

Probabilità discreta

2.1 Introduzione

La parola “probabilità” è ampiamente utilizzata nel linguaggio comune, a volte anche a sproposito. In questo capitolo vogliamo cercare di dare una presentazione più precisa e meno ambigua di questo concetto, illustrandone gli aspetti matematici e indicandone alcuni utilizzi in ambito scientifico e biologico. Per esempio, saremo in grado di spiegare come rispondere a domande del tipo:

- *Se tiro un dado non truccato, con che probabilità ottengo 6?*
- *Se tiro due dadi non truccati, con che probabilità ottengo 6?*
- *Se Leonardo e Martina hanno entrambi sangue con fattore Rh positivo, con che probabilità avranno un figlio con fattore Rh negativo?*
- *Qual è la probabilità che un bimbo italiano pesi alla nascita più di 1500 g?*
- *Cosa vuol dire la frase “un esame del DNA ha dato una probabilità del 97.5% che il sangue trovato sul luogo del delitto sia dell’accusato”?*
- *Se ho vissuto fino a 60 anni, che probabilità ho di raggiungere i 70?*
- *Cosa vuol dire la frase “questo test diagnostico ha un valore predittivo della malattia di oltre il 69%”?*

Detto con parole un po’ vaghe, la probabilità è una misura di quanto sia plausibile che avvenga un dato evento fra tutti quelli possibili. Per studiare con tecniche probabilistiche un dato fenomeno possiamo allora procedere con i seguenti passi:

- a) identificare quali sono i possibili “eventi” che si possono verificare;
- b) attribuire, in seguito a ipotesi teoriche o a misurazioni di laboratorio, delle probabilità ad alcuni di questi eventi, scelti fra i più semplici, soddisfacendo alcune regole basilari;
- c) calcolare, partendo dalle probabilità degli eventi semplici, dalle regole basilari e da ipotesi sullo sviluppo del fenomeno, le probabilità degli eventi più complessi, in modo da effettuare predizioni teoriche sul fenomeno studiato;
- d) verificare sperimentalmente le predizioni, con l’obiettivo di confermare o smentire (e quindi modificare) le ipotesi fatte nei passi b) e c).

Nelle prossime sezioni discuteremo più in dettaglio, con numerosi esempi, questa procedura. Vale la pena però di osservare fin da ora due cose. Prima di tutto, questa procedura è tipica del metodo scientifico: si parte da ipotesi esplicitamente dichiarate sul fenomeno che si vuole studiare; si deducono, tramite ragionamenti precisi e spesso matematici, predizioni sullo sviluppo del fenomeno; si verificano sperimentalmente queste predizioni; e se le predizioni non sono corrette si modificano di conseguenza le ipotesi iniziali, e si ripete il procedimento. Il processo di validazione di ipotesi esplicite tramite verifica sperimentale di predizioni ottenute con ragionamenti logici e matematici è il cuore del metodo scientifico contemporaneo.

La seconda osservazione è che nella procedura sopra descritta si vedono bene i ruoli complementari svolti dalla matematica e dalla biologia (o altra scienza coinvolta). La biologia fornisce gli oggetti di studio e le ipotesi sul fenomeno; la matematica fornisce le regole formali da soddisfare e le tecniche di ragionamento da applicare. Inutile dire che una parte non irrilevante del lavoro consiste nel far comunicare questi due aspetti: esprimere i concetti biologici in modo da poterne dare una trattazione matematica, e, viceversa, riuscire a tradurre le conclusioni matematiche in termini biologicamente significativi. In questo capitolo vedremo qualche esempio (estremamente semplice...), soprattutto in ambito genetico.

Infine, un ultimo commento. Quando è necessario usare un approccio di tipo probabilistico a un problema? Generalizzando molto, si può dire che è necessario farlo quando non si ha una conoscenza completa del modo con cui si sviluppa il fenomeno. Questo può essere dovuto all'eccessiva complessità del fenomeno (dovuta a una sua complessità intrinseca o all'eccessivo numero di fattori da tenere in considerazione), agli errori necessariamente introdotti da qualsiasi strumento di misura, o anche banalmente alla nostra ignoranza. Quale che sia il motivo, la mancanza di conoscenza completa ci costringe ad abbandonare un approccio deterministico e a perseguire un approccio probabilistico, che fornisce appunto una misura della maggiore o minore plausibilità di certi risultati rispetto ad altri, senza però negare che talvolta anche l'improbabile possa verificarsi.

2.2 Eventi

Come accennato prima, il primo passo necessario per lo studio probabilistico di un fenomeno è stabilire quali sono gli eventi possibili. Matematicamente parlando, lo *spazio¹ degli eventi* è un insieme Ω , che può essere qualsiasi. Un *evento semplice* (o *evento elementare*) è un elemento dello spazio degli eventi; un *evento composto* è un sottoinsieme dello spazio degli eventi.

Alcuni esempi chiariranno meglio l'uso di questa terminologia.

ESEMPIO 2.1 Lo spazio degli eventi Ω del fenomeno “lancio di un dado a sei facce” è l'insieme dei possibili risultati del lancio del dado, cioè $\Omega = \{1, 2, 3, 4, 5, 6\}$.

¹ In questo contesto, “spazio” è semplicemente sinonimo di “insieme”, senza nessuna connotazione geometrica.

L'elemento "6" è un evento semplice; il sottoinsieme "pari" (cioè $\{2, 4, 6\}$) è un evento composto.

ESEMPIO 2.2 Nello studio del peso alla nascita dei bimbi italiani, una scelta dello spazio degli eventi potrebbe essere l'intervallo $[200, 5000]$ della retta reale, misurando i pesi in grammi. Così facendo, però, stiamo facendo un'ipotesi: stiamo escludendo a priori che possano esservi neonati di peso inferiore ai 200 grammi o superiore ai 5 chili. Per non introdurre ipotesi limitative a questo stadio, possiamo invece scegliere come spazio degli eventi l'intera semiretta \mathbb{R}^+ dei numeri reali positivi, in modo da assumere soltanto che un neonato pesi qualcosa...

Osservazione 2.1 Mentre lo spazio degli eventi dell'Esempio 2.1 era finito, lo spazio degli eventi in questo esempio è infinito. Lo studio delle probabilità nel caso di spazi degli eventi infiniti è più delicato che nel caso di spazi degli eventi finiti, e richiede tecniche più avanzate. Per questo motivo, in questo capitolo tranne rare eccezioni ci limiteremo a studiare spazi degli eventi finiti, rimandando lo studio della *probabilità continua* (probabilità su insiemi infiniti; la probabilità su insiemi finiti è detta *probabilità discreta*) a un capitolo successivo.

ESEMPIO 2.3 Un gene è una zona di un cromosoma con una funzione precisa, quale dirigere la codifica di una data proteina. Il numero esatto di geni nell'uomo non è ancora noto, ma si stima siano fra i venti e i trentamila. Anche la lunghezza di ciascun gene è molto variabile, da qualche centinaio a diverse decine di migliaia di basi; in ogni caso, i geni coprono meno del 10% dell'intero DNA umano. L'intero DNA umano è composto da circa 3.3 miliardi di (paia di) basi; meno dell'1 per mille (circa 3 milioni di basi) può variare da individuo a individuo. Le possibili varianti di quei geni (o, in generale, di quelle sezioni del DNA) che possono effettivamente variare sono dette *alleli* (o *varianti alleliche*). Supponiamo che un dato gene possa avere solo due alleli, che indicheremo con le lettere A e a . Il corredo genetico di ciascun individuo è composto da coppie di cromosomi, per cui ciascun gene compare due volte. Il *genotipo* di un individuo (relativo a quel gene) è la coppia di alleli di quel gene presente nel suo corredo cromosomico. Se entrambi gli alleli sono uguali, per cui sono la coppia AA oppure la coppia aa , diremo che l'individuo è *omozigote* (rispetto a quel gene); se invece sono diversi, diremo che è *eterozigote*. Nota che siccome non c'è modo di distinguere un cromosoma dall'altro, non possiamo distinguere la coppia Aa dalla coppia aA , per cui il genotipo eterozigote viene indicato con Aa . Riassumendo, per studiare il genotipo di un individuo relativo a un gene con due alleli lo spazio degli eventi adatto è $\Omega = \{AA, Aa, aa\}$. L'evento "omozigote" è l'evento composto $\{AA, aa\}$, mentre l'evento "eterozigote" è l'evento $\{Aa\}$.

ESEMPIO 2.4 Nel caso di un gene con due soli alleli, l'evento "eterozigote" è un evento semplice; questo non è più vero se gli alleli sono tre o più. Per esempio, nel caso di un gene con tre alleli A_1 , A_2 e A_3 , lo spazio degli eventi è $\Omega = \{A_1A_1, A_1A_2, A_1A_3, A_2A_2, A_2A_3, A_3A_3\}$, e l'evento "eterozigote" è l'evento composto $\{A_1A_2, A_1A_3, A_2A_3\}$.

Osservazione 2.2 Seguendo le definizioni alla lettera, anche un sottoinsieme $\{a\}$ composto da un unico elemento a di uno spazio degli eventi Ω sarebbe un evento composto, cosa francamente ridicola. Per questo motivo, con un lieve abuso di terminologia chiameremo eventi semplici anche gli eventi “composti” formati da un unico elemento.

ESEMPIO 2.5 Lo spazio degli eventi del fenomeno “lancio di 2 dadi distinti a sei facce” è l’insieme delle coppie *ordinate* (in quanto sappiamo distinguere un dado dall’altro, per esempio perché uno è rosso e uno verde, per cui abbiamo un primo risultato e un secondo risultato) dei possibili esiti dei lanci di ciascun dado, cioè

$$\begin{aligned}\Omega = \{ & (1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), \\ & (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6), \\ & (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6) \} .\end{aligned}$$

Usando la terminologia introdotta nella Sezione 1.8, possiamo dire che Ω è il prodotto cartesiano dello spazio degli eventi del lancio del primo dado con lo spazio degli eventi del lancio del secondo dado:

$$\Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\} .$$

ESEMPIO 2.6 Come nel caso del genotipo, lo spazio degli eventi del fenomeno “lancio di 2 dadi indistinguibili a sei facce” non è costituito dalle coppie ordinate, in quanto una volta effettuato il lancio non siamo più in grado di distinguere un dado dall’altro. Se indichiamo una coppia non ordinata semplicemente giustapponendo i due elementi della coppia (come fatto negli Esempi 2.3 e 2.4) lo spazio degli eventi in questo caso è

$$\Omega = \{11, 12, 13, 14, 15, 16, 22, 23, 24, 25, 26, 33, 34, 35, 36, 44, 45, 46, 55, 56, 66\} .$$

Osservazione 2.3 La coppia non ordinata “23” può essere rappresentata dall’insieme A costituito da due coppie ordinate, le coppie che corrispondono alla coppia non ordinata “23”: $A = \{(2, 3), (3, 2)\}$. Invece la coppia ordinata “11” può essere rappresentata dall’unica coppia ordinata corrispondente $\{(1, 1)\}$. In altre parole, all’evento semplice coppia non ordinata “ ij ” con $i \neq j$ dell’Esempio 2.6 possiamo associare l’evento composto $\{(i, j), (j, i)\}$ dell’Esempio 2.5, mentre all’evento semplice coppia non ordinata “ ii ” dell’Esempio 2.6 associamo l’evento semplice $\{(i, i)\}$ dell’Esempio 2.5.

ESEMPIO 2.7 Lo spazio degli eventi del fenomeno “somma del lancio di 2 dadi” è invece $\Omega = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ (dove stavolta “12” è il numero dodici e non la coppia non ordinata composta da 1 e 2, e così via). Nota che otteniamo lo stesso spazio degli eventi sia che i due dadi siano distinguibili sia che non lo siano.

Osservazione 2.4 Anche in questo caso possiamo associare a ciascun evento semplice k un evento (semplice o composto) dell’Esempio 2.5: il sottoinsieme delle

coppie ordinate la cui somma sia uguale a k . Per esempio, l'evento "somma 2" è associato all'evento semplice $\{(1, 1)\}$, mentre l'evento "somma 5" è associato all'evento composto $\{(1, 4), (2, 3), (3, 2), (4, 1)\}$.

Osservazione 2.5 Il fatto che lo spazio degli eventi negli ultimi due esempi è chiaramente derivato da quello dell'Esempio 2.5 fa sospettare che ci sarà un modo per assegnare delle probabilità agli eventi degli Esempi 2.6 e 2.7 partendo da probabilità assegnate agli eventi dell'Esempio 2.5.

L'interpretazione degli elementi e dei sottoinsiemi di un insieme come eventi ha suggerito l'introduzione in teoria della probabilità di una nomenclatura che è una semplice traduzione della terminologia insiemistica usuale. I termini più usati sono:

- lo spazio degli eventi Ω è chiamato *evento certo* (qualcosa deve accadere, Ω contiene tutto ciò che può accadere, e quindi Ω accade di certo);
- l'insieme vuoto \emptyset è chiamato *evento impossibile* (qualche evento deve accadere, \emptyset non contiene eventi, e quindi non può accadere);
- dati due eventi A e B , l'evento che si verifica se succede A oppure se succede B (o entrambi) è l'*evento unione* $A \cup B$, e coincide effettivamente con l'unione insiemistica di A e B . In alcuni testi, l'evento unione è indicato con " $A + B$ " oppure con " $A \text{ o } B$ ".
- dati due eventi A e B , l'evento che si verifica se succedono sia A sia B è l'*evento intersezione* $A \cap B$, e coincide effettivamente con l'intersezione insiemistica di A e B . In alcuni testi, l'evento intersezione è indicato con " $A \cdot B$ " oppure con " $A \text{ e } B$ ".
- dato un evento A , l'evento che si verifica se e solo se A non succede è l'*evento complementare* $\Omega \setminus A$, che coincide effettivamente con la differenza insiemistica di Ω e A . In alcuni testi, l'evento complementare è indicato con A^c o con \bar{A} , o ancora con *non A*.

ESEMPIO 2.8 Nel caso del lancio di un dado a sei facce, possiamo considerare gli eventi "pari" $A = \{2, 4, 6\}$ e "maggiore di 3" $B = \{4, 5, 6\}$. Allora l'evento unione $A \cup B = \{2, 4, 5, 6\}$ corrisponde proprio all'evento "pari oppure maggiore di 3", e l'evento intersezione $A \cap B = \{4, 6\}$ corrisponde proprio all'evento "pari e maggiore di 3". Infine, l'evento complementare di "pari" è (ovviamente) "dispari", cioè $\Omega \setminus A = \{1, 3, 5\}$.

Usando unione e intersezione, possiamo introdurre altri termini di uso comune in questo contesto. Diremo che due eventi A e B sono *incompatibili* (o *mutuamente esclusivi*) se sono disgiunti, cioè se $A \cap B = \emptyset$. In altre parole, dire che A e B sono incompatibili vuol dire che se si verifica A allora B non può verificarsi, e viceversa.

Invece diremo che A e B sono *esaustivi* se $A \cup B = \Omega$. In altre parole, dire che A e B sono esaustivi vuol dire che siamo sicuri che uno dei due si verificherà (e magari tutti e due).

ESEMPIO 2.9 Sempre nel solito caso del lancio di un dado a sei facce, gli eventi

“pari” $A = \{2, 4, 6\}$ e “dispari” $B = \{1, 3, 5\}$ sono sia incompatibili (intersezione vuota: $A \cap B = \emptyset$) sia esaustivi (l’unione è tutto lo spazio degli eventi: $A \cup B = \{1, 2, 3, 4, 5, 6\}$). Gli eventi “minore o uguale di 5” $C = \{1, 2, 3, 4, 5\}$ e “maggiore di 4” $D = \{5, 6\}$ sono esaustivi (ogni numero fra 1 e 6 è minore o uguale a 5 oppure maggiore di 4: $C \cup D = \{1, 2, 3, 4, 5, 6\}$) ma non incompatibili (5 è un numero sia minore o uguale a 5 sia maggiore di 4: $C \cap D = \{5\}$). Gli eventi “maggiore di 5” $E = \{6\}$ e “minore di 3” $F = \{1, 2\}$ sono incompatibili (nessun numero è contemporaneamente maggiore di 5 e minore di 3: $E \cap F = \emptyset$) ma non esaustivi (esistono numeri che non sono maggiori di 5 o minori di 3: per esempio, $4 \notin E \cup F$). Infine, gli eventi “pari” e “minore di 3” non sono né incompatibili né esaustivi: infatti $A \cap F = \{2\}$ e $5 \notin A \cap F$.

Osservazione 2.6 Un evento e il suo complementare sono sempre incompatibili ed esaustivi: infatti A e $\Omega \setminus A$ sono sempre disgiunti e danno come unione tutto Ω . Viceversa, due eventi sono contemporaneamente incompatibili ed esaustivi se e solo se (perché?) sono uno il complementare dell’altro.

Osservazione 2.7 Due eventi semplici distinti sono sempre incompatibili (perché?).

Più in generale, diremo che n eventi A_1, \dots, A_n sono *esaustivi* se la loro unione è tutto lo spazio degli eventi, cioè $\Omega = A_1 \cup \dots \cup A_n$. Se inoltre sono anche a 2 a 2 incompatibili (cioè $A_i \cap A_j = \emptyset$ per qualsiasi coppia di indici i e j distinti compresi fra 1 ed n) allora diremo che formano una *partizione* dello spazio degli eventi. In altre parole, una partizione è una famiglia di eventi di cui siamo certi che se ne verificherà *esattamente* uno: almeno uno (esaustivi) e non più di uno (a 2 a 2 incompatibili).

ESEMPIO 2.10 Gli eventi $A_1 = \{1, 2\}$, $A_2 = \{3, 4\}$ e $A_3 = \{5, 6\}$ sono una partizione del solito spazio degli eventi $\Omega = \{1, 2, 3, 4, 5, 6\}$. Infatti, $\Omega = A_1 \cup A_2 \cup A_3$, e $A_1 \cap A_2 = A_1 \cap A_3 = A_2 \cap A_3 = \emptyset$.

2.3 Distribuzioni di probabilità

Adesso che sappiamo cosa sono gli eventi, possiamo passare al passo successivo e cercare di capire come attribuire delle probabilità agli eventi semplici.

Ci sono tipicamente due metodi per attribuire una probabilità agli eventi semplici:

- *a priori*, in funzione di ipotesi teoriche sulla natura degli eventi semplici (e a volte si parla di *probabilità teorica* o *stimata*);
- *a posteriori*, in funzione di misurazioni effettuate sul fenomeno che si sta studiando (e a volte si parla di *probabilità misurata*).

In entrambi i casi (e questa sezione e la prossima sono dedicate a presentare numerosi esempi dei due metodi, in modo da darti un’idea di come si procede) l’attribuzione delle probabilità deve sottostare ad alcune semplici condizioni, senza le quali non sarebbe corretto parlare di probabilità. Le condizioni sono solo tre:

- (1) *l'evento certo ha probabilità 1, e l'evento impossibile probabilità 0.* Questa è una convenzione, ma utile. Avremmo potuto assegnare probabilità 27 all'evento certo e -12 all'evento impossibile, ma ci saremmo complicati la vita inutilmente.
- (2) *tutti gli eventi hanno una probabilità compresa fra 0 e 1 inclusi.* Puro buon senso: stiamo chiedendo che nulla possa essere più probabile dell'evento certo o più improbabile dell'evento impossibile. Nota che potrebbero esserci eventi con probabilità 1 distinti dall'intero spazio degli eventi, ed eventi con probabilità 0 che non sono l'insieme vuoto. Per esempio, possiamo ragionevolmente assumere che la probabilità che un neonato pesi meno di 5 tonnellate sia 1, e che pesi meno di 5 grammi sia 0.
- (3) *la probabilità che avvenga uno di due eventi incompatibili è la somma della probabilità dei due eventi².* Di nuovo, è una richiesta molto ragionevole. Se due eventi sono incompatibili, o accade uno o accade l'altro (o nessuno dei due); quindi richiedere che la probabilità che accada uno qualsiasi dei due sia la somma della probabilità è in accordo con la nostra idea intuitiva di probabilità. In particolare, questo si applica agli eventi semplici, che sono sempre incompatibili. La probabilità che tirando un dado venga 2 o 5 sarà la somma della probabilità che venga 2 e della probabilità che venga 5.

