la funzione di partizione di un sistema di $n-k_1$ particelle, ciascuna delle quali può assumere $r-1$ stati differenti, con energie associate $\varepsilon_2<\dots<\varepsilon_r$ (cf. (2.28)), e nell'ultima uguaglianza abbiamo fatto il cambio di variabili $m:=n-k_1$. In analogia con (2.29), per ogni $x\in\mathbb{C}$ con $|x|\le e^{\beta\varepsilon_2}$ si ha

$$\tilde{\Gamma}_{r-1,\beta}(x) := \sum_{m\in\mathbb{N}_0} x^m \tilde{Z}^{BE}_{m,r-1,\beta} = \frac{1}{(1-xe^{-\beta\varepsilon_2})\cdots(1-xe^{-\beta\varepsilon_r})}\,. \tag{2.32}$$

Ritornando a (2.31), possiamo stimare dal basso il denominatore $Z^{BE}_{n,r,\beta}\ge e^{-\beta\varepsilon_1 n}$, come è chiaro restringendo la somma in (2.28) alla singola configurazione $(k_1,\dots,k_r)=(0,\dots,0)$. Per quanto riguarda il numeratore, fissiamo $\eta>0$ sufficientemente piccolo in modo che $(\varepsilon_2-\varepsilon_1)-\eta>0$ e notiamo che per ogni $m\ge\ell$ possiamo stimare dall'alto

$$e^{-\beta(n-m)\varepsilon_1} = e^{-\beta\varepsilon_1 n} e^{\beta(\varepsilon_2-\eta)m} e^{-\beta((\varepsilon_2-\varepsilon_1)-\eta)m} \le e^{-\beta\varepsilon_1 n} e^{\beta(\varepsilon_2-\eta)m} e^{-\beta((\varepsilon_2-\varepsilon_1)-\eta)\ell}\,.$$

Di conseguenza, dalla relazione (2.31) otteniamo

$$P^{BE}_{n,r,\beta}(A_\ell) \le e^{-\beta((\varepsilon_2-\varepsilon_1)-\eta)\ell}\left(\sum_{m=\ell}^{\infty} e^{\beta(\varepsilon_2-\eta)m}\tilde{Z}^{BE}_{m,r-1,\beta}\right) = C_{\beta,r}\, e^{-\beta\, c\, \ell}, \qquad (2.33)$$

dove abbiamo posto $c := (\varepsilon_2-\varepsilon_1)-\eta$ e $C_{r,\beta} := \tilde{\Gamma}_{r-1,\beta}(e^{\beta(\varepsilon_2-\eta)})$, cf. (2.32). Si noti che c e $C_{\beta,r}$ sono costanti positive che non dipendono dal numero di particelle n.

La relazione (2.33) mostra che tipicamente il numero di particelle al di fuori dello stato fondamentale è uniformemente limitato da una costante, che dipende da β ma non dal numero totale n di particelle. Più precisamente, fissiamo una soglia di probabilità $\delta > 0$ (per esempio $\delta = 0.1\%$) e definiamo $\ell_0 = \ell_0^{BE}(n,r,\beta)$ come il più piccolo valore di ℓ tale che $P^{BE}_{n,r,\beta}(A_\ell) \le \delta$. La relazione (2.33) mostra allora che

$$\ell_0^{BE}(n,r,\beta,\delta) \le \hat{\ell}_0^{BE}(r,\beta,\delta) := \frac{1}{c\beta}(\log C_{\beta,r} + \log(1/\delta)) \qquad (2.34)$$

In altri termini, per ogni valore di β, r, δ, con grande probabilità (maggiore di $1-\delta$) *ci sono al massimo* $\hat{\ell}_0^{BE}(r,\beta,\delta) < \infty$ *particelle al di fuori dello stato fondamentale, non importa quanto sia elevato il numero totale n di particelle del sistema*. Quando questo effetto diventa macroscopicamente rilevante, ossia quando $\ell_0 \ll n$, si ha la *condensazione di Bose-Einstein*.

Come mostriamo nell'Osservazione 2.1, per $\beta \to \infty$ si ha $C_{\beta,r} \to 1$ uniformemente in r e dunque $\hat{\ell}_0^{BE}(r,\beta,\delta) \sim \frac{\log(1/\delta)}{c}\frac{1}{\beta} \to 0$ uniformemente in r. Per β sufficientemente grande (essenzialmente per $\beta \gg 1/c \approx 1/(\varepsilon_2-\varepsilon_1)$) si ha dunque $\hat{\ell}_0^{BE}(r,\beta,\delta) \ll n$ e la condensazione di Bose-Einstein è macroscopicamente osservabile.

Osservazione 2.1. Studiamo la dipendenza della costante $C_{r,\beta}$ dall'inverso della temperatura β e dal numero r di stati possibili per ciascuna particella. Supponiamo che ciascuna particella possa assumere infiniti stati, indicizzati da $\mathbb{N}$ e con energie $\{\varepsilon_i\}_{i\in\mathbb{N}}$ (dove $\varepsilon_1 < \varepsilon_2 < \ldots$) e facciamo l'ipotesi che la successione ε_i diverga più che logaritmicamente, più precisamente $\sum_{i\in\mathbb{N}} e^{-\beta\varepsilon_i} < \infty$ per ogni $\beta > 0$. Allora, segue dalla relazione (2.32) che per ogni $x \in \mathbb{C}$ con $|x| \le e^{\beta\varepsilon_2}$ esiste finito il limite (crescente)

$$\tilde{\Gamma}_{\infty,\beta}(x) := \lim_{r\to\infty}\tilde{\Gamma}_{r-1,\beta}(x) = \prod_{i=2}^{\infty}\frac{1}{1-xe^{-\beta\varepsilon_i}} = \exp\left(\sum_{i=2}^{\infty} -\log(1-xe^{-\beta\varepsilon_i})\right).$$

Di conseguenza, ricordando che $C_{r,\beta} = \tilde{\Gamma}_{r-1,\beta}(e^{\beta(\varepsilon_2-\eta)})$, esiste finito il limite

$$\lim_{r\to\infty} C_{r,\beta} = C_{\infty,\beta} := \tilde{\Gamma}_{\infty,\beta}(e^{\beta(\varepsilon_2-\eta)}) = \exp\left(\sum_{i=2}^{\infty} -\log(1-e^{-\beta\eta}e^{-\beta(\varepsilon_i-\varepsilon_2)})\right). \qquad (2.35)$$

In effetti, dato che $\tilde{\Gamma}_{r-1,\beta}(x)$ è crescente in r, si ha $1 \le C_{r,\beta} \le C_{\infty,\beta}$ per ogni β e r, e si verifica direttamente da (2.35) che la funzione $C_{\infty,\beta}$ è decrescente di β e tende a

1 per $\beta \to +\infty$ (esercizio). Questo mostra che $\lim_{\beta\to\infty} C_{r,\beta} = 1$ uniformemente in r (in particolare, $\log C_{r,\beta}$ è uniformemente limitata).

