CAPITOLO

14

Probabilità e calcolo combinatorio

Contare bene è metà del lavoro; l'altra metà è decidere che cosa contare. (testo della redazione)

OBIETTIVI

1. Riconoscere gli assiomi di Kolmogorov come quadro comune alle definizioni di probabilità e dedurne i primi teoremi.
2. Calcolare probabilità classiche con il calcolo combinatorio: urne con e senza reimmissione, carte, lotto, compleanni.
3. Usare la probabilità composta di più eventi e l'evento contrario per i problemi del tipo «almeno uno».
4. Ricavare e applicare la formula di Bernoulli per le prove ripetute e individuare il numero più probabile di successi.
5. Collegare probabilità e frequenza con la legge dei grandi numeri e riconoscere le ipotesi di un modello probabilistico.

INDICE DEL CAPITOLO

14.1 Dalle tre definizioni agli assiomi 2

14.2 Probabilità classica e calcolo combinatorio 6

14.3 Probabilità composta ed evento contrario 10

14.4 Le prove ripetute: lo schema di Bernoulli 13

14.5 Probabilità e frequenza 18

Matematica 4 · Parte V · Calcolo combinatorio e probabilità · versione 1.0
Capitolo 14: Probabilità e calcolo combinatorio · pagine 391–416 del volume completo · estratto per lo studio

In *Matematica 2* abbiamo calcolato le prime probabilità contando a mano: le facce di un dado, le coppie di risultati di due dadi, i rami di un albero con tre lanci di una moneta. Bastava elencare i casi possibili e quelli favorevoli. Ma quanto vale la probabilità di ricevere una coppia in una mano di cinque carte da poker? I casi possibili sono più di due milioni e mezzo, e nessuno li elencherebbe uno per uno. Qui entra in gioco il calcolo combinatorio del capitolo precedente: le disposizioni, le permutazioni e le combinazioni ci permettono di *contare senza elencare*, e quindi di calcolare probabilità che a mano sarebbero irraggiungibili.

In questo capitolo riprendiamo le regole della probabilità e le ordiniamo intorno a pochi assiomi, dai quali seguono tutte le altre; poi le applichiamo alle estrazioni da un'urna, ai giochi di carte, al lotto e al celebre problema dei compleanni. Infine studiamo le **prove ripetute**: un esperimento con due soli esiti, ripetuto più volte nelle stesse condizioni, come una serie di tiri liberi, il controllo dei pezzi prodotti da una macchina o la nascita dei figli di una coppia. La formula che ne ricaveremo, dovuta a Jakob Bernoulli, unisce il coefficiente binomiale e la regola del prodotto, e ci porterà a rileggere il legame tra probabilità e frequenza.

14.1 Dalle tre definizioni agli assiomi

14.1.1 Che cosa sappiamo già

Un **esperimento aleatorio** è un esperimento il cui risultato non si può prevedere con certezza; l'insieme dei suoi esiti è lo **spazio campionario** Ω e ogni sottoinsieme di Ω è un **evento**. Sugli eventi si opera come sugli insiemi, e alla probabilità di un evento si dà un valore con una delle tre definizioni studiate in *Matematica 2* (capitolo 14): classica, frequentista o soggettiva. La Tabella 14.1 riassume ciò che useremo; se qualche riga non ti è chiara, rileggi il capitolo indicato prima di proseguire.

Tabella 14.1 Richiamo dei concetti di base (vedi *Matematica 2*, capitolo 14).

Concetto	Notazione o regola	Esempio (lancio di un dado)
evento contrario	$\overline{E}$: si verifica se e solo se non si verifica E	E = «esce 6», $\overline{E}$ = «esce un numero da 1 a 5»
unione, intersezione	$A \cup B$: almeno uno dei due; $A \cap B$: entrambi	«pari» $\cup$ «maggiore di 3» = {2, 4, 5, 6}
eventi incompatibili	$A \cap B = \varnothing$	«esce 1» e «esce un numero pari»
definizione classica	$p(E) = \dfrac{\text{casi favorevoli}}{\text{casi possibili}}$, esiti equiprobabili	$p(\text{«pari»}) = \dfrac{3}{6} = \dfrac{1}{2}$
definizione frequentista	$p(E) \approx$ frequenza relativa su molte prove	su 6000 lanci, circa 1000 volte 6
definizione soggettiva	prezzo equo di una scommessa che paga 1 se E si verifica	scommessa sulla vittoria di una squadra
probabilità condizionata	$p(A\|B) = \dfrac{p(A \cap B)}{p(B)}$, con $p(B) > 0$	$p(6 \mid \text{«pari»}) = \dfrac{1}{3}$
probabilità del prodotto	$p(A \cap B) = p(A) \cdot p(B\|A)$	due estrazioni successive da un'urna
eventi indipendenti	$p(A \cap B) = p(A) \cdot p(B)$	un dado e una moneta lanciati insieme

Le formule della tabella ci basteranno per quasi tutto il capitolo. Prima però vale la pena di chiederci *perché* valgono per tutte e tre le definizioni, che nascono da idee così diverse.

14.1.2 Gli assiomi di Kolmogorov

La definizione classica richiede esiti equiprobabili, quella frequentista richiede un esperimento ripetibile, quella soggettiva un'opinione coerente. Sono tre risposte diverse alla domanda «che cos'è la probabilità?». Nel 1933 il matematico russo Andrej N. Kolmogorov propose di cambiare domanda: invece di dire che cosa la probabilità *sia*, stabiliamo quali *regole* deve rispettare. Le regole scelte come punto di partenza, senza dimostrazione, sono gli **assiomi della probabilità**; tutto il resto si deduce da essi, come nella geometria di Euclide i teoremi si deducono dai postulati. Questa è l'**impostazione assiomatica** del calcolo delle probabilità. Nella forma adatta agli spazi campionari finiti, che sono gli unici che useremo, essa dice:

IN SINTESI

Una **probabilità** su uno spazio campionario finito Ω è una funzione che associa a ogni evento $E \subseteq \Omega$ un numero reale $p(E)$ in modo che:

- **assioma 1**: $p(E) \geq 0$ per ogni evento E;
- **assioma 2**: $p(\Omega) = 1$;
- **assioma 3**: se $A \cap B = \varnothing$, allora $p(A \cup B) = p(A) + p(B)$.

Rispetto all'elenco visto in *Matematica 2* abbiamo tolto la condizione $p(E) \leq 1$: non serve assumerla, perché la dimostreremo. È una caratteristica del metodo assiomatico: si chiede il minimo indispensabile e si dimostra tutto il resto.

Le tre definizioni rispettano gli assiomi. Per quella classica è facile verificarlo: se Ω ha n esiti equiprobabili ed E ne contiene m, allora $p(E) = m/n \geq 0$, $p(\Omega) = n/n = 1$, e se A e B sono incompatibili gli esiti di $A \cup B$ sono quelli di A più quelli di B, senza ripetizioni, quindi le frazioni si sommano. Lo stesso ragionamento vale per le frequenze relative, che sono anch'esse rapporti fra conteggi; per la definizione soggettiva, la coerenza delle scommesse impone proprio queste tre regole. Gli assiomi sono dunque il *quadro comune* che contiene le tre definizioni: ciò che dimostriamo a partire da essi vale qualunque sia il modo in cui abbiamo assegnato le probabilità.

14.1.3 I primi teoremi

Vediamo come dagli assiomi si ricavano le regole che conosciamo. Ogni dimostrazione usa solo i tre assiomi e le proprietà degli insiemi.

Evento contrario. E ed $\overline{E}$ sono incompatibili e $E \cup \overline{E} = \Omega$. Per l'assioma 3 e l'assioma 2, $p(E) + p(\overline{E}) = p(\Omega) = 1$, quindi $p(\overline{E}) = 1 - p(E)$. In particolare, poiché $\overline{\Omega} = \varnothing$, si ha $p(\varnothing) = 1 - 1 = 0$: l'**evento impossibile** ha probabilità nulla. c.v.d.

Probabilità al più 1. Per l'assioma 1 applicato a $\overline{E}$, $1 - p(E) = p(\overline{E}) \geq 0$, cioè $p(E) \leq 1$. c.v.d.

Monotonia. Se $A \subseteq B$, l'evento B si scompone nei due eventi incompatibili A e $B \setminus A$ (gli esiti di B che non stanno in A). Per l'assioma 3, $p(B) = p(A) + p(B \setminus A)$, e poiché $p(B \setminus A) \geq 0$ si ottiene $p(A) \leq p(B)$. È la **monotonia** della probabilità: se ogni volta che si verifica A si verifica anche B, allora B non può essere meno probabile di A. c.v.d.

Unione di due eventi qualsiasi. Se A e B sono compatibili, l'assioma 3 non si applica direttamente. Scomponiamo allora, come nella Figura 14.1, l'unione in tre parti incompatibili a due a due: $A \setminus B$, $A \cap B$ e $B \setminus A$. Si ha

$$p(A \cup B) = p(A \setminus B) + p(A \cap B) + p(B \setminus A).$$

Ma anche A è unione dei due eventi incompatibili $A \setminus B$ e $A \cap B$, quindi $p(A \setminus B) = p(A) - p(A \cap B)$; allo stesso modo $p(B \setminus A) = p(B) - p(A \cap B)$. Sostituendo,

$$p(A \cup B) = p(A) - p(A \cap B) + p(A \cap B) + p(B) - p(A \cap B) = p(A) + p(B) - p(A \cap B).$$

La probabilità dell'intersezione si sottrae una volta perché, sommando $p(A)$ e $p(B)$, gli esiti comuni sono stati contati due volte. c.v.d.

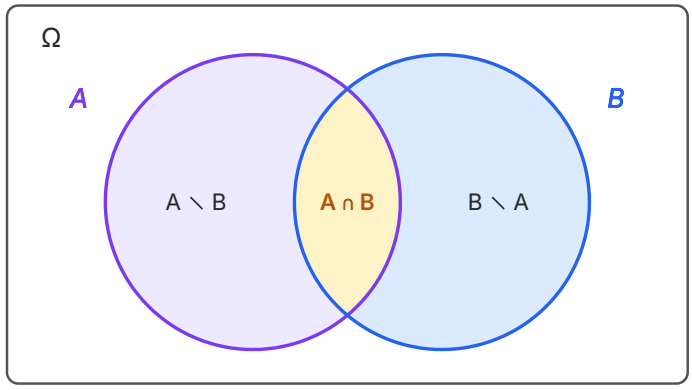


Figura 14.1 L'unione $A \cup B$ scomposta in tre eventi incompatibili: $A \setminus B$ (solo A), $A \cap B$ (entrambi) e $B \setminus A$ (solo B). Sommando $p(A)$ e $p(B)$, la zona centrale viene contata due volte.