Vediamo ora come mescolando queste condizioni a ipotesi teoriche opportune possiamo attribuire a priori delle probabilità.

ESEMPIO 2.11 Il lancio di un dado a 6 facce non truccato. Nell'Esempio 2.1 abbiamo visto che lo spazio degli eventi è $\Omega = \{1, 2, 3, 4, 5, 6\}$. Dire che il dado non è truccato vuol dire che tutte le facce sono uguali; non c'è nulla che favorisca una faccia rispetto alle altre. Quindi è naturale attribuire a ciascun risultato la stessa probabilità, che indicheremo con p . Ma quant'è p ? Abbiamo detto che, per la proprietà (1), la probabilità dell'evento certo dev'essere 1. Siccome ci sono 6 risultati possibili, ognuno dei quali ha probabilità p , dobbiamo avere $6p = 1$, cioè $p = 1/6$ (attento che qui stiamo usando la proprietà (3); ne ripareremo nella Sezione 2.5). Quindi l'ipotesi "dado non truccato" si traduce nell'assegnare la stessa probabilità di $1/6$ a tutti i possibili risultati del lancio di un dado.

ESEMPIO 2.12 Il lancio di due dadi distinti a 6 facce non truccati. Nell'Esempio 2.5 abbiamo visto che lo spazio degli eventi è composto da 36 possibili risultati. Nel caso di 2 dadi, la frase "non truccati" non significa solo che le facce di ciascun dado sono tutte uguali, ma anche che i due dadi sono *indipendenti*, nel senso che il risultato di un dado non influenza in alcun modo il risultato dell'altro (ripareremo del concetto di indipendenza nella Sezione 2.6). Questo vuol dire che non c'è nulla che favorisca un risultato rispetto agli altri; tutti i 36 risultati sono *equiprobabili*. Ragionando come nell'esempio precedente dobbiamo quindi assegnare una probabilità di $1/36$ a ciascun risultato possibile.

² Nel caso di spazio degli eventi infinito potrebbe servire una proprietà più forte, che discuteremo nella Curiosità 2.1.

Osservazione 2.8 Quando assegnamo la stessa probabilità p a tutti gli elementi di uno spazio degli eventi Ω , diremo che abbiamo una *distribuzione di probabilità uniforme*. Nota che se Ω contiene N elementi allora si deve avere (perché?) $p = 1/N$.

ESEMPIO 2.13 Il lancio di due dadi a 6 facce indistinguibili non truccati. Nell'Esempio 2.6 abbiamo visto che lo spazio degli eventi è composto da 21 eventi semplici; ma, contrariamente al caso dei due dadi distinti, i vari eventi non sono equiprobabili. Il punto è che abbiamo due tipi di eventi semplici: le coppie non ordinate composte da due numeri diversi (per esempio, “23” o “46”), e le coppie ordinate composte da due copie dello stesso numero (per esempio, “44” o “11”). Le nostre ipotesi teoriche (dadi indistinguibili non truccati) ci portano ad attribuire ai risultati del primo tipo una probabilità doppia rispetto a quelli del secondo tipo. Infatti, nell'Osservazione 2.3 abbiamo visto che il risultato “23” corrisponde ai due eventi semplici (2, 3) e (3, 2) dello spazio degli eventi dei dadi distinti. Siccome abbiamo visto che ciascuno di questi due eventi ha probabilità $1/36$, usando la proprietà (3) vediamo che l'evento “23” deve avere probabilità $1/36 + 1/36 = 2/36 = 1/18$. Invece il risultato “11” si può ottenere in un solo modo: entrambi i dadi devono dare 1. Siccome questo evento ha probabilità $1/36$, anche l'evento “11” ha probabilità $1/36$. Riassumendo, in questo caso otteniamo probabilità $1/18$ per le coppie non ordinate “ ij ” con $i \neq j$ che variano da 1 a 6, e probabilità $1/36$ per le coppie non ordinate “ ii ”, con i che varia da 1 a 6; la distribuzione di probabilità non è uniforme. Nota che ci sono 15 coppie del primo tipo e 6 del secondo, e che

$$15 \cdot \frac{1}{18} + 6 \cdot \frac{1}{36} = 1 ,$$

come dev'essere (perché?).

ESEMPIO 2.14 La somma del lancio di due dadi a 6 facce non truccati. Nell'Osservazione 2.4 abbiamo visto come associare a ciascun evento “somma uguale a k ” l'evento composto dell'Esempio 2.5 costituito da tutte le coppie ordinate con somma k . Siccome ogni coppia ordinata ha probabilità $1/36$, l'evento “somma uguale a k ” deve avere probabilità uguale a $1/36$ per il numero di coppie ordinate con somma k . Per esempio, siccome 2 si può ottenere solo come somma di (1, 1), l'evento “somma uguale a 2” ha probabilità $1/36$. Invece, “5” si può ottenere come somma in quattro modi diversi, per cui l'evento “somma uguale a 5” ha probabilità $4/36 = 1/9$. Ti lascio il compito di verificare la correttezza della Tabella 2.1, che riassume la distribuzione di probabilità in questo esempio.

ESEMPIO 2.15 Come già detto nell'Esempio 2.3, il corredo genetico di ogni individuo è composto da coppie di cromosomi. Un figlio riceve un cromosoma da ciascun genitore; quindi in caso di geni con più alleli il genotipo del figlio comprende un allele proveniente dal padre e un allele proveniente dalla madre. Un'importante ipotesi teorica che sottende la trasmissione del genotipo da genitori a figli è la *legge di disgiunzione di Mendel: le possibili trasmissioni di alleli da genitori a figli sono tutte equiprobabili*. Vediamo come questa legge ci permetta di stabilire la probabilità che un figlio abbia un certo genotipo conoscendo i genotipi dei genitori.

<i>Somma</i>	<i>Coppie ordinate</i>	<i>Probabilità</i>
2	(1, 1)	1/36
3	(1, 2), (2, 1)	1/18
4	(1, 3), (2, 2), (3, 1)	1/12
5	(1, 4), (2, 3), (3, 2), (4, 1)	1/9
6	(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)	5/36
7	(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)	1/6
8	(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)	5/36
9	(3, 6), (4, 5), (5, 4), (6, 3)	1/9
10	(4, 6), (5, 5), (6, 4)	1/12
11	(5, 6), (6, 6)	1/18
12	(6, 6)	1/36

TABELLA 2.1

Prendiamo un gene con due possibili alleli, che indicheremo come al solito A e a . Supponiamo che il padre abbia genotipo AA e la madre genotipo Aa . Allora il figlio può ricevere gli alleli dai genitori in quattro modi diversi: il primo allele A del padre e il primo allele A della madre (genotipo del figlio AA); il primo allele A del padre e il secondo allele a della madre (genotipo del figlio Aa); il secondo allele A del padre e il primo allele A della madre (genotipo del figlio AA); il secondo allele A del padre e il secondo allele a della madre (genotipo del figlio Aa). La legge di disgiunzione di Mendel ci dice che queste quattro eventualità sono equiprobabili; quindi ciascuna di esse ha probabilità $1/4$. Siccome otteniamo come genotipo del figlio AA in due casi e Aa negli altri due, possiamo concludere che se i genitori hanno genotipo AA e Aa allora il figlio ha genotipo AA con probabilità $1/4 + 1/4 = 1/2$ e genotipo Aa con probabilità $1/2$. La Tabella 2.2 (che avrai cura di verificare) riassume le varie possibilità: ciascuna casella centrale contiene le probabilità che il figlio abbia un certo genotipo sapendo che il padre ha il genotipo indicato in quella colonna e la madre il genotipo indicato nella riga.

	AA	Aa	aa
AA	$AA : 1$ $Aa : 0$ $aa : 0$	$AA : 1/2$ $Aa : 1/2$ $aa : 0$	$AA : 0$ $Aa : 1$ $aa : 0$
Aa	$AA : 1/2$ $Aa : 1/2$ $aa : 0$	$AA : 1/4$ $Aa : 1/2$ $aa : 1/4$	$AA : 0$ $Aa : 1/2$ $aa : 1/2$
aa	$AA : 0$ $Aa : 1$ $aa : 0$	$AA : 0$ $Aa : 1/2$ $aa : 1/2$	$AA : 0$ $Aa : 0$ $aa : 1$

TABELLA 2.2

ESEMPIO 2.16 Un tipico esempio di gene con due alleli è quello che determina la

presenza o meno del fattore Rh nel sangue. Inoltre, abbiamo un allele dominante e un allele recessivo. Un allele è *dominante* quando la sua presenza nel genotipo causa la comparsa di una certa caratteristica (detta *fenotipo*) nell'individuo indipendentemente da quale sia l'altro allele; è *recessivo* altrimenti. Per esempio, l'allele Rh^+ è dominante sull'altro allele Rh^- : la sua presenza nel genotipo implica la presenza nel sangue del fattore Rh (fenotipo Rh^+), mentre l'unico modo perché nel sangue non ci sia il fattore Rh (fenotipo Rh^-) è che entrambi gli alleli del genotipo siano Rh^- . Se Leonardo e Martina hanno fenotipo Rh^+ , possono avere genotipo omozigote Rh^+Rh^+ oppure eterozigote Rh^+Rh^- . Se anche uno solo dei due ha genotipo omozigote Rh^+Rh^+ , la tabella precedente ci dice che la probabilità che abbiano un figlio con fenotipo Rh^- (e quindi genotipo Rh^-Rh^-) è zero. Invece, se entrambi hanno genotipo eterozigote Rh^+Rh^- , la probabilità che abbiano un figlio con fenotipo Rh^- è $1/4$.

Osservazione 2.9 L'esempio precedente ancora non risponde del tutto alla domanda 3 della Sezione 2.1, in quanto non sappiamo ancora con che probabilità un individuo con fenotipo Rh^+ ha genotipo omozigote o eterozigote. Riprenderemo l'argomento nella Sezione 2.7.

2.4 Frequenze relative

Se non abbiamo alcuna ipotesi teorica a cui aggrapparci per attribuire a priori delle probabilità ai possibili eventi del fenomeno che stiamo studiando, dobbiamo rimboccarci le maniche e metterci a misurare. Ripetiamo lo stesso esperimento un numero (possibilmente) grande di volte: diremo *frequenza assoluta* (o *numero di successi*) di un evento il numero di volte che l'evento si verifica ripetendo l'esperimento, e *frequenza relativa* il rapporto fra la frequenza assoluta e il numero degli esperimenti (o *numero di tentativi*) effettuati:

$$\text{frequenza relativa} = \frac{\text{numero di successi}}{\text{numero di tentativi}}.$$

L'attribuzione a posteriori della probabilità consiste allora nell'*assegnare come probabilità a ciascun evento la sua frequenza relativa*.

ESEMPIO 2.17 Supponi di avere un dado a 6 facce, e di non sapere se è truccato o meno, per cui non puoi attribuire le probabilità a priori. Incarichi il tuo assistente di passare una notte insonne a tirare 1000 volte il dado. Il giorno dopo, lui ti consegna la Tabella 2.3.

Usando la frequenza relativa per calcolare la probabilità, possiamo quindi dire che, in base ai dati disponibili, la probabilità che un lancio di questo dado abbia come risultato 1 è 0.12, che abbia come risultato 2 è 0.412, e così via. In particolare, in assenza di altre informazioni, siamo portati a concludere che questo dado è truccato, in quanto favorisce in maniera pesante il 2 e sfavorisce il 3 e il 6.

Risultato	Frequenza assoluta	Frequenza relativa
1	120	$120/1000 = 0.12$
2	412	$412/1000 = 0.412$
3	66	$66/1000 = 0.066$
4	222	$222/1000 = 0.222$
5	127	$127/1000 = 0.127$
6	53	$53/1000 = 0.053$

TABELLA 2.3

Osservazione 2.10 Attenzione: qui c'è un punto delicato. Lo stesso esperimento effettuato con una dado non truccato *avrebbe potuto dare lo stesso risultato*. Infatti, come abbiamo visto, qualsiasi sequenza di risultati di lanci di un dado non truccato ha la stessa probabilità *non nulla* di verificarsi. Però, come vedremo nella Sezione 2.9, la probabilità di ottenere quella tabella tirando 1000 volte un dado non truccato è circa $4.15 \cdot 10^{-112}$, mentre la probabilità di ottenere tirando 1000 volte un dado non truccato una tabella in cui i 6 risultati hanno una frequenza relativa di $1/6$ è di circa $6.85 \cdot 10^{-8}$, estremamente più grande (anche se in assoluto piccola!). In altre parole, è possibile ma estremamente improbabile che un dado non truccato dia il risultato che abbiamo ottenuto.

Osservazione 2.11 La distinzione sottile da tenere presente è se stiamo usando le frequenze relative per *verificare una probabilità a priori* oppure per *attribuire una probabilità a posteriori*. Ne riparlamo fra poco.

L'esempio precedente illustra una possibilità per l'attribuzione delle probabilità a posteriori tramite la ripetizione dello stesso esperimento. Un altro caso possibile è invece tramite l'esame più o meno esaustivo di una popolazione.

ESEMPIO 2.18 Vogliamo vedere qual è la probabilità che una cellula epiteliale di ratto subisca una mutazione quando sottoposta a una radioattività di 1 Röntgen (1 Röntgen — abbreviato R — corrisponde a $2.58 \cdot 10^{-4}$ C/kg, dove C sta per Coulomb, la misura di carica elettrica, e $1 \text{ C} = 1 \text{ A} \cdot \text{s}$). Sottoponiamo 1 000 di cellule epiteliali di ratto ciascuna a una radiazione di 1 R; se ne sono mutate 5, possiamo dire che la probabilità di mutazione è $5/1000 = 0.005$.

ESEMPIO 2.19 In un esperimento vengono pesate 15 cavia, ottenendo i seguenti pesi in grammi:

28 32 37 29 31 30 32 26 32 27 29 30 28 31 31 .

Qual è la probabilità che una cavia presa a caso fra queste pesi meno di 29 grammi? Per rispondere a questa domanda contiamo: 4 cavia su 15 pesano meno di 29 grammi; quindi la probabilità è $4/15$. Analogamente, 6 cavia su 15 pesano almeno 31 grammi, per cui la probabilità che una cavia presa a caso fra queste pesi almeno 31 grammi è $6/15$.

ESEMPIO 2.20 All'Ospedale S. Camillo di Roma, nel periodo 16/4–15/10 2002 hanno registrato il peso dei bambini nati nel reparto maternità. I risultati sono riassunti nella Tabella 2.4.

<i>Peso alla nascita</i>	<i>Frequenza assoluta</i>	<i>Frequenza relativa</i>
≤ 1000 g	9	$9/116 \simeq 8\%$
1001–1500 g	21	$21/116 \simeq 18\%$
1501–2500 g	50	$50/116 \simeq 43\%$
> 2500 g	36	$36/116 \simeq 31\%$
Totale	116	$116/116 = 100\%$

TABELLA 2.4

Qual è la probabilità che un neonato preso a caso fra questi 116 pesi più di 1500 g? Siccome 86 neonati su 116 pesano più di 1500 g, la risposta è $86/116 \simeq 74\%$.

Osservazione 2.12 Siccome le probabilità sono numeri compresi fra 0 e 1, sono spesso indicate come percentuali (cioè come frazioni con denominatore uguale a 100).

Osservazione 2.13 Attenzione: l'Esempio 2.20 non ci permette di concludere che la probabilità che un neonato italiano pesi più di 1500 g è il 74% circa. Il fatto è che i 116 neonati studiati in questo esperimento sono solo un *campione* dell'intera *popolazione* di neonati italiani. Per essere certi che le frequenze relative ci diano l'esatta probabilità per l'intera popolazione dovremmo misurare il peso di tutti i neonati italiani. In molte situazioni sperimentali misurare tutti i membri di una popolazione è assolutamente impossibile; l'unica possibilità è studiare un campione che speriamo *rappresentativo* dell'intera popolazione, dove “rappresentativo” vuol dire che ha una distribuzione delle probabilità simile a quella dell'intera popolazione. Se il campione è scelto bene, le frequenze relative del campione saranno probabilmente vicine alle frequenze relative dell'intera popolazione. Uno dei principali compiti della Statistica è esattamente misurare quale errore si compie misurando solo un campione invece dell'intera popolazione; ne riparleremo.

Osservazione 2.14 Nei due esempi precedenti abbiamo usato la frase “preso a caso”, che vuol dire che ogni membro della popolazione ha la stessa probabilità di essere scelto. Se questa condizione non è soddisfatta, le frequenze relative non danno la probabilità dell'evento. Per esempio, la probabilità che un neonato prematuro fra i 116 esaminati pesi più di 1500 g è sicuramente molto inferiore al 74%; e infatti, imporre la condizione che il neonato sia prematuro vuol dire che non stiamo scegliendo un neonato a caso, in quanto stiamo escludendo a priori i neonati non prematuri. La scelta a caso deve avvenire tramite un meccanismo che non dipenda da caratteristiche specifiche dei componenti della popolazione. Per esempio, per ottenere una scelta a caso possiamo associare a ciascun bambino un numero fra 1 e 116 (numeri diversi a bambini diversi, s'intende!) e poi chiedere a un computer

di generare un numero a caso³ fra 1 e 116.

Osservazione 2.15 La definizione di probabilità tramite la frequenza relativa soddisfa le proprietà (1)–(3). Infatti, il numero dei successi è (non negativo e) minore o uguale al numero dei tentativi, per cui le frequenze relative sono numeri compresi fra 0 e 1. Il numero di successi dell'evento impossibile è zero, per cui la sua frequenza relativa è zero; analogamente, la frequenza relativa dell'evento certo è 1. Infine, il numero di volte in cui si verifica l'uno o l'altro di due eventi incompatibili è la somma del numero di volte in cui si verifica uno con il numero di volte in cui si verifica l'altro; dividendo per il numero dei tentativi troviamo che la frequenza relativa dell'unione dei due eventi è uguale alla somma delle frequenze relative.

Osservazione 2.16 Se lo spazio degli eventi è finito con una distribuzione uniforme della probabilità, allora la frequenza relativa (calcolata esaminando tutti gli eventi!) coincide con la probabilità a priori. Infatti, in questo caso (vedi l'Osservazione 2.8) la probabilità a priori di un evento composto è uguale al numero di eventi semplici che lo compongono (cioè al numero di successi) divisa per il numero totale di eventi semplici (cioè il numero di tentativi).

ESEMPIO 2.21 In un esperimento, hai stabilito di sottoporre la cavia A al trattamento a , la cavia B al trattamento b , e la cavia C al trattamento c . Sfortunatamente il tuo assistente (che non si è ancora ripreso dalla notte insonne passata a tirare il dado) si sbaglia e attribuisce a caso i tre trattamenti alle tre cavie. Qual è la probabilità che a nessuna cavia venga attribuito il trattamento giusto? Siccome i tre trattamenti sono attribuiti a caso, le possibili attribuzioni sono equiprobabili. Ci sono 6 possibili attribuzioni — (a, b, c) , (a, c, b) , (b, a, c) , (b, c, a) , (c, a, b) , (c, b, a) , dove (x, y, z) vuol dire trattamento x alla cavia A , trattamento y alla cavia B , trattamento z alla cavia C —, per cui ogni attribuzione ha probabilità $1/6$. Il sottoinsieme delle attribuzioni in cui nessuna cavia ha il trattamento giusto è costituito da 2 elementi, (b, c, a) e (c, a, b) . Il ragionamento a priori ci dice allora che la probabilità cercata è $2 \cdot \frac{1}{6}$; il ragionamento a posteriori ci dice che la probabilità cercata è $2/6$, e coerentemente otteniamo $1/3$ in entrambi i casi.

Come accennato sopra, è importante sapere se stiamo utilizzando le frequenze relative per *definire* una distribuzione di probabilità o per *verificare* una distribuzione di probabilità. Nel primo caso non c'è nessun problema: facciamo le nostre misure, calcoliamo le frequenze relative e siccome queste soddisfano le proprietà (1)–(3) (vedi l'Osservazione 2.15 e la prossima sezione) otteniamo dei numeri che hanno tutto il diritto di essere chiamati probabilità. Ma nella maggior parte dei casi non è questo che ci interessa.

Nella maggior parte dei casi noi vogliamo usare le frequenze relative per misurare un'altra distribuzione di probabilità. Infatti, tipicamente siamo in una delle seguenti due situazioni:

³ Ovviamente, questo apre il problema di come si fa a generare un numero a caso con un computer. . . problema, per fortuna, già risolto da matematici e informatici.