Osservazione 2.2. La condensazione di Bose-Einstein è un effetto quantistico che non ha un analogo classico. Come abbiamo già ricordato, se consideriamo un sistema di n particelle distinguibili, lo spazio naturale delle configurazioni è $\Omega_{n,r}$, cf. (1.17), quindi la probabilità naturale sullo spazio dei numeri di occupazione $\widehat{\Omega}_{n,r}$, quando tutti gli stati hanno la stessa energia, è data dalla statistica classica di Maxwell-Boltzmann P_{MB}, definita in (2.19).

Consideriamo ora il caso in cui ogni stato $i \in \{1,\dots,r\}$ della singola particella abbia un'energia ε_i, con $\varepsilon_1 < \ldots < \varepsilon_r$, e le particelle siano debolmente interagenti, così che l'energia totale di una configurazione $\omega = (k_1,\dots,k_r) \in \widehat{\Omega}_{n,r}$ del sistema di n particelle sia data da $H(\omega) = k_1\varepsilon_1 + \ldots + k_r\varepsilon_r$. La misura di Gibbs $P^{MB}_{n,r,\beta}$ in questo caso è allora definita per ogni $(k_1,\dots,k_r) \in \widehat{\Omega}_{n,r}$ da

$$P^{MB}_{n,r,\beta}\big(\{(k_1,\dots,k_r)\}\big) \;:=\; \frac{e^{-\beta(k_1\varepsilon_1+\ldots+k_r\varepsilon_r)}}{Z^{MB}_{n,r,\beta}}\,\frac{n!}{k_1!\cdots k_r!}\,, \tag{2.36}$$

dove

$$Z^{MB}_{n,r,\beta} \;:=\; \sum_{(k_1,\dots,k_r)\in(\mathbb{N}_0)^r:\,k_1+\ldots+k_r=n} \frac{n!}{k_1!\cdots k_r!}\,e^{-\beta(k_1\varepsilon_1+\ldots+k_r\varepsilon_r)}\,. \tag{2.37}$$

(Rispetto all'equazione (2.19) abbiamo tralasciato il fattore $1/r^n$, che non dipende da $(k_1,\dots,k_r)$ e dunque si semplificherebbe tra numeratore e denominatore in (2.36).) È possibile esprimere $Z^{MB}_{n,r,\beta}$ in forma chiusa. Infatti, in analogia con il binomio di Newton, per ogni $n,r \in \mathbb{N}$ e per ogni scelta di $a_1,\dots,a_r \in \mathbb{R}^+$ si ha

$$\sum_{(k_1,\dots,k_r)\in(\mathbb{N}_0)^r:\,k_1+\ldots+k_r=n} \frac{n!}{k_1!\cdots k_r!}\,(a_1)^{k_1}\cdots(a_r)^{k_r} = (a_1+\ldots+a_r)^n\,,$$

come si può facilmente dimostrare per induzione su r. Di conseguenza possiamo scrivere

$$Z^{MB}_{n,r,\beta} \;=\; \big(e^{-\beta\varepsilon_1}+\ldots+e^{-\beta\varepsilon_r}\big)^n\,. \tag{2.38}$$

Studiamo ora la probabilità dell'evento A_ℓ definito in (2.30). In analogia con (2.31), ricordando (2.37) possiamo scrivere

$$\begin{aligned} P^{MB}_{n,r,\beta}(A_\ell) &= \sum_{(k_1,\dots,k_r)\in A_\ell} P^{MB}_{n,r,\beta}(\{(k_1,\dots,k_r)\}) \\ &= \frac{1}{Z^{MB}_{n,r,\beta}} \sum_{k_1=0}^{n-\ell} e^{-\beta k_1\varepsilon_1}\,\frac{n!}{k_1!} \sum_{\substack{(k_2,\dots,k_r)\in(\mathbb{N}_0)^{r-1}\\ k_2+\ldots+k_r=n-k_1}} \frac{e^{-\beta(k_2\varepsilon_2+\ldots+k_r\varepsilon_r)}}{k_2!\cdots k_r!} \\ &= \frac{1}{Z^{MB}_{n,r,\beta}} \sum_{k_1=0}^{n-\ell} e^{-\beta k_1\varepsilon_1}\,\frac{n!}{k_1!(n-k_1)!}\,\tilde{Z}^{MB}_{n-k_1,r-1,\beta}\,. \end{aligned} \tag{2.39}$$

Usando la formula (2.38) e cambiando variabile $m = n - k_1$, otteniamo

$$\begin{aligned} P^{MB}_{n,r,\beta}(A_\ell) &= \sum_{m=\ell}^{n} \frac{n!}{m!(n-m)!} \frac{(e^{-\beta\varepsilon_2} + \ldots + e^{-\beta\varepsilon_r})^m \, (e^{-\beta\varepsilon_1})^{n-m}}{(e^{-\beta\varepsilon_1} + \ldots + e^{-\beta\varepsilon_r})^n} \\ &= \sum_{m=\ell}^{n} \frac{n!}{m!(n-m)!} \, p^m \, (1-p)^{n-m} , \end{aligned} \tag{2.40}$$

dove

$$p = p_{r,\beta} := 1 - \frac{e^{-\beta\varepsilon_1}}{e^{-\beta\varepsilon_1} + \ldots + e^{-\beta\varepsilon_r}} \in (0,1) .$$

Analogamente all'Osservazione 2.1, per studiare la dipendenza di $p_{r,\beta}$ da r facciamo l'ipotesi standard che $\sum_{i\in\mathbb{N}} e^{-\beta\varepsilon_i} < \infty$ per ogni $\beta > 0$. Allora esiste finito il limite $\lim_{r\to\infty} p_{r,\beta} =: p_{\infty,\beta}$, dove

$$p_{\infty,\beta} = 1 - \frac{e^{-\beta\varepsilon_1}}{\sum_{i=1}^{\infty} e^{-\beta\varepsilon_i}} \in (0,1) .$$

e $p_{r,\beta} \le p_{\infty,\beta}$ per ogni r, β, perché $p_{r,\beta}$ è crescente in r. Per $\beta \to +\infty$ si ha $p_{\infty,\beta} \to 0$, tuttavia per ogni $\beta < \infty$ si ha $p_{r,\beta} \in (0,1)$ per ogni $r \in \mathbb{N}$.

Ritornando a (2.40), fissiamo una soglia di probabilità $\delta \in (0,1)$ e chiediamoci qual è il più piccolo valore di $\ell_0 = \ell_0^{MB}(n,r,\beta,\delta)$ affinché $P^{MB}_{n,r,\beta}(A_{\ell_0}) \le \delta$. Applicando la legge dei grandi numeri, è possibile mostrare che per $n \to \infty$

$$\ell_0^{MB}(n,r,\beta,\delta) = p_{r,\beta} \cdot n + o(n) .$$

In altri termini, per un sistema a cui si applica la statistica di Maxwell-Boltzmann *il numero ℓ_0 di particelle al di fuori dello stato fondamentale è sempre una frazione propria $p_{r,\beta} \in (0,1)$ del numero totale n di particelle*, in forte contrasto con quanto accade per la statistica di Bose-Einstein.