IN SINTESI

Teoremi (conseguenze degli assiomi). Per eventi qualsiasi A, B, E di uno spazio finito Ω:

- $p(\overline{E}) = 1 - p(E)$ e $p(\varnothing) = 0$;
- $0 \le p(E) \le 1$;
- se $A \subseteq B$, allora $p(A) \le p(B)$ (monotonia);
- $p(A \cup B) = p(A) + p(B) - p(A \cap B)$ (teorema dell'unione).

ESEMPIO SVOLTO

Si lanciano due dadi. Calcoliamo la probabilità che la somma sia 7 oppure che esca almeno un 6.

Lo spazio campionario ha $6 \cdot 6 = 36$ coppie ordinate equiprobabili. Sia A = «la somma è 7»: le coppie sono (1; 6), (2; 5), (3; 4), (4; 3), (5; 2), (6; 1), quindi $p(A) = \frac{6}{36}$. Sia B = «esce almeno un 6»: conviene passare al contrario, «nessun 6», che ha $5 \cdot 5 = 25$ coppie, dunque $p(B) = 1 - \frac{25}{36} = \frac{11}{36}$.

Gli eventi sono compatibili: $A \cap B = \{(1; 6), (6; 1)\}$, con $p(A \cap B) = \frac{2}{36}$. Per il teorema dell'unione

$$p(A \cup B) = \frac{6}{36} + \frac{11}{36} - \frac{2}{36} = \frac{15}{36} = \frac{5}{12} \approx 0{,}417.$$

Sommando soltanto $\frac{6}{36} + \frac{11}{36} = \frac{17}{36}$ avremmo contato due volte le coppie (1; 6) e (6; 1).

14.2 Probabilità classica e calcolo combinatorio

14.2.1 Contare i casi senza elencarli

Con la definizione classica una probabilità è un rapporto fra due conteggi. Quando gli esiti sono pochi li elenchiamo; quando sono molti, li *contiamo* con gli strumenti del capitolo precedente. La scelta dello strumento dipende da come descriviamo gli esiti, e la regola d'oro è una sola: **casi favorevoli e casi possibili vanno contati con lo stesso modello**. Se negli esiti conta l'ordine, contiamo con le disposizioni sia gli uni sia gli altri; se l'ordine non conta, usiamo le combinazioni per entrambi. Mescolare i due modelli, per esempio favorevoli ordinati e possibili non ordinati, dà risultati sbagliati.

Il caso tipico è l'**urna**: un contenitore con palline uguali al tatto, di colori diversi, da cui si estrae «a caso», cioè in modo che ogni pallina abbia la stessa probabilità di essere scelta. L'urna è un modello astratto per molte situazioni reali: un campione di pezzi da una scatola, un gruppo di studenti sorteggiati, le carte servite da un mazzo ben mescolato.

14.2.2 Estrazioni senza reimmissione

Si ha un'**estrazione senza reimmissione** quando la pallina estratta non viene rimessa nell'urna, così che la composizione dell'urna cambia dopo ogni estrazione. Le palline si possono estrarre in due modi: tutte insieme, con un'**estrazione contemporanea** (una «manciata»), oppure una alla volta, con un'**estrazione successiva**. Nel primo caso un esito è un *gruppo* di palline, senza

ordine, e si conta con le combinazioni; nel secondo un esito è una *sequenza* ordinata, e si conta con le disposizioni.

ESEMPIO SVOLTO

Un'urna contiene 10 palline: 4 bianche e 6 nere. Se ne estraggono 3 senza reimmissione. Qual è la probabilità di ottenere esattamente 2 palline bianche?

Estrazione contemporanea. Immaginiamo le palline numerate da 1 a 10, così che siano tutte distinguibili. I casi possibili sono i gruppi di 3 palline scelte fra 10: $C_{10,3} = \dfrac{10 \cdot 9 \cdot 8}{3!} = 120$. Un gruppo favorevole si forma scegliendo 2 bianche fra 4 e 1 nera fra 6; per il principio fondamentale del calcolo combinatorio i gruppi favorevoli sono $C_{4,2} \cdot C_{6,1} = 6 \cdot 6 = 36$. Quindi

$$p = \frac{C_{4,2} \cdot C_{6,1}}{C_{10,3}} = \frac{36}{120} = \frac{3}{10}.$$

Estrazione successiva. Ora gli esiti sono le terne ordinate di palline distinte: $D_{10,3} = 10 \cdot 9 \cdot 8 = 720$. Per contare quelle favorevoli scegliamo prima la posizione della pallina nera (3 modi), poi riempiamo le due posizioni rimanenti con bianche diverse, in $4 \cdot 3$ modi, e la posizione della nera in 6 modi: $3 \cdot 4 \cdot 3 \cdot 6 = 216$ terne. Quindi $p = \dfrac{216}{720} = \dfrac{3}{10}$.

Il risultato è lo stesso. Non è una coincidenza: ogni gruppo di 3 palline corrisponde a esattamente $3! = 6$ terne ordinate, sia tra i casi possibili ($720 = 120 \cdot 6$) sia tra quelli favorevoli ($216 = 36 \cdot 6$), e nel rapporto il fattore 6 si semplifica. La Figura 14.2 mette a confronto i conteggi.

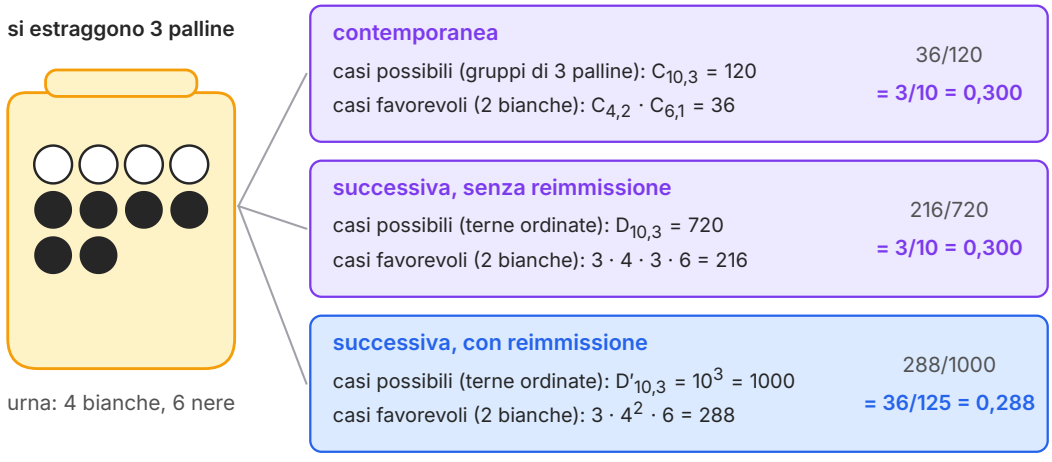


Figura 14.2 Tre modi di estrarre 3 palline da un'urna con 4 bianche e 6 nere, con i conteggi dei casi possibili e dei casi favorevoli all'evento «esattamente 2 bianche». Senza reimmissione, contemporanea e successiva danno la stessa probabilità; con reimmissione la probabilità cambia.

Il ragionamento dell'esempio vale in generale. Se un'urna contiene n palline, di cui b bianche e $n - b$ nere, e se ne estraggono m senza reimmissione, la probabilità di ottenere esattamente k bianche è

$$p(k \text{ bianche}) = \frac{C_{b,k} \cdot C_{n-b,m-k}}{C_{n,m}}, \quad \text{con } 0 \le k \le b \text{ e } 0 \le m - k \le n - b.$$

La formula vale sia per l'estrazione contemporanea sia per quella successiva: conviene quasi sempre usare le combinazioni, che danno numeri più piccoli. Nell'esempio, per $k = 0, 1, 2, 3$ si ottengono le probabilità $\frac{1}{6}$, $\frac{1}{2}$, $\frac{3}{10}$, $\frac{1}{30}$, la cui somma è 1, come deve essere, perché i quattro eventi sono incompatibili e la loro unione è Ω.

14.2.3 Estrazioni con reimmissione

Si ha un'**estrazione con reimmissione** quando ogni pallina, dopo essere stata osservata, viene rimessa nell'urna e le palline vengono rimescolate. Ora l'urna è identica prima di ogni estrazione, e la stessa pallina può uscire più volte. Gli esiti sono sequenze ordinate con eventuali ripetizioni, e si contano con le disposizioni con ripetizione.

Riprendiamo l'urna con 4 bianche e 6 nere ed estraiamo 3 palline con reimmissione. I casi possibili sono $D'_{10,3} = 10^3 = 1000$. Per le terne con esattamente 2 bianche: 3 modi di scegliere la posizione della nera, $4 \cdot 4$ modi per le due bianche (che ora possono anche essere la stessa pallina) e 6 modi per la nera, in tutto $3 \cdot 4^2 \cdot 6 = 288$. La probabilità è $\frac{288}{1000} = 0{,}288$, un po' minore di 0,3: senza reimmissione, dopo aver estratto una bianca, la nera successiva diventa leggermente più probabile, e il risultato «misto» ne guadagna. Torneremo su questo calcolo nel paragrafo sulle prove ripetute, dove lo ritroveremo come caso particolare di una formula generale.

14.2.4 Giochi di carte

Un mazzo di carte ben mescolato è un'urna in cui le «palline» sono le carte, e una mano servita è un'estrazione contemporanea senza reimmissione. In Italia si usano soprattutto due mazzi: il mazzo da 40 carte (4 semi di 10 carte ciascuno: asso, dal 2 al 7, e tre figure) e il mazzo francese da 52 carte (4 semi di 13 carte: asso, dal 2 al 10, fante, donna e re).

ESEMPIO SVOLTO

Si ricevono 5 carte da un mazzo di 52. Calcoliamo la probabilità di avere **a)** una coppia, cioè esattamente due carte dello stesso valore e le altre tre di valori diversi fra loro e dal primo; **b)** un full, cioè tre carte di un valore e due di un altro.

