Capitolo 5

La probabilità

Il concetto di probabilità è importante in quasi tutti i campi della scienza, compresa la biologia. Sulla probabilità si basa anche l'analisi dei dati. Nel capitolo precedente abbiamo visto che si deve ricorrere al concetto di probabilità per quantificare l'incertezza dei risultati; vedremo in tutto il libro che ciò è vero anche in altri contesti.
La probabilità è essenziale per la biologia perché quasi sempre studiamo il mondo naturale mediante un campione e, come abbiamo visto, la probabilità può svolgere un ruolo importante nell'analisi delle proprietà dei campioni. In questo capitolo presenteremo i principi e i calcoli fondamentali della probabilità e nel capitolo seguente cominceremo ad applicare questi concetti all'analisi dei dati.

■ 5.1 La probabilità di un evento

Immaginate di avere sul vostro lettore di musica digitale 1000 canzoni, ciascuna registrata esattamente una volta. Quando premete il pulsante «shuffle», il dispositivo riproduce una canzone in modo casuale prelevandola dalla lista dei 1000 brani. La probabilità che la canzone riprodotta sia proprio quella che preferite è 1/1000, ossia 0,001. La probabilità che la canzone riprodotta non sia quella che preferite è 999/1000, ossia 0,999. Che cosa significano esattamente questi numeri?

Il concetto di probabilità si basa su quello di **prova casuale** ripetuta molte volte. Una prova casuale è un processo o un esperimento che ha due o più risultati possibili che non possono essere predetti. Nell'esempio del lettore musicale, una prova casuale consiste nel premere una volta il pulsante shuffle. Un risultato specifico è «*la riproduzione della vostra canzone preferita*», che è uno dei 1000 risultati possibili. Altri esempi di prova casuale sono i seguenti:

- Lanciare una moneta e verificare se il risultato è testa oppure croce
- Lanciare una coppia di dadi e guardare quale sia la somma dei numeri che si presentano
- Campionare casualmente un individuo da una popolazione di salmone rosso (*Oncorhynchus nerka*) e determinare quale sia la sua massa corporea
- Campionare casualmente 10 individui da una popolazione di granciporro del Pacifico (*Cancer magister*), un crostaceo decapode, e determinare la proporzione di femmine

Per descrivere una probabilità, si devono definire l'**evento** e lo **spazio campionario**. Lo spazio campionario è la lista di tutti i risultati possibili di una prova ca-

 © 978-88-08-06297-0

suale. Un evento è qualsiasi sottoinsieme potenziale dello spazio campionario. Per esempio, se lanciamo un dado a 6 facce, sono possibili 6 risultati: i numeri da 1 a 6. Questi 6 differenti numeri rappresentano nel loro insieme lo spazio campionario. Possiamo definire molti eventi che potrebbero interessare, quali «*il risultato è un numero pari*», «*il risultato è un numero maggiore di 3*» o anche il semplice evento «*il risultato è 4*».

La **probabilità** di un evento è la *proporzione* di tutte le prove casuali con cui si verifica l'evento specificato quando la stessa prova casuale viene ripetuta moltissime volte, in modo indipendente e nelle stesse condizioni.[1] Se campioniamo casualmente un individuo da una popolazione che è composta per 2/3 da maschi, la probabilità di campionare un maschio sarà pari a 2/3.

- La **probabilità** di un evento è il numero di volte che l'evento si verificherebbe se si ripetesse una prova casuale moltissime volte nelle stesse condizioni.

Un'utile e rapida definizione è la seguente:

Pr[A] significa «la probabilità dell'evento A»

Perciò, volendo enunciare la probabilità di «*ottenere 4*» lanciando un dado a 6 facce, possiamo scrivere

$$\Pr[\textit{ottenere } 4] = 1/6$$

Le probabilità, essendo proporzioni, devono essere sempre comprese tra 0 e 1, estremi inclusi: un evento ha probabilità 0 se *non si verifica mai* e ha probabilità 1 se *si verifica sempre*.