2.5 Il modello di Ising in meccanica statistica

Descriviamo ora un celebre modello in meccanica statistica, il *modello di Ising* per un materiale ferromagnetico. Sia Λ un sottoinsieme finito di $\mathbb{Z}^d$. I punti di Λ vanno interpretati come i *nodi* (detti anche *siti*) di un cristallo regolare. Ogni nodo è occupato da un atomo, il moto dei cui elettroni produce un campo magnetico. In questo modello semplificato, si assume che tale campo magnetico, che chiameremo *spin*, assuma solo due valori, $+1$ e -1. Una *configurazione* per tale sistema è

$$\sigma = (\sigma_x)_{x\in\Lambda} ,$$

dove $\sigma_x = \pm 1$ è lo spin nel nodo $x \in \Lambda$. In altre parole $\Omega = \{-1,1\}^\Lambda$ è l'insieme di tutte le configurazioni.

L'energia (potenziale) associata ad una configurazione è dovuta all'interazione tra gli spin dei nodi in Λ e all'interazione con l'esterno. In questa presentazione, assumiamo che l'interazione sia *locale*: l'interazione tra gli spin in Λ avviene solo tra siti *primi vicini*, la cui distanza è pari a 1, mentre l'interazione con l'esterno riguarda solo i nodi del "bordo" di Λ, cioè $\partial\Lambda = \{x \in \Lambda : \exists y \in \Lambda^c \text{ tale che } |x-y| = 1\}$. Più precisamente, l'energia (o Hamiltoniana) di una configurazione σ è data da

$$H_\Lambda^\tau(\sigma) := -\sum_{\substack{x,y\in\Lambda \\ |x-y|=1}} \sigma_x\sigma_y - \sum_{x\in\partial\Lambda} \tau_x\sigma_x\,. \tag{2.41}$$

Si noti che il primo termine in H_Λ^τ descrive l'interazione tra gli spin primi vicini in Λ, mentre il secondo termine può essere interpretato come risultante da campi magnetici di valore τ_x agenti sugli spin del bordo di Λ. Assumiamo per semplicità che $\tau_x = \pm 1$, cioè $\tau \in \{-1,1\}^{\partial\Lambda}$.

Si noti che se fosse $\tau_x \equiv 0$, l'energia H_Λ^τ avrebbe esattamente due minimi assoluti, dati rispettivamente da $\sigma_x \equiv 1$ e $\sigma_x \equiv -1$; più in generale, l'energia di una configurazione σ sarebbe uguale a quella della configurazione $-\sigma$. La presenza di un campo magnetico al bordo $\tau \in \{-1,1\}^{\partial\Lambda}$, rompe tale simmetria: in particolare, se $\tau_x \equiv +1$, l'unico minimo di H_Λ^τ è la configurazione con $\sigma_x \equiv +1$. In ogni caso, una configurazione ha un valore tanto più basso dell'energia quanto più gli spin della configurazione sono allineati tra di loro.

Una quantità che gioca un ruolo fisico fondamentale è la temperatura. Il "moto termico" degli atomi si traduce in un "disturbo aleatorio" sugli spin: il sistema ha una "preferenza" per le configurazioni a bassa energia, ma tale preferenza è tanto più dobole tanto più è alta la temperatura. Queste considerazioni intuitive hanno una traduzione precisa in meccanica statistica considerando la *misura di Gibbs* associata all'Hamiltoniana H_Λ, descritta nell'Esempio 1.3, che ora riprendiamo in dettaglio. Se T è la temperatura assoluta, indichiamo con $\beta = 1/(k_B T)$ la temperatura inversa, dove k_B è la costante di Boltzmann. È conveniente semplificare le notazioni ponendo $k_B = 1$ (il che equivale a misurare la temperatura in unità di k_B), di modo che $\beta = 1/T$. Secondo l'ipotesi di Gibbs, se il sistema è in equilibrio ad una temperatura inversa $\beta > 0$ con campo magnetico al bordo τ, la probabilità di osservare una configurazione di spin σ è data da

$$\mu_{\Lambda,\beta}^\tau(\{\sigma\}) := \frac{1}{Z_\Lambda^\tau} \exp[-\beta H_\Lambda^\tau]\,,$$

dove

$$Z_{\Lambda,\beta}^\tau := \sum_{\sigma\in\Omega} \exp[-\beta H_\Lambda^\tau]\,.$$

In questo modo $\sum_{\sigma\in\Omega} \mu_{\Lambda,\beta}^\tau(\{\sigma\}) = 1$. Dunque, come in (1.4), $\mu_{\Lambda,\beta}^\tau$ si può estendere ad una probabilità su Ω ponendo, per $A \subset \Omega$:

$$\mu_{\Lambda,\beta}^\tau(A) := \sum_{\sigma\in A} \mu_\Lambda^\tau(\{\sigma\})\,.$$

Come abbiamo visto nell'Esempio 1.3, per ogni $\sigma \in \Omega$ si ha che

$$\lim_{\beta \to 0} \mu^{\tau}_{\Lambda,\beta}(\{\sigma\}) \;=\; \frac{1}{|\Omega|}\,,$$

cioè, nel limite di temperatura infinita, tutte le configurazioni diventano equiprobabili. Inoltre, per ogni $\sigma \in \Omega$ che non sia un minimo assoluto di H^{τ}_{Λ} si ha

$$\lim_{\beta \to +\infty} \mu^{\tau}_{\Lambda}(\{\sigma\}) \;=\; 0\,,$$

cioè, nel limite di temperatura zero, il sistema tende a "congelarsi" nelle configurazioni che minimizzano l'energia.