I casi possibili sono le mani di 5 carte: $C_{52,5} = \dfrac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}{5!} = 2\,598\,960$.

a) Costruiamo una mano favorevole per scelte successive: il valore della coppia (13 modi); i due semi della coppia ($C_{4,2} = 6$ modi); i tre valori diversi delle altre carte, scelti fra i 12 rimasti ($C_{12,3} = 220$ modi); il seme di ciascuna di esse ($4^3 = 64$ modi). In tutto $13 \cdot 6 \cdot 220 \cdot 64 = 1\,098\,240$ mani, e

$$p(\text{coppia}) = \frac{1\,098\,240}{2\,598\,960} = \frac{352}{833} \approx 0{,}423.$$

b) Valore del tris (13 modi) e suoi semi ($C_{4,3} = 4$); valore della coppia, diverso dal primo (12 modi), e suoi semi ($C_{4,2} = 6$): $13 \cdot 4 \cdot 12 \cdot 6 = 3744$ mani. Quindi $p(\text{full}) = \dfrac{3744}{2\,598\,960} = \dfrac{6}{4165} \approx 0{,}0014$: circa una mano ogni 700.

Nel punto b) i valori del tris e della coppia si scelgono con $13 \cdot 12$, non con $C_{13,2}$: i due valori hanno ruoli diversi (un full di re e assi non è un full di assi e re), quindi qui l'ordine conta. Nel punto a), invece, i tre valori «spaiati» hanno lo stesso ruolo e si scelgono con $C_{12,3}$.

14.2.5 Il lotto

Nel gioco del lotto, su ciascuna «ruota» si estraggono 5 numeri fra i 90 numeri da 1 a 90, senza reimmissione. Gli esiti sono le cinquine non ordinate: $C_{90,5} = 43\,949\,268$. Qual è la probabilità che un numero fissato, per esempio il 17, sia fra gli estratti? Le cinquine che contengono il 17 si ottengono scegliendo gli altri 4 numeri fra gli 89 rimasti, quindi

$$p(\text{estratto}) = \frac{C_{89,4}}{C_{90,5}} = \frac{5}{90} = \frac{1}{18},$$

come è ragionevole: 5 numeri su 90. Per un **ambo**, cioè due numeri fissati entrambi estratti, le cinquine favorevoli sono $C_{88,3}$ e

$$p(\text{ambo}) = \frac{C_{88,3}}{C_{90,5}} = \frac{5 \cdot 4}{90 \cdot 89} = \frac{2}{801} \approx 0{,}0025.$$

Per un terno si trova $\dfrac{C_{87,2}}{C_{90,5}} = \dfrac{1}{11\,748}$. Queste probabilità restano le stesse a ogni estrazione, qualunque cosa sia uscita nelle settimane precedenti: le estrazioni sono indipendenti, e i numeri «ritardatari» non hanno alcun vantaggio (è la fallacia del giocatore di *Matematica 2*).

14.3 Probabilità composta ed evento contrario

14.3.1 La regola del prodotto per più eventi

In *Matematica 2* la probabilità del prodotto di due eventi era $p(A \cap B) = p(A) \cdot p(B|A)$. Applicando la stessa regola più volte si ottiene la **probabilità composta** di n eventi:

$$p(A_1 \cap A_2 \cap \ldots \cap A_n) = p(A_1) \cdot p(A_2 \mid A_1) \cdot p(A_3 \mid A_1 \cap A_2) \cdot \ldots \cdot p(A_n \mid A_1 \cap \ldots \cap A_{n-1}).$$

Ogni fattore è la probabilità di un evento *sapendo* che si sono verificati tutti i precedenti. Sul diagramma ad albero, è il prodotto delle probabilità scritte sui rami di un percorso. Per esempio, estraendo successivamente 3 carte da un mazzo di 40 senza reimmissione, la probabilità che siano tutte assi è

$$\frac{4}{40} \cdot \frac{3}{39} \cdot \frac{2}{38} = \frac{24}{59\,280} = \frac{1}{2470},$$

perché dopo ogni asso estratto ne restano uno in meno, in un mazzo con una carta in meno. Con le combinazioni si ritrova lo stesso numero: $\frac{C_{4,3}}{C_{40,3}} = \frac{4}{9880} = \frac{1}{2470}$. I due metodi sono equivalenti: con la regola del prodotto si ragiona per estrazioni successive, con il calcolo combinatorio si contano le mani; conviene scegliere quello che rende il conteggio più semplice.

14.3.2 Eventi indipendenti

Se ciascun evento non è influenzato dai precedenti, le probabilità condizionate coincidono con quelle semplici e la regola del prodotto diventa $p(A_1 \cap \ldots \cap A_n) = p(A_1) \cdot \ldots \cdot p(A_n)$. Si dice che n eventi sono **indipendenti** quando questa uguaglianza vale per *tutti* i gruppi che si possono formare con essi (a due a due, a tre a tre, …, tutti insieme). Non basta che siano indipendenti a due a due. Lanciamo due monete e consideriamo A = «la prima dà testa», B = «la seconda dà testa», E = «le due monete danno lo stesso risultato». Ciascuno ha probabilità $\frac{1}{2}$, e ogni intersezione di due di essi coincide con l'esito (T; T), di probabilità $\frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2}$: i tre eventi sono indipendenti a due a due. Però $A \cap B \cap E = \{(\text{T; T})\}$ ha probabilità $\frac{1}{4}$, non $\frac{1}{8}$: se si verificano A e B, l'evento E è certo. I tre eventi, presi insieme, *non* sono indipendenti.

Nella pratica l'indipendenza di più prove è quasi sempre un'**ipotesi del modello**: la assumiamo per i lanci successivi di un dado o di una moneta, per le estrazioni con reimmissione e, con qualche cautela, per eventi fisicamente separati, come il guasto di due componenti costruiti e montati in modo indipendente.

14.3.3 «Almeno uno»: passare al contrario

Molte domande chiedono la probabilità che un evento si verifichi *almeno una volta* in n prove. Calcolarla direttamente richiede di sommare i casi «una volta», «due volte», … fino a «n volte». È molto più semplice passare all'evento contrario, «mai». Se le prove sono indipendenti e in ciascuna l'evento ha probabilità p, la probabilità che non si verifichi mai è $(1-p)^n$, e quindi

$$p(\text{almeno un successo in } n \text{ prove}) = 1 - (1-p)^n.$$

È il calcolo che il cavaliere de Méré chiedeva a Pascal (lo abbiamo raccontato in *Matematica 2*): la probabilità di almeno un 6 in 4 lanci di un dado è $1 - (\frac{5}{6})^4 \approx 0{,}518$, mentre quella di almeno un doppio 6 in 24 lanci di due dadi è $1 - (\frac{35}{36})^{24} \approx 0{,}491$. La prima scommessa è favorevole, la seconda no, anche se un'analogia ingenua le farebbe sembrare equivalenti.

ESEMPIO SVOLTO

Quante volte bisogna lanciare un dado perché la probabilità di ottenere almeno un 6 superi il 90%?

Con n lanci la probabilità è $1 - (\frac{5}{6})^n$. Imponiamo

$$1 - \left(\frac{5}{6}\right)^n > 0{,}9 \quad \Leftrightarrow \quad \left(\frac{5}{6}\right)^n < 0{,}1.$$

Passiamo ai logaritmi naturali (*Matematica 3*): $n \ln \frac{5}{6} < \ln 0{,}1$. Poiché $\ln \frac{5}{6} < 0$, dividendo per questo numero il verso della disuguaglianza si inverte:

$$n > \frac{\ln 0{,}1}{\ln (5/6)} \approx \frac{-2{,}3026}{-0{,}1823} \approx 12{,}63.$$

Servono dunque almeno **13 lanci**. Controllo: con 12 lanci la probabilità è circa 0,888, con 13 è circa 0,907.

ERRORE FREQUENTE

Sommare le probabilità di eventi compatibili. «In un lancio la probabilità di un 6 è $\frac{1}{6}$, quindi in 6 lanci è $6 \cdot \frac{1}{6} = 1$: il 6 esce di sicuro.» Il ragionamento è sbagliato, perché gli eventi «6 al primo lancio», «6 al secondo lancio», … non sono incompatibili: possono verificarsi insieme, e l'assioma 3 non si applica. La probabilità corretta è

$$1 - \left(\frac{5}{6}\right)^6 = \frac{31\,031}{46\,656} \approx 0{,}665.$$

Con lo stesso errore, in 7 lanci si otterrebbe una probabilità $\frac{7}{6} > 1$, che contraddice il teorema $p(E) \le 1$. Ricorda: la somma delle probabilità è lecita solo per eventi incompatibili; per eventi compatibili si usa il teorema dell'unione o, per «almeno uno», l'evento contrario.

14.3.4 Il problema dei compleanni

In un gruppo di n persone, qual è la probabilità che almeno due compiano gli anni lo stesso giorno? Adottiamo il modello più semplice: un anno di 365 giorni (trascuriamo il 29 febbraio) e compleanni indipendenti ed equiprobabili. Un esito è la sequenza ordinata dei compleanni delle n persone: i casi possibili sono $D'_{365,n} = 365^n$. L'evento «almeno due compleanni coincidono» è difficile da contare direttamente; il suo contrario, «i compleanni sono tutti diversi», corrisponde alle sequenze senza ripetizioni, cioè alle disposizioni semplici $D_{365,n} = 365 \cdot 364 \cdot \ldots \cdot (365 - n + 1)$. Quindi

$$p(n) = 1 - \frac{D_{365,n}}{365^n} = 1 - \frac{365}{365} \cdot \frac{364}{365} \cdot \ldots \cdot \frac{365 - n + 1}{365}.$$

La Figura 14.3 mostra $p(n)$ per n da 1 a 60. La probabilità cresce in fretta: vale circa 0,117 per 10 persone, 0,411 per 20, e supera $\frac{1}{2}$ per la prima volta con $n = 23$, dove $p(23) \approx 0{,}5073$. Con 30 persone è circa 0,706, con 40 circa 0,891, con 57 supera 0,99. Il risultato sorprende perché istintivamente pensiamo alla probabilità che qualcuno abbia il *nostro* compleanno, che con 22 altre persone è appena $1 - \left(\frac{364}{365}\right)^{22} \approx 0{,}059$. Ma le coppie di persone che possono avere lo stesso compleanno sono $C_{23,2} = 253$, e ciascuna è un'occasione di coincidenza.