- non siamo in grado di trattare l'intera popolazione, per cui vogliamo misurare la distribuzione di probabilità dell'intera popolazione (definita per esempio tramite le frequenze relative dell'intera popolazione, che però non possiamo ottenere) usando le frequenze relative di un campione (come nel caso del peso dei neonati);
- tramite ipotesi teoriche abbiamo dedotto una distribuzione di probabilità a priori, e vogliamo usare le frequenze relative per verificare se questa è l'effettiva distribuzione di probabilità. Per esempio, per verificare la legge di disgiunzione di Mendel potremmo calcolare le frequenze relative dei genotipi di un gran numero di figli per vedere se coincidono con quelle predette nell'Esempio 2.16 (oppure per verificare se un dado non è truccato potremmo chiedere al nostro assistente di passare un'altra notte insonne. . .)

In entrambi i casi, perché le frequenze relative misurino davvero la distribuzione (teorica o incognita) di probabilità a cui siamo interessati devono essere verificate tre condizioni importanti:

1. Dev'essere possibile ripetere la misura o l'esperimento *un gran numero di volte* e in condizioni *pressoché costanti*. Solo due lanci non ci potranno mai dire se un dado è truccato o meno; e un miliardo di lanci effettuati cambiando dado ogni volta non servono a stabilire se uno specifico dado è truccato o non lo è.
2. I tentativi effettuati — la scelta del campione — devono essere davvero *casuali*, e non dipendere da fattori estranei che possano falsare i risultati. Anche se al tuo assistente è antipatico il numero 6, non può escludere dall'esperimento i lanci il cui risultato sia 6.
3. Ciascuna misura o esperimento dev'essere *indipendente* dalle precedenti. Al tuo assistente non è permesso di appoggiare il dado sul tavolo senza tirarlo se il risultato precedente era 1, e di tirarlo negli altri casi; deve comportarsi nello stesso modo indipendentemente dai risultati precedenti.

Vediamo un paio di esempi per chiarire queste idee.

ESEMPIO 2.22 Vogliamo calcolare la probabilità che a uno studente dell'Università di Pisa piaccia andare in discoteca. Nell'A.A. 2005/06 gli iscritti erano 51130; chiaramente non possiamo aspettarci che basti intervistare 5 studenti per avere risultati significativi. Le interviste devono essere effettuate tutte nello stesso periodo di tempo (se ci mettiamo dieci anni, gli studenti sono cambiati, le mode pure, e le risposte ottenute all'inizio non sono più confrontabili con quelle ottenute alla fine), e dobbiamo porre a ciascun intervistato le stesse domande. La scelta degli intervistati è cruciale: devono essere scelti a caso. Per esempio, scegliere il campione solo fra gli studenti incontrati in discoteca un sabato sera darebbe dei risultati probabilmente piuttosto falsati. Infine, la scelta di chi intervistare non può dipendere da chi abbiamo già intervistato. Se, una volta intervistato uno studente, passiamo a intervistare i suoi amici, probabilmente otterremo risposte più rappresentative dei suoi gusti (in quanto gli amici tendono ad avere gusti simili) che dei gusti degli studenti in generale.

ESEMPIO 2.23 Vuoi misurare le reazioni di alcune cavie (umane o animali) a un certo stimolo. Hai a disposizione un certo numero di (più o meno) volontari, e il

tuo assistente (che la notte prima ha dormito) ne sceglie a caso uno. Effettuato l'esperimento, riaffidi la cavia al tuo assistente che (non avendo letto queste dispense) la rimette assieme alle altre. Questo è un errore. Le cavie (sicuramente se umane, ma anche se animali) comunicano fra di loro; e quindi la reazione della cavia successiva potrebbe dipendere dalle informazioni ricevute dalla cavia precedente. Questo è evidente se si tratta di esseri umani, ma vale anche per gli animali: se la cavia precedente è spaventata, potrebbe trasmettere paura (o per lo meno preoccupazione) alle altre cavie, e questo potrebbe falsare la reazione delle cavie successive allo stimolo.

Siamo al cuore del metodo scientifico: si fanno delle ipotesi sul fenomeno in studio; si deducono delle probabilità a priori; si misurano, con un esperimento condotto con le regole di cui sopra, le frequenze relative. Se queste coincidono con le probabilità a priori, siamo felici e scriviamo un articolo; se invece differiscono, vuol dire che dobbiamo modificare le nostre ipotesi, perché ci sono in gioco dei fattori che non abbiamo considerato.

ESEMPIO 2.24 Hai fatto l'ipotesi che un certo dado non sia truccato, e ne hai dedotto una distribuzione uniforme della probabilità. Hai fatto passare una notte insonne al tuo assistente a tirare il dado, che ha ottenuto la Tabella 2.3. Per quanto detto nell'Osservazione 2.10, sei costretto a dedurre che molto probabilmente il tuo dado è truccato.

ESEMPIO 2.25 Com'è noto, il sesso del nascituro dipende da quale cromosoma della coppia "XY" viene trasmesso dal padre. Come prime ipotesi, potremmo assumere che il numero degli spermatozoi col cromosoma X sia uguale al numero degli spermatozoi col cromosoma Y; e che la possibilità di fecondazione non dipenda dal cromosoma portato dallo spermatozoo. Queste ipotesi portano a concludere una distribuzione di probabilità uniforme: la probabilità a priori che nasca un maschio è uguale alla probabilità che nasca una femmina, il 50% in entrambi i casi. Invece, misure effettuate sul campo mostrano che nascono sistematicamente più maschi che femmine. Per esempio, in Italia nel 2005 sono nati 284 472 maschi su un totale di 554 951 bambini; quindi la probabilità di un maschio in Italia nel 2005 è di circa il 51.3%, mentre quella di una femmina è di circa il 48.7%. Anche se con cifre lievemente diverse, il fenomeno si ripete analogo negli altri anni e in altre popolazioni. Quindi siamo costretti a concludere che le possibilità di fecondazione dipendano dal cromosoma portato dallo spermatozoo (attenzione: qui stiamo anche supponendo che gli aborti, spontanei o meno, non siano legati al sesso del nascituro. In alcune popolazioni, questa ipotesi non è verificata, gli aborti femminili sono più frequenti degli aborti maschili). In effetti, sembra che spermatozoi con cromosomi diversi abbiano mobilità diversa a seconda della concentrazione di certi ormoni femminili, concentrazione che cambia nel corso del ciclo mestruale, e che la mobilità influenzi le possibilità di fecondazione; ma non c'è ancora una spiegazione completa del fenomeno.

Osservazione 2.17 Attenzione: quando abbiamo detto che usiamo le frequenze re-

lative per misurare una distribuzione di probabilità a priori teorica o incognita, abbiamo sottinteso un fatto cruciale. Perché il ragionamento funzioni, occorre sapere che *effettuando un gran numero di misure/esperimenti indipendenti le frequenze relative approssimano sempre meglio le probabilità a priori*. Questo, se ci pensi un attimo, non è affatto ovvio, soprattutto se la distribuzione di probabilità a priori non è uniforme. Per fortuna, esiste un risultato matematico, il *Teorema di Glivenko-Cantelli*, conseguenza della *legge dei grandi numeri forte di Kolmogorov*, che ci assicura che le cose stanno proprio così. Sfortunatamente, le tecniche che permettono di enunciare precisamente e dimostrare questo risultato sono ben oltre lo scopo di queste dispense. Devi tenere però ben presente che questo teorema *non* dice due cose: non dice *quante* misure sono necessarie per arrivare a una buona approssimazione delle probabilità a priori, e non dice che le frequenze *assolute* devono approssimare sempre meglio il prodotto numero di tentativi per probabilità a priori.

ESEMPIO 2.26 Facendo tirare un dado 100 volte non truccato al tuo (ormai eshausto) assistente ottieni 15 volte 1; con 1000 lanci ottieni 161 volte 1; con 10000 lanci ottieni 1650 volte 1. Le frequenze relative sono $15/100 = 0.15$, $161/1000 = 0.161$ e $1650/10000 = 0.165$, che sono approssimazioni sempre migliori della probabilità a priori $1/6 = 0.\overline{16}$. Invece, la differenza fra il numero dei tentativi diviso 6 e le frequenze assolute *aumenta*, passando da $1.\overline{6}$ a $5.\overline{6}$ a $16.\overline{6}$.

2.5 Assiomi della probabilità

I discorsi fatti finora appartengono probabilmente più alla filosofia della scienza che alla matematica propriamente detta. Per il matematico conta solo fino a un certo punto come sono state definite le probabilità; quello che importa sono le proprietà di cui godono, e che conseguenze se ne possono trarre.

Nelle due sezioni precedenti abbiamo visto come per lavorare con le probabilità fossero fondamentali le proprietà (1)–(3). Il matematico allora parte da qui: una distribuzione di probabilità è qualunque cosa soddisfi le proprietà (1)–(3).

Vediamo di precisare questa affermazione. Dato uno spazio degli eventi Ω , indichiamo con \mathcal{A} la famiglia di tutti⁴ i sottoinsiemi di Ω . Una *distribuzione* (o *misura*) di probabilità p su Ω è un modo di associare un numero reale $p(E)$ a ogni evento $E \in \mathcal{A}$ in modo che

- (P1) $p(\emptyset) = 0$ e $p(\Omega) = 1$;
- (P2) $p(E) \in [0, 1]$ per ogni $E \in \mathcal{A}$;
- (P3) se $E_1 \cap E_2 = \emptyset$ allora $p(E_1 \cup E_2) = p(E_1) + p(E_2)$.

Dovrebbe esserti evidente (se non lo è ripensaci finché non te ne convinci) che gli assiomi (P1)–(P3) sono la traduzione matematica delle proprietà (1)–(3) enunciate nella Sezione 2.3. Per il matematico, qualunque cosa che soddisfa queste proprietà ha diritto di essere chiamata distribuzione di probabilità.

⁴ Se Ω è finito; se Ω è infinito, spesso conviene non considerarli tutti. Vedi la Curiosità 2.1

Un buon esempio da tenere a mente (anche per capire meglio alcune cose che faremo fra poco) è il seguente:

ESEMPIO 2.27 *Il rapporto di aree come probabilità.* Prendiamo come Ω un sottoinsieme del piano di area positiva, per esempio un cerchio, e sia \mathcal{A} la famiglia dei sottoinsiemi di Ω (di cui possiamo calcolare l'area; vedi la Curiosità 2.1). Allora a ogni sottoinsieme $E \in \mathcal{A}$ possiamo associare il rapporto fra la sua area e l'area di tutto Ω :

$$p(E) = \frac{\text{Area}(E)}{\text{Area}(\Omega)}.$$

In questo modo abbiamo definito una distribuzione di probabilità su Ω . Infatti, che si abbia $p(\Omega) = 1$ è ovvio; se poniamo, per convenzione, che l'insieme vuoto ha area nulla, allora $p(\emptyset) = 0$, e (P1) è soddisfatta. Siccome l'area è non negativa $p(E)$ è sempre maggiore di zero; inoltre, siccome $E \subseteq \Omega$ l'area di E non può essere maggiore dell'area di Ω , per cui $p(E) \leq 1$ sempre, e anche (P2) è soddisfatta. Infine, se E_1 ed E_2 sono insiemi disgiunti, l'area dell'unione è chiaramente la somma delle aree: $\text{Area}(E_1 \cup E_2) = \text{Area}(E_1) + \text{Area}(E_2)$. Dividendo per $\text{Area}(\Omega)$ otteniamo proprio $p(E_1 \cup E_2) = p(E_1) + p(E_2)$, e anche (P3) è verificata.

CURIOSITÀ 2.1 Una scoperta decisamente inaspettata della matematica contemporanea è che esistono dei sottoinsiemi del piano di cui non è possibile misurare l'area, in alcun modo (ma non ti preoccupare; sono talmente orrendi che se non sei un matematico non ti capiterà mai di incontrarli). Questo vuol dire che la probabilità come rapporto di aree non può essere definita su tutti i possibili sottoinsiemi di Ω , ma solo su alcuni, più decenti, di cui si può calcolare l'area. Per superare questo problema, si è deciso che per parlare di distribuzione di probabilità basta poter associare una probabilità a una famiglia di sottoinsiemi, e non necessariamente a tutti. Questa famiglia \mathcal{A} di sottoinsiemi non può essere qualsiasi, però. Deve contenere sia Ω che l'insieme vuoto \emptyset ; se contiene un insieme E deve contenere anche il suo complementare $\Omega \setminus E$; e se contiene una successione finita o infinita di insiemi E_1, E_2, E_3, \dots allora deve contenere anche l'unione $E_1 \cup E_2 \cup E_3 \cup \dots$. Data una famiglia \mathcal{A} di sottoinsiemi di Ω con queste proprietà (in termini tecnici \mathcal{A} si chiama *σ -algebra*), una *misura di probabilità* p su \mathcal{A} è un modo di associare a ciascun elemento $E \in \mathcal{A}$ un numero reale $p(E)$ in modo che siano verificate (P1)–(P3) e la proprietà (P4): data una successione infinita E_1, E_2, E_3, \dots di elementi di \mathcal{A} a due a due disgiunti, $p(E_1 \cup E_2 \cup \dots)$ è uguale alla somma infinita $p(E_1) + p(E_2) + \dots$ (parleremo delle somme infinite in un prossimo capitolo). Con questo trucco si riesce a parlare in modo efficace di probabilità anche nel caso di spazi degli eventi infiniti particolarmente complicati.

Una volta stabilito con precisione quali sono le proprietà di cui deve godere una distribuzione di probabilità, il matematico comincia a vedere che conseguenze può trarne. Per esempio, l'assioma (P3) si applica a due eventi incompatibili. E se ne abbiamo tre, a due a due incompatibili? Vediamo. Supponiamo di avere tre eventi E_1, E_2, E_3 a due a due incompatibili nel senso che $E_1 \cap E_2 = E_1 \cap E_3 = E_2 \cap E_3 = \emptyset$. L'assioma (P3) ci dice che

$$p(E_1 \cup E_2) = p(E_1) + p(E_2).$$

D'altra parte, noi sappiamo che

$$(E_1 \cup E_2) \cap E_3 = (E_1 \cap E_3) \cup (E_2 \cap E_3) = \emptyset \cup \emptyset = \emptyset;$$

quindi l'assioma (P3) ci dice anche che

$$p((E_1 \cup E_2) \cup E_3) = p(E_1 \cup E_2) + p(E_3) .$$

Mettendo il tutto insieme, e ricordando che l'unione è associativa, otteniamo

$$p(E_1 \cup E_2 \cup E_3) = p(E_1) + p(E_2) + p(E_3) .$$

Dunque anche la probabilità dell'unione di tre eventi a due a due incompatibili è uguale alla somma delle probabilità. In maniera analoga (convincitene!) si riesce a far vedere che questo vale per un qualsiasi numero finito di eventi: *se E_1, \dots, E_k sono degli eventi a due a due incompatibili, allora*

$$p(E_1 \cup \dots \cup E_k) = p(E_1) + \dots + p(E_k) . \quad (2.1)$$

Una conseguenza di questo risultato è un fatto che abbiamo implicitamente usato più volte nelle sezioni precedenti. Supponiamo che $E = \{x_1, \dots, x_k\}$ sia un evento costituito da un numero finito di eventi semplici. Siccome gli eventi semplici sono sempre a due a due incompatibili, possiamo applicare la formula (2.1) e ricavare

$$p(E) = p(x_1) + \dots + p(x_k) , \quad (2.2)$$

cioè la probabilità di un evento composto da un numero finito di elementi è sempre uguale alla somma delle probabilità dei suoi elementi.

Se conosciamo la probabilità di un evento, riusciamo a calcolare la probabilità del suo complementare? La risposta è affermativa, grazie agli assiomi (P1) e (P3). Infatti, se E è un evento e $\Omega \setminus E$ il suo complementare, noi sappiamo che $E \cap (\Omega \setminus E) = \emptyset$ e che $E \cup (\Omega \setminus E) = \Omega$. Allora gli assiomi (P1) e (P3) implicano

$$1 = p(\Omega) = p(E \cup (\Omega \setminus E)) = p(E) + p(\Omega \setminus E) ,$$

per cui abbiamo ottenuto la formula

$$p(\Omega \setminus E) = 1 - p(E) . \quad (2.3)$$

L'assioma (P3) ci fornisce la probabilità dell'unione di due eventi quando gli eventi sono incompatibili. Riusciamo a trovare una formula per la probabilità dell'unione di due eventi E_1 ed E_2 qualsiasi? Anche stavolta la risposta è affermativa, ma occorre ragionare un pochino. Ovviamente, vogliamo applicare l'assioma (P3). Per farlo, dobbiamo scrivere l'unione $E_1 \cup E_2$ come unione di due insiemi disgiunti. Per esempio, possiamo scrivere $E_1 \cup E_2$ come unione di E_2 con $E_1 \setminus E_2$; allora l'assioma (P3) ci dà

$$p(E_1 \cup E_2) = p(E_1 \setminus E_2) + p(E_2) .$$

Dobbiamo ora trovare una formula per calcolare la probabilità di $E_1 \setminus E_2$. Ma basta osservare che E_1 si può scrivere come unione degli eventi disgiunti $E_1 \setminus E_2$ ed $E_1 \cap E_2$; quindi $p(E_1) = p(E_1 \setminus E_2) + p(E_1 \cap E_2)$, ovvero

$$p(E_1 \setminus E_2) = p(E_1) - p(E_1 \cap E_2) . \quad (2.4)$$

Mettendo tutto insieme otteniamo

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2) , \quad (2.5)$$

formula che esprime la probabilità dell'unione di eventi in termini della probabilità dei singoli eventi e della probabilità dell'intersezione.

Osservazione 2.18 Prova a interpretare le formule (2.3), (2.4) ed (2.5) nel caso della probabilità vista come rapporto di aree (Esempio 2.27), magari facendo un disegno, e vedrai come sono del tutto naturali.

ESEMPIO 2.28 Sia $\Omega = \{1, 2, 3, 4, 5, 6\}$ il solito spazio degli eventi del tiro di un dado non truccato, con la solita distribuzione di probabilità uniforme, e consideriamo gli eventi “minore o uguale a 3” $E_1 = \{1, 2, 3\}$, e “pari” $E_2 = \{2, 4, 6\}$. Entrambi questi eventi hanno probabilità $1/2$. L'evento intersezione è “pari minore o uguale a 3” $E_1 \cap E_2 = \{2\}$, e ha probabilità $1/6$. Allora la (2.5) ci dice che l'evento unione “pari oppure minore o uguale a 3” deve avere probabilità

$$p(E_1 \cup E_2) = \frac{1}{2} + \frac{1}{2} - \frac{1}{6} = \frac{5}{6} ;$$

e infatti $E_1 \cup E_2 = \{1, 2, 3, 4, 6\}$.

2.6 Eventi indipendenti

La formula (2.5) esprime la probabilità dell'evento unione in funzione della probabilità dei due eventi e dell'evento intersezione. Cosa possiamo dire sulla probabilità dell'evento intersezione? In generale poco. È la probabilità che avvengano *entrambi* gli eventi; dipenderà quindi da che relazione c'è fra i due eventi.

ESEMPIO 2.29 Sia $\Omega = \{1, 2, 3, 4, 5, 6\}$ il solito spazio degli eventi del tiro di un dado a 6 facce non truccato, con la distribuzione uniforme di probabilità, e consideriamo gli eventi “pari” $E_1 = \{2, 4, 6\}$, “minore o uguale di 3” $E_2 = \{1, 2, 3\}$ e “dispari” $E_3 = \{1, 3, 5\}$. La probabilità di questi tre eventi è $3/6 = 1/2$ in ciascun caso. Le probabilità delle tre intersezioni sono invece diverse. Infatti, $E_1 \cap E_2 = \{2\}$, per cui $p(E_1 \cap E_2) = 1/6$, mentre $E_1 \cap E_3 = \emptyset$ per cui $p(E_1 \cap E_3) = 0$, e $E_2 \cap E_3 = \{1, 3\}$, per cui $p(E_2 \cap E_3) = 1/3$.

La conseguenza di questo esempio è che non può esistere una formula generale, sempre valida, che dia la probabilità dell'intersezione solo in funzione della probabilità degli eventi. Ma osserviamo i seguenti esempi.