Supponiamo ora di fissare il reticolo $\Lambda = \Lambda_n = \{-n, -n+1, \ldots, 0, \ldots, n-1, n\}^d$ e le *condizioni al bordo* $\tau_x \equiv 1$. Poniamo $\Omega_n := \{-1,1\}^{\Lambda_n}$ e scriveremo $\mu^{+}_{n,\beta}$ in luogo di $\mu^{\tau}_{\Lambda_n,\beta}$, $Z^{+}_{n,\beta}$ in luogo di $Z^{\tau}_{\Lambda_n,\beta}$ e H^{+}_{n} in luogo di $H^{\tau}_{\Lambda_n}$. Introduciamo l'evento

$$A \;:=\; \{\sigma \in \Omega_n : \sigma_0 = +1\}\,, \tag{2.42}$$

dove $0 = (0,0,\ldots,0)$ indica l'origine in $\mathbb{Z}^d$. Quindi $\mu^{+}_{n,\beta}(A)$ è la probabilità che lo spin nell'origine sia positivo. Come suggerito in precedenza, la presenza al bordo di un campo magnetico positivo "favorisce" gli spin positivi rispetto a quelli negativi: di conseguenza, è intuitivamente plausibile che si abbia

$$\mu^{+}_{n,\beta}(A) \;>\; 1/2\,.$$

Questa disuguaglianza è effettivamente vera, per ogni valore fissato di $n \in \mathbb{N}$ e $\beta > 0$, ma ne omettiamo per brevità la dimostrazione. Essa esprime il fatto che lo spin nell'origine "risente" del campo magnetico al bordo, e, con probabilità $> 1/2$, si allinea allo stesso. Facendo crescere n, aumenta la distanza tra l'origine e il campo magnetico al bordo, la cui influenza, si può congetturare, diventi sempre meno rilevante. In altre parole, potrebbe accadere che

$$\lim_{n \to +\infty} \mu^{+}_{n,\beta}(A) = 1/2. \tag{2.43}$$

Viceversa, potrebbe accadere che l'influenza del campo magnetico al bordo sull'origine sia rilevante anche per n grande, ossia che esista $\varepsilon > 0$ tale che, *per ogni* n,

$$\mu^{+}_{n,\beta}(A) > \frac{1}{2} + \varepsilon. \tag{2.44}$$

Se accade (2.44), si dice che (per il valore di β dato) si ha *magnetizzazione spontanea*.

Per i ferromagneti reali, la magnetizzazione spontanea è un fenomeno effettivamente osservato, purché la temperatura sia non troppo elevata. Il problema che ci poniamo è di stabilire se il modello di Ising, almeno per questo aspetto, è un buon modello per un ferromagnete reale. Il risultato interessante è che la risposta dipende

dalla dimensione d dello spazio: per $d \geq 2$ si ha magnetizzazione spontanea a basse temperature, mentre per $d = 1$ non si ha magnetizzazione spontanea per nessun valore della temperatura. Non dovrebbe sorprendere il fatto che l'analisi del modello diventa via via più difficile al crescere della dimensione d. Benchè la dimensione "fisica" sia $d = 3$, ci occuperemo per semplicità solo dei casi $d = 1$ e $d = 2$.

2.5.1 Il caso $d = 1$

Per $d = 1$ si ha $\Lambda_n = \{-n, -n+1, \ldots, n-1, n\}$ e $\Omega_n = \{-1, 1\}^{\Lambda_n}$. Con un conto esplicito possiamo scrivere:

$$\begin{aligned}
\mu^+_{n,\beta}(A) &= \sum_{\sigma \in A} \mu^+_{n,\beta}(\{\sigma\}) = \frac{1}{Z^+_{n,\beta}} \sum_{\sigma \in \Omega_n : \sigma_0 = 1} \exp\left[\beta\left(\sigma_{-n} + \sum_{k=-n}^{n-1} \sigma_k \sigma_{k+1} + \sigma_n\right)\right] \\
&= \frac{1}{Z^+_{n,\beta}} \sum_{\substack{\sigma_{-n},\ldots,\sigma_{-1} \\ \sigma_1,\ldots,\sigma_n}} \exp\left[\beta\left(\sigma_{-n} + \sum_{k=-n}^{-2} \sigma_k \sigma_{k+1} + \sigma_{-1}\right)\right] \exp\left[\beta\left(\sigma_1 + \sum_{k=1}^{n-1} \sigma_k \sigma_{k+1} + \sigma_n\right)\right] \\
&= \frac{1}{Z^+_{n,\beta}} \left(\sum_{\sigma_{-n},\ldots,\sigma_{-1}} e^{\beta\left(\sigma_{-n} + \sum_{k=-n}^{-2} \sigma_k \sigma_{k+1} + \sigma_{-1}\right)} \right) \left(\sum_{\sigma_1,\ldots,\sigma_n} e^{\beta\left(\sigma_1 + \sum_{k=1}^{n-1} \sigma_k \sigma_{k+1} + \sigma_n\right)} \right).
\end{aligned}$$

Si noti ora che le due somme contenute in quest'ultima espressione sono uguali, cambiando solo i nomi delle variabili sommate ($\sigma_i \leftrightarrow \sigma_{-i}$). Dunque:

$$\mu^+_{n,\beta}(A) = \frac{1}{Z^+_{n,\beta}} \left\{ \sum_{\sigma_1,\ldots,\sigma_n} \exp\left[\beta\left(\sigma_1 + \sum_{k=1}^{n-1} \sigma_k \sigma_{k+1} + \sigma_n\right)\right] \right\}^2 . \tag{2.45}$$

Per semplificare questa espressione, introduciamo un operatore lineare T, che agisce sullo spazio vettoriale delle funzioni f da $\{-1, +1\}$ in $\mathbb{R}$ nel modo seguente: la funzione Tf, sempre da $\{-1, +1\}$ in $\mathbb{R}$, è definita da

$$(Tf)(s) := \sum_{s'=\pm 1} e^{\beta s s'} f(s') = e^{\beta s} f(1) + e^{-\beta s} f(-1).$$

Una funzione $f : \{-1, +1\} \to \mathbb{R}$ può essere identificata con il vettore colonna $\binom{f(-1)}{f(+1)}$. In questo modo, la trasformazione $f \to Tf$ corrisponde alla trasformazione lineare sui vettori di dimensione due data dalla matrice $T_{s,s'} := e^{\beta s s'}$, cioè

$$T = \begin{pmatrix} e^{\beta} & e^{-\beta} \\ e^{-\beta} & e^{\beta} \end{pmatrix}. \tag{2.46}$$

Posto $\varphi(s) = e^{\beta s}$, possiamo riscrivere la relazione (2.45) come

$$\mu^+_{n,\beta}(A) \;=\; \frac{1}{Z^+_{n,\beta}} \left[\sum_{\sigma_1,\ldots,\sigma_n} T_{1,\sigma_1} \cdot T_{\sigma_1,\sigma_2} \cdots T_{\sigma_{n-1},\sigma_n} \cdot \varphi(\sigma_n) \right]^2 \;=\; \frac{1}{Z^+_{n,\beta}} \big[(T^n\varphi)(1)\big]^2. \tag{2.47}$$

Si osservi che, se avessimo voluto calcolare $\mu^+_n(A^c)$, avremmo dovuto sommare sulle configurazioni per le quali $\sigma_0 = -1$. L'unica differenza, rispetto all'espressione in (2.45), è che l'addendo σ_1 nell'esponenziale sarebbe stato sostituito da $-\sigma_1$. Usando le notazioni or ora introdotte, possiamo concludere che

$$\mu^+_{n,\beta}(A^c) \;=\; \frac{1}{Z^+_{n,\beta}} \big[(T^n\varphi)(-1)\big]^2.$$