5.2 I diagrammi di Venn

Un modo utile per ragionare sulle probabilità degli eventi è usare uno strumento grafico, detto **diagramma di Venn**. L'area del diagramma di Venn rappresenta tutti i risultati possibili di una prova casuale, e si possono rappresentare vari eventi come aree del diagramma.

La Figura 5.2-1 mostra un diagramma di Venn per il lancio di un dado a 6 facce non truccato. I 6 possibili risultati riempiono il diagramma, indicando che questi rappresentano tutte le possibilità. In questo caso, il box per ciascun risultato ha la stessa area e ciò significa che questi risultati sono equiprobabili: ciascuno di essi contiene 1/6 dell'area del diagramma di Venn.

Possiamo usare i diagrammi di Venn per visualizzare anche eventi più complessi. Nella Figura 5.2-2, per esempio, è visualizzato l'evento «*il risultato è maggiore di 2*».

5.3 Eventi incompatibili

Quando due eventi sono mutuamente esclusivi, ovvero non possono verificarsi simultaneamente, si parla di **eventi incompatibili**. Per esempio, un singolo dado lanciato una sola volta non può presentare simultaneamente sia 1 sia 6: gli eventi «*1*» e «*6*» sono incompatibili.

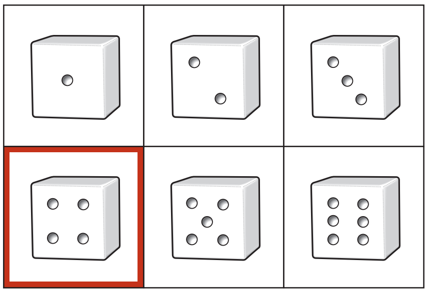


Figura 5.2-1
Diagramma di Venn per i risultati del lancio di un dado a 6 facce. L'area corrispondente all'evento «*il risultato è 4*» è messa in evidenza in rosso.

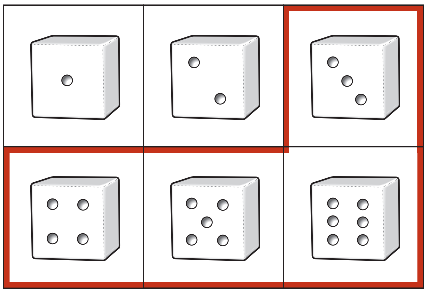


Figura 5.2-2
Diagramma di Venn che rappresenta, evidenziato in rosso, l'evento «*il risultato è maggiore di 2*». La probabilità di questo evento è 4/6 = 2/3, uguale all'area evidenziata dalla linea rossa.

- Due eventi sono **incompatibili** se non possono verificarsi simultaneamente.

I vincoli fisici spiegano talvolta perché certi eventi siano incompatibili, come nel caso dei dadi: è impossibile che da un singolo lancio si ottenga più di un numero. Certi eventi sono talvolta incompatibili perché non possono verificarsi mai in natura. Per esempio, «ha i denti» e «ha le penne» sono eventi incompatibili quando si sceglie casualmente una singola specie animale vivente, perché nessun animale ha sia i denti sia le penne. Se si sceglie a caso una specie animale vivente, la probabilità che abbia sia i denti sia le penne è zero, anche se molte specie animali hanno i denti *o* le penne.

In termini matematici, due eventi A e B sono incompatibili se

$$\Pr[A \text{ e } B] = 0$$

In questo caso, Pr[A e B] significa la probabilità che si verifichi sia A sia B.

[1] Il termine «probabilità» ha altre definizioni, anche nell'ambito della statistica. In una definizione alternativa, la «probabilità» identifica una sorta di convinzione soggettiva, da parte del ricercatore, rispetto alla verità. Per esempio, «C'è una probabilità del 95% che io abbia chiuso il rubinetto del gas prima di partire per le vacanze.» In questo libro, però, la probabilità è sempre definita come una proporzione.