ESEMPIO 2.30 Sia Ω lo spazio degli eventi del lancio di due dadi a 6 facce distinti, con la distribuzione di probabilità uniforme, e consideriamo gli eventi “primo dado 4” $E_1 = \{(4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6)\}$ e “secondo dado 6” $E_2 = \{(1, 6), (2, 6), (3, 6), (4, 6), (5, 6), (6, 6)\}$. Abbiamo $p(E_1) = p(E_2) = 1/6$, mentre l’evento intersezione $E_1 \cap E_2 = \{(4, 6)\}$ ha probabilità $p(E_1 \cap E_2) = 1/36$, che è anche uguale al prodotto $p(E_1) \cdot p(E_2)$.

ESEMPIO 2.31 Consideriamo stavolta il lancio di due dadi “a 3 facce” non truccati distinti. Un modo per ottenere un dado a 3 facce non truccato consiste nell’usare un dado a 6 facce non truccato e nel dichiarare che un risultato di 4 equivale a un 1, un 5 a un 2, e un 6 a un 3. In altre parole, lo spazio degli eventi del lancio di un dado a 3 facce non truccato è $\Omega = \{\{1, 4\}, \{2, 5\}, \{3, 6\}\}$, con la distribuzione uniforme di probabilità. Lo spazio degli eventi del lancio di due dadi a 3 facce distinti ha quindi $3 \cdot 3 = 9$ elementi, con distribuzione della probabilità uniforme; l’evento “primo dado 2, secondo dado 3” ha quindi probabilità $1/9$. Ora, possiamo giungere a questo risultato anche in un altro modo. Lanciando il primo dado a 6 facce, si ottiene 2 col corrispondente dado a tre facce in 2 casi su 6. Analogamente, lanciando il secondo dado a 6 facce, si ottiene 3 col corrispondente dado a tre facce in 2 casi su 6. Siccome i due dadi sono *indipendenti*, cioè non si influenzano a vicenda, il numero di casi in cui lanciando i due dadi a 6 facce si ottiene la coppia $(2, 3)$ coi corrispondenti dadi a 3 facce è uguale al prodotto del numero casi con cui si ottiene 2 col primo dado per il numero dei casi con cui si ottiene 3 col secondo dado, cioè $4 = 2 \cdot 2$. Il numero totale dei casi possibili del lancio dei due dadi a sei facce è $6 \cdot 6 = 36$. Siccome stiamo usando la distribuzione di probabilità uniforme, l’Osservazione 2.16 ci assicura che la probabilità di ottenere $(2, 3)$ con la coppia di dadi a 3 facce (l’intersezione dei due eventi) è $4/36 = (2/6) \cdot (2/6)$. Abbiamo ottenuto nuovamente $1/9$, ma in questo modo vediamo chiaramente come il risultato ottenuto sia uguale al prodotto delle probabilità dei due eventi che abbiamo intersecato.

Quello che questi esempi sembrano suggerire è che se due eventi non si influenzano a vicenda allora la probabilità dell’evento intersezione è il prodotto delle probabilità dei due eventi.

Trasformiamo questo suggerimento in una definizione: diremo che due eventi E_1 ed E_2 sono *indipendenti* se e solo se $p(E_1 \cap E_2) = p(E_1)p(E_2)$.

Osservazione 2.19 Attenzione a non fare confusione. La formula (2.5) è una formula sempre vera, valida per qualsiasi coppia di eventi. La formula

$$p(E_1 \cap E_2) = p(E_1)p(E_2)$$

invece è una *condizione* sui due eventi; se è verificata diciamo che gli eventi sono indipendenti, ma può benissimo non essere verificata (e in tal caso gli eventi *non* sono indipendenti).

ESEMPIO 2.32 Gli eventi E_1 ed E_2 dell’Esempio 2.29 (come pure gli eventi E_1 ed E_3 , e gli eventi E_2 ed E_3) non sono indipendenti, perché la probabilità dell’inter-

sezione non è uguale al prodotto delle probabilità. Questo è anche intuitivamente ragionevole: sapere che è accaduto l'evento E_1 (risultato pari) chiaramente influisce sull'evento E_2 (risultato minore o uguale di 3 con lo stesso tiro dello stesso dado), perché riduce il numero di modi in cui E_2 può essersi verificato. Invece, gli eventi E_1 ed E_2 dell'Esempio 2.30 sono indipendenti. Anche questo è intuitivamente sensato: il risultato del lancio del primo dado non influisce in alcun modo sul risultato del lancio del secondo dado.

ESEMPIO 2.33 *Un cassetto contiene 12 calzini bianchi e 4 calzini neri. Qual è la probabilità di estrarne un paio spaiato?* Per rispondere a questa domanda chiariamo prima come vogliamo effettuare l'estrazione. Il modo naturale di procedere se vogliamo indossarli consiste nell'estrarne prima uno e poi l'altro; in questo caso si parla di *estrazione senza rimbussolamento*. Se invece vogliamo limitarci ad annotare il colore dei calzini senza indossarli, possiamo estrarne uno, segnarcene il colore, rimmetterlo dentro ed estrarre il secondo; in questo caso si parla di *estrazione con rimbussolamento*. Vediamo come cambia la probabilità. Nel primo caso (assumendo che i calzini siano indistinguibili, per cui ciascuno abbia la stessa probabilità di essere estratto), alla prima estrazione abbiamo una probabilità di $12/16 = 3/4$ di estrarre un calzino bianco, e di $4/16 = 1/4$ di estrarre un calzino nero. Se abbiamo estratto un calzino bianco, la probabilità di estrarre poi un calzino nero è $4/15$ (ricorda che ora il cassetto contiene un calzino in meno). Essendo le due estrazioni indipendenti, la probabilità di estrarre prima un calzino bianco e poi uno nero è⁵ $(3/4) \cdot (4/15) = 1/5$. In modo analogo si vede che la probabilità di estrarre prima un calzino nero e poi uno bianco è $(1/4) \cdot (12/15) = 1/5$; quindi l'assioma (P3) ci dice che la probabilità di estrarre (senza rimbussolamento) un paio spaiato è $1/5 + 1/5 = 2/5$.

Nel caso con rimbussolamento, la probabilità di estrarre prima un calzino bianco e poi un calzino nero è $(3/4) \cdot (1/4) = 3/16$, dove di nuovo abbiamo usato l'indipendenza delle due estrazioni. Analogamente la probabilità di estrarre prima un calzino nero e poi uno bianco è $(1/4) \cdot (3/4) = 3/16$, per cui la probabilità di estrarre con rimbussolamento un paio spaiato è $3/16 + 3/16 = 3/8$, lievemente minore rispetto a quella senza rimbussolamento.

CURIOSITÀ 2.2 La definizione precisa di indipendenza per più di 2 eventi è un attimo complicata.

Si dice che n eventi E_1, \dots, E_n sono *indipendenti* fra loro se e solo se per ogni $2 \leq k \leq n$ e ogni scelta di k eventi E_{i_1}, \dots, E_{i_k} fra E_1, \dots, E_n si ha

$$p(E_{i_1} \cap \dots \cap E_{i_k}) = p(E_{i_1}) \cdots p(E_{i_k}) .$$

Attenzione: il fatto che tre eventi E_1, E_2, E_3 siano a due a due indipendenti non implica che i tre eventi assieme siano indipendenti. Per esempio, nel caso del lancio di due dadi a 6 facce distinti, consideriamo gli eventi "primo dado 1" $E_1 = \{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6)\}$,

⁵ Un altro modo per vederlo, che conferma l'indipendenza delle due estrazioni, è il seguente: abbiamo 12 modi (su 16) per estrarre un calzino bianco la prima volta, e 4 modi (su 15) di estrarre un calzino nero la seconda volta, per cui abbiamo $12 \cdot 4$ modi (su $16 \cdot 15$) per estrarre prima un calzino bianco e poi un calzino nero.

“secondo dado 2” $E_2 = \{(1, 2), (2, 2), (3, 2), (4, 2), (5, 2), (6, 2)\}$, e “somma dei dadi uguale a 7” $E_3 = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$. Allora si verifica subito che i tre eventi sono a due a due indipendenti (cioè $p(E_1 \cap E_2) = p(E_1)p(E_2)$ e così via), ma i tre eventi assieme non sono indipendenti: $p(E_1 \cap E_2 \cap E_3) = 0 \neq (1/6)^3 = p(E_1)p(E_2)p(E_3)$.

Analogamente, sapere che $p(F_1 \cap F_2 \cap F_3) = p(F_1)p(F_2)p(F_3)$ non implica che i tre eventi siano indipendenti a due a due (e quindi non implica che siano indipendenti tutti assieme). Per esempio, prendiamo $F_1 = \Omega \setminus \{(1, 1), (2, 1), (3, 1), (4, 1), (5, 1), (1, 2), (2, 2), (3, 2), (3, 4)\}$, $F_2 = \{(1, 3), (2, 3), (3, 3), (4, 3), (5, 3), (6, 3), (6, 2), (6, 1)\}$ ed $F_3 = E_3$. Allora si vede subito che $F_1 \cap F_2 \cap F_3 = \{(6, 1)\}$ per cui $p(F_1 \cap F_2 \cap F_3) = 1/36 = (27/36)(8/36)(6/36) = p(F_1)p(F_2)p(F_3)$ ma $p(F_2 \cap F_3) = 1/18 \neq (8/36)(6/36) = p(F_2)p(F_3)$.

Il succo di questo ragionamento è che n eventi E_1, \dots, E_n sono indipendenti se e solo se la probabilità dell'intersezione di qualsiasi di essi è uguale al prodotto delle probabilità.

2.7 La legge di Hardy-Weinberg

Vogliamo ora discutere alcune applicazioni alla genetica del concetto di eventi indipendenti.

Supponiamo di avere una popolazione composta da molti individui, e consideriamo un gene con due possibili alleli A e a ; vogliamo studiare come varia la distribuzione dei genotipi da una generazione all'altra. Indichiamo con p_{AA} (rispettivamente, p_{Aa} , p_{aa}) la probabilità che un individuo preso a caso nella popolazione abbia genotipo AA (rispettivamente, Aa , aa); in particolare, $p_{AA} + p_{Aa} + p_{aa} = 1$. Supponiamo anche che la distribuzione dei genotipi sia la stessa per i due sessi; che tutti gli individui della popolazione siano fertili; e che il genotipo non influenzi gli accoppiamenti, per cui il genotipo del padre e quello della madre sono indipendenti. Il nostro primo obiettivo è calcolare la distribuzione di probabilità dei genotipi della generazione dei figli in funzione di quella dei genitori, applicando la legge di disgiunzione di Mendel.

Supponiamo che il padre abbia genotipo AA (evento che ha probabilità p_{AA}) e la madre abbia genotipo AA (evento che ha probabilità p_{AA}). Per ipotesi, i due eventi sono indipendenti; quindi la probabilità che entrambi i genitori abbiano genotipo AA è il prodotto delle probabilità, cioè p_{AA}^2 . Questa coppia di genitori ha necessariamente un figlio con genotipo AA , per cui la probabilità di trovare una coppia con entrambi i genitori e il figlio di genotipo AA è p_{AA}^2 . Supponiamo invece che il padre abbia genotipo Aa (evento che ha probabilità p_{Aa}) e la madre ancora genotipo AA (evento che ha probabilità p_{AA}). Per ipotesi, i due eventi sono indipendenti; quindi la probabilità che i genitori abbiano questi genotipi è il prodotto delle probabilità, cioè $p_{AA}p_{Aa}$. Per la legge di disgiunzione di Mendel, questa coppia di genitori ha un figlio con genotipo AA con probabilità $1/2$, e un figlio con genotipo Aa con probabilità $1/2$. Quindi la probabilità di trovare una coppia di genitori con quei genotipi e il figlio di genotipo AA è $p_{AA}p_{Aa}/2$, e la probabilità di trovare una coppia di genitori con quei genotipi e il figlio di genotipo Aa è ancora $p_{AA}p_{Aa}/2$.

Ragionando in questo modo costruiamo la Tabella 2.5, in cui le colonne corrispondono al genotipo del padre, le righe a quello della madre, e le caselle centrali contengono le probabilità di trovare una coppia di genitori con i genotipi della riga e colonna corrispondenti, e un figlio con il genotipo indicato.

	$AA: p_{AA}$	$Aa: p_{Aa}$	$aa: p_{aa}$
$AA: p_{AA}$	$AA: p_{AA}^2$ $Aa: 0$ $aa: 0$	$AA: \frac{1}{2}p_{AA}p_{Aa}$ $Aa: \frac{1}{2}p_{AA}p_{Aa}$ $aa: 0$	$AA: 0$ $Aa: p_{AA}p_{aa}$ $aa: 0$
$Aa: p_{Aa}$	$AA: \frac{1}{2}p_{AA}p_{Aa}$ $Aa: \frac{1}{2}p_{AA}p_{Aa}$ $aa: 0$	$AA: \frac{1}{4}p_{Aa}^2$ $Aa: \frac{1}{2}p_{Aa}^2$ $aa: \frac{1}{4}p_{Aa}^2$	$AA: 0$ $Aa: \frac{1}{2}p_{Aa}p_{aa}$ $aa: \frac{1}{2}p_{Aa}p_{aa}$
$aa: p_{aa}$	$AA: 0$ $Aa: p_{AA}p_{aa}$ $aa: 0$	$AA: 0$ $Aa: \frac{1}{2}p_{Aa}p_{aa}$ $aa: \frac{1}{2}p_{Aa}p_{aa}$	$AA: 0$ $Aa: 0$ $aa: p_{aa}^2$

TABELLA 2.5

Siccome gli eventi contenuti nelle caselle centrali della Tabella 2.5 sono a due a due mutuamente esclusivi, per trovare la probabilità p_{AA}^F che un figlio abbia genotipo AA basta sommare tutte le probabilità delle caselle centrali corrispondenti al genotipo AA :

$$p_{AA}^F = p_{AA}^2 + 2 \cdot \frac{1}{2}p_{AA}p_{Aa} + \frac{1}{4}p_{Aa}^2 = (p_{AA} + \frac{1}{2}p_{Aa})^2. \quad (2.6)$$

In modo analogo troviamo la probabilità p_{Aa}^F che un figlio abbia genotipo Aa , e la probabilità p_{aa}^F che un figlio abbia genotipo aa :

$$\begin{aligned} p_{Aa}^F &= 2\frac{1}{2}p_{AA}p_{Aa} + 2p_{AA}p_{aa} + \frac{1}{2}p_{Aa}^2 + 2\frac{1}{2}p_{Aa}p_{aa} \\ &= 2(p_{AA} + \frac{1}{2}p_{Aa})(p_{aa} + \frac{1}{2}p_{Aa}), \\ p_{aa}^F &= \frac{1}{4}p_{Aa}^2 + 2 \cdot \frac{1}{2}p_{Aa}p_{aa} + p_{aa}^2 = (\frac{1}{2}p_{Aa} + p_{aa})^2. \end{aligned} \quad (2.7)$$

Non è finita qui. Le varie probabilità p_{AA} , p_{Aa} e p_{aa} non sono completamente indipendenti fra di loro, in quanto anche i genitori sono figli di qualcuno. Infatti, le possiamo ottenere tutte a partire da un unico dato: la probabilità p_A della presenza dell'allele A nel genotipo di un individuo preso a caso nella popolazione⁶. Chiaramente, se indichiamo con p_a la probabilità della presenza dell'allele a nel genotipo di un individuo preso a caso nella popolazione, abbiamo $p_a = 1 - p_A$, ovvero

$$p_A + p_a = 1.$$

La legge di disgiunzione di Mendel implica (perché?) che la presenza di uno o l'altro allele in un cromosoma è indipendente da quale allele è presente nell'altro

⁶ Per il discorso successivo non è importante come sia stata calcolata p_A , se come frequenza relativa o in altro modo.

cromosoma. Quindi la probabilità p_{AA} che un individuo abbia genotipo AA è il prodotto della probabilità p_A che abbia un allele A con la probabilità p_A che abbia un secondo allele A ; quindi $p_{AA} = p_A^2$. Analogamente, $p_{aa} = p_a^2$. Infine, siccome ci sono due modi per ottenere un genotipo Aa , abbiamo $p_{Aa} = 2p_A p_a$. Dunque otteniamo

$$p_{AA} = p_A^2, \quad p_{Aa} = 2p_A p_a, \quad p_{aa} = p_a^2. \quad (2.8)$$

Diremo che una popolazione con distribuzione di genotipi che soddisfa la (2.8) per opportuni valori p_A e $p_a = 1 - p_A$ ha una *distribuzione* (o *equilibrio*) di *Hardy-Weinberg*; a volte, la chiameremo semplicemente *popolazione di Hardy-Weinberg*.

Siamo all'ultimo passaggio. Mettiamo la (2.8) nelle (2.6) e (2.7); usando l'identità $p_A + p_a = 1$ otteniamo

$$\begin{aligned} p_{AA}^F &= (p_A^2 + \tfrac{1}{2}2p_A p_a)^2 = p_A^2(p_A + p_a)^2 = p_A^2 = p_{AA}, \\ p_{Aa}^F &= 2(p_A^2 + \tfrac{1}{2}2p_A p_a)(p_a^2 + \tfrac{1}{2}2p_A p_a) = 2p_A p_a(p_A + p_a)^2 = 2p_A p_a = p_{Aa}, \\ p_{aa}^F &= (\tfrac{1}{2}2p_A p_a + p_a^2)^2 = p_a^2(p_A + p_a)^2 = p_a^2 = p_{aa}. \end{aligned}$$

In altre parole, *se una popolazione ha una distribuzione di Hardy-Weinberg, la distribuzione di probabilità dei genotipi fra i figli è identica alla distribuzione di probabilità dei genotipi fra i genitori!* In altre parole, *la distribuzione dei genotipi in una popolazione di Hardy-Weinberg non cambia da una generazione all'altra.*

Questo risultato è detto *legge di Hardy-Weinberg*, ed è stato scoperto indipendentemente dal matematico inglese G.H. Hardy e dal medico tedesco W. Weinberg nel 1908.

Osservazione 2.20 I conti che abbiamo fatto hanno un'ulteriore conseguenza interessante. Supponiamo che la popolazione di partenza non abbia una distribuzione di Hardy-Weinberg, per esempio perché ottenuta mescolando (a causa di una migrazione) due popolazioni con differente distribuzione degli alleli. In questo caso, p_{AA} , p_{Aa} e p_a^2 non soddisfano (2.8). Nonostante ciò, la popolazione dei figli è di Hardy-Weinberg! Infatti, guardando (2.6) e (2.7) si vede subito che basta porre $p_A^F = p_{AA} + \tfrac{1}{2}p_{Aa}$ e $p_a^F = p_{aa} + \tfrac{1}{2}p_{Aa}$ per avere $p_{AA}^F = (p_A^F)^2$, $p_{Aa}^F = 2p_A^F p_a^F$, $p_{aa}^F = (p_a^F)^2$, e $p_A^F + p_a^F = 1$, come richiesto dalla (2.8). In altre parole, *la popolazione dei figli di una popolazione qualsiasi è sempre di Hardy-Weinberg*, e quindi *la distribuzione dei genotipi in una popolazione qualsiasi si stabilizza dalla seconda generazione in poi.*

Osservazione 2.21 Vale la pena ribadire che la legge di Hardy-Weinberg e le sue conseguenze si applicano solo in situazioni in cui il genotipo dei padri e quello delle madri sono indipendenti e scelti a caso. In particolare, non si applicano per alleli che favoriscono l'accoppiamento, né in popolazioni piccole con molti accoppiamenti fra famigliari. Ma su grandi popolazioni e per buona parte dei geni la legge di Hardy-Weinberg è ampiamente rispettata.

ESEMPIO 2.34 Vediamo come calcolare la probabilità che Leonardo e Martina abbiano un figlio con fenotipo Rh^- senza sapere nulla sui loro genotipi. Ovviamente, ci serve un qualche dato per partire. Il dato più semplice da recuperare in

questo caso è la distribuzione dei fenotipi; in Italia, il 15.1% della popolazione ha fenotipo Rh^- , e l'84.9% ha fenotipo Rh^+ . Il fenotipo Rh^- è generato soltanto dal genotipo Rh^-Rh^- , per cui possiamo dire che la probabilità p_{--} che un italiano preso a caso abbia genotipo Rh^-Rh^- è 0.151. Invece, il fenotipo Rh^+ è generato dai genotipi Rh^+Rh^+ e Rh^+Rh^- , per cui possiamo dire che $p_{++} + p_{+-} = 0.849$, dove p_{++} (rispettivamente, p_{+-}) è la probabilità del genotipo Rh^+Rh^+ (rispettivamente, Rh^+Rh^-).