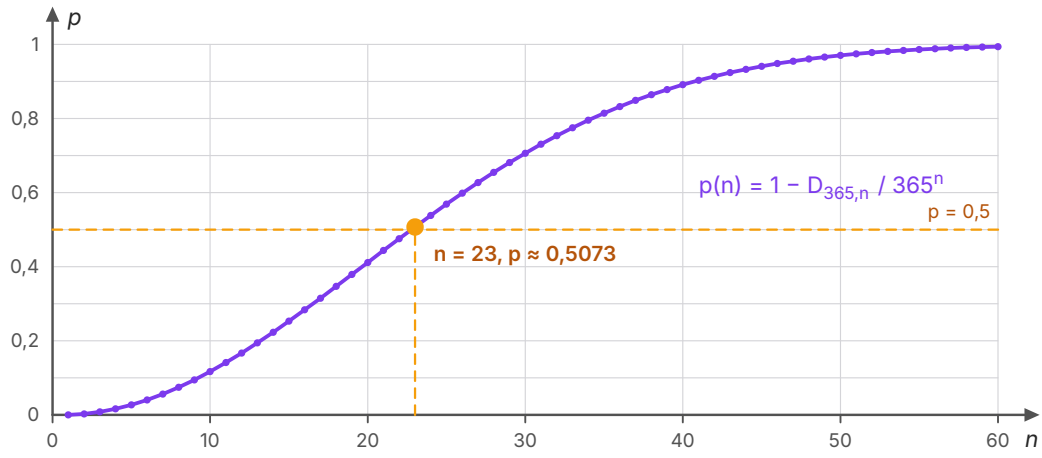


Figura 14.3 Probabilità $p(n)$ che, in un gruppo di n persone, almeno due compiano gli anni lo stesso giorno (365 giorni equiprobabili). La curva supera 0,5 per la prima volta a $n = 23$.

Come ogni modello, anche questo semplifica: le nascite reali non sono distribuite in modo perfettamente uniforme nel corso dell'anno. Si può dimostrare che una distribuzione non uniforme rende le coincidenze ancora *più* probabili, quindi il numero 23 è, se mai, una stima per eccesso.

14.4 Le prove ripetute: lo schema di Bernoulli

14.4.1 Un esempio: rispondere a caso

Un test ha 3 domande, ciascuna con 4 risposte di cui una sola corretta. Uno studente impreparato risponde a caso. Qual è la probabilità che indovini esattamente 2 risposte?

In ogni domanda la probabilità di rispondere correttamente (C) è $\frac{1}{4}$, quella di sbagliare (S) è $\frac{3}{4}$, e le risposte sono indipendenti. L'albero della Figura 14.4 ha $2^3 = 8$ percorsi. Quelli con esattamente 2 risposte corrette sono 3: CCS, CSC, SCC. Ciascuno ha probabilità $\frac{1}{4} \cdot \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{64}$, perché è il prodotto di due fattori $\frac{1}{4}$ e di un fattore $\frac{3}{4}$, in un ordine che non cambia il prodotto. I tre percorsi sono incompatibili, quindi

$$p(\text{2 risposte corrette}) = 3 \cdot \left(\frac{1}{4}\right)^2 \cdot \frac{3}{4} = \frac{9}{64} \approx 0{,}141.$$

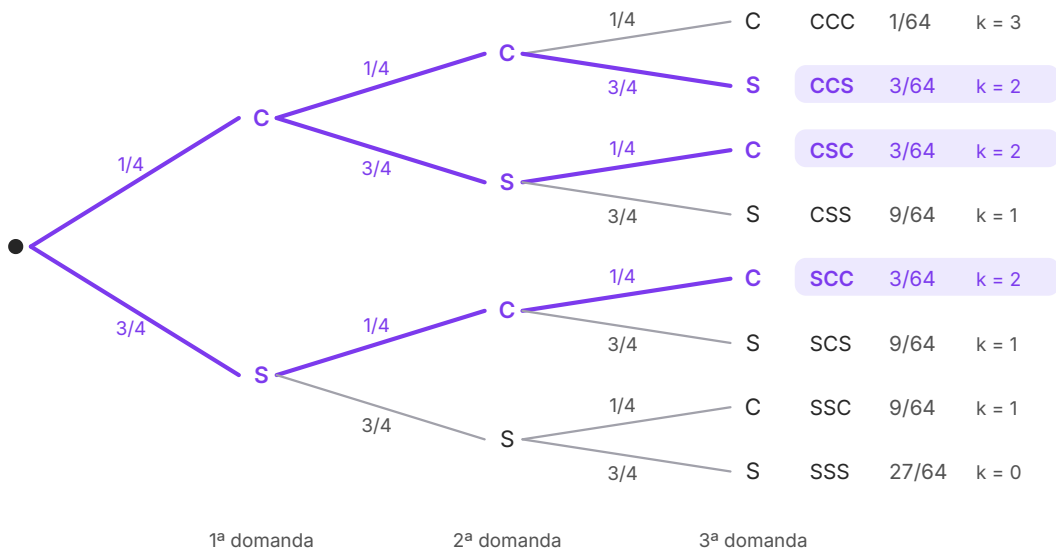


Figura 14.4 Diagramma ad albero di 3 prove ripetute con probabilità di successo 1/4 (C = risposta corretta, S = risposta sbagliata). Sono evidenziati i 3 percorsi con esattamente 2 successi, ciascuno di probabilità 3/64; a destra, la probabilità di ogni percorso e il numero k di successi.

Il numero 3 che moltiplica non è casuale: è il numero dei modi di scegliere, fra le 3 domande, le 2 in cui lo studente indovina, cioè $C_{3,2}$. Sommando le probabilità dei percorsi per ogni valore di k si trovano $\frac{27}{64}$, $\frac{27}{64}$, $\frac{9}{64}$ e $\frac{1}{64}$ per $k = 0, 1, 2, 3$, con somma 1.

14.4.2 La formula di Bernoulli

L'esempio contiene tutti gli ingredienti di una situazione che si presenta molto spesso. Si parla di **prove ripetute**, o di **schema di Bernoulli**, quando:

- si eseguono n prove, ciascuna con due soli esiti possibili, che chiamiamo convenzionalmente **successo** e **insuccesso** (senza alcun giudizio: il «successo» può essere un pezzo difettoso);
- la probabilità di successo è la stessa, p, in ogni prova;
- le prove sono indipendenti.

IN SINTESI

Teorema (formula di Bernoulli).

Ipotesi: n prove indipendenti, ciascuna con probabilità di successo p; $0 \le k \le n$.

Tesi: la probabilità di ottenere esattamente k successi è

$$p_n(k) = C_{n,k} \cdot p^k (1-p)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k}.$$

Dimostrazione. Un esito delle n prove è una sequenza di n lettere S o I (successo, insuccesso). Consideriamo una sequenza con esattamente k lettere S. Per l'indipendenza, la sua probabilità è il prodotto delle probabilità delle singole prove: k fattori uguali a p e $n-k$ fattori uguali a $1-p$, cioè $p^k (1-p)^{n-k}$, qualunque sia la posizione delle S. Le sequenze con k successi sono tante quanti i modi di scegliere le k posizioni delle S fra le n posizioni, cioè $C_{n,k}$. Esse sono incompatibili a due a due (sequenze diverse non possono verificarsi insieme), quindi le loro probabilità si sommano: $C_{n,k}$ addendi uguali a $p^k(1-p)^{n-k}$. c.v.d.

La formula ritrova i risultati già ottenuti. Per lo studente del test, $n = 3$, $p = \frac{1}{4}$, $k = 2$: $C_{3,2} \left(\frac{1}{4}\right)^2 \frac{3}{4} = \frac{9}{64}$. Per l'urna con reimmissione del paragrafo precedente, ogni estrazione dà una bianca con probabilità $\frac{4}{10} = \frac{2}{5}$, e $C_{3,2} \left(\frac{2}{5}\right)^2 \frac{3}{5} = \frac{36}{125} = 0{,}288$, come avevamo contato.

14.4.3 Le probabilità hanno somma 1

I valori $k = 0, 1, \ldots, n$ esauriscono tutti i casi e sono incompatibili, quindi la somma delle probabilità $p_n(k)$ deve essere 1. Lo conferma il binomio di Newton del capitolo precedente: posto $q = 1 - p$,

$$C_{n,0}q^n + C_{n,1}pq^{n-1} + C_{n,2}p^2q^{n-2} + \ldots + C_{n,n}p^n = (p+q)^n = 1^n = 1.$$

È il motivo per cui si parla di probabilità «binomiali»: i termini dello sviluppo di $(p+q)^n$ sono, uno per uno, le probabilità di 0, 1, ..., n successi.

14.4.4 Il numero più probabile di successi

La Tabella 14.2 e la Figura 14.5 mostrano le probabilità $p_{10}(k)$ per $n = 10$ prove, nei casi $p = 0{,}5$ (dieci lanci di una moneta) e $p = 0{,}2$. Con $p = 0{,}5$ la distribuzione è simmetrica, perché $C_{10,k} = C_{10,10-k}$ e $p = 1 - p$; il valore più probabile è $k = 5$, con probabilità di circa 0,246. Con $p = 0{,}2$ le probabilità si concentrano sui valori piccoli e il massimo è a $k = 2$, con circa 0,302. In entrambi i casi il **numero più probabile di successi** coincide con $n \cdot p$: 5 e 2. In generale non è sempre

così, perché $n \cdot p$ può non essere intero, ma il numero più probabile è sempre uno dei due interi più vicini a $n \cdot p$ (un esercizio della sezione «Dimostrare e argomentare» mostra come trovarlo con precisione).

Tabella 14.2 Probabilità $p_{10}(k)$ di k successi in 10 prove (valori arrotondati a 4 cifre decimali).

k	0	1	2	3	4	5	6	7	8	9	10
$p = 0{,}5$	0,0010	0,0098	0,0439	0,1172	0,2051	0,2461	0,2051	0,1172	0,0439	0,0098	0,0010
$p = 0{,}2$	0,1074	0,2684	0,3020	0,2013	0,0881	0,0264	0,0055	0,0008	0,0001	0,0000	0,0000

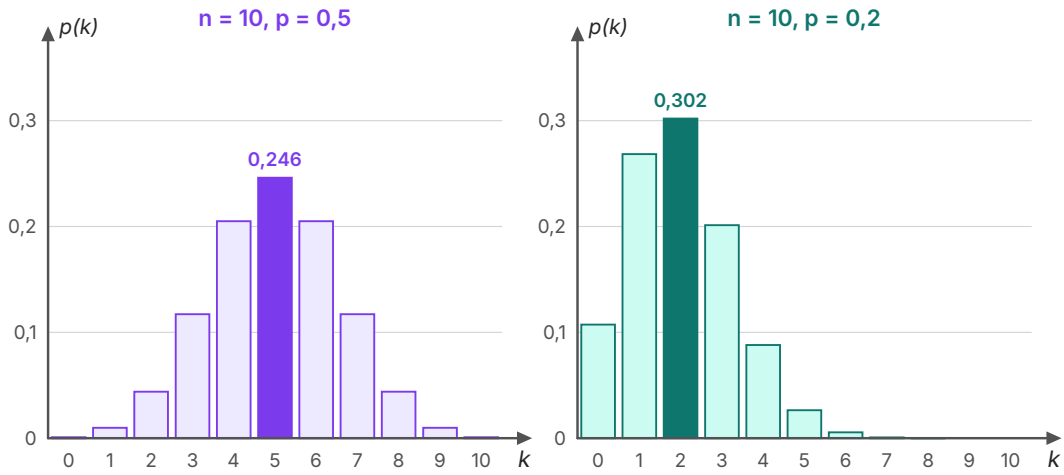


Figura 14.5 Probabilità di k successi in $n = 10$ prove ripetute, con $p = 0{,}5$ (a sinistra) e $p = 0{,}2$ (a destra). La barra piena è il numero più probabile di successi; la somma delle altezze è 1 in entrambi i grafici.