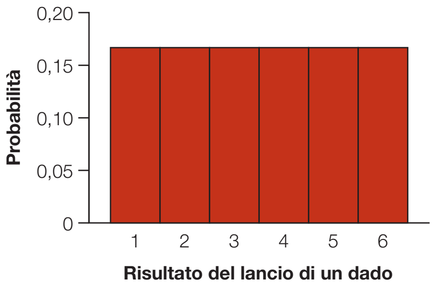


Figura 5.4-1
La distribuzione di probabilità dei risultati del lancio di un singolo dado a 6 facce non truccato. La probabilità di ogni risultato possibile è 1/6 = 0,167.

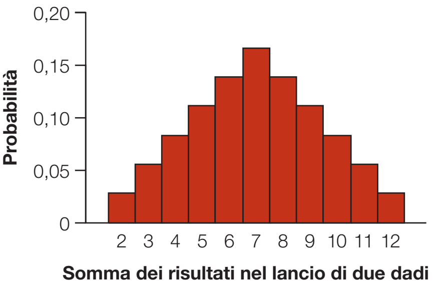


Figura 5.4-2
La distribuzione di probabilità della somma dei punteggi ottenuti lanciando due dadi a 6 facce non truccati.

5.4 Le distribuzioni di probabilità

Nel Paragrafo 1.4 abbiamo definito la *distribuzione di probabilità* come la distribuzione di una variabile nell'intera popolazione. La distribuzione di probabilità di una prova casuale è la probabilità di ciascuno dei risultati incompatibili della prova. Alcune distribuzioni di probabilità possono essere descritte matematicamente con una funzione. Si parla, in questo caso, di distribuzione teoriche di probabilità. In altri casi non esiste una funzione matematica, e la distribuzione di probabilità è semplicemente una lista dei risultati e delle loro probabilità. Il significato preciso di distribuzione di probabilità dipende dal fatto che la variabile sia discreta oppure continua.

◆ Una **distribuzione di probabilità** è una lista delle probabilità di tutti i risultati incompatibili (ovvero mutuamente esclusivi) di una prova casuale.

Distribuzioni di probabilità discrete

Una distribuzione di probabilità discreta fornisce la probabilità di ogni possibile valore o risultato che può assumere una variabile discreta. Per esempio, la distribuzione di probabilità dei risultati del singolo lancio di un dado non truccato è presentata nella Figura 5.4-1. In questo caso, tutti i numeri interi compresi tra 1 e 6 sono risultati equiprobabili (probabilità = 1/6 = 0,167). L'istogramma nella Figura 5.4-2 visualizza invece la distribuzione di probabilità di un'altra variabile, la somma dei due numeri risultanti dal lancio di due dadi. In questo caso i differenti risultati *non* sono equiprobabili.

Per definizione, la somma di tutte le probabilità in una distribuzione di probabilità deve essere uguale a 1.

Distribuzioni di probabilità continue

A differenza delle variabili discrete, le variabili continue possono assumere qualsiasi valore reale entro un certo intervallo. Tra due valori qualsiasi di una variabile continua (denotata con Y) è possibile un numero infinito di altri valori. La distribuzione viene descritta con una curva la cui altezza è detta **densità di probabilità**. La densità di probabilità permette di descrivere la probabilità di qualsiasi intervallo di valori.

La distribuzione normale (o distribuzione di Gauss), introdotta nel Paragrafo 1.4, è una distribuzione teorica di probabilità continua. È la curva a campana mostrata nella Figura 5.4-3. Esamineremo questa distribuzione in modo più particolareggiato nel Capitolo 10.

A differenza delle distribuzioni di probabilità discrete, l'altezza della curva in una distribuzione di probabilità continua in corrispondenza di un certo valore, per esempio $Y = 2{,}4$, non fornisce la probabilità di ottenere $Y = 2{,}4$ quando viene estratta casualmente una singola misura dalla popolazione. Poiché una distribuzione di probabilità continua copre un numero infinito di possibili risultati, la probabilità di ottenere un qualsiasi risultato particolare è infinitesima.