A questo punto possiamo procedere in modo furbo o in modo pedante. Il modo furbo è il seguente: la popolazione italiana considerata soddisfa le ipotesi della legge di Hardy-Weinberg; quindi, la distribuzione dei genotipi nei figli è uguale a quella nei genitori; quindi, la probabilità che il figlio abbia fenotipo Rh^- (cioè genotipo Rh^-Rh^-) è $p_{--} = 0.151 = 15.1\%$.

Il modo pedante consiste nel fare tutti i conti. Indichiamo con p_+ (rispettivamente, p_-) la probabilità di presenza dell'allele Rh^+ (rispettivamente, Rh^-) in un italiano preso a caso. Siccome la popolazione italiana è di Hardy-Weinberg, sappiamo che $p_{--} = p_-^2$; quindi ricaviamo $p_- = \sqrt{0.151} \simeq 0.389$ e $p_+ = 1 - p_- \simeq 0.611$. Otteniamo allora $p_{++} = p_+^2 \simeq 0.374$ e $p_{+-} = 2p_+p_- \simeq 0.475$. Nota che i valori ricavati di p_{++} e di p_{+-} sono tali che $p_{++} + p_{+-} \simeq 0.849$, che è proprio la frequenza relativa misurata nella popolazione italiana del fenotipo Rh^+ .

Adesso calcoliamo la probabilità che Leonardo e Martina abbiano un certo genotipo e loro figlio sia Rh^- . Usando il fatto che i genotipi di padre e madre sono indipendenti e usando nuovamente la legge di disgiunzione di Mendel otteniamo la Tabella 2.6, in cui abbiamo elencato solo le situazioni in cui Leonardo e Martina possono avere un figlio con fenotipo Rh^- (negli altri casi la probabilità che possano avere un figlio con fenotipo Rh^- è 0).

<i>Genotipi</i>	<i>Probabilità</i>
Leonardo Rh^+Rh^- , Martina Rh^+Rh^- , Figlio Rh^-Rh^-	$\frac{1}{4}p_{+-}^2 \simeq 5.6\%$
Leonardo Rh^+Rh^- , Martina Rh^-Rh^- , Figlio Rh^-Rh^-	$\frac{1}{2}p_{+-}p_{--} \simeq 3.6\%$
Leonardo Rh^-Rh^- , Martina Rh^+Rh^- , Figlio Rh^-Rh^-	$\frac{1}{2}p_{--}p_{+-} \simeq 3.6\%$
Leonardo Rh^-Rh^- , Martina Rh^-Rh^- , Figlio Rh^-Rh^-	$p_{--}^2 \simeq 2.3\%$

TABELLA 2.6

Siccome questi eventi sono a due a due incompatibili, la probabilità che il figlio di Leonardo e Martina abbia fenotipo Rh^- è la somma di queste probabilità, che è proprio 15.1%, come predetto dalla legge di Hardy-Weinberg.

Osservazione 2.22 Esiste una legge di Hardy-Weinberg anche per geni con più di due alleli. Supponiamo che un gene abbia n alleli A_1, \dots, A_n , con relative probabilità p_{A_1}, \dots, p_{A_n} , soddisfacenti ovviamente $p_{A_1} + \dots + p_{A_n} = 1$. Se valgono le solite ipotesi della legge di Hardy-Weinberg, la legge di disgiunzione di Mendel ci dice che la probabilità di trovare nella popolazione un genotipo omozigota A_iA_i

è

$$p_{A_i A_i} = p_{A_i}^2, \quad (2.9)$$

mentre la probabilità di trovare un genotipo eterozigota $A_i A_j$ con $i \neq j$ è

$$p_{A_i A_j} = 2p_{A_i} p_{A_j}. \quad (2.10)$$

Anche stavolta chiameremo di *Hardy-Weinberg* una popolazione che soddisfa (2.9) e (2.10). Allora procedendo come prima si può verificare che la distribuzione dei genotipi in una popolazione di Hardy-Weinberg è costante da una generazione all'altra, e che ogni popolazione diventa di Hardy-Weinberg dopo una generazione.

Esercizio 2.1 *La distribuzione dei gruppi sanguigni nella popolazione italiana è: gruppo A 44%, gruppo B 11%, gruppo AB 5%, gruppo 0 40%. Il gruppo sanguigno è determinato da un gene con tre alleli, che indicheremo con A, B e 0. Il fenotipo A corrisponde ai genotipi AA e A0, il fenotipo B corrisponde ai genotipi BB e B0, il fenotipo AB corrisponde al genotipo AB, e il fenotipo 0 corrisponde al genotipo 00.*

- (i) *Supponendo che la popolazione italiana sia di Hardy-Weinberg, calcola le probabilità di ogni singolo allele e di ogni singolo genotipo.*
- (ii) *Verifica che i risultati ottenuti nel punto precedente siano in accordo con l'ipotesi di distribuzione di Hardy-Weinberg.*
- (iii) *Trova in modo furbo e in modo pedante la probabilità che Leonardo e Martina abbiano un figlio con gruppo sanguigno 0.*

Un'altra applicazione del concetto di probabilità indipendenti e dell'equilibrio di Hardy-Weinberg è il test del DNA. La situazione è la seguente: la polizia è in possesso di un campione del DNA (ricavato da sangue, sperma o altro materiale organico) del colpevole di un reato. Viene prelevato un campione di DNA di un sospetto, e si verifica che alcuni geni (o, più in generale, marcatori genetici, cioè zone del DNA variabili nella popolazione⁷) del sospetto sono uguali a quelli del colpevole. Qual è la probabilità che si tratti di una coincidenza? In altri termini, *qual è la probabilità che un individuo preso a caso nella popolazione abbia gli stessi marcatori genetici del colpevole?*

Ormai dovrebbe essere chiaro come si procede. Si parte dalla probabilità della presenza nella popolazione di ciascuno degli alleli dei marcatori genetici coinvolti. Supponendo che la popolazione sia di Hardy-Weinberg, si usano (2.9) e (2.10) per calcolare la probabilità del genotipo di ciascuno dei marcatori genetici del colpevole. Infine, supponendo che i vari marcatori genetici siano scelti in modo da essere indipendenti fra loro (ipotesi non sempre verificata; in alcuni casi la presenza di un certo allele in un dato gene favorisce la presenza di un allele specifico in un altro gene), si moltiplicano le probabilità dei vari genotipi e si ottiene la probabilità che un individuo preso a caso abbia gli stessi genotipi del colpevole in quei marcatori genetici. Ma vediamo un esempio numerico.

⁷ I marcatori genetici nell'uomo costituiscono meno dello 0.1% dell'intero DNA umano; il 99.9% del DNA è lo stesso per tutti gli individui.

ESEMPIO 2.35 Due marcatori genetici spesso usati nei test del DNA sono il CSF1P0 e il TPOX; è noto che sono indipendenti fra loro. Uno studio effettuato nella Polonia meridionale ha trovato 8 alleli diversi per CSF1P0, e 5 alleli diversi per TPOX, con le distribuzioni di probabilità riportate nella Tabella 2.7.

	Allele	Probabilità		Allele	Probabilità
TPOX	1	0.556	CSF1P0	1	0.008
	2	0.125		2	0.032
	3	0.081		3	0.355
	4	0.214		4	0.266
	5	0.024		5	0.262
				6	0.060
				7	0.004
				8	0.013

TABELLA 2.7

Il DNA del colpevole è risultato essere eterozigote con alleli 2 e 4 nel marcatore genetico TPOX, e omozigota di allele 5 nel marcatore genetico CSF1P0. La Tabella 2.7 ci dice (perché?) che la probabilità che un individuo preso a caso nella popolazione abbia genotipo eterozigote “24” in TPOX è $2 \cdot 0.125 \cdot 0.214 = 0.0535$, mentre la probabilità che un individuo preso a caso nella popolazione abbia genotipo omozigote “55” in CSF1P0 è $(0.262)^2 = 0.068644$. Quindi la probabilità che un individuo preso a caso nella popolazione sia eterozigote “24” in TPOX e omozigote “55” in CSF1P0 è $0.0535 \cdot 0.068644 \simeq 0.00367 = 0.367\%$. In altre parole, meno di 4 individui su 1000 della popolazione studiata può avere il DNA del colpevole.

Osservazione 2.23 Bisogna fare attenzione a come si usa il test del DNA nel decidere se un sospettato è colpevole o innocente. Il test del DNA ci permette di *assolvere* con certezza un sospettato: se il DNA è diverso da quello del colpevole, non è stato lui. Ma *non* ci permette di condannare con certezza un sospettato, anche se la probabilità che un individuo preso a caso nella popolazione abbia lo stesso DNA del colpevole è minima. Ci sono sia motivi matematici che motivi extramatematici per questo. Per esempio, nella nostra analisi abbiamo *supposto* che il colpevole faccia parte della stessa popolazione del sospettato; se non è così, la distribuzione di probabilità del DNA del colpevole è diversa da quella usata per i nostri conti, e quindi il risultato che abbiamo ottenuto non è significativo. Se poi (e questo è un motivo extramatematico) il sospettato è in grado di dimostrare che era in un altro continente quando è avvenuto lo stupro, allora è innocente anche in presenza di un test del DNA che indica una probabilità su 6 miliardi. Questo non deve stupirti: le probabilità forniscono una stima della *plausibilità* di un evento, e non un’indicazione certa che quell’evento sia o meno avvenuto.

Esercizio 2.2 Usando i dati della Tabella 2.7, calcola la probabilità che un individuo preso a caso sia omozigote “11” in TPOX ed eterozigote “34” in CSF1P0.

2.8 Probabilità condizionata

Confronta le tre domande seguenti:

- (a) Qual è la probabilità che Leonardo e Martina abbiano un figlio con fenotipo Rh^- ?
- (b) Qual è la probabilità che Leonardo e Martina abbiano fenotipo Rh^+ e loro figlio abbia fenotipo Rh^- ?
- (c) Qual è la probabilità che Leonardo e Martina abbiano un figlio con fenotipo Rh^- , sapendo che Leonardo e Martina hanno fenotipo Rh^+ ?

La prima domanda chiede semplicemente di calcolare la probabilità dell'evento “figlio con fenotipo Rh^- ”, indipendentemente da cosa succeda e da cosa sappiamo sui genitori — e abbiamo visto nella sezione precedente come usare la legge di Hardy-Weinberg per rispondere a domande del genere (Esempio 2.34), ottenendo una probabilità di circa il 15.1%. La seconda domanda chiede di calcolare la probabilità che avvengano contemporaneamente i due eventi “Leonardo e Martina con fenotipo Rh^+ ”, e “figlio con fenotipo Rh^- ”. Se Leonardo o Martina hanno genotipo omozigote Rh^+Rh^+ , non possono avere un figlio con fenotipo Rh^- ; quindi la risposta alla seconda domanda è uguale alla probabilità che Leonardo e Martina abbiano genotipo Rh^+Rh^- e il figlio abbia genotipo Rh^-Rh^- , probabilità che nell'Esempio 2.34 abbiamo visto essere circa 5.6%.

La terza domanda è di natura un po' diversa: ci chiede di calcolare la probabilità dell'evento “figlio con fenotipo Rh^- ” *conoscendo* il fenotipo dei genitori. Rispetto alla prima domanda, abbiamo un'informazione in più (il fenotipo dei genitori); rispetto alla seconda domanda, non stiamo chiedendo con quale probabilità avvengano due eventi assieme, ma bensì con quale probabilità avviene un evento sapendo che l'altro è già avvenuto.

In altre parole, stiamo cercando un modo di calcolare la probabilità di un evento B sapendo che l'evento A è già avvenuto. Questa informazione chiaramente cambia lo spazio degli eventi (dobbiamo escludere tutti gli eventi incompatibili con A); quindi è ragionevole che cambi anche la probabilità. Ma come?

Per avere un'idea di quale può essere una risposta ragionevole a questa domanda, torniamo all'esempio di probabilità data come rapporto di aree (Esempio 2.27). Sapere che l'evento A è accaduto corrisponde a considerare solo il sottoinsieme A nell'intero spazio degli eventi Ω . In questo “sottospazio degli eventi”, l'evento B è rappresentato dall'intersezione $B \cap A$. Quindi se la probabilità di B nell'intero spazio degli eventi corrispondeva al rapporto fra l'area di B e l'area di Ω , la probabilità di B nel sottospazio degli eventi dato da A corrisponderà al rapporto fra l'area di $B \cap A$ e l'area di A . Ma possiamo esprimere questo rapporto in maniera più furba (e generalizzabile in tutti gli altri casi). Infatti,

$$\frac{\text{Area}(B \cap A)}{\text{Area}(A)} = \frac{\text{Area}(B \cap A)/\text{Area}(\Omega)}{\text{Area}(A)/\text{Area}(\Omega)} = \frac{p(A \cap B)}{p(A)}.$$

Questa formula ci suggerisce la seguente definizione. Diremo *probabilità condizio-*

nata (o condizionale) $p(B|A)$ dell'evento B rispetto all'evento A il quoziente

$$p(B|A) = \frac{p(B \cap A)}{p(A)} ; \quad (2.11)$$

ovviamente, $p(B|A)$ è definito solo se l'evento A ha probabilità non nulla. A volte si trova scritto $p(B \text{ dato } A)$ invece di $p(B|A)$.

Come detto prima, $p(B|A)$ dev'essere interpretata come la probabilità che sia avvenuto l'evento B sapendo che l'evento A è accaduto. Ma vediamo alcuni esempi.

ESEMPIO 2.36 Vogliamo calcolare la probabilità che la somma del lancio di 2 dadi a sei facce non truccati sia maggiore o uguale a 8 sapendo che il risultato del primo dado è pari. L'evento A “risultato del primo dado pari” contiene 18 elementi, per cui $p(A) = 18/36 = 1/2$. Fra questi, solo 9 appartengono anche all'evento B “somma maggiore o uguale a 8” (sono gli elementi dell'evento $B \cap A$); quindi la probabilità che la somma sia maggiore o uguale a 8 sapendo che il primo dado è pari è uguale a $9/18$ (numero di successi diviso numero di tentativi). D'altra parte, siccome l'evento $B \cap A$ contiene 9 elementi, abbiamo $p(B \cap A) = 9/36 = 1/4$, per cui otteniamo $p(B|A) = p(B \cap A)/p(A) = (1/4)/(1/2) = 1/2$, in accordo con l'interpretazione intuitiva della probabilità condizionata.

ESEMPIO 2.37 Abbiamo il solito cassetto contenente 12 calzini bianchi e 4 calzini neri. Vogliamo calcolare la probabilità di estrarre (senza rimbussolamento) un paio spaiato sapendo che il primo calzino estratto è nero. Dobbiamo calcolare $p(B|A)$, dove A è l'evento “primo calzino estratto nero” e B è l'evento “i calzini estratti sono spaiati”. Nell'Esempio 2.33 abbiamo visto che la probabilità dell'evento A è $p(A) = 1/4$. L'evento $B \cap A$ corrisponde a estrarre prima un calzino nero e poi un calzino bianco, e sempre nell'Esempio 2.33 abbiamo calcolato $p(B \cap A) = 1/5$; quindi otteniamo

$$p(B|A) = \frac{1/5}{1/4} = \frac{4}{5}.$$

D'altra parte, è chiaro che la probabilità di estrarre (senza rimbussolamento) un paio spaiato se il primo calzino estratto è nero coincide con la probabilità di estrarre un calzino bianco da un cassetto contenente 12 calzini bianchi e 3 calzini neri. Questa probabilità è $12/15 = 4/5$, in accordo con quanto appena trovato.

ESEMPIO 2.38 Vogliamo finalmente calcolare la probabilità che Leonardo e Martina abbiano un figlio con fenotipo Rh^- negativo sapendo che sia Leonardo sia Martina hanno fenotipo Rh^+ . Indicando con A l'evento “Leonardo e Martina hanno fenotipo Rh^+ ” e con B l'evento “il figlio ha fenotipo Rh^- ”, dobbiamo calcolare la probabilità condizionata $p(B|A)$, partendo ovviamente dai dati visti nell'Esempio 2.34. Siccome i fenotipi di Leonardo e Martina sono indipendenti, la probabilità che entrambi abbiano fenotipo Rh^+ è il quadrato della probabilità che uno di loro due abbia fenotipo Rh^+ , cioè $p(A) = (84.9/100)^2 \simeq 0.72$. All'inizio di questa sezione abbiamo anche calcolato la probabilità che Leonardo e Martina abbiano fenotipo Rh^+ e il figlio abbia fenotipo Rh^- , ottenendo $p(B \cap A) \simeq 0.056$. Quindi

$p(B|A) \simeq 0.056/0.72 \simeq 0.078$. Quindi la probabilità che Leonardo e Martina abbiano un figlio con fenotipo Rh^- negativo sapendo che sia Leonardo sia Martina hanno fenotipo Rh^+ è di circa il 7.8%.

Esercizio 2.3 Calcola (usando i dati dell'Esempio 2.34) la probabilità condizionata che Leonardo e Martina abbiano un figlio con fenotipo Rh^- negativo sapendo che Leonardo ha fenotipo Rh^+ e Martina ha fenotipo Rh^- .

Esercizio 2.4 Calcola (usando i dati dell'Esempio 2.34) la probabilità condizionata che Leonardo e Martina abbiano un figlio con fenotipo Rh^- negativo sapendo che Leonardo e Martina hanno genotipo eterozigote Rh^+Rh^- , e confronta il risultato che hai ottenuto con la legge di disgiunzione di Mendel. In particolare, per rispondere a questa domanda ti serve davvero conoscere la distribuzione di probabilità dei vari genotipi o puoi farne a meno?

ESEMPIO 2.39 Supponiamo che un uomo abbia felicemente compiuto sessant'anni. Come si calcola la probabilità che sopravviva fino ai settant'anni? Ovviamente, ci servono dei dati. Per i maschi italiani nel 2002, la probabilità di morte in un certo intervallo d'età (calcolata come rapporto fra il numero di morti in quell'intervallo d'età nel 2002 e il numero totale di morti nel 2002) è riassunta nella Tabella 2.8.

Intervallo d'età	Probabilità di morte
0-9	0.616%
10-19	0.377%
20-29	0.88%
30-39	1.029%
40-49	2.024%
50-59	5.203%
60-69	12.665%
70-79	27.126%
80-89	35.57%
90-999	14.51%

TABELLA 2.8

La probabilità dell'evento A “sopravvivere fino a sessant'anni” è uguale alla probabilità di morire dopo i sessant'anni, per cui

$$p(A) = 0.12665 + 0.27126 + 0.3557 + 0.1451 = 89.871\% .$$

Siccome l'evento B “sopravvivere fino a settant'anni” è chiaramente un sottoinsieme di A , abbiamo $B \cap A = B$. La probabilità di B è uguale alla probabilità di morire dopo i settant'anni, per cui

$$p(B \cap A) = p(B) = 0.27126 + 0.3557 + 0.1451 = 77.206\% .$$

Dunque la probabilità di arrivare a settant'anni avendone compiuti sessanta è

$$p(B|A) = \frac{0.77206}{0.89871} \simeq 86\% .$$

Osservazione 2.24 Se gli eventi A e B sono indipendenti, abbiamo

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

In altre parole, il fatto che sia accaduto l'evento A non muta le probabilità dell'evento B , in perfetto accordo con l'idea intuitiva di cosa vuol dire che due eventi sono indipendenti.

Riscrivendo in maniera diversa la definizione di probabilità condizionata, otteniamo la seguente identità:

$$p(A)p(B|A) = p(B \cap A) = p(A \cap B) = p(B)p(A|B) . \quad (2.12)$$

Di conseguenza otteniamo la *formula di Bayes*

$$p(A|B) = \frac{p(A)p(B|A)}{p(B)} , \quad (2.13)$$

che collega la probabilità condizionata che sia avvenuto B sapendo A alla probabilità condizionata che sia avvenuto A sapendo B .