ESEMPIO SVOLTO

a) Si lancia un dado 5 volte: qual è la probabilità di ottenere esattamente due 6? **b)** Una macchina produce pezzi difettosi con probabilità del 3%, in modo indipendente da un pezzo all'altro (dati inventati). Si controllano 20 pezzi: qual è la probabilità di trovarne al più uno difettoso? **c)** Una giocatrice di pallacanestro segna un tiro libero con probabilità 0,75 (dato inventato). Qual è la probabilità che su 8 tiri ne segni almeno 6?

a) $n = 5$, $p = \frac{1}{6}$, $k = 2$: $C_{5,2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3 = 10 \cdot \frac{1}{36} \cdot \frac{125}{216} = \frac{625}{3888} \approx 0{,}161$.

b) Il «successo» è il pezzo difettoso: $n = 20$, $p = 0{,}03$. «Al più uno» significa 0 oppure 1:

$$p_{20}(0) + p_{20}(1) = 0{,}97^{20} + 20 \cdot 0{,}03 \cdot 0{,}97^{19} \approx 0{,}5438 + 0{,}3364 \approx 0{,}880.$$

Nel 12% circa dei controlli si troveranno dunque 2 o più pezzi difettosi, anche se la macchina funziona come previsto: è un dato di cui chi organizza il controllo deve tenere conto prima di dare l'allarme.

c) $n = 8$, $p = 0{,}75$, «almeno 6» significa $k = 6$, 7 o 8:

$$C_{8,6} \cdot 0{,}75^6 \cdot 0{,}25^2 + C_{8,7} \cdot 0{,}75^7 \cdot 0{,}25 + 0{,}75^8 \approx 0{,}3115 + 0{,}2670 + 0{,}1001 \approx 0{,}679.$$

Il valore esatto è $\frac{44\,469}{65\,536}$. Il numero più probabile di canestri è $k = 6 = 8 \cdot 0{,}75$.

ERRORE FREQUENTE

Trattare come indipendenti le estrazioni senza reimmissione. Da un'urna con 4 bianche e 6 nere si estraggono 3 palline *senza* reimmissione. Applicare la formula di Bernoulli con $p = \frac{2}{5}$ dà 0,288 per «esattamente 2 bianche», mentre il valore corretto è $\frac{3}{10}$. Senza reimmissione la probabilità di estrarre una bianca cambia dopo ogni estrazione, e le prove non sono indipendenti: manca un'ipotesi dello schema di Bernoulli. Si usa invece la formula con le combinazioni (o la regola del prodotto con le probabilità condizionate).

La differenza fra i due modelli diventa trascurabile solo quando l'urna è molto grande rispetto al numero di estrazioni: togliere 3 palline da un'urna di 10 000 cambia pochissimo la sua composizione. Per questo nei controlli di qualità su grandi lotti si usa spesso lo schema di Bernoulli come approssimazione.

APPROFONDIMENTO · VARIABILE ALEATORIA E VALORE ATTESO

Nelle prove ripetute il numero X di successi è un numero che dipende dal caso: prima di eseguire le prove sappiamo solo che sarà uno dei valori 0, 1, ..., n, ciascuno con la sua probabilità $p_n(k)$. Una grandezza numerica di questo tipo si chiama **variabile aleatoria**. Come per il guadagno di un gioco studiato in *Matematica 2*, il suo **valore atteso** è la somma dei valori possibili, ciascuno moltiplicato per la sua probabilità:

$$0 \cdot p_n(0) + 1 \cdot p_n(1) + 2 \cdot p_n(2) + \ldots + n \cdot p_n(n) = n \cdot p.$$

Per lo studente che risponde a caso alle 3 domande: $0 \cdot \frac{27}{64} + 1 \cdot \frac{27}{64} + 2 \cdot \frac{9}{64} + 3 \cdot \frac{1}{64} = \frac{48}{64} = \frac{3}{4} = 3 \cdot \frac{1}{4}$. Il valore atteso non è il risultato «più probabile», né un risultato possibile (nessuno indovina $\frac{3}{4}$ di risposta): è la media dei successi su moltissime ripetizioni dell'intero test. Lo studio delle variabili aleatorie e delle loro distribuzioni è uno degli argomenti del quinto anno.

COLLEGAMENTI · GENETICA E INDUSTRIA

Genetica. Due genitori portatori sani di un carattere recessivo (genotipo Aa) hanno, per le leggi di Mendel, probabilità $\frac{1}{4}$ di avere un figlio con genotipo aa, che manifesta il carattere. I figli successivi sono prove indipendenti con la stessa probabilità: è uno schema di Bernoulli. Con 3 figli, la probabilità che esattamente 2 manifestino il carattere è $C_{3,2}\left(\frac{1}{4}\right)^2 \frac{3}{4} = \frac{9}{64} \approx 0{,}14$, lo stesso numero dello studente che risponde a caso, perché il modello matematico è identico.

Industria. Nel controllo statistico della qualità non si esamina ogni pezzo, ma un campione. La formula di Bernoulli dice quanti difettosi ci si deve aspettare nel campione se il processo funziona come previsto; un numero di difettosi molto improbabile secondo il modello è il segnale che il processo va controllato.

14.5 Probabilità e frequenza

14.5.1 La legge dei grandi numeri, rivista

In *Matematica 2* abbiamo incontrato la legge empirica del caso: su molte prove ripetute nelle stesse condizioni, la frequenza relativa di un evento tende ad avvicinarsi alla sua probabilità. Abbiamo anche ricordato che Jakob Bernoulli ne dimostrò una versione rigorosa, oggi chiamata **legge dei grandi numeri**: in n prove indipendenti con probabilità di successo p, fissato un

margine piccolo quanto si vuole, la probabilità che la frequenza relativa dei successi differisca da p più di quel margine tende a 0 al crescere di n.

La Figura 14.6 mostra il comportamento descritto dalla legge con una **simulazione**: un calcolatore ha prodotto tre serie di 2000 lanci di un dado, usando un generatore di numeri pseudocasuali, e per ogni n ha calcolato la frequenza relativa f dell'esito 6 nei primi n lanci. Nei primi lanci le tre curve oscillano molto e sono lontane fra loro; al crescere di n si stringono intorno a 1/6. Dopo 2000 lanci le frequenze sono 340/2000 = 0,170, 341/2000 ≈ 0,171 e 303/2000 ≈ 0,152: vicine a 1/6 ≈ 0,167, ma non uguali. È ciò che la legge prevede: la frequenza si avvicina alla probabilità, ma non la raggiunge esattamente, né si avvicina in modo regolare.

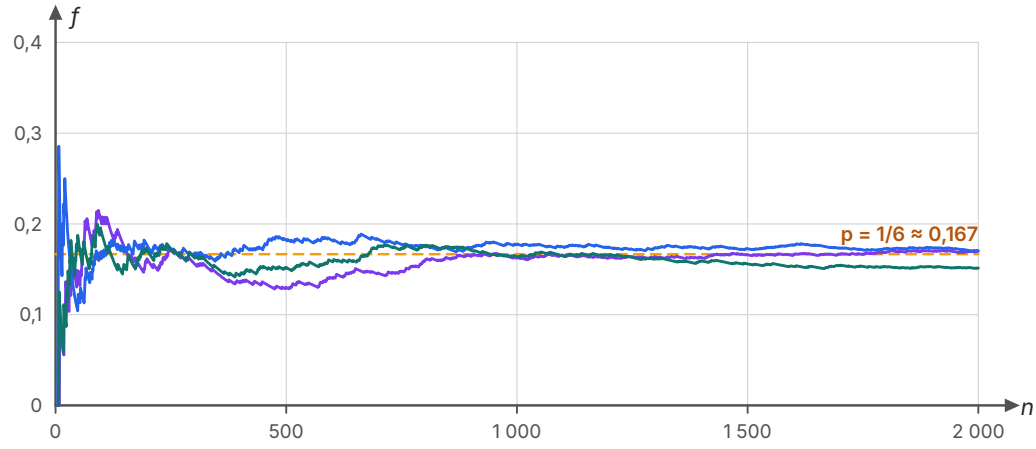


Figura 14.6 Simulazione al calcolatore: frequenza relativa f dell'esito 6 nei primi n lanci di un dado, in tre serie indipendenti di 2000 lanci. Al crescere di n le frequenze si stabilizzano intorno a p = 1/6. I dati sono simulati, non osservati.

14.5.2 Quanto vicina? Un calcolo con la formula di Bernoulli

La formula di Bernoulli permette di misurare *quanto* è probabile che la frequenza sia vicina a p. Lanciamo una moneta n volte e chiediamoci con quale probabilità la frequenza delle teste è compresa fra 0,45 e 0,55, cioè dista da 0,5 al più 0,05. Basta sommare le probabilità $p_n(k)$ per i valori di k con $0{,}45\,n \le k \le 0{,}55\,n$. I risultati, nella Tabella 14.3, sono calcolati esattamente con la formula e arrotondati solo alla fine.

Tabella 14.3 Probabilità che la frequenza relativa delle teste in n lanci di una moneta sia compresa fra 0,45 e 0,55 (calcolo esatto con la formula di Bernoulli).

Numero di lanci n	Valori di k ammessi	Probabilità
10	5	0,246
100	da 45 a 55	0,729
1000	da 450 a 550	0,9986

Con 10 lanci, una frequenza compresa fra 0,45 e 0,55 si ottiene solo con 5 teste esatte, e succede meno di una volta su quattro. Con 1000 lanci, uscire da quell'intervallo è un evento rarissimo, circa una volta su 700. La legge dei grandi numeri, che la tabella illustra in un caso particolare, giustifica la definizione frequentista: quando non sappiamo calcolare una probabilità, la frequenza su molte prove ne è una stima tanto più affidabile quanto più grande è il numero delle prove.