L'altezza della curva, invece, indica la densità di misure in corrispondenza di $Y = 2{,}4$. Nel caso delle distribuzioni di probabilità continue, come la gaussiana, è più ragionevole parlare della probabilità di ottenere un valore di Y entro un certo intervallo. Questa probabilità è data dall'area della regione sottesa dalla curva tra gli estremi dell'intervallo. Per esempio, la probabilità che un singolo individuo scelto casualmente abbia una misura compresa tra due numeri a e b è uguale all'area della regione sottesa dalla curva tra a e b (Figura 5.4-4). L'utilizzo delle aree, e non delle altezze (che qui sono densità), ci permette di definire le probabilità nelle distribuzioni di variabili continue. Le aree sottese da queste distribuzioni sono quindi rilevanti, non le altezze. Questo concetto era già stato introdotto nel Capitolo 2 durante la trattazione degli istogrammi.

L'area della regione sottesa dalla curva tra a e b si cal-

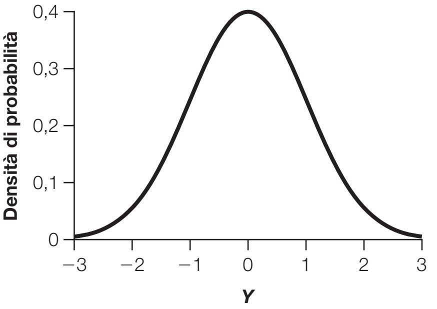


Figura 5.4-3
Una distribuzione normale (o gaussiana).

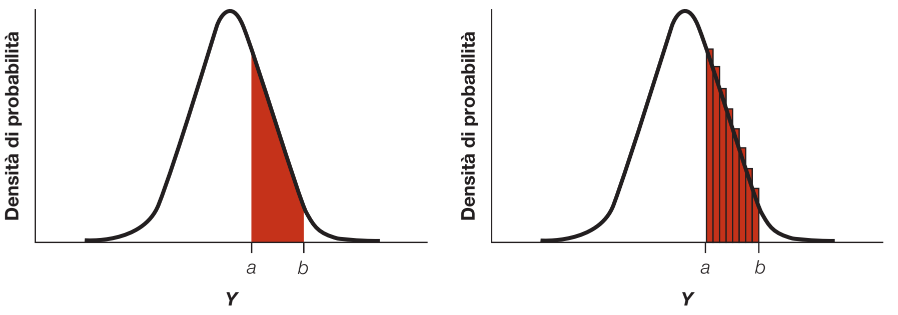


Figura 5.4-4
La probabilità che una misura *Y* scelta casualmente sia compresa tra *a* e *b* è l'area della regione sottesa dalla curva di densità di probabilità tra *a* e *b* (parte sinistra). Nella parte destra la stessa area è stata approssimata con barre discrete.

cola integrando[2] la funzione di densità di probabilità tra i valori *a* e *b*. L'integrazione è l'analogo, per le variabili continue, della somma, quindi integrare la funzione di densità di probabilità tra *a* e *b* è analogo a sommare la probabilità di *Y* per tutti i valori compresi tra *a* e *b* (si veda la parte destra della Figura 5.4-4).

Per ogni distribuzione di probabilità, l'area della regione sottesa dall'intera curva di una funzione di densità di probabilità continua è sempre uguale a 1. Infine, poiché la probabilità di ogni singolo valore *Y* è infinitesima in una distribuzione di densità di probabilità continua, $\Pr[a \leq Y \leq b]$ è uguale a $\Pr[a < Y < b]$.