Un'altra formula utile, la *legge delle alternative*, segue dal fatto che B è l'unione disgiunta di $B \cap A$ e $B \cap (\Omega \setminus A)$. Infatti, da (2.12) segue

$$\begin{aligned} p(B) &= p((B \cap A) \cup (B \cap \Omega \setminus A)) = p(B \cap A) + p(B \cap \Omega \setminus A) \\ &= p(A)p(B|A) + p(\Omega \setminus A)p(B|\Omega \setminus A) . \end{aligned} \quad (2.14)$$

In parole povere, la legge delle alternative dice che *la probabilità di A è uguale alla somma del prodotto fra la probabilità condizionata di A sapendo che è avvenuto B e la probabilità di B , con il prodotto fra la probabilità condizionata di A sapendo che non è avvenuto B e la probabilità che non sia avvenuto B .*

Un'ultima formula utile permette di calcolare $p(\Omega \setminus B|A)$ conoscendo $p(B|A)$. Infatti

$$p(\Omega \setminus B|A) = \frac{p((\Omega \setminus B) \cap A)}{p(A)} = \frac{p(A) - p(B \cap A)}{p(A)} = 1 - p(B|A) . \quad (2.15)$$

ESEMPIO 2.40 Il *tasso d'incidenza* (o *prevalenza*) di una malattia in una popolazione è la probabilità che un individuo scelto a caso nella popolazione sia malato. Supponiamo di avere una popolazione composta da un 15% di soggetti a rischio (cioè con caratteristiche che li rendono più attaccabili dalla malattia), e di sapere che il tasso d'incidenza della malattia è 0.2 fra i soggetti a rischio, e 0.006 fra i

soggetti non a rischio. Vogliamo trovare: l'incidenza della malattia nell'intera popolazione; qual è la probabilità che un individuo malato sia un soggetto a rischio; e qual è la probabilità che un individuo sano sia un soggetto a rischio.

Indichiamo con Ω l'intera popolazione, con A l'evento "soggetto a rischio", e con B l'evento "individuo malato". I dati a nostra disposizione sono $p(A) = 0.15$, la probabilità che un individuo sia a rischio; $p(B|A) = 0.2$, la probabilità condizionata che un soggetto a rischio sia malato; e $p(B|\Omega \setminus A) = 0.006$, la probabilità condizionata che un soggetto non a rischio sia malato. Da questi dati vogliamo ricavare: $p(B)$, la probabilità che un individuo sia malato; $p(A|B)$, la probabilità condizionata che un individuo malato sia a rischio; e $p(A|\Omega \setminus B)$, la probabilità condizionata che un individuo sano sia a rischio.

Per calcolare $p(B)$ conviene usare la legge delle alternative. L'unico dato che ci manca è la probabilità $p(\Omega \setminus A)$ che un individuo non sia a rischio; ma la (2.3) ci dà subito

$$p(\Omega \setminus A) = 1 - p(A) = 0.85 .$$

Quindi

$$p(B) = p(A)p(B|A) + p(\Omega \setminus A)p(B|\Omega \setminus A) = 0.15 \cdot 0.2 + 0.85 \cdot 0.006 = 0.0351 ,$$

per cui l'incidenza della malattia sull'intera popolazione è del 3.51%.

Per calcolare $p(A|B)$ chiaramente conviene usare la formula di Bayes:

$$p(A|B) = \frac{p(A)p(B|A)}{p(B)} = \frac{0.15 \cdot 0.2}{0.0351} \simeq 0.855 ,$$

cioè più di 8 malati su 10 sono soggetti a rischio. Infine, calcoliamo $p(A|\Omega \setminus B)$ usando la formula di Bayes e la (2.15):

$$p(A|\Omega \setminus B) = \frac{p(A)p(\Omega \setminus B|A)}{p(\Omega \setminus B)} = \frac{p(A)(1 - p(B|A))}{1 - p(B)} = \frac{0.15(1 - 0.2)}{1 - 0.0351} \simeq 0.124 ,$$

cioè solo circa 1 individuo sano su 8 è a rischio.

Un *test diagnostico* è un esame il cui scopo è determinare se un dato individuo soffre di una certa malattia. La probabilità condizionata permette di definire alcune caratteristiche di un test diagnostico: la specificità, la sensibilità, e i valori predittivi. Se indichiamo con M^+ l'evento "malato", con $M^- = \Omega \setminus M^+$ l'evento "sano", con T^+ l'evento "risultato del test positivo", e con $T^- = \Omega \setminus T^+$ l'evento "risultato del test negativo", allora la *specificità* Sp del test è la probabilità condizionata $p(T^-|M^-)$, cioè la probabilità di risultato negativo in individui sani; maggiore è la specificità minori sono i falsi positivi (cioè i casi in cui il test dà risultato positivo su un soggetto sano). La *sensibilità* Se è la probabilità condizionata $p(T^+|M^+)$, cioè la probabilità di risultato positivo in individui malati; maggiore è la sensibilità minori sono i falsi negativi (cioè i casi in cui il test dà risultato negativo su un soggetto malato). Specificità e sensibilità sono tipicamente

calcolati sottoponendo al test una popolazione campione di cui si conosca già, per altre vie, il numero di malati e di sani (vedi l'esempio successivo).

Per il medico che vuole invece usare il test per stabilire se un suo paziente è ammalato o meno, sono ben più interessanti i valori predittivi. Il *valore predittivo di esito negativo* Vp^- è la probabilità condizionata $p(M^-|T^-)$, cioè la probabilità che un individuo con test negativo sia effettivamente sano, mentre il *valore predittivo di esito positivo* Vp^+ è la probabilità condizionata $p(M^+|T^+)$, cioè la probabilità che un individuo con test positivo sia effettivamente malato.

La formula di Bayes e la legge delle alternative ci permettono di ricavare i valori predittivi a partire da specificità, sensibilità e tasso d'incidenza $p(M^+)$. Infatti, abbiamo

$$Vp^- = p(M^-|T^-) = \frac{p(M^-)p(T^-|M^-)}{p(T^-)} = \frac{1 - p(M^+)}{p(T^-)} Sp.$$

La legge delle alternative e la (2.15) ci permettono di calcolare $p(T^-)$:

$$\begin{aligned} p(T^-) &= p(M^-)p(T^-|M^-) + p(M^+)p(T^-|M^+) \\ &= (1 - p(M^+))Sp + p(M^+)(1 - p(T^+|M^+)) \\ &= (1 - p(M^+))Sp + p(M^+)(1 - Se), \end{aligned}$$

per cui

$$Vp^- = \frac{1 - p(M^+)}{(1 - p(M^+))Sp + p(M^+)(1 - Se)} Sp.$$

Analogamente,

$$\begin{aligned} Vp^+ &= p(M^+|T^+) = \frac{p(M^+)p(T^+|M^+)}{p(T^+)} = \frac{p(M^+)}{1 - p(T^-)} Se \\ &= \frac{p(M^+)}{p(M^+)Se + (1 - p(M^+))(1 - Sp)} Se. \end{aligned}$$

ESEMPIO 2.41 Hai effettuato un test diagnostico su un campione di 128 individui, di cui 86 malati e 42 sani. Il tuo assistente ti comunica che 72 malati sono risultati positivi al test, e che 30 individui sani sono risultati negativi al test; con questi dati vuoi calcolare specificità, sensibilità e valori predittivi del test. La specificità è la probabilità che un individuo sano risulti negativo al test; quindi $Sp = 30/42 \simeq 0.71$. La sensibilità è la probabilità che un individuo malato risulti positivo al test; quindi $Se = 72/86 \simeq 0.84$. L'incidenza della malattia è la probabilità che un individuo nel campione sia malato; quindi $p(M^+) = 86/128 \simeq 0.67$. Da questi dati ricaviamo i valori predittivi:

$$\begin{aligned} Vp^+ &\simeq \frac{0.67}{0.67 \cdot 0.84 + (1 - 0.67)(1 - 0.71)} 0.84 \simeq 0.85, \\ Vp^- &\simeq \frac{1 - 0.67}{(1 - 0.67) \cdot 0.71 + 0.67(1 - 0.84)} 0.71 \simeq 0.69, \end{aligned}$$

Quindi con questo test abbiamo una probabilità di circa l'85% che un soggetto positivo al test sia effettivamente malato, e una probabilità di circa il 69% che un soggetto negativo al test sia effettivamente sano.

Esercizio 2.5 *Nell'esempio precedente, 14 malati sono risultati negativi al test, per cui il numero totale di risultati negativi è 44. Ricordando che il numero dei sani con risultato del test negativo è 30, verifica che il valore esatto del valore predittivo di esito negativo è $Vp^- = 30/44$. In modo analogo, verifica che il valore esatto del valore predittivo di esito positivo è $Vp^+ = 72/84$. Secondo te, come mai abbiamo ottenuto valori lievemente diversi da quelli calcolati nell'esempio?*

Bisogna stare attenti nell'interpretare le probabilità condizionate, in quanto ogni tanto portano a risultati apparentemente paradossali.

ESEMPIO 2.42 È passato qualche anno, e ora Leonardo e Martina hanno due figli. Qual è la probabilità che siano entrambi maschi, sapendo che uno dei due lo è? E qual è la probabilità che siano entrambi maschi, sapendo che il primo lo è? Le due domande sembrano praticamente identiche, ma non lo sono: il risultato è diverso. Lo spazio degli eventi è $\Omega = \{(M, M), (M, F), (F, M), (F, F)\}$, dove (M, F) vuol dire “primo figlio maschio, secondo femmina”, e così via. Nella prima domanda, l'evento A è “Leonardo e Martina hanno almeno un figlio maschio”, cioè $A = \{(M, M), (M, F), (F, M)\}$, e l'evento B è “entrambi i figli maschi”, cioè $B = \{(M, M)\}$. Abbiamo visto nell'Esempio 2.7 che la probabilità in Italia della nascita di un figlio maschio è circa il 51.3%. Supponendo che il sesso nella seconda nascita sia indipendente dal sesso nella prima, troviamo $p(B) = 0.513 \cdot 0.513 \simeq 26.3\%$. L'evento A si può realizzare in due modi: il primogenito è maschio (evento $A_1 = \{(M, M), (M, F)\}$), oppure il secondogenito è maschio (evento $A_2 = \{(M, M), (F, M)\}$). Ora $A = A_1 \cup A_2$, ma A_1 e A_2 non sono incompatibili: $A_1 \cap A_2 = B$. Siccome $p(A_1) = p(A_2) = 0.513$, otteniamo

$$p(A) = p(A_1) + p(A_2) - p(B) \simeq 0.513 + 0.513 - 0.263 = 0.763 = 76.3\% .$$

Siccome $B \subset A$, abbiamo $p(B \cap A) = p(B)$; quindi la risposta alla prima domanda è

$$p(B|A) \simeq \frac{0.263}{0.763} \simeq 34.5\% .$$

La seconda domanda chiede invece di calcolare $p(B|A_1)$; siccome $B \subset A_1$ otteniamo

$$p(B|A_1) \simeq \frac{0.263}{0.513} = 51.3\% \neq 34.5\% .$$

La morale di questo esempio è che *ogni informazione, per quanto apparentemente insignificante, può modificare le probabilità condizionate*.

Esercizio 2.6 *Se p è la probabilità di avere un figlio maschio, verifica che, avendo due figli, la probabilità che siano entrambi maschi sapendo che uno è maschio è $p/(2-p)$, mentre la probabilità che siano entrambi maschi sapendo che il primo è maschio è p .*

CURIOSITÀ 2.3 La probabilità condizionata è una probabilità. Infatti, supponiamo di avere uno spazio degli eventi Ω con una distribuzione di probabilità p , e un evento $A \subseteq \Omega$ con $p(A) > 0$. Se associamo a ogni evento $B \in \mathcal{A}$ il numero $p_A(B) = p(B|A)$ si vede subito che questa associazione soddisfa gli assiomi (P1)–(P3), per cui è una distribuzione di probabilità.

2.9 Calcolo combinatorio

Più di una volta nelle sezioni precedenti abbiamo avuto bisogno di contare gli elementi di un insieme (il numero dei possibili risultati di due dadi, distinti o identici; il numero di possibili attribuzioni di trattamenti a cavie; eccetera). Questo tipo di problemi è materia del *calcolo combinatorio*.

Un principio che permette di risolvere diversi problemi di calcolo combinatorio è il seguente *principio base*: *Se devi compiere k scelte, e per la scelta i -esima hai n_i alternative possibili, allora il numero totale di alternative possibili per le k scelte è $n_1 \cdot n_2 \cdots n_k$.*

Per esempio, supponi di dover fare 3 scelte ($k = 3$), e di avere 2 alternative per la prima scelta ($n_1 = 2$), 4 alternative per la seconda scelta ($n_2 = 4$), e 3 alternative per la terza scelta ($n_3 = 3$). La prima scelta può essere effettuata in due modi diversi. Dato che la seconda scelta può essere effettuata in 4 modi diversi, abbiamo due possibilità per la prima e 4 per la seconda, per un totale di $2 \cdot 4 = 8$ possibilità per la coppia (ordinata) delle prime due scelte. Analogamente, se abbiamo 8 possibilità per la coppia delle prime due scelte, e 3 alternative per la terza scelta, otteniamo $8 \cdot 3 = 24$ possibilità per la tripla di scelte.

Vediamo qualche esempio molto semplice di applicazione di questo principio.

ESEMPIO 2.43 *Quante triple di risultati distinti possiamo ottenere tirando un dado a 6 facce, uno a 8 facce e uno a 20 facce?* Il dado a 6 facce ha 6 risultati possibili, quello a 8 facce ne ha 8, e quello a 20 facce ne ha 20. Il principio base ci dice allora che il numero totale di possibili triple ottenibili è $6 \cdot 8 \cdot 20 = 960$.

ESEMPIO 2.44 *Quante parole (sensate o meno) di 4 lettere possiamo scrivere con l'alfabeto italiano?* Siccome l'alfabeto italiano è composto da 21 lettere, possiamo scegliere ciascuna lettera in 21 modi diversi. Quindi si tratta di effettuare 4 scelte, ognuna con 21 alternative; la risposta quindi è $21 \cdot 21 \cdot 21 \cdot 21 = 21^4 = 194\,481$.

ESEMPIO 2.45 *Supponiamo di avere un'urna contenente 90 palline numerate da 1 a 90. Quante possibili sequenze ordinate di numeri possiamo ottenere estraendo (con rimbussolamento) 5 palline?* Siccome dopo ogni estrazione rimettiamo la pallina estratta nell'urna, ogni estrazione ha 90 risultati possibili. Quindi la risposta è $90^5 = 5\,904\,900\,000$.

Questi ultimi due esempi sono chiaramente casi dello stesso fenomeno. Una *disposizione con ripetizione di n oggetti k alla volta* consiste nell'effettuare k scelte consecutive fra n oggetti; ogni oggetto può venire scelto anche più di una volta. Dunque abbiamo k scelte, e n alternative per ogni scelta. Il principio base quindi ci dice che il numero $D_{n,k}$ di possibili disposizioni con ripetizione di n oggetti k

alla volta è

$$D_{n,k} = n^k . \quad (2.16)$$

La situazione diventa un poco più interessante se escludiamo le ripetizioni.

ESEMPIO 2.46 *Quante parole (sensate o meno) di 4 lettere tutte distinte possiamo scrivere con l'alfabeto italiano?* Come prima, possiamo scegliere la prima lettera in 21 modi diversi. Ma per seconda non possiamo usare la lettera scelta prima; quindi per la seconda lettera abbiamo solo 20 alternative. La terza lettera dev'essere diversa dalle prime due; quindi può essere scelta solo in 19 modi diversi. Analogamente, la quarta lettera dev'essere diversa dalle altre tre, e quindi abbiamo 18 alternative. Il principio base ci fornisce allora la risposta $21 \cdot 20 \cdot 19 \cdot 18 = 143\,640$.

ESEMPIO 2.47 *Supponiamo di avere un'urna contenente 90 palline numerate da 1 a 90. Quante possibili sequenze ordinate di numeri possiamo ottenere estraendo (senza rimbussolamento) 5 palline?* Siccome dopo ogni estrazione non rimettiamo la pallina estratta nell'urna, ogni estrazione ha un risultato possibile in meno rispetto alla precedente. Quindi la risposta è $90 \cdot 89 \cdot 88 \cdot 87 \cdot 86 = 5\,273\,912\,160$.

Anche in questo caso è chiaro cosa sta succedendo. Una *disposizione senza ripetizioni* di n oggetti k alla volta consiste nell'effettuare k scelte consecutive fra n oggetti, ma ogni oggetto può venire scelto al massimo una volta sola. Dunque abbiamo n alternative per la prima scelta, $n - 1$ per la seconda, $n - 2$ per la terza, e così via, fino ad avere $n - k + 1$ alternative per la k -esima scelta. Il principio base quindi ci dice che il numero $P_{n,k}$ di possibili disposizioni senza ripetizioni di n oggetti k alla volta è

$$P_{n,k} = n(n-1) \cdots (n-k+1) . \quad (2.17)$$

Un caso particolarmente importante è $n = k$. Una *permutazione* su n oggetti è una disposizione senza ripetizioni di n oggetti presi n alla volta. Siccome non abbiamo ripetizioni, ogni oggetto viene scelto esattamente una volta, per cui una permutazione non è altro che un modo di mettere in sequenza gli n oggetti. La formula (2.17) ci dice che il numero $P_n = P_{n,n}$ di permutazioni su n oggetti è

$$P_n = n(n-1) \cdots 2 \cdot 1 . \quad (2.18)$$

Molto spesso, il numero $1 \cdot 2 \cdots n$ viene indicato con $n!$, che si legge *n fattoriale*. Per esempio,

$$1! = 1 , \quad 2! = 1 \cdot 2 = 2 , \quad 3! = 1 \cdot 2 \cdot 3 = 6 , \quad 4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24 , \quad 5! = 4! \cdot 5 = 120 .$$

Usando questo simbolo, la formula (2.18) diventa

$$P_n = n! .$$

Inoltre, si vede subito che per $n > k$ la (2.17) diventa

$$P_{n,k} = \frac{n!}{(n-k)!} = \frac{P_n}{P_{n-k}}. \quad (2.19)$$

Osservazione 2.25 Per convenzione si pone $0! = 1$. In questo modo la (2.19) è vera anche per $n = k$.

ESEMPIO 2.48 *Supponendo che le nascite siano distribuite in modo casuale su tutto l'anno, e trascurando il 29 febbraio, qual è il minimo k per cui la probabilità che in un gruppo di k persone scelte a caso ce ne siano 2 con lo stesso compleanno sia maggiore di $1/2$?* Per rispondere, calcoliamo la probabilità che k persone prese a caso abbiano tutte compleanno diverso. Per quanto visto, ci sono $D_{365,k} = 365^k$ possibili disposizioni di 365 compleanni k alla volta. Fra queste, solo $P_{365,k} = 365!/(365-k)!$ sono composte da compleanni tutti diversi. Quindi la probabilità che k persone prese a caso abbiano tutte compleanno diverso è $P_{365,k}/D_{365,k}$, e la probabilità che fra k persone prese a caso almeno due abbiano lo stesso compleanno è $1 - P_{365,k}/D_{365,k}$. Ora,

$$\frac{P_{365,1}}{D_{365,1}} = \frac{365}{365} = 1, \quad \frac{P_{365,2}}{D_{365,2}} = \frac{365}{365} \cdot \frac{364}{365}, \quad \frac{P_{365,3}}{D_{365,3}} = \left(\frac{365}{365} \cdot \frac{364}{365} \right) \cdot \frac{363}{365},$$

e, in generale,

$$\frac{P_{365,k}}{D_{365,k}} = \frac{P_{365,k-1}}{D_{365,k-1}} \cdot \frac{365-k+1}{365} < \frac{P_{365,k-1}}{D_{365,k-1}}.$$

Quindi all'aumentare di k il quoziente $P_{365,k}/D_{365,k}$ decresce, e la probabilità $1 - P_{365,k}/D_{365,k}$ cresce. Calcolando si trova facilmente che il minimo valore di k per cui questa probabilità è maggiore di $1/2$ è $k = 23$, quando vale circa 0.507. Quindi in un gruppo di 23 persone prese a caso c'è una probabilità maggiore del 50% che due di loro abbiano lo stesso compleanno. Se poi il gruppo contiene almeno $k = 41$ persone, allora la probabilità è maggiore del 90%.