14.5.3 Che cos'è un modello probabilistico

In tutto il capitolo abbiamo calcolato probabilità a partire da ipotesi: palline «ben mescolate», compleanni equiprobabili, prove indipendenti con la stessa probabilità p. L'insieme di queste ipotesi è un **modello probabilistico** del fenomeno. Il calcolo, una volta scelto il modello, è rigoroso; ma la scelta del modello non è un teorema, è una decisione che va giustificata e, quando possibile, confrontata con i dati. Lo schema di Bernoulli descrive bene i lanci di una moneta; descrive meno bene i tiri liberi di una giocatrice reale, la cui probabilità di segnare può cambiare con la stanchezza o con la tensione della partita.

Il confronto con le frequenze osservate è il modo per mettere alla prova un modello: se in 1000 lanci di una moneta otteniamo 600 teste, il modello «moneta equa» rende quell'esito estremamente improbabile, e abbiamo buone ragioni per dubitare che la moneta sia equa. Ragionamenti di questo tipo sono alla base della statistica inferenziale e saranno ripresi negli studi successivi; nel prossimo capitolo vedremo come la formula di Bayes permetta di *aggiornare* una probabilità quando arrivano nuove informazioni.

SINTESI

- Gli **assiomi di Kolmogorov** (1933) per uno spazio finito: $p(E) \geq 0$; $p(\Omega) = 1$; se $A \cap B = \varnothing$, $p(A \cup B) = p(A) + p(B)$. Le definizioni classica, frequentista e soggettiva li rispettano.
- Dagli assiomi seguono $p(\overline{E}) = 1 - p(E)$, $p(\varnothing) = 0$, $0 \leq p(E) \leq 1$, la monotonia ($A \subseteq B \Rightarrow p(A) \leq p(B)$) e il teorema dell'unione $p(A \cup B) = p(A) + p(B) - p(A \cap B)$.
- Nella definizione classica casi favorevoli e possibili si contano **con lo stesso modello** combinatorio: entrambi ordinati o entrambi non ordinati.
- Senza reimmissione, la probabilità di k bianche in m estrazioni da un'urna con b bianche su n palline è $\dfrac{C_{b,k} \cdot C_{n-b,m-k}}{C_{n,m}}$; estrazioni contemporanee e successive danno la stessa probabilità.
- Con reimmissione la composizione dell'urna non cambia e le estrazioni sono indipendenti.
- Probabilità composta: $p(A_1 \cap \ldots \cap A_n)$ è il prodotto delle probabilità condizionate lungo un percorso dell'albero; per eventi indipendenti è il prodotto delle probabilità. L'indipendenza a due a due non basta.
- «Almeno un successo» in n prove indipendenti: $1 - (1 - p)^n$. Le probabilità di eventi compatibili non si sommano.
- Problema dei compleanni: $p(n) = 1 - D_{365,n}/365^n$, che supera $\dfrac{1}{2}$ da $n = 23$ ($p \approx 0{,}5073$).
- **Schema di Bernoulli**: n prove indipendenti, due esiti, probabilità di successo costante p. La probabilità di k successi è $C_{n,k}\, p^k\, (1-p)^{n-k}$; per il binomio di Newton la somma su k è 1.
- Il numero più probabile di successi è uno dei due interi più vicini a $n \cdot p$; il valore atteso del numero di successi è $n \cdot p$.
- Legge dei grandi numeri: al crescere di n diventa sempre più improbabile che la frequenza relativa si discosti da p più di un margine fissato.
- Ogni calcolo dipende da un **modello** (equiprobabilità, indipendenza): il modello va giustificato e confrontato con i dati.

ESERCIZI

Comprendere

1. ● Se $A \subseteq B$, dagli assiomi della probabilità segue che
   - **A.** $p(A) \geq p(B)$
   - **B.** $p(A) \leq p(B)$
   - **C.** $p(A) + p(B) = 1$
   - **D.** $p(B \setminus A) = p(A) + p(B)$

2. ● Da un'urna con 5 palline rosse e 5 blu se ne estraggono contemporaneamente 2. La probabilità che siano entrambe rosse è

   A. $\frac{1}{4}$

   B. $\frac{1}{2}$

   C. $\frac{2}{9}$

   D. $\frac{1}{5}$

3. ● La probabilità di ottenere almeno una testa in 4 lanci di una moneta è

   A. $\frac{1}{2}$

   B. $\frac{1}{16}$

   C. 1

   D. $\frac{15}{16}$

4. ● In 5 prove indipendenti, ciascuna con probabilità di successo $\frac{1}{3}$, la probabilità di esattamente 2 successi è

   A. $C_{5,2}\,(\frac{1}{3})^2\,(\frac{2}{3})^3$

   B. $(\frac{1}{3})^2\,(\frac{2}{3})^3$

   C. $2 \cdot (\frac{1}{3})^2\,(\frac{2}{3})^3$

   D. $C_{5,2}\,(\frac{1}{3})^2\,(\frac{2}{3})^2$

5. ● Quale delle seguenti condizioni *non* è richiesta dallo schema di Bernoulli?

   A. le prove sono indipendenti

   B. la probabilità di successo è la stessa in ogni prova

   C. ogni prova ha due soli esiti, successo o insuccesso

   D. la probabilità di successo è $\frac{1}{2}$

6. ● Vero o falso? Correggi le affermazioni false. **a)** Se $p(A) = 0{,}3$ e $p(B) = 0{,}5$, allora $p(A \cup B) = 0{,}8$. **b)** Estraendo 2 carte da un mazzo senza reimmissione, la probabilità di due assi è la stessa sia che le carte si estraggano insieme, sia che si estraggano una dopo l'altra. **c)** Lanciando un dado 6 volte si è certi di ottenere almeno un 6. **d)** Nello schema di Bernoulli la somma delle probabilità di 0, 1, …, n successi è 1.

7. ● Completa. **a)** Secondo gli assiomi di Kolmogorov, se $A \cap B = \varnothing$ allora $p(A \cup B) =$ _____ . **b)** In n prove indipendenti con probabilità di successo p, la probabilità di non ottenere alcun successo è _____ e quella di ottenerne almeno uno è _____ . **c)** Per eventi qualsiasi, $p(A \cup B) =$ _____ . **d)** Un'estrazione si dice _____ quando la pallina estratta viene rimessa nell'urna prima dell'estrazione successiva.

8. ● Spiega perché nelle estrazioni senza reimmissione le prove successive non sono indipendenti, mentre lo sono in quelle con reimmissione. Fai un esempio con un'urna di 3 palline bianche e 2 nere.

Calcolare

9. ● Si lanciano due dadi. Calcola la probabilità che **a)** la somma sia 8 oppure esca almeno un 5; **b)** la somma sia pari oppure entrambi i dadi diano un numero maggiore di 4.

10. ● Di due eventi si sa che $p(A) = 0{,}6$, $p(B) = 0{,}5$ e $p(A \cap B) = 0{,}3$. Calcola **a)** $p(A \cup B)$; **b)** la probabilità che non si verifichi nessuno dei due; **c)** $p(A \setminus B)$; **d)** stabilisci se A e B sono indipendenti.

11. ● Un'urna contiene 6 palline bianche e 4 rosse; se ne estraggono 4 contemporaneamente. Calcola la probabilità che **a)** siano tutte bianche; **b)** esattamente 2 siano rosse; **c)** almeno una sia rossa.

12. ● Dalla stessa urna dell'esercizio precedente si estraggono 4 palline *con* reimmissione. Calcola la probabilità di ottenere esattamente 2 rosse e confrontala con il risultato del punto b) dell'esercizio precedente.

13. ●● Da un mazzo di 40 carte (4 semi di 10 carte) se ne estraggono 3. Calcola la probabilità che **a)** siano tutte dello stesso seme; **b)** siano di tre semi diversi; **c)** ci sia almeno un asso.

14. ●● Si ricevono 5 carte da un mazzo francese di 52. Calcola la probabilità di avere **a)** un poker, cioè quattro carte dello stesso valore; **b)** cinque carte dello stesso seme.

15. ●● Al lotto, su una ruota si estraggono 5 numeri fra 90. Calcola la probabilità che **a)** esca un terno fissato; **b)** nessuno dei numeri da 1 a 10 sia estratto.

16. ●● Quante volte bisogna lanciare un dado perché la probabilità di ottenere almeno un 6 superi il 99%?

17. ●● Si lancia una moneta 8 volte. Calcola la probabilità di ottenere **a)** esattamente 4 teste; **b)** almeno 7 teste.

18. ●● Si lancia un dado 4 volte. Calcola la probabilità di ottenere **a)** esattamente un 6; **b)** al più un 6.

19. ●● In 6 prove indipendenti con probabilità di successo 0,3 calcola le probabilità di 0, 1, ..., 6 successi (4 cifre decimali), verifica che la somma è 1 e indica il numero più probabile di successi.

Dimostrare e argomentare

20. ●● Dimostra, usando soltanto i tre assiomi di Kolmogorov, che $p(\varnothing) = 0$ e che $p(\overline{E}) = 1 - p(E)$ per ogni evento E.

21. ●●● Dimostra la formula dell'unione di tre eventi: $p(A \cup B \cup E) = p(A) + p(B) + p(E) - p(A \cap B) - p(A \cap E) - p(B \cap E) + p(A \cap B \cap E)$. Verificala poi con il lancio di un dado e A = «pari», B = «multiplo di 3», E = «maggiore di 4».

22. ●● Dimostra che, per un'urna con b palline bianche su n, la probabilità di k bianche in m estrazioni senza reimmissione è la stessa per estrazioni contemporanee e successive, cioè che $\frac{C_{b,k} \cdot C_{n-b,m-k}}{C_{n,m}} = \frac{C_{m,k} \cdot D_{b,k} \cdot D_{n-b,m-k}}{D_{n,m}}$.

23. ●● Trova l'errore: «La probabilità di ottenere almeno un 6 in 3 lanci di un dado è $\frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$». Calcola il valore corretto.

24. ●●● Nello schema di Bernoulli, con $0 < p < 1$ e $1 \le k \le n$, dimostra che $\frac{p_n(k)}{p_n(k-1)} = \frac{(n-k+1)\,p}{k\,(1-p)}$. Deduci che $p_n(k) > p_n(k-1)$ se e solo se $k < (n+1)\,p$, e spiega perché il numero più probabile di successi è uno dei due interi più vicini a $n \cdot p$.