Quest'ultima uguaglianza, assieme a (2.47) e al fatto che $\mu^+_{n,\beta}(A) + \mu^+_{n,\beta}(A^c) = 1$, ci dà

$$Z^+_{n,\beta} = \big[(T^n\varphi)(1)\big]^2 + \big[(T^n\varphi)(-1)\big]^2.$$

Mettendo tutto assieme:

$$\mu^+_{n,\beta}(A) = \frac{\big[(T^n\varphi)(1)\big]^2}{\big[(T^n\varphi)(1)\big]^2 + \big[(T^n\varphi)(-1)\big]^2}. \tag{2.48}$$

Usiamo ora un po' di algebra lineare. La matrice T definita in (2.46) ha come autovalori $\lambda_1 = 2\cosh(\beta)$, $\lambda_2 = 2\sinh(\beta)$, corrispondenti agli autovettori $v_1 = \binom{1}{1}$ e $v_2 = \binom{1}{-1}$. Identificando vettori e funzioni come sopra indicato, possiamo esprimere la funzione φ come

$$\varphi \;=\; \cosh(\beta)\, v_1 - \sinh(\beta)\, v_2\,,$$

da cui, usando la linearità di T, si ottiene

$$(T^n\varphi)(\pm 1) \;=\; 2^n \cosh^{n+1}(\beta) \pm 2^n \sinh^{n+1}(\beta)\,.$$

In conclusione, abbiamo calcolato l'espressione *esatta* di $\mu^+_{n,\beta}(A)$:

$$\mu^+_{n,\beta}(A) \;=\; \frac{\big[\cosh^{n+1}(\beta) + \sinh^{n+1}(\beta)\big]^2}{\big[\cosh^{n+1}(\beta) + \sinh^{n+1}(\beta)\big]^2 + \big[\cosh^{n+1}(\beta) - \sinh^{n+1}(\beta)\big]^2}\,.$$

Si noti che effettivamente $\mu^+_{n,\beta}(A) > \frac{1}{2}$, per ogni valore fissato di $n \in \mathbb{N}$ e $\beta > 0$ (mentre $\mu^+_{n,\beta}(A) = \frac{1}{2}$ per ogni $n \in \mathbb{N}$, se $\beta = 0$). Dato che $\cosh(\beta) > \sinh(\beta) > 0$ per ogni $\beta > 0$, lasciamo al lettore il compito di dedurre dalla formula precedente che, per ogni $\beta > 0$, si ha

$$\lim_{n\to+\infty} \mu^+_{n,\beta}(A) \;=\; \frac{1}{2}\,.$$

Questo mostra che in dimensione 1 non c'è magnetizzazione spontanea per nessun $\beta > 0$.

2.5.2 Il caso $d = 2$

In dimensione due non tenteremo di effettuare calcoli esatti con il modello di Ising. In realtà, molti calcoli esatti sono possibili: la loro complessità va però al di là del livello di questo libro. Dimostreremo l'esistenza di magnetizzazione spontanea in $d = 2$ (a basse temperature) mediante un argomento geometrico-combinatorio semplice ed efficace, generalizzabile a molti modelli più complessi: si tratta del celebre *argomento di Peierls*.

In questo caso il reticolo $\Lambda_n = \{-n, -n+1, \dots, n-1, n\}^2$ è formato dai punti a coordinate intere del quadrato di lato $2n$ avente l'origine al centro. Poniamo $\Omega_n := \{-1, 1\}^{\Lambda_n}$. Per comodità di calcolo, conviene modificare leggermente la definizione dell'energia, ponendo

$$\widetilde{H}_n^+(\sigma) := -\sum_{\substack{x,y\in\Lambda_n \\ |x-y|=1}} (\sigma_x\sigma_y - 1) - \sum_{x\in\partial\Lambda_n} (\sigma_x - 1). \tag{2.49}$$

Si noti che, con riferimento all'energia originale H_n^+ definita in (2.41), per ogni $\sigma \in \Omega$ si ha $\widetilde{H}_n^+(\sigma) = H_n^+(\sigma) + c_n$, dove $c_n = |\{x, y \in \Lambda_n : |x-y| = 1\}| + |\partial\Lambda_n|$ è una costante che non dipende da σ. Di conseguenza possiamo scrivere

$$\mu_{n,\beta}^+(\sigma) = \frac{1}{Z_{n,\beta}^+} \exp\left[-\beta H_n^+(\sigma)\right] = \frac{1}{\widetilde{Z}_{n,\beta}^+} \exp\left[-\beta \widetilde{H}_n^+(\sigma)\right],$$

dove

$$\widetilde{Z}_{n,\beta}^+ := \sum_{\sigma\in\Omega_n} \exp\left[-\beta \widetilde{H}_n^+(\sigma)\right]. \tag{2.50}$$

In altre parole, la nuova energia $\widetilde{H}_n^+$ è fisicamente equivalente a quella originale, cioè determina *la stessa misura di Gibbs*. La ragione per introdurre l'energia $\widetilde{H}_n^+$ è che essa si può riscrivere come

$$\widetilde{H}_n^+(\sigma) = 2\,|\{(x,y) : |x-y| = 1 \text{ e } \sigma_x \neq \sigma_y\}| + 2\,|\{x \in \partial\Lambda_n : \sigma_x = -1\}|. \tag{2.51}$$

Vedremo tra poco l'utilità di tale espressione.

Data una configurazione di spin σ su Λ_n, completiamola ad una configurazione di spin su tutto $\mathbb{Z}^2$ assegnando spin $+1$ a tutti i punti esterni a Λ_n (in realtà, per la costruzione che segue, è sufficiente assegnare spin $+1$ ai punti di Λ_n^c che distano 1 da qualche punto di Λ_n). Per ogni coppia di punti x, y tali che $|x-y| = 1$ e $\sigma_x \neq \sigma_y$ disegniamo quindi nel piano un segmento di lunghezza 1, ortogonale al segmento congiungente x e y, il cui punto medio sia $\frac{x+y}{2}$. La figura ottenuta dall'unione di tutti i segmenti disegnati è detta *contour* (si veda la Figura 2.1 per un esempio).