Adesso vediamo di risolvere un problema un poco più complicato.

ESEMPIO 2.49 *Quanti sottoinsiemi di 4 lettere distinte si possono formare con l'alfabeto italiano?* Abbiamo visto che con l'alfabeto italiano possiamo formare $P_{21,4}$ parole con 4 lettere distinte. Due parole diverse danno lo stesso sottoinsieme se e solo se sono composte dalle stesse lettere, cioè se e solo se sono una permutazione dell'altra. Dunque possiamo ottenere tutte le parole composte da 4 lettere distinte scegliendo prima un sottoinsieme di 4 lettere distinte, e poi scegliendo una permutazione di queste 4 lettere. Il principio base ci dice allora che $P_{21,4}$ dev'essere uguale al prodotto del numero che cerchiamo con P_4 , per cui la risposta è $P_{21,4}/P_4 = 5985$.

ESEMPIO 2.50 *Supponiamo di avere un'urna contenente 90 palline numerate da 1 a 90. Quante possibili cinquine possiamo ottenere estraendo (senza rimbussolamento) 5 palline?* Una cinquina è il risultato dell'estrazione di 5 palline distinte ordinate in ordine crescente. Ogni cinquina può venire ottenuta estraendo una sua permutazione qualsiasi; ci sono $P_{90,5}$ estrazioni possibili e P_5 permutazioni possibili. Come nell'esempio precedente, otteniamo quindi $P_{90,5}/P_5 = 43\,949\,268$ cinquine possibili.

Una *combinazione senza ripetizioni* di n oggetti k alla volta è un sottoinsieme degli n oggetti composto da k elementi. Ognuna delle P_k permutazioni dei k oggetti forma una disposizione senza ripetizione di n oggetti k alla volta; viceversa, ognuna delle $P_{n,k}$ disposizione senza ripetizione di n oggetti k alla volta si ottiene in un solo modo con questa tecnica. Quindi se indichiamo con $C_{n,k}$ il numero delle possibili combinazioni senza ripetizione di n oggetti k alla volta otteniamo $C_{n,k}P_k = P_{n,k}$, cioè

$$C_{n,k} = \frac{P_{n,k}}{P_k} = \frac{n!}{(n-k)!k!}. \quad (2.20)$$

ESEMPIO 2.51 *Quanti genotipi può avere un gene con 7 alleli?* I genotipi omozigoti sono chiaramente 7. I genotipi eterozigoti sono invece combinazioni senza ripetizioni dei 7 alleli presi 2 alla volta, per cui sono $C_{7,2} = 7!/5!2! = (7 \cdot 6)/2 = 21$. Quindi in totali ci sono 28 genotipi possibili.

Il numero $C_{n,k}$ è spesso chiamato *coefficiente binomiale*, ed è indicato col simbolo $\binom{n}{k}$:

$$\binom{n}{k} = \frac{n!}{(n-k)!k!},$$

ed è definito per $0 \leq k \leq n$; nota che la convenzione $0! = 1$ implica

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

Inoltre,

$$\binom{n}{1} = n = \binom{n}{n-1}, \quad \binom{n}{2} = \frac{n(n-1)}{2} = \binom{n}{n-2},$$

e

$$\binom{n}{k} = \frac{n!}{(n-k)!k!} = \frac{n!}{(n-(n-k))!(n-k)!} = \binom{n}{n-k} \quad (2.21)$$

per ogni $0 \leq k \leq n$.

Il nome “coefficiente binomiale” deriva dal fatto che compare nello sviluppo delle potenze del binomio. Se ci pensi un attimo, il quadrato e il cubo del binomio si possono scrivere in questo modo:

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

In generale, nello sviluppo di $(a+b)^n$ il termine $a^{n-k}b^k$ è ottenuto scegliendo b in k dei fattori $a+b$ del prodotto, ed a nei restanti $n-k$ fattori. Quindi il coefficiente di $a^{n-k}b^k$ è uguale al numero di modi con cui possiamo scegliere k fattori distinti fra gli n disponibili, e l'ordine non conta. In altre parole, il coefficiente di $a^{n-k}b^k$ è esattamente $C_{n,k} = \binom{n}{k}$, per cui abbiamo la formula generale per la potenza del binomio:

$$(a+b)^n = \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \cdots + \binom{n}{k}a^{n-k}b^k + \cdots + \binom{n}{n-1}ab^{n-1} + \binom{n}{n}b^n. \quad (2.22)$$

Osservazione 2.26 Per esprimere somme come questa è stato introdotto l'utilissimo simbolo di sommatoria \sum . Una somma del tipo $a_1 + \cdots + a_n$, in cui dobbiamo sommare n termini a_i dipendenti da un indice i che varia da 1 a n , si può scrivere col simbolo di sommatoria come segue:

$$a_1 + \cdots + a_n = \sum_{i=1}^n a_i = \sum_{1 \leq i \leq n} a_i.$$

L'indice può variare anche in altri intervalli, e può assumere anche altri nomi; per esempio, la formula (2.22) può essere scritta in questo modo:

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^{n-k} b^k. \quad (2.23)$$

Il vantaggio del simbolo di sommatoria è che, oltre a essere più compatto, evidenzia bene la dipendenza del singolo addendo dall'indice di sommatoria. Altre utili informazioni sul simbolo di sommatoria le troverai nei complementi a questo capitolo; qui mi preme ricordare che esiste anche il simbolo \prod di *produttoria*, che è l'equivalente della sommatoria per il prodotto. Per esempio, la definizione di fattoriale si può scrivere in questo modo:

$$n! = \prod_{j=1}^n j.$$

I coefficienti binomiali godono di una quantità incredibile di proprietà. Una delle più note viene rivelata se costruiamo il cosiddetto *triangolo di Tartaglia* (o di *Pascal*) mettendo i coefficienti binomiali in riga uno sopra l'altro come nella Tabella 2.9.

Si nota subito che ogni numero in questo triangolo è uguale alla somma dei due numeri subito sopra di lui (e questo vale anche per i numeri sul bordo se immaginiamo di bordare il triangolo con degli zeri). In altre parole, il triangolo di Tartaglia ci suggerisce la seguente identità:

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}, \quad (2.24)$$

TABELLA 2.9

La formula per lo sviluppo del binomio fornisce altre due identità che a volte sono utili:

Infatti.

e

Verificale sulle righe del triangolo di Tartaglia.

Siccome le caviè sono scelte a caso, tutte le possibili selezioni di 4 caviè hanno la stessa probabilità di verificarsi. Quindi se indichiamo con C il numero delle possibili selezioni di 4 caviè, la probabilità che le caviè scelte siano la 2, la 7, la 11

e la 12 è $1/C$. Ora, siccome l'ordine in cui vengono scelte le cavi è irrilevante, C è uguale al numero di combinazioni di 15 oggetti a 4 alla volta, cioè

$$C = C_{15,4} = \binom{15}{4} = \frac{15!}{11!4!} = \frac{15 \cdot 14 \cdot 13 \cdot 12}{4 \cdot 3 \cdot 2 \cdot 1} = 1365.$$

Dunque la risposta alla domanda (a) è $1/1365 \simeq 0.07\%$. Per rispondere alla domanda (b), basta notare che il numero di possibili selezioni di 4 cavi fra le prime 8 è

$$C_{8,4} = \binom{8}{4} = \frac{8!}{4!4!} = 70,$$

per cui la risposta alla seconda domanda è $70/1365 \simeq 5.1\%$. Infine, la domanda (c) chiede di calcolare una probabilità condizionata. Indichiamo con A l'evento "la selezione contiene la cavia 1", e con B l'evento "la selezione contiene solo cavi scelte fra le prime 8". Dobbiamo calcolare $p(B|A) = p(B \cap A)/p(A)$. Siccome stiamo usando una distribuzione di probabilità uniforme, la probabilità $p(E)$ di un qualsiasi evento E è data da $p(E) = \#E/1365$, dove $\#E$ indica il numero di elementi di E . Quindi $p(B|A) = \#(B \cap A)/\#A$, per cui dobbiamo contare il numero di elementi di A e il numero di elementi di $B \cap A$. Ora, il numero di selezioni di 4 cavi che contengono la cavia 1 è chiaramente uguale al numero di selezioni di 3 cavi fra le rimanenti 14; quindi $\#A = C_{14,3} = \binom{14}{3} = 364$. Invece, il numero di selezioni di 4 cavi che contengono la cavia 1 e che sono scelte fra le prime 8 è chiaramente uguale al numero di selezioni di 3 cavi fra le 7 numerate da 2 a 8; quindi $\#(B \cap A) = C_{7,3} = \binom{7}{3} = 35$. La risposta alla domanda (c) è quindi $p(B|A) = 35/364 \simeq 9.6\%$.

ESEMPIO 2.53 Siamo ora in grado di calcolare le probabilità usate nell'Osservazione 2.10. Vogliamo per esempio calcolare la probabilità che tirando 1000 volte un dado a 6 facce non truccato si ottenga 1 per $a_1 = 120$ volte, 2 per $a_2 = 412$ volte, 3 per $a_3 = 66$ volte, 4 per $a_4 = 222$ volte, 5 per $a_5 = 127$ volte, e 6 per $a_6 = 53$ volte. Siccome il dado non è truccato, e ci sono $D_{6,1000} = 6^{1000}$ possibili sequenze di 1000 lanci di dado, la probabilità di una specifica sequenza è $1/6^{1000}$. Dobbiamo contare quante sequenze contengono a_1 volte 1, a_2 volte 2, e così via. Ora, ci sono $\binom{1000}{a_1}$ modi diversi di selezionare gli a_1 lanci che devono fare 1. Una volta fatta questa scelta, ci sono

$$\binom{1000 - a_1}{a_2} = \binom{a_2 + a_3 + a_4 + a_5 + a_6}{a_2}$$

modi diversi di selezionare altri a_2 lanci che devono fare 2. Proseguendo in questo modo (e ricordando il principio base) si vede che il numero di possibili sequenze con a_1 volte 1, a_2 volte 2 e così via è

$$\binom{1000}{a_1} \binom{a_2 + a_3 + a_4 + a_5 + a_6}{a_2} \binom{a_3 + a_4 + a_5 + a_6}{a_3} \binom{a_4 + a_5 + a_6}{a_4} \binom{a_5 + a_6}{a_5},$$

dove ovviamente non abbiamo scritto l'ultimo fattore $\binom{a_6}{a_6} = 1$. Quindi la probabilità che un dado non truccato produca la Tabella 2.3 è

$$\frac{1}{6^{1000}} \binom{1000}{120} \binom{880}{412} \binom{468}{66} \binom{402}{222} \binom{180}{127} \simeq 4.148 \cdot 10^{-112}.$$

Invece, la probabilità che 1000 lanci dello stesso dado si ottenga $a_1 = 167$ volte 1, $a_2 = 166$ volte 2, $a_3 = 167$ volte 3, $a_4 = 166$ volte 4, $a_5 = 167$ volte 5, e $a_6 = 167$ volte 6 è

$$\frac{1}{6^{1000}} \binom{1000}{167} \binom{833}{166} \binom{667}{167} \binom{500}{166} \binom{334}{167} \simeq 6.855 \cdot 10^{-8}.$$

ESEMPIO 2.54 Possiamo fare un conto analogo anche con un dado truccato. Supponiamo che un dado sia truccato in modo da dare risultato j con probabilità p_j , per $j = 1, \dots, 6$. Siccome i lanci sono tutti indipendenti, la probabilità di ottenere una specifica sequenza di risultati che contenga a_j risultati uguali a j , per $j = 1, \dots, 6$, è uguale (perché?) al prodotto $p_1^{a_1} \cdots p_6^{a_6}$. Contando le sequenze di lanci come nell'esempio precedente vediamo che la probabilità che lanciando $a_1 + \cdots + a_6$ volte questo dado si ottenga a_j volte j per $j = 1, \dots, 6$ è

$$p_1^{a_1} \cdots p_6^{a_6} \binom{a_1 + \cdots + a_6}{a_1} \times \binom{a_2 + a_3 + a_4 + a_5 + a_6}{a_2} \binom{a_3 + a_4 + a_5 + a_6}{a_3} \binom{a_4 + a_5 + a_6}{a_4} \binom{a_5 + a_6}{a_5}.$$

Per esempio, supponiamo che il nostro dado sia truccato con le probabilità date dalle frequenze relative della Tabella 2.3. Allora la probabilità che lanciando 1000 volte questo dado si ottengano i risultati della Tabella 2.3 è

$$0.12^{120} 0.412^{412} 0.066^{66} 0.222^{222} 0.127^{127} 0.053^{53} \binom{1000}{120} \times \binom{880}{412} \binom{468}{66} \binom{402}{222} \binom{180}{127} \simeq 1.440 \cdot 10^{-7},$$

mentre la probabilità che si ottenga $a_1 = 167$ volte 1, $a_2 = 166$ volte 2, $a_3 = 167$ volte 3, $a_4 = 166$ volte 4, $a_5 = 167$ volte 5, e $a_6 = 167$ volte 6 è

$$0.12^{167} 0.412^{166} 0.066^{167} 0.222^{166} 0.127^{167} 0.053^{167} \binom{1000}{167} \times \binom{833}{166} \binom{667}{167} \binom{500}{166} \binom{334}{167} \simeq 8.535 \cdot 10^{-116}.$$

Ci rimane da introdurre un ultimo concetto. Una *combinazione con ripetizioni di n oggetti k alla volta* consiste nella scelta di k oggetti fra gli n disponibili, ma ogni oggetto può venire scelto più di una volta (come nelle disposizioni con ripetizioni)

e non è importante l'ordine delle scelte (come nelle combinazioni senza ripetizioni). Anche il numero $\tilde{C}_{n,k}$ di combinazioni con ripetizioni di n oggetti k alla volta può essere espresso con un coefficiente binomiale; per l'esattezza si ha

$$\tilde{C}_{n,k} = \binom{n+k-1}{k} = \frac{(n+k-1)!}{(n-1)!k!}. \quad (2.26)$$

Se sei interessato a come si ottiene questa formula, leggi la Curiosità 2.4.

ESEMPIO 2.55 *Quanti possibili risultati diversi possiamo ottenere tirando 2 dadi a 6 facce indistinguibili? E tirando 3 dadi indistinguibili a 8 facce?* Il fatto che i due (o tre) dadi siano indistinguibili significa che consideriamo il risultato a meno dell'ordine. Quindi nel primo caso ci stiamo chiedendo in quanti modi diversi possiamo ottenere 2 risultati fra i 6 possibili, con ripetizioni ammesse e senza considerare l'ordine; la risposta è quindi $\tilde{C}_{6,2} = \binom{6+2-1}{2} = 7!/(5!2!) = 21$, in accordo con l'Esempio 2.13. Analogamente, la risposta alla seconda domanda è $\tilde{C}_{8,3} = \binom{10}{3} = 120$.

ESEMPIO 2.56 *Quanti genotipi può avere un gene con 7 alleli?* Chiaramente la risposta è $\tilde{C}_{7,2} = \binom{8}{2} = 28$, in accordo con l'Esempio 2.51.

CURIOSITÀ 2.4 Vediamo come si può ottenere la formula (2.26). Il numero $\tilde{C}_{n,k}$ conta in quanti modi possiamo scegliere k numeri fra 1 ed n , con ripetizioni e a meno dell'ordine. Siccome l'ordine non è rilevante, possiamo metterli in ordine crescente, e chiederci in quanti modi possiamo scegliere i numeri a_1, \dots, a_k in modo che si abbia

$$1 \leq a_1 \leq a_2 \leq \dots \leq a_k \leq n.$$

Ora, sommiamo 1 ad a_2 , 2 ad a_3 , e così via; otteniamo

$$1 \leq a_1 < a_2 + 1 < a_3 + 2 < \dots < a_k + (k-1) \leq n + k - 1,$$

cioè una lista ordinata di k numeri *distinti* scelti fra 1 ed $n + k - 1$. In altre parole, una combinazione con ripetizioni di n oggetti k alla volta ci ha fornito una combinazione senza ripetizioni di $n + k - 1$ oggetti k alla volta. Viceversa, supponiamo di avere una combinazione senza ripetizioni di $n + k - 1$ oggetti k alla volta, cioè di aver scelto k numeri distinti b_1, \dots, b_k fra 1 e $n + k - 1$. Siccome l'ordine non conta, possiamo scrivere

$$1 \leq b_1 < b_2 < \dots < b_k \leq n + k - 1.$$

Sottraiamo ora 1 a b_2 , 2 a b_3 e così via; otteniamo

$$1 \leq b_1 \leq b_2 - 1 \leq b_3 - 2 \leq \dots \leq b_k - (k-1) \leq n,$$

cioè una lista ordinata di k numeri, eventualmente ripetuti, scelti fra 1 ed n . In altre parole, invertendo il processo precedente abbiamo trovato che una combinazione senza ripetizioni di $n + k - 1$ oggetti k alla volta fornisce una combinazione con ripetizioni di n oggetti k alla volta. Di conseguenza possiamo concludere che il numero delle combinazioni con ripetizioni di n oggetti k alla volta è uguale al numero delle combinazioni senza ripetizioni di $n + k - 1$ oggetti k alla volta, cioè che

$$\tilde{C}_{n,k} = C_{n+k-1,k} = \binom{n+k-1}{k}.$$

2.10 Distribuzione binomiale

Nell'ultima sezione di questo capitolo vediamo un'applicazione dei coefficienti binomiali alla probabilità. Cominciamo con un paio di esempi.

ESEMPIO 2.57 *Leonardo e Martina ora hanno tre figli. Qual è la probabilità che siano due femmine e un maschio?* Indichiamo con p la probabilità che un figlio sia maschio, e con $q = 1 - p$ la probabilità che un figlio sia femmina; nell'Esempio 2.25 abbiamo visto che in Italia $p \simeq 0.513$ e $q \simeq 0.487$. Supponiamo inoltre che le nascite siano indipendenti, nel senso che il sesso dei figli già nati non influisce su quello dei nascituri. L'ipotesi di indipendenza ci dice allora che la probabilità che il primogenito sia un maschio e gli altri femmine è $pq^2 \simeq 0.122$. Ma questo non è l'unico modo per ottenere un maschio e due femmine; il maschio potrebbe essere il secondogenito, o il terzogenito. La probabilità che si verifichi ciascuno di questi casi è sempre pq^2 , per cui la risposta finale è $3pq^2 \simeq 0.365$.

ESEMPIO 2.58 *Più in generale, se Leonardo e Martina hanno n figli, qual è la probabilità che siano k maschi ed $n - k$ femmine?* Riesaminiamo l'esempio precedente (con le stesse notazioni). La probabilità di una specifica sequenza di nascite con $n - k$ femmine e k maschi è, per l'indipendenza, uguale a $p^k q^{n-k}$. Ma a noi non interessa l'ordine in cui sono avvenute le nascite, ma solo quanti figli sono maschi e quante sono femmine. Quante sono le sequenze di nascite con esattamente k maschi? Sono tante quante il numero di modi con cui possiamo scegliere k nascite fra le n totali, indipendentemente dall'ordine, per cui è il coefficiente binomiale $C_{n,k}$. Ne segue che la risposta alla nostra domanda è

$$\binom{n}{k} p^k q^{n-k}.$$

Vediamo di generalizzare la situazione. Supponiamo di stare studiando un fenomeno in cui un certo evento E può verificarsi con probabilità $p = p(E)$, e quindi non verificarsi con probabilità $q = p(\Omega \setminus E) = 1 - p$. Ripetiamo l'esperimento n volte, con ogni ripetizione indipendente da tutte le precedenti, e ci chiediamo qual è la probabilità che l'evento E sia avvenuto esattamente k volte. Ragionando esattamente come nell'esempio precedente, otteniamo

$$p(k \text{ volte}) = \binom{n}{k} p^k q^{n-k}. \quad (2.27)$$

La distribuzione di probabilità data da questa formula sullo spazio degli eventi $\{1, \dots, n\}$ è detta *distribuzione di Bernoulli*, e si applica in tutti i casi in cui si studiano esperimenti con risposte dicotomiche del tipo “sì/no”, “successo/fallimento”, “positivo/negativo”, eccetera; esperimenti di questo tipo vengono detti *dicotomici*, o *di Bernoulli*.