Problemi

25. ●● Due genitori portatori sani di un carattere recessivo (Aa × Aa) hanno 4 figli. Calcola la probabilità che **a)** almeno un figlio manifesti il carattere; **b)** esattamente uno lo manifesti.

26. ●● Una macchina produce pezzi difettosi con probabilità 2%, in modo indipendente (dati inventati). In una scatola ci sono 50 pezzi. Calcola la probabilità che **a)** nessuno sia difettoso; **b)** al più 2 siano difettosi.

27. ●● Un giocatore segna un tiro libero con probabilità 0,8 (dato inventato) e ne tira 5, in modo indipendente. Calcola la probabilità che **a)** li segni tutti; **b)** ne segni almeno 4. **c)** Quale ipotesi del modello potrebbe non valere in una partita vera?

28. ●● In una classe ci sono 25 studenti (365 giorni equiprobabili, nessuno nato il 29 febbraio). Calcola la probabilità che **a)** almeno due studenti compiano gli anni lo stesso giorno; **b)** almeno uno studente compia gli anni lo stesso giorno della professoressa. Spiega perché i due risultati sono così diversi.

29. ●● In una classe ci sono 12 ragazze e 10 ragazzi. Si sorteggia una commissione di 4 studenti. Calcola la probabilità che sia formata da 2 ragazze e 2 ragazzi, e quella che contenga almeno un ragazzo.

30. ●●● Un test ha 10 domande, ciascuna con 4 risposte di cui una sola corretta. Per superarlo bisogna rispondere correttamente ad almeno 6 domande. Qual è la probabilità di superarlo rispondendo a caso? Quante risposte corrette ci si deve aspettare in media?

Con il computer

31. ●● In un foglio di calcolo scrivi in una colonna i valori $k = 0, 1, \ldots, 20$ e nella colonna accanto la probabilità di k successi in 20 prove con $p = 0{,}1$, usando la funzione che calcola le combinazioni (o la funzione binomiale del programma). **a)** Verifica che la somma della colonna è 1. **b)** Trova il numero più probabile di successi. **c)** Rappresenta i valori con un istogramma.

32. ●● Simula il problema dei compleanni con un foglio di calcolo: in una riga genera 23 numeri interi casuali fra 1 e 365 (funzione «casuale tra»), e con una formula controlla se fra essi ci sono ripetizioni. Copia la riga 1000 volte e calcola la frequenza relativa delle righe con almeno una ripetizione. Confrontala con la probabilità teorica.

Verso la verifica

33. ●● Un'urna contiene 3 palline rosse e 5 verdi. **a)** Si estraggono 2 palline senza reimmissione: calcola la probabilità che siano entrambe verdi e quella che siano di colori diversi. **b)** Si estraggono 4 palline con reimmissione: calcola la probabilità di ottenere esattamente una rossa.

34. ●● Anna lancia un dado 10 volte. Calcola la probabilità che ottenga **a)** esattamente due 6; **b)** almeno due 6. **c)** Qual è il numero più probabile di 6?

35. ●●● Sfida. Quante persone devono esserci in una stanza, oltre a te, perché la probabilità che almeno una compia gli anni il tuo stesso giorno superi $\frac{1}{2}$? Confronta la risposta con il numero 23 del problema dei compleanni e spiega la differenza.

PER APPROFONDIRE

- FEL — il classico manuale di probabilità: i primi capitoli trattano lo spazio campionario, il calcolo combinatorio applicato alle urne e alle carte, il problema dei compleanni e le prove di Bernoulli, con moltissimi esempi.
- BOY — i capitoli sul Seicento e sul Settecento raccontano la nascita del calcolo delle probabilità, da Pascal e Fermat a Jakob Bernoulli.
- POL — per allenarsi a scegliere la strategia: il passaggio all'evento contrario e la ricerca di un problema più semplice equivalente sono esempi delle euristiche descritte nel libro.

Glossario del capitolo

Assiomi della probabilità
Le tre regole (non negatività, probabilità 1 dello spazio campionario, additività per eventi incompatibili) da cui Kolmogorov dedusse tutto il calcolo delle probabilità.

Estrazione con reimmissione
Estrazione in cui l'oggetto estratto viene rimesso nell'urna prima della successiva: le estrazioni sono indipendenti.

Estrazione contemporanea
Estrazione di più oggetti tutti insieme; gli esiti sono gruppi non ordinati e si contano con le combinazioni.

Estrazione senza reimmissione
Estrazione in cui l'oggetto estratto non viene rimesso nell'urna: la composizione dell'urna cambia e le estrazioni non sono indipendenti.

Estrazione successiva
Estrazione di più oggetti uno dopo l'altro; gli esiti sono sequenze ordinate e si contano con le disposizioni.

Evento impossibile
L'evento $\varnothing$, che non contiene alcun esito; ha probabilità 0.

Formula di Bernoulli
Probabilità di k successi in n prove ripetute: $C_{n,k}\, p^k (1-p)^{n-k}$.

Impostazione assiomatica
Modo di fondare il calcolo delle probabilità a partire dagli assiomi, senza stabilire che cosa la probabilità «sia»; contiene le definizioni classica, frequentista e soggettiva.

Indipendenza di n eventi
Proprietà di n eventi: la probabilità dell'intersezione di ogni loro gruppo (a due a due, a tre a tre, ..., tutti insieme) è il prodotto delle probabilità; l'indipendenza a due a due non basta.

Legge dei grandi numeri
Teorema di Jakob Bernoulli: al crescere del numero delle prove diventa sempre più improbabile che la frequenza relativa si discosti dalla probabilità più di un margine fissato.

Monotonia della probabilità
Proprietà della probabilità: se $A \subseteq B$, allora $p(A) \leq p(B)$.

Numero più probabile di successi
Il valore di k per cui la probabilità di k successi è massima; è uno dei due interi più vicini a $n \cdot p$.

Probabilità composta
Probabilità dell'intersezione di più eventi, prodotto delle probabilità condizionate lungo un percorso dell'albero.

Problema dei compleanni
Calcolo della probabilità che in un gruppo di n persone almeno due compiano gli anni lo stesso giorno; supera 1/2 da $n = 23$.

Schema di Bernoulli
Modello di n prove indipendenti, ciascuna con due esiti e con la stessa probabilità p di successo.

Teorema dell'unione
$p(A \cup B) = p(A) + p(B) - p(A \cap B)$, valido per eventi qualsiasi.

Valore atteso
Somma dei valori di una variabile aleatoria moltiplicati per le loro probabilità; per il numero di successi in n prove vale $n \cdot p$.

Variabile aleatoria
Grandezza numerica il cui valore dipende dall'esito di un esperimento aleatorio, come il numero di successi in n prove.

Soluzioni

1. B. Per la monotonia: B è unione dei due eventi incompatibili A e $B \setminus A$, quindi $p(B) = p(A) + p(B \setminus A) \geq p(A)$.

2. C. $\frac{C_{5,2}}{C_{10,2}} = \frac{10}{45} = \frac{2}{9}$. Oppure, per estrazioni successive, $\frac{5}{10} \cdot \frac{4}{9} = \frac{2}{9}$.

3. D. Contrario: «nessuna testa», di probabilità $(\frac{1}{2})^4 = \frac{1}{16}$; quindi $1 - \frac{1}{16} = \frac{15}{16}$.

4. A. Formula di Bernoulli con $n = 5$, $k = 2$, $p = \frac{1}{3}$: gli insuccessi sono $5 - 2 = 3$ e le sequenze sono $C_{5,2} = 10$.

5. D. La probabilità di successo p può essere qualsiasi numero fra 0 e 1, purché sia la stessa in ogni prova.

6. a) Falso: vale solo se A e B sono incompatibili. In generale $p(A \cup B) = 0{,}8 - p(A \cap B)$, quindi $0{,}5 \leq p(A \cup B) \leq 0{,}8$. **b)** Vero: contemporanea e successiva senza reimmissione danno la stessa probabilità. **c)** Falso: la probabilità è $1 - (\frac{5}{6})^6 \approx 0{,}665$. **d)** Vero: per il binomio di Newton la somma è $(p + 1 - p)^n = 1$.

7. a) $p(A) + p(B)$. **b)** $(1 - p)^n$; $1 - (1 - p)^n$. **c)** $p(A) + p(B) - p(A \cap B)$. **d)** con reimmissione.

8. Senza reimmissione l'urna cambia dopo ogni estrazione, quindi la probabilità di un colore alla seconda estrazione dipende dall'esito della prima. Nell'urna con 3 bianche e 2 nere, $p(2^a \text{ bianca} \mid 1^a \text{ bianca}) = \frac{2}{4} = \frac{1}{2}$, mentre $p(2^a \text{ bianca} \mid 1^a \text{ nera}) = \frac{3}{4}$: le condizionate sono diverse, le prove sono dipendenti. Con reimmissione l'urna è sempre la stessa e la probabilità di bianca è $\frac{3}{5}$ a ogni estrazione, qualunque sia l'esito precedente.

9. a) «Somma 8»: (2; 6), (3; 5), (4; 4), (5; 3), (6; 2), 5 coppie; «almeno un 5»: $36 - 25 = 11$ coppie; in comune (3; 5) e (5; 3). $p = \frac{5 + 11 - 2}{36} = \frac{14}{36} = \frac{7}{18}$. **b)** «Somma pari»: 18 coppie; «entrambi maggiori di 4»: (5; 5), (5; 6), (6; 5), (6; 6), cioè 4 coppie, di cui 2 a somma pari. $p = \frac{18 + 4 - 2}{36} = \frac{20}{36} = \frac{5}{9}$.

10. a) $0{,}6 + 0{,}5 - 0{,}3 = 0{,}8$. **b)** $1 - 0{,}8 = 0{,}2$. **c)** $p(A) - p(A \cap B) = 0{,}3$. **d)** Sì: $p(A) \cdot p(B) = 0{,}3 = p(A \cap B)$.

11. Casi possibili: $C_{10,4} = 210$. **a)** $\frac{C_{6,4}}{210} = \frac{15}{210} = \frac{1}{14}$. **b)** $\frac{C_{6,2} \cdot C_{4,2}}{210} = \frac{15 \cdot 6}{210} = \frac{3}{7}$. **c)** $1 - \frac{1}{14} = \frac{13}{14}$.