5.5 Va bene sia questo che quello: sommare le probabilità

Molto spesso, sia nella vita di tutti i giorni che nella statistica, siamo interessati a conoscere la probabilità che si verifichi un evento *oppure* un altro. Nel gioco del «craps»,[3] per esempio, si vince se, nel primo lancio di due dadi, la somma dei numeri che si presentano è 7 o 11. Qual è la probabilità di vincere al primo lancio?

La regola della somma

Se gli eventi che tentiamo di combinare sono incompatibili, allora il calcolo della probabilità che si verifichi l'uno o l'altro evento è facile e piuttosto intuitivo. La probabilità di ottenere l'uno o l'altro è semplicemente la somma delle probabilità per ciascuno di quegli eventi considerati separatamente. Ottenere 7 e ottenere 11 lanciando due dadi sono eventi incompatibili, perciò, la probabilità di ottenere 7 o 11 lanciando due dadi è la probabilità che esca 7 più la probabilità che esca 11:

$$\Pr[\textit{ottenere 7} \text{ o } \textit{ottenere 11}] =$$
$$= \Pr[\textit{ottenere 7}] + \Pr[\textit{ottenere 11}]$$

Il diagramma di Venn per questo evento è presentato nella Figura 5.2-1.Questa proprietà additiva delle probabilità di eventi incompatibili è detta **regola della somma** (un caso particolare del teorema dell'addizione delle probabilità).

◆ **Regola della somma**: Se due eventi *A* e *B* sono incompatibili, allora

$$\Pr[A \text{ o } B] = \Pr[A] + \Pr[B]$$

La regola della somma può essere estesa a più di due eventi, purché essi siano tutti incompatibili. Supponiamo di volere conoscere la probabilità di ottenere 3 o più di 3 con un singolo lancio di un dado. «3 o più di 3» comprende quattro differenti risultati: «*3*», «*4*», «*5*» e «*6*». Queste quattro possibilità sono incompatibili, perché non si può ottenere 4 e 5 (o qualsiasi altra combinazione delle quattro possibilità) simultaneamente con lo stesso dado. Possiamo quindi calcolare la probabilità di ottenere 3 o più di tre usando la regola della somma, come segue:

$$\begin{aligned}\Pr[\textit{ottenere 3 o più di 3}] &= \Pr[3] + \Pr[4] + \Pr[5] + \Pr[6]\\ &= 1/6 + 1/6 + 1/6 + 1/6\\ &= 2/3\end{aligned}$$

(Per mantenere breve l'equazione, abbiamo usato la notazione abbreviata Pr[3] per designare la probabilità di ottenere 3, ecc.). È importante notare che la somma delle aree (delle probabilità) dei quattro risultati incompatibili è uguale all'area (alla probabilità) dell'evento cercata.

La regola della somma riguarda enunciati che implicano «oppure». Se due eventi sono incompatibili e vogliamo conoscere la probabilità che si verifichi l'uno *oppure* l'altro, possiamo usare la regola della somma. Questa proprietà è essenziale per l'analisi dei dati, perché permette di calcolare le probabilità di differenti risultati del campionamento quando essi sono incompatibili.

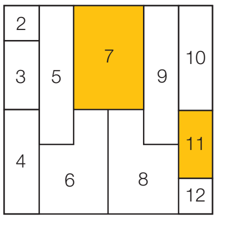


Figura 5.5-1
La probabilità di ottenere 7 o 11 lanciando due dadi è uguale alla somma della probabilità di ottenere 7 e della probabilità di ottenere 11 perché i due eventi sono incompatibili. (Questo diagramma di Venn rappresenta la stessa distribuzione di probabilità che è rappresentata dall'istogramma nella Figura 5.4-2.)

[2] Niente panico! In questo libro non vi chiederemo di svolgere integrali.
[3] Il «craps» è un gioco d'azzardo che implica il lancio di due dadi a 6 facce.