Osservazione 2.27 Se $p = 1/2$ allora $q = 1/2$, e la (2.27) diventa

$$p(k \text{ volte}) = \binom{n}{k} \frac{1}{2^n}.$$

ESEMPIO 2.59 *Supponiamo che la probabilità che un paziente sottoposto a una data operazione muoia entro un mese sia del 12%. Qual è la probabilità che su 4 operazioni almeno 3 pazienti sopravvivano più di un mese?* Se indichiamo con E l'evento "paziente muore entro un mese", l'evento complementare $\Omega \setminus E$ è "paziente sopravvive più di un mese". Sappiamo che $p = p(E) = 0.12$, per cui $q = p(\Omega \setminus E) = 0.88$. Supponiamo che ciascuna operazione sia indipendente dall'altra, per cui siamo in presenza di un fenomeno di Bernoulli. Dire che almeno 3 pazienti sopravvivono più di un mese equivale a dire che al massimo 1 paziente muore entro un mese. La (2.27) ci dice che la probabilità che esattamente un paziente su 4 muoia entro un mese è

$$\binom{4}{1} p q^3 = 4 \cdot 0.12 \cdot (0.88)^3 \simeq 0.327,$$

mentre la probabilità che nessun paziente muoia entro un mese è

$$\binom{4}{0} p^0 q^4 = 1 \cdot 1 \cdot (0.88)^4 \simeq 0.6.$$

Quindi la probabilità che su 4 operazioni almeno 3 pazienti sopravvivano più di un mese è la somma di questi due valori, cioè $0.927 = 92.7\%$.

ESEMPIO 2.60 *Supponiamo che la probabilità che un gene di ratto subisca una mutazione se sottoposto a una radiazione di 1 Röntgen sia $2.5 \cdot 10^{-7}$. Qual è la probabilità di avere almeno 2 mutazioni in un campione di 10 000 geni sottoposti a una radiazione di 1 Röntgen?* Ovviamente, supponiamo che il mutare o meno di un gene sia indipendente da cosa succede agli altri. Allora siamo in presenza di un fenomeno di Bernoulli, con $p = 2.5 \cdot 10^{-7}$ e $q = 1 - 2.5 \cdot 10^{-7}$. Per calcolare la probabilità che ci siano almeno 2 mutazioni potremmo calcolare la probabilità che ce ne siano esattamente 2, poi che ce ne siano esattamente 3, e così via fino a 10 000, e sommare; compito improbo. È molto più semplice calcolare la probabilità dell'evento complementare, cioè la probabilità che ci sia al massimo 1 mutazione. La probabilità che non ce ne sia nessuna è

$$\binom{10\,000}{0} p^0 q^{10\,000} = (1 - 2.5 \cdot 10^{-7})^{10\,000} \simeq 0.99750312,$$

mentre la probabilità che ce ne sia esattamente 1 è

$$\binom{10\,000}{1} p q^{9\,999} = 10\,000 \cdot (2.5 \cdot 10^{-7}) \cdot (1 - 2.5 \cdot 10^{-7})^{9\,999} \simeq 0.00249376.$$

Quindi la probabilità che ci sia al più una mutazione è circa

$$0.99750312 + 0.00249376 = 0.99999688 ,$$

e la probabilità che su 10 000 geni ci siano almeno due mutazioni è circa

$$1 - 0.99999688 = 31.2 \cdot 10^{-7} .$$

ESEMPIO 2.61 In Sassonia nel 1889 è stato calcolato che la probabilità di avere un figlio maschio era $p \simeq 0.519215$, mentre la probabilità di avere una figlia femmina era $q \simeq 0.480785$. Supponendo, come abbiamo fatto finora, che il sesso del nascituro sia indipendente da quello degli altri fratelli, possiamo calcolare la probabilità delle varie combinazioni di maschi e femmine in famiglie con 12 figli. Un'indagine effettuata su quella popolazione ha trovato 6115 famiglie con 12 figli⁸, e ha misurato le frequenze relative delle varie combinazioni di maschi e femmine fra i figli. I risultati sono contenuti nella Tabella 2.10.

Dalla tabella risulta evidente che le frequenze osservate sono sistematicamente maggiori delle probabilità teoriche per le famiglie con almeno 9 figli dello stesso sesso, e sistematicamente minori per le famiglie con al più 7 figli dello stesso sesso. Questo risultato fa sospettare che in realtà il sesso del nascituro *non* sia indipendente dal sesso dei fratelli già nati, ma che ci sia in atto un qualche meccanismo che favorisce il formarsi di famiglie con molti figli dello stesso sesso.

COMPLEMENTI

2C.1 Sommatorie

In queste dispense capiterà spesso di dover usare somme del tipo $a_1 + \cdots + a_n$, dove a_j è un numero (o un vettore, o una matrice, o altro ancora) dipendente da un indice j che, di solito, è un numero intero variabile in un certo insieme, per esempio da 1 a n . Per trattare questo tipo di somme è stata introdotta una notazione abbreviata molto comoda: la *sommatoria* \sum .

Nella sua versione più semplice, \sum si utilizza come segue:

$$\sum_{j=1}^n a_j = \sum_{1 \leq j \leq n} a_j = a_1 + \cdots + a_n .$$

⁸ In Italia nel 2001 c'erano solo 10 001 famiglie con più di 5 figli, contro 4 216 946 famiglie con un solo figlio...

<i>Numero maschi</i>	<i>Numero femmine</i>	<i>Probabilità teorica</i>	<i>Frequenza relativa osservata</i>
12	0	$\binom{12}{0} p^{12} q^0 \simeq 0.000\,384$	$\frac{7}{6115} \simeq 0.001\,145$
11	1	$\binom{12}{1} p^{11} q^1 \simeq 0.004\,263$	$\frac{45}{6115} \simeq 0.007\,359$
10	2	$\binom{12}{2} p^{10} q^2 \simeq 0.021\,725$	$\frac{181}{6115} \simeq 0.029\,599$
9	3	$\binom{12}{3} p^9 q^3 \simeq 0.067\,041$	$\frac{478}{6115} \simeq 0.078\,168$
8	4	$\binom{12}{4} p^8 q^4 \simeq 0.139\,703$	$\frac{829}{6115} \simeq 0.135\,568$
7	5	$\binom{12}{5} p^7 q^5 \simeq 0.206\,977$	$\frac{1112}{6115} \simeq 0.181\,848$
6	6	$\binom{12}{6} p^6 q^6 \simeq 0.223\,590$	$\frac{1343}{6115} \simeq 0.219\,624$
5	7	$\binom{12}{7} p^5 q^7 \simeq 0.177\,459$	$\frac{1033}{6115} \simeq 0.168\,929$
4	8	$\binom{12}{8} p^4 q^8 \simeq 0.102\,707$	$\frac{670}{6115} \simeq 0.109\,567$
3	9	$\binom{12}{9} p^3 q^9 \simeq 0.042\,280$	$\frac{286}{6115} \simeq 0.046\,770$
2	10	$\binom{12}{10} p^2 q^{10} \simeq 0.011\,743$	$\frac{104}{6115} \simeq 0.017\,007$
1	11	$\binom{12}{11} p^1 q^{11} \simeq 0.001\,975$	$\frac{24}{6115} \simeq 0.003\,925$
0	12	$\binom{12}{12} p^0 q^{12} \simeq 0.000\,153$	$\frac{3}{6115} \simeq 0.000\,491$

TABELLA 2.10.

Ma sono possibili anche costruzioni più complesse, come

$$\sum_{\substack{0 \leq j \leq 2n \\ j \text{ pari}}} a_j = a_0 + a_2 + a_4 + \cdots + a_{2n} .$$

Questa stessa somma poteva anche essere scritta in modo più compatto come

$$\sum_{h=0}^n a_{2h} = \sum_{0 \leq h \leq n} a_{2h} .$$

In questa sezione cercheremo di descrivere brevemente come utilizzare al meglio il simbolo di sommatoria, indicandone le principali proprietà e studiando qualche esempio. Cominciamo con la definizione generale del simbolo di sommatoria.

Se $R(j)$ è una condizione che coinvolge il numero intero j (qualcosa del tipo “ $-3 \leq j \leq 3$ ”, oppure “ $0 \leq j \leq r$ e j pari”), allora il simbolo

$$\sum_{R(j)} a_j$$

indica la somma di tutti gli a_j per cui l'intero j soddisfa la condizione $R(j)$. Per convenzione, se nessun intero j soddisfa la condizione $R(j)$ si assume che la sommatoria valga zero. A volte scriveremo anche $\sum_{R(j)} a_j$; un'altra possibilità, se $R(j)$ è della forma $\alpha \leq j \leq \beta$, è

$$\sum_{j=\alpha}^{\beta} a_j .$$

L'indice j è chiamato *indice muto* della sommatoria. L'indice muto serve soltanto all'interno del simbolo di sommatoria, e non ha alcun significato intrinseco — vedi la proprietà (B) più oltre. Tipici indici muti sono h, i, j, k, l , talvolta accentati o con indici a loro volta.

Osservazione 2C.1 Per il momento useremo il simbolo di sommatoria solo per indicare somme finite, cioè in situazioni in cui la condizione $R(j)$ è soddisfatta soltanto da un numero finito di interi j . Il simbolo di sommatoria serve anche per indicare somme infinite (ovvero le *serie*), di cui parleremo più oltre.

Le principali proprietà del simbolo di sommatoria sono quattro.

(A) *Proprietà distributiva*, che riguarda il prodotto di sommatorie:

$$\left(\sum_{R(h)} a_h \right) \left(\sum_{S(k)} b_k \right) = \sum_{R(h)} \left(\sum_{S(k)} a_h b_k \right) . \quad (2C.1)$$

Per esempio,

$$\begin{aligned} \left(\sum_{h=1}^2 a_h \right) \left(\sum_{k=1}^3 b_k \right) &= (a_1 + a_2)(b_1 + b_2 + b_3) \\ &= (a_1 b_1 + a_1 b_2 + a_1 b_3) + (a_2 b_1 + a_2 b_2 + a_2 b_3) \\ &= \sum_{h=1}^2 \left(\sum_{k=1}^3 a_h b_k \right) . \end{aligned}$$

Di solito, le parentesi nelle “somme multiple” come quella sul lato destro di (2C.1) vengono tralasciate: invece di $\sum_{R(h)} \left(\sum_{S(k)} a_{hk} \right)$ si scrive $\sum_{R(h)} \sum_{S(k)} a_{hk}$.

(B) *Cambiamento di variabile*. L'indice muto di una sommatoria è, appunto, muto, per cui non si lamenta se gli viene cambiato il nome. In altri termini, si ha

$$\sum_{R(h)} a_h = \sum_{R(k)} a_k = \sum_{R(s)} a_s , \quad (2C.2)$$

e così via, dove k , s e simili sono tutti indici non altrimenti presenti nella condizione R o nei termini a_k (per intenderci: se $a_k = n^k$, oppure se $R(k)$ è la relazione “ $1 \leq k \leq n$ ”, non possiamo usare n come indice muto). Ma si può fare di più: data una condizione $R(h)$, indichiamo con A l'insieme degli $h \in \mathbb{Z}$ che soddisfano questa condizione. Una *permutazione degli indici* è un modo di associare a ciascun elemento k di un insieme $B \subset \mathbb{Z}$ un elemento $p(k) \in A$ in modo che per ogni $h \in A$ esista uno e un solo $k \in \mathbb{Z}$ tale che $p(k) = h$. Allora se p è una permutazione degli indici per la condizione $R(h)$ si ha

$$\sum_{R(h)} a_h = \sum_{R(p(k))} a_{p(k)} . \quad (2C.3)$$

Infatti, su entrambi i lati di (2C.3) sono presenti gli stessi addendi. I casi più comuni di permutazione degli indici sono del tipo $p(k) = \pm k \pm c$, dove c è un intero fissato che non dipende da k . Per esempio, se $p(k) = k - 1$ abbiamo

$$\sum_{1 \leq h \leq n} a_h = \sum_{1 \leq k-1 \leq n} a_{k-1} = \sum_{2 \leq k \leq n+1} a_{k-1} = \sum_{2 \leq h \leq n+1} a_{h-1} ,$$

dove nell'ultimo passaggio abbiamo utilizzato (2C.2).

(C) *Proprietà commutativa.* Quando effettuiamo una somma, l'ordine in cui consideriamo gli addendi è irrilevante. Quindi

$$\sum_{R(h)} \sum_{S(k)} a_{hk} = \sum_{S(k)} \sum_{R(h)} a_{hk} . \quad (2C.4)$$

Per esempio,

$$\sum_{R(h)} \sum_{1 \leq k \leq 2} a_{hk} = \sum_{R(h)} (a_{h1} + a_{h2}) = \sum_{R(h)} a_{h1} + \sum_{R(h)} a_{h2} = \sum_{1 \leq k \leq 2} \sum_{R(h)} a_{hk} .$$

Potrebbe capitare anche un caso più generale, in cui la condizione $S(k)$ dipenda anche da h — e quindi sarà scritta $S(h, k)$. In tal caso la (2C.4) diventa qualcosa del tipo

$$\sum_{R(h)} \sum_{S(h, k)} a_{hk} = \sum_{S'(k)} \sum_{R'(h, k)} a_{hk} , \quad (2C.5)$$

dove $S'(k)$ è la condizione “esiste un intero h tale che sia $R(h)$ sia $S(h, k)$ sono verificate”, e $R'(h, k)$ è la condizione “sia $R(h)$ sia $S(h, k)$ sono verificate”.

Un esempio per chiarire. Se $R(h)$ è la solita condizione $1 \leq h \leq n$, ed $S(h, k)$ è la condizione $1 \leq k \leq h$, allora la condizione $S'(k)$ è “esiste un intero h tale che $1 \leq h \leq n$ e $1 \leq k \leq h$ ”, cioè semplicemente “ $1 \leq k \leq n$ ”, mentre la condizione $R'(h, k)$ è “devono valere sia $1 \leq h \leq n$ che $1 \leq k \leq h$ ”, cioè “ $k \leq h \leq n$ ”. Quindi

$$\sum_{1 \leq h \leq n} \sum_{1 \leq k \leq h} a_{hk} = \sum_{1 \leq k \leq n} \sum_{k \leq h \leq n} a_{hk} . \quad (2C.6)$$

(D) *Manipolazione del dominio.* Se $R(h)$ e $S(h)$ sono due condizioni, abbiamo

$$\sum_{R(h)} a_h + \sum_{S(h)} a_h = \sum_{R(h) \text{ o } S(h)} a_h + \sum_{R(h) \text{ e } S(h)} a_h, \quad (2C.7)$$

dove “ $R(h)$ o $S(h)$ ” vuol dire che h soddisfa $R(h)$ oppure soddisfa $S(h)$, mentre “ $R(h)$ e $S(h)$ ” vuol dire che h soddisfa sia $R(h)$ che $S(h)$. Per esempio, se $m \leq n$ si ha

$$\sum_{1 \leq k \leq m} a_k + \sum_{m \leq k \leq n} a_k = \left(\sum_{1 \leq k \leq n} a_k \right) + a_m.$$

Se poi non capita mai che $R(h)$ ed $S(h)$ siano simultaneamente verificate, allora la seconda sommatoria sul lato destro di (2C.7) semplicemente scompare.

Vediamo ora qualche esempio di applicazione di queste regole.

ESEMPIO 2C.1 *La somma di una progressione aritmetica.* Prendiamo $a, b \in \mathbb{R}$ ed $n \in \mathbb{N}$. Allora:

$$\begin{aligned} a + (a+b) + \cdots + (a+nb) &= \sum_{0 \leq h \leq n} (a+hb) \\ &= \sum_{0 \leq n-h \leq n} (a+(n-h)b) = \sum_{0 \leq n-h \leq n} ((2a+nb) - (a+hb)) \\ &= \sum_{0 \leq h \leq n} (2a+nb) - \sum_{0 \leq h \leq n} (a+hb) \\ &= (n+1)(2a+nb) - \sum_{0 \leq h \leq n} (a+hb), \end{aligned}$$

grazie a: (2C.3) assieme a (2C.2); il fatto che le condizioni “ $0 \leq n-h \leq n$ ” e “ $0 \leq h \leq n$ ” sono verificate dagli stessi interi h ; e il fatto che la prima sommatoria nella penultima riga è semplicemente la somma di $n+1$ addendi che non dipendono da h . Ma ora se confrontiamo la prima e l’ultima riga e dividiamo per 2 otteniamo la formula cercata:

$$\sum_{0 \leq h \leq n} (a+hb) = a(n+1) + \frac{bn(n+1)}{2}.$$

ESEMPIO 2C.2 *La somma di una progressione geometrica.* Prendiamo $a, x \in \mathbb{R}$

con $x \neq 1$, ed $n \in \mathbb{N}$. Allora

$$\begin{aligned}
 a + ax + \cdots + ax^n &= \sum_{0 \leq k \leq n} ax^k \\
 &= a + \sum_{1 \leq k \leq n} ax^k \\
 &= a + x \sum_{1 \leq k \leq n} ax^{k-1} \\
 &= a + x \sum_{0 \leq k \leq n-1} ax^k \\
 &= a + x \sum_{0 \leq k \leq n} ax^k - ax^{n+1},
 \end{aligned}$$

grazie a, nell'ordine: (2C.7); un caso molto particolare di (2C.1); (2C.3); (2C.7). Confrontando la prima riga con l'ultima otteniamo

$$(1-x) \sum_{0 \leq k \leq n} ax^k = a(1-x^{n+1}),$$

e quindi la formula cercata:

$$\sum_{0 \leq k \leq n} ax^k = a \frac{1-x^{n+1}}{1-x}.$$

ESEMPIO 2C.3 Poniamo

$$S_1 = \sum_{0 \leq h \leq n} \sum_{0 \leq k \leq h} a_h a_k.$$

Allora si ha

$$S_1 = \sum_{0 \leq h \leq n} \sum_{0 \leq k \leq h} a_h a_k = \sum_{0 \leq k \leq n} \sum_{k \leq h \leq n} a_h a_k = \sum_{0 \leq h \leq n} \sum_{h \leq k \leq n} a_h a_k,$$

dove abbiamo usato (2C.6) e (2C.2). Se chiamiamo S_2 quest'ultima sommatoria,

abbiamo $S_1 = S_2$ e quindi

$$\begin{aligned}
 2S_1 = S_1 + S_2 &= \sum_{0 \leq h \leq n} \left(\sum_{0 \leq k \leq h} a_h a_k + \sum_{h \leq k \leq n} a_h a_k \right) \\
 &= \sum_{0 \leq h \leq n} \left(\sum_{0 \leq k \leq n} a_h a_k + a_h a_h \right) = \sum_{0 \leq h \leq n} \sum_{0 \leq k \leq n} a_h a_k + \sum_{0 \leq h \leq n} a_h a_h \\
 &= \left(\sum_{0 \leq h \leq n} a_h \right) \left(\sum_{0 \leq k \leq n} a_k \right) + \left(\sum_{0 \leq h \leq n} a_h^2 \right) \\
 &= \left(\sum_{0 \leq h \leq n} a_h \right)^2 + \left(\sum_{0 \leq h \leq n} a_h^2 \right),
 \end{aligned}$$

grazie a, nell'ordine, (2C.7), (2C.1), e (2C.2). Abbiamo dunque ricavato la seguente identità:

$$\sum_{0 \leq h \leq n} \sum_{0 \leq k \leq h} a_h a_k = \frac{1}{2} \left\{ \left(\sum_{0 \leq h \leq n} a_h \right)^2 + \left(\sum_{0 \leq h \leq n} a_h^2 \right) \right\}.$$

ESEMPIO 2C.4 Un altro simbolo che si utilizza spesso assieme alle sommatorie è il *delta di Kronecker* δ_{ij} , definito da

$$\delta_{ij} = \begin{cases} 0 & \text{se } i \neq j, \\ 1 & \text{se } i = j. \end{cases}$$

In particolare si ha

$$\sum_{j=1}^n b_j \delta_{ij} = b_i,$$

e quindi otteniamo formule del tipo

$$\sum_{i=1}^n \sum_{j=1}^n a_i b_j \delta_{ij} = \sum_{i=1}^n a_i \sum_{j=1}^n b_j \delta_{ij} = \sum_{i=1}^n a_i b_i.$$