Introduciamo una notazione importante: definiamo *poligonale chiusa autoevitante* (p.c.a.) l'unione $\bigcup_{i=1}^{k-1} \overline{P_i P_{i+1}}$ dei segmenti che congiungono in successione k punti $P_1, \dots, P_k$ del piano, dove $P_k = P_1$ (poligonale *chiusa*), $P_i \neq P_j$ se $\{i, j\} \neq \{1, k\}$ (poligonale *autoevitante*) e inoltre $P_i = (x_i \pm \frac{1}{2}, y_i \pm \frac{1}{2})$ con $(x_i, y_i) \in \Lambda_n$ e $|P_{i+1} - P_i| = 1$,

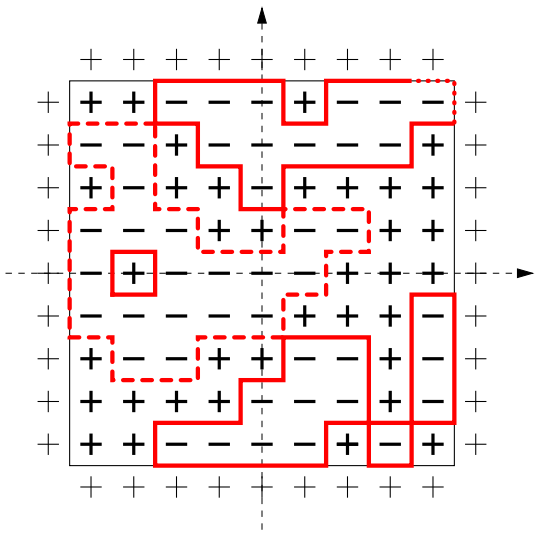


Figura 2.1 Una configurazione di spin σ per il modello di Ising nel piano, sul reticolo $\Lambda_n = \{-n,\dots,n\}^2$ con $n = 4$, con condizioni al bordo positive. In rosso è tracciato il contour $\mathscr{C}$ corrispondente. La parte di contour tratteggiata è una poligonale chiusa autoevitante che contiene l'origine. I due segmenti a puntini, nell'angolo in alto a destra, danno contributo 1 alla lunghezza $\ell(\mathscr{C})$ del contour (che in questo caso è pari a 83).

per ogni $i = 1,\dots,k$. Un esempio di poligonale chiusa autoevitante è tratteggiato in rosso nella Figura 2.1.

Il punto fondamentale è che ogni contour $\mathscr{C}$ si può sempre scrivere come unione $\bigcup_{i=1}^m \gamma_i$ di p.a.c. *disgiunte*, dove, con leggero abuso di notazione, intendiamo disgiunte anche due poligonali che si intersecano in un numero finito di punti. Viceversa, un'unione di p.a.c. disgiunte è sempre un contour ammissibile, cioè esiste una configurazione di spin $\sigma \in \Omega_n$ che lo determina, e tale configurazione è unica[1]. Una dimostrazione formale di queste affermazioni non è difficile ma è piuttosto lunga e noiosa e sarà pertanto omessa[2]. Abbiamo ottenuto una caratterizzazione esplicita dell'insieme dei contour ammissibili, che indicheremo con Ξ_n, che è in corrispondenza biunivoca con lo spazio delle configurazioni Ω_n. Sottolineiamo che la decomposizione $\mathscr{C} = \bigcup_{i=1}^m \gamma_i$ con γ_i p.c.a. disgiunte in generale non è unica (si veda ad esempio la Figura 2.1).

Da qui in poi identificheremo una configurazione σ di spin con il contour $\mathscr{C}$ corrispondente: in particolare, scriveremo $\mu^+_{n,\beta}(\{\mathscr{C}\})$, $\widetilde{H}^+_n(\mathscr{C})$, ecc. Definiamo la

[1] La configurazione si costruisce assegnando il valore $+$ (a causa delle condizioni al bordo positive) agli spin "esterni", cioè che non sono racchiusi da alcuna poligonale; quindi assegnando il valore $-$ agli spin dentro le poligonali che "confinano" con spin di valore $+$, e così via.

[2] Il punto fondamentale è il seguente: per come è costruito un contour, da ogni punto $(x \pm \frac{1}{2}, y \pm \frac{1}{2})$ con $(x,y) \in \Lambda_n$ partono necessariamente 0, 2 oppure 4 segmenti del contour. È questa proprietà che permette di decomporre il contour in poligonali chiuse autoevitanti.

lunghezza $\ell(\mathscr{C})$ di un contour $\mathscr{C}$ come il numero dei segmenti di lunghezza 1 che lo compongono (lunghezza geometrica), eccetto nel caso in cui una o più coppie di segmenti del contour descrivano uno dei 4 angoli del reticolo Λ_n: in questo caso conveniamo che ciascuna coppia di tali segmenti dia contributo 1 alla lunghezza $\ell(\mathscr{C})$ (si veda la Figura 2.1). Con queste convenzioni, ricordando la relazione (2.51) si ottiene la rappresentazione basilare

$$\widetilde{H}_n^+(\mathscr{C}) \;=\; 2\,\ell(\mathscr{C})\,.$$

Possiamo quindi scrivere

$$\mu_{n,\beta}^+(\{\mathscr{C}\}) \;=\; \frac{1}{\widetilde{Z}_n^+} e^{-2\beta\,\ell(\mathscr{C})}\,, \qquad \widetilde{Z}_{n,\beta}^+ \;=\; \sum_{\mathscr{C}'\in\Xi_n} e^{-2\beta\,\ell(\mathscr{C}')}\,. \tag{2.52}$$

Sia ora $\mathscr{C}$ un contour corrispondente a una configurazione di spin σ in cui $\sigma_0 = -1$. Dato che l'"isola" di spin -1 a cui l'origine appartiene dev'essere separata dal "mare" di spin $+1$ che ricopre tutto Λ_n^c, il contour $\mathscr{C}$ dev'essere della forma $\gamma\cup\mathscr{C}'$, dove γ è una p.c.a. avente l'origine al suo interno e $\mathscr{C}'$ è un contour tale che $\gamma\cap\mathscr{C}' = \emptyset$ (o, più precisamente, l'intersezione deve consistere di un numero finito di punti). Ricordando la definizione (2.42) dell'evento A, in termini dei contour vale la seguente inclusione:

$$A^c \;\subseteq\; \big\{\mathscr{C}\in\Xi_n :\; \mathscr{C} \,=\, \gamma\cup\mathscr{C}'\,,\; \gamma \text{ p.c.a. che racchiude } 0\,,\; \mathscr{C}'\in\Xi_n\,,\; \gamma\cap\mathscr{C}' = \emptyset\big\}\,.$$

Si noti che in generale la decomposizione $\mathscr{C} = \gamma\cup\mathscr{C}'$ non è unica (si veda ancora la Figura 2.1) ma questo non sarà un problema. Osserviamo anche che, essendo γ e $\mathscr{C}'$ disgiunti, si ha chiaramente $\ell(\gamma\cup\mathscr{C}') = \ell(\gamma) + \ell(\mathscr{C}')$. Possiamo dunque scrivere

$$\begin{aligned}\mu_{n,\beta}^+(A^c) \;&=\; \frac{1}{\widetilde{Z}_{n,\beta}^+} \sum_{\mathscr{C}\in A^c} e^{-2\beta\,\ell(\mathscr{C})} \;\le\; \frac{1}{\widetilde{Z}_{n,\beta}^+} \sum_{\substack{\gamma \text{ p.c.a.}\\ \text{che racchiude } 0}} e^{-2\beta\,\ell(\gamma)} \sum_{\substack{\mathscr{C}'\in\Xi_n\\ \gamma\cap\mathscr{C}'=\emptyset}} e^{-2\beta\ell(\mathscr{C}')} \\ &\le \sum_{\substack{\gamma \text{ p.c.a.}\\ \text{che racchiude } 0}} e^{-2\beta\,\ell(\gamma)} \left(\frac{1}{\widetilde{Z}_{n,\beta}^+} \sum_{\mathscr{C}'\in\Xi_n} e^{-2\beta\ell(\mathscr{C}')}\right) \;=\; \sum_{\substack{\gamma \text{ p.c.a.}\\ \text{che racchiude } 0}} e^{-2\beta\,\ell(\gamma)}\,,\end{aligned}$$

dove si è usata la seconda relazione in (2.52). Disintegrando rispetto ai possibili valori di $\ell(\gamma)$ otteniamo

$$\mu_n^+(A^c) \le \sum_{m=1}^{\infty} K_m\, e^{-2\beta\, m}\,, \qquad \text{dove } K_m := \big|\{\gamma \text{ p.c.a. che racchiude } 0 :\; \ell(\gamma) = m\}\big|\,.$$