12. Bernoulli con $n = 4$, $p = \frac{2}{5}$: $C_{4,2} (\frac{2}{5})^2 (\frac{3}{5})^2 = 6 \cdot \frac{4}{25} \cdot \frac{9}{25} = \frac{216}{625} = 0{,}3456$, minore di $\frac{3}{7} \approx 0{,}4286$.

13. Casi possibili: $C_{40,3} = 9880$. **a)** $\frac{4 \cdot C_{10,3}}{9880} = \frac{480}{9880} = \frac{12}{247} \approx 0{,}049$. **b)** Scelta dei 3 semi ($C_{4,3} = 4$) e di una carta per seme (10^3): $\frac{4000}{9880} = \frac{100}{247} \approx 0{,}405$. **c)** $1 - \frac{C_{36,3}}{9880} = 1 - \frac{7140}{9880} = \frac{137}{494} \approx 0{,}277$.

14. Casi possibili $C_{52,5} = 2\,598\,960$. **a)** Valore del poker (13 modi) e quinta carta (48 modi): $\frac{13 \cdot 48}{2\,598\,960} = \frac{624}{2\,598\,960} = \frac{1}{4165} \approx 0{,}00024$. **b)** $\frac{4 \cdot C_{13,5}}{2\,598\,960} = \frac{5148}{2\,598\,960} = \frac{33}{16\,660} \approx 0{,}0020$.

15. a) $\frac{C_{87,2}}{C_{90,5}} = \frac{3741}{43\,949\,268} = \frac{1}{11\,748}$. **b)** I 5 numeri vanno scelti fra gli 80 rimanenti: $\frac{C_{80,5}}{C_{90,5}} = \frac{24\,040\,016}{43\,949\,268} \approx 0{,}547$.

16. $1 - (\frac{5}{6})^n > 0{,}99 \Leftrightarrow n > \frac{\ln 0{,}01}{\ln(5/6)} \approx 25{,}26$: almeno 26 lanci (con 25 la probabilità è circa 0,9895, con 26 circa 0,9913).

17. a) $C_{8,4} \cdot \frac{1}{2^8} = \frac{70}{256} = \frac{35}{128} \approx 0{,}273$. **b)** $\frac{C_{8,7} + C_{8,8}}{256} = \frac{9}{256} \approx 0{,}035$.

18. a) $C_{4,1} \cdot \frac{1}{6} \cdot (\frac{5}{6})^3 = \frac{500}{1296} = \frac{125}{324} \approx 0{,}386$. **b)** $(\frac{5}{6})^4 + \frac{125}{324} = \frac{625}{1296} + \frac{500}{1296} = \frac{1125}{1296} = \frac{125}{144} \approx 0{,}868$.

19. $p_6(k)$ per $k = 0, \ldots, 6$: 0,1176; 0,3025; 0,3241; 0,1852; 0,0595; 0,0102; 0,0007 (somma 0,9998 per gli arrotondamenti; il valore esatto è 1). Il numero più probabile è $k = 2$; infatti $n \cdot p = 1{,}8$.

20. Traccia. E ed $\overline{E}$ sono incompatibili e $E \cup \overline{E} = \Omega$; per gli assiomi 3 e 2, $p(E) + p(\overline{E}) = 1$. Con $E = \Omega$, $\overline{E} = \varnothing$ e $p(\varnothing) = 1 - 1 = 0$ (vedi «I primi teoremi»).

21. Traccia. Si applica due volte il teorema dell'unione: $p((A \cup B) \cup E) = p(A \cup B) + p(E) - p((A \cup B) \cap E)$; per la proprietà distributiva $(A \cup B) \cap E = (A \cap E) \cup (B \cap E)$, la cui probabilità è $p(A \cap E) + p(B \cap E) - p(A \cap B \cap E)$. Sostituendo si ottiene la tesi. Verifica: $A = \{2, 4, 6\}$, $B = \{3, 6\}$, $E = \{5, 6\}$; $A \cup B \cup E = \{2, 3, 4, 5, 6\}$, probabilità $\frac{5}{6}$. Formula: $\frac{3 + 2 + 2 - 1 - 1 - 1 + 1}{6} = \frac{5}{6}$.

22. Traccia. Nelle estrazioni successive si scelgono le k posizioni delle bianche ($C_{m,k}$ modi), le bianche in ordine ($D_{b,k}$) e le nere in ordine ($D_{n-b,m-k}$). Poiché $D_{b,k} = C_{b,k} \cdot k!$, $D_{n-b,m-k} = C_{n-b,m-k} \cdot (m-k)!$ e $D_{n,m} = C_{n,m} \cdot m!$, il secondo membro diventa $\frac{C_{m,k} \cdot k!\,(m-k)!}{m!} \cdot \frac{C_{b,k} \cdot C_{n-b,m-k}}{C_{n,m}}$, e il primo fattore vale $C_{m,k} \cdot \frac{1}{C_{m,k}} = 1$.

23. Gli eventi «6 al lancio k» non sono incompatibili, quindi non si possono sommare (si contano più volte i casi con due o tre 6). Corretto: $1 - (\frac{5}{6})^3 = \frac{91}{216} \approx 0{,}421$.

24. Traccia. $\frac{C_{n,k}}{C_{n,k-1}} = \frac{n-k+1}{k}$ e $\frac{p^k (1-p)^{n-k}}{p^{k-1} (1-p)^{n-k+1}} = \frac{p}{1-p}$. Il rapporto è maggiore di 1 $\Leftrightarrow (n-k+1)\,p > k\,(1-p) \Leftrightarrow (n+1)\,p > k$. Le probabilità crescono finché $k < (n+1)\,p$ e poi decrescono: il massimo si ha nel massimo intero $k \le (n+1)\,p$ (con due massimi uguali se $(n+1)\,p$ è intero). Poiché $n\,p - 1 < (n+1)\,p - 1 \le k \le (n+1)\,p = n\,p + p < n\,p + 1$, il valore più probabile dista da $n \cdot p$ meno di 1.

25. Prove ripetute con $p = \frac{1}{4}$. **a)** $1 - (\frac{3}{4})^4 = \frac{175}{256} \approx 0{,}684$. **b)** $C_{4,1} \cdot \frac{1}{4} \cdot (\frac{3}{4})^3 = \frac{108}{256} = \frac{27}{64} \approx 0{,}422$.

26. a) $0{,}98^{50} \approx 0{,}364$. **b)** $0{,}98^{50} + 50 \cdot 0{,}02 \cdot 0{,}98^{49} + C_{50,2} \cdot 0{,}02^2 \cdot 0{,}98^{48} \approx 0{,}364 + 0{,}372 + 0{,}186 \approx 0{,}922$.

27. a) $0{,}8^5 = 0{,}32768 \approx 0{,}328$. **b)** $C_{5,4} \cdot 0{,}8^4 \cdot 0{,}2 + 0{,}8^5 = 0{,}4096 + 0{,}32768 = 0{,}73728 \approx 0{,}737$. **c)** La probabilità costante e l'indipendenza: stanchezza, fiducia dopo un canestro o tensione possono cambiare la probabilità da un tiro all'altro.

28. a) $1 - \frac{D_{365,25}}{365^{25}} \approx 0{,}569$. **b)** $1 - (\frac{364}{365})^{25} \approx 0{,}066$. Nel caso a) le coincidenze possibili sono tutte le coppie di studenti, $C_{25,2} = 300$; nel caso b) solo le 25 coppie formate da uno studente e la professoressa.

29. $\frac{C_{12,2} \cdot C_{10,2}}{C_{22,4}} = \frac{66 \cdot 45}{7315} = \frac{54}{133} \approx 0{,}406$. Almeno un ragazzo: $1 - \frac{C_{12,4}}{7315} = 1 - \frac{495}{7315} = \frac{6820}{7315} = \frac{124}{133} \approx 0{,}932$.

30. Bernoulli con $n = 10$, $p = 1/4$: $p_{10}(6) + p_{10}(7) + \ldots + p_{10}(10) = \frac{10\,343}{524\,288} \approx 0{,}0197$: circa 2 volte su 100. In media $n \cdot p = 10 \cdot \frac{1}{4} = 2{,}5$ risposte corrette.

31. a) La somma vale 1 (a meno degli arrotondamenti). **b)** $p_{20}(0) \approx 0{,}1216$, $p_{20}(1) \approx 0{,}2702$, $p_{20}(2) \approx 0{,}2852$, $p_{20}(3) \approx 0{,}1901$, poi i valori decrescono: il più probabile è $k = 2 = 20 \cdot 0{,}1$. **c)** L'istogramma è asimmetrico, concentrato sui valori piccoli, come quello di destra della Figura 14.5.

32. La frequenza ottenuta varia da una simulazione all'altra, ma per la legge dei grandi numeri è in genere vicina a 0,507 (con 1000 righe, scarti di qualche centesimo sono normali). Una formula possibile conta, per ogni numero della riga, quante volte compare: se il massimo è maggiore di 1 c'è una ripetizione.

33. a) $\frac{5}{8} \cdot \frac{4}{7} = \frac{5}{14}$; colori diversi: $\frac{C_{3,1} \cdot C_{5,1}}{C_{8,2}} = \frac{15}{28}$. **b)** $C_{4,1} \cdot \frac{3}{8} \cdot (\frac{5}{8})^3 = \frac{1500}{4096} = \frac{375}{1024} \approx 0{,}366$.

34. a) $C_{10,2} (\frac{1}{6})^2 (\frac{5}{6})^8 = \frac{1\,953\,125}{6\,718\,464} \approx 0{,}291$. **b)** $1 - (\frac{5}{6})^{10} - 10 \cdot \frac{1}{6} (\frac{5}{6})^9 \approx 1 - 0{,}1615 - 0{,}3230 \approx 0{,}515$.

c) $(n + 1)\, p = \frac{11}{6} \approx 1{,}83$: il più probabile è $k = 1$ ($p_{10}(1) \approx 0{,}323$, contro $p_{10}(0) \approx 0{,}162$ e $p_{10}(2) \approx 0{,}291$).

35. $1 - (\frac{364}{365})^n > \frac{1}{2} \Leftrightarrow n > \frac{\ln 0{,}5}{\ln(364/365)} \approx 252{,}65$: servono almeno 253 persone (con 252 la probabilità è circa 0,4991, con 253 circa 0,5005). Nel problema dei compleanni basta una coincidenza fra *due persone qualsiasi*, e con 23 persone le coppie sono già 253; qui conta solo la coincidenza con un giorno fissato, e servono 253 persone per avere 253 «coppie» con te.