Ci resta da stimare K_m. Se γ è una poligonale chiusa autoevitante con $\ell(\gamma) = m$, la sua lunghezza geometrica è compresa tra m e $m+4$; se γ racchiude 0 al suo interno, necessariamente γ è interamente contenuta nel quadrato $Q := [-\frac{m+4}{2}, \frac{m+4}{2}] \times [-\frac{m+4}{2}, \frac{m+4}{2}]$. Costruiamo ora una curva nel modo seguente:

- scegliamo un punto in Q della forma $(x\pm\frac{1}{2}, y\pm\frac{1}{2})$, con $x, y\in\mathbb{Z}$ (per questa scelta

abbiamo al massimo $(\frac{m+4}{2}+\frac{m+4}{2}+1)^2=(m+5)^2$ possibilità);
- scegliamo una delle quattro direzioni possibili e tracciamo un segmento di lunghezza uno in quella direzione;
- a questo punto, per il tratto successivo, scegliamo una delle tre direzioni che non ci fanno tornare al punto da cui proveniamo;
- iteriamo la procedura per l passi.

Tra le curve costruite in questo modo in un numero di passi compreso tra m e $m+4$ ci sono in particolare tutte le possibili p.c.a. γ con $\ell(\gamma)=m$. Di conseguenza

$$K_m \;\le\; \sum_{l=m}^{m+4}(m+5)^2\,4\,3^{l-1} \;\le\; 5\cdot((m+5)^2\,4\,3^{m+3})\,.$$

Semplifichiamo questa espressione con la stima (molto rozza) $x^2\le 3^x$ per ogni $x\in\mathbb{N}$:

$$K_m \;\le\; C\cdot 9^m\,, \qquad \text{dove } C \;:=\; 3^8\cdot 20\,.$$

In questo modo otteniamo.

$$\mu^+_{n,\beta}(A^c) \;\le\; C\sum_{m=1}^{+\infty} 9^m\,e^{-2\beta m} \;=\; C\sum_{m=1}^{+\infty} e^{-c(\beta)m} \;=\; C\,\frac{e^{-c(\beta)}}{1-e^{-c(\beta)}}\,,$$

dove abbiamo posto $c(\beta):=2\beta-\log 9$. Si noti che la stima ottenuta *non dipende da n*. Visto che $\lim_{\beta\to\infty}c(\beta)=+\infty$ e dato che C è una costante fissata, segue che[3] esiste $\beta_0\in(0,\infty)$ tale che per ogni $\beta>\beta_0$ e per ogni $n\in\mathbb{N}$ si ha

$$\mu^+_{n,\beta}(A^c) \;\le\; \frac{1}{4}\,,$$

ovvero

$$\mu^+_{n,\beta}(A) \;\ge\; \frac{3}{4}\,.$$

Abbiamo dunque mostrato che, per grandi valori di β (cioè a temperatura sufficientemente bassa) nel modello di Ising in dimensione 2 ha luogo il fenomeno della magnetizzazione spontanea. È possibile mostrare (non lo faremo) che, al contrario, per valori piccoli di β *non* si ha magnetizzazione spontanea.

2.6 Il modello di Hardy-Weinberg in genetica

Le nozioni di probabilità condizionata e indipendenza permettono, come vedremo in questo paragrafo, la formulazione e l'analisi rigorosa di un noto modello per la trasmissione ereditaria dei caratteri, in *modello di Hardy*. Consideriamo una popolazione d una determinata specie, e supponiamo di essere interessati ad uno specifico

[3] In effetti abbiamo dimostrato un'affermazione più forte: per ogni $\eta>0$ esiste $\beta_0(\eta)\in(0,\infty)$ tale che per ogni $\beta>\beta_0(\eta)$ e per ogni $n\in\mathbb{N}$ si ha $\mu^+_n(A)>1-\eta$.

carattere. Assumiamo che questo carattere si manifesto in solo due modi, che chiameremo *dominante* e *recessivo*, e che esso sia caratterizzato da una coppia (non ordinata) di *geni*, ognuno dei quali può assumere uno dei due valori a, A. Dunque, ogni individuo della specie possiede una delle tre possibili coppie di geni AA, aA e aa. Nel terzo caso il carattere manifestato sarà quello recessivo, negli altri due quello dominante. Le tre coppie AA, aA e aa sono chiamate *genotipi*.

Il modello di Hardy ha per oggetto l'evoluzione della frequenza dei genotipi in popolazioni sessuate. In un determinato istante consideriamo gli individui della popolazione, che chiameremo di *generazione* 0. Assumiamo che tale popolazione sia numerosa, usiamo le seguenti notazioni:

- $u \in (0,1)$ è la *frazione* di individui della generazione 0 di genotipo AA;
- $2v \in (0,1)$ è la *frazione* di individui della generazione 0 di genotipo aA
- $w \in (0,1)$ è la *frazione* di individui della generazione 0 di genotipo aa;

Evidentemente $u + 2v + w = 1$. Tali numeri hanno un'evidente interpretazione probabilistica: se si sceglie *a caso* un individuo della generazione 0, la probabilità che esso sia di genotipo AA (risp. aA, aa) è u (risp. $2v$, w).

Consideriamo in un istante successivo, gli individui che sono stati generati da una coppia della generazione 0; l'insieme di essi verrà chiamata *generazione* 1. Supponiamo di scegliere a caso un individuo della generazione 1, e consideriamo, per $g \in \{AA, aA, aa\}$, gli eventi

$$D_g := \text{“l'individuo è di genotipo } g\text{“}.$$

Siano

$$\begin{aligned} u_1 &= P(D_{AA}) \\ 2v_1 &= P(D_{aA}) \\ w_1 &= P(D_{aa}). \end{aligned}$$

In altre parole, $u_1, 2v_1$ e w_1 sono le frequenze dei tre genotipi nella generazione 1. Introduciamo anche i seguenti eventi, per $g \in \{AA, aA, aa\}$ e $x \in \{A, a\}$:

$F_{g,x}$:= “l'individuo scelto ha padre di genotipo g, e ha da lui ereditato il gene x”
$M_{g,x}$:= “l'individuo scelto ha madre di genotipo g, e ha da lei ereditato il gene x”.

Inoltre poniamo

$$F_g := \text{“l'individuo scelto ha padre di genotipo } g\text{”} = F_{g,A} \cup F_{g,a},$$

e analogamente per M_g. Il modello di Hardy è definito dalle seguenti ipotesi.

(i) Per ogni scelta di $g, h \in \{AA, aA, aa\}$ e $x, y \in \{A, a\}$, gli eventi $F_{g,x}$ e $M_{h,y}$ sono indipendenti.

(ii)

$$P(F_{AA}) = P(M_{AA}) = u \; P(F_{Aa}) = P(M_{Aa}) = 2v \; P(F_{aa}) = P(M_{aa}) = w.$$

(iii)

$$P(F_{AA,A}|F_{AA}) = P(M_{AA,A}|M_{AA}) = 1 \; P(F_{aA,A}|F_{aA}) = P(M_{aA,A}|M_{aA}) = \tfrac{1}{2} \; P(F_{aa,A}) = 0$$

Le condizioni (i) e (ii) sono dette di *accoppiamento casuale*, e implicano in particolare che la fertilità è indipendente dal sesso e dal genotipo, e che la frequenza dei genotipi non dipende dal sesso. La condizione (iii) è invece una semplice istanza delle *Leggi di Mendel*.

Lo scopo di quanto segue è quello di mostrare che le ipotesi precedenti implicano una relazione funzionale tra le frequenze dei genotipi nella generazione 0 e quelle nella generazione 1. Si noti anzitutto che

$$D_{AA} = [F_{AA,A} \cap M_{AA,A}] \cup [F_{aA,A} \cap M_{AA,A}] \cup [F_{AA,A} \cap M_{aA,A}] \cup [F_{aA,A} \cap M_{aA,A}] \, .$$

Evidentemente, le unioni precedenti sono tra eventi disgiunti. Inoltre ognuna delle intersezioni nella formula precedente sono, per l'ipotesi (i), tra eventi indipendenti. Pertanto:

$$u_1 = P(D_{AA}) = P(F_{AA,A})P(M_{AA,A}) + P(F_{aA,A})P(M_{AA,A}) + P(F_{AA,A})P(M_{aA,A}) + P(F_{aA,A})P(M_{aA,A}). \tag{2.53}$$

Inoltre, per l'ipotesi (iii)

$$P(F_{AA,A}) = P(M_{AA,A}) = P(F_{AA}) = u \; P(F_{aA,A}) = P(M_{aA,A}) = \tfrac{1}{2}P(F_{aA}) = v.$$

Perciò

$$u_1 = u^2 + 2uv + v^2 = (u+v)^2.$$

Per simmetria

$$w_1 = (v+w)^2.$$

Essendo

$$u_1 + w_1 = (u+v)^2 + (v+w)^2 = (u+2v+w)^2 - 2(u+v)(v+w) = 1 - 2(u+v)(v+w)$$

si ha

$$2v_1 = 1 - u_1 - w_1 = 2(u+v)(v+w).$$

Quindi, se definiamo la funzione di tre variabili

$$T(u,v,w) = ((u+v)^2, (u+v)(v+w), (v+w)^2),$$

abbiamo ottenuto la relazione

$$(u_1, v_1, w_1) = T(u,v,w).$$

In altre parole, la mappa T fornisce la relazione tra le frequenze dei genotipi di una generazione e quelle della successiva. A questo punto possiamo anche affermare che le frequenze dei genotipi nella generazione successiva alla generazione 1, diciamo la generazione 2, sono date da

$$(u_2, v_2, w_2) = T(u_1, v_1, w_1) = T(T(u,v,w))$$

Si osservi che la prima componente di $T(T(u,v,w))$ è

$$[(u+v)^2+(u+v)(v+w)]^2=[(u+v)(u+2v+w)]^2=(u+v)^2,$$

essendo $u+2v+w=1$, e quindi è uguale alla prima componente di $T(u,v,w)$. In modo analogo si mostra che la terza componente di $T(T(u,v,w))$ coincide con la terza componente di $T(u,v,w)$. Pertanto, essendo $u_2+2v_2+w_2=u_1+2v_1+w_1=1$, possiamo concludere che

$$T(T(u,v,w))=T(u,v,w).$$

Questo implica che *la frequenza dei genotipi dalla generazione* 1 *in poi rimane costante*: la popolazione raggiunge un equilibrio dopo una sola generazione!

Quello che abbiamo fin qui descritto è la versione standard del modello di Hardy-Weinberg. Sono state proposte numerose modifiche e generalizzazioni, per tener conto di fenomeni quali la dipendenza dal sesso dei caratteri e la selezione naturale. Vedremo ora brevemente una di queste versioni modificate, che rappresenta un semplice modello per la *selezione naturale*. Questo modello si ottiene dal precedente modificando l'ipotesi (ii) come segue:

(ii)')

$$P(F_{AA})=P(M_{AA})=\tfrac{u}{u+2v}\ P(F_{Aa})=P(M_{Aa})=\tfrac{2v}{u+2v}\ P(F_{aa})=P(M_{aa})=0.$$

Il significato di (ii)' è evidente: gli individui che manifestano il carattere recessivo, cioè di genotipo *aa* non si riproducono. Lasciamo al lettore controllare che la relazione in (2.53) diventa:

$$u_1=\left(\frac{u}{u+2v}\right)^2+\left(\frac{u}{u+2v}\right)\left(\frac{v}{u+2v}\right)+\left(\frac{v}{u+2v}\right)^2=\left(\frac{u+v}{u+2v}\right)^2.$$

Analogamente

$$w_1=\left(\frac{v}{u+2v}\right)^2,\ \ v_1=\frac{v(u+v)}{(u+2v)^2}.$$

Posto, come prima, $(u_1,v_1,w_1)=:T(u,v,w)$, possiamo definire induttivamente

$$(u_n,v_n,w_n)=T(u_{n-1},v_{n-1},w_{n-1}). \tag{2.54}$$

Ne segue che $u_n, 2v_n$ e w_n sono le frequenze dei tre genotipi nell'n-esima generazione. Con un calcolo paziente ma elementare, si verifica per induzione che

$$\begin{aligned} u_n &= \left(\tfrac{u+nv}{u+(n+1)v}\right)^2 \\ v_n &= \tfrac{v(u+nv)}{(u+(n+1)v)^2} \\ w_n &= \tfrac{v^2}{(u+(n+1)v)^2}. \end{aligned}$$

Notare che asintoticamente in n

$$v_n \simeq \frac{1}{n+1} \quad w_n \simeq \frac{1}{(n+1)^2}.$$

Abbiamo dunque quantificato, come effetto della selezione naturale, la progressiva diminuzione dei genotipi che contengono il gene a.