La somma delle probabilità di tutti i possibili eventi incompatibili è pari a 1

La somma delle probabilità di tutti i possibili risultati incompatibili di una prova casuale deve essere uguale a 1. Per esempio, nel caso del singolo lancio di un dado a 6 facce non truccato, sono possibili 6 risultati incompatibili (i numeri da 1 a 6), ciascuno dei quali ha probabilità 1/6. Perciò, la somma delle probabilità di tutti i risultati è

$$\begin{aligned}\Pr[1 \text{ o } 2 \text{ o } 3 \text{ o } 4 \text{ o } 5 \text{ o } 6] &= \Pr[1] + \Pr[2] + \Pr[3] + \Pr[4] + \Pr[5] \\ &\quad + \Pr[6] \\ &= 1/6 + 1/6 + 1/6 + 1/6 + 1/6 + 1/6 \\ &= 1\end{aligned}$$

Ciò significa che la probabilità che un evento *non* si verifichi è semplicemente 1 meno la probabilità che si verifichi. Per esempio, la probabilità di *non* ottenere 2 quando si lancia un dado non truccato è

$$\Pr[\text{non } \textit{ottenere } 2] = 1 - \Pr[\textit{ottenere } 2] = 5/6$$

Questo calcolo è molto più facile che sommare le probabilità di tutti i risultati diversi da 2.

La regola della somma generalizzata (teorema della somma delle probabilità)

Non tutti gli eventi sono incompatibili. Nel lancio di due dadi, per esempio, potremmo essere interessati alla probabilità di ottenere 1 nel primo lancio oppure 6 nel secondo, ma i due risultati non sono incompatibili. Un animale può avere sia le penne sia la capacità di volare, anche se non tutti gli animali che hanno le penne sono capaci di volare e non tutti gli animali che sono capaci di volare hanno le penne. Se due eventi non sono incompatibili, come si calcola la probabilità che l'uno oppure l'altro evento possa verificarsi?

Nella notazione matematica, la regola della somma generalizzata può essere scritta come

$$\Pr[A \text{ o } B] = \Pr[A] + \Pr[B] - \Pr[A \text{ e } B]$$

Quando gli eventi A e B sono incompatibili $\Pr[A \text{ e } B] = 0$, quindi la regola della somma generalizzata si riduce alla regola della somma per eventi incompatibili introdotta precedentemente. Il motivo per cui si deve sottrarre la probabilità di sovrapposizione parziale è illustrato nella figura 5.5-2: se *non* si sottrae la probabilità che si verifichino sia A che B, allora si conteranno due volte i risultati in cui gli eventi si verificano contemporaneamente.

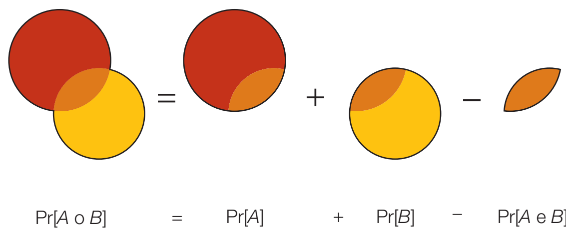


Figura 5.5-2
La regola generale della somma. La Pr[*A* e *B*] viene sottratta da Pr[*a*] + Pr[*B*] in modo tale che i risultati in cui si verificano sia *A* che *B* (le aree ombreggiate in colore arancione) non siano contate due volte.

5.6 L'indipendenza e la regola del prodotto

La scienza si occupa di studiare i pattern, e i pattern sono generati da relazioni tra eventi. Rispetto alle donne, gli uomini hanno una maggiore tendenza a essere di alta statura, ad avere la barba, a morire giovani e ad andare in prigione. In altre parole, nella popolazione umana la statura, la barba, l'età alla morte e la condotta criminale non sono indipendenti dal sesso.

Talvolta, però, la probabilità di un evento *non* dipende da un altro evento. Per esempio, se lanciamo due dadi, il numero che compare su un dado non influenza il numero che compare sull'altro. Se la conoscenza di un evento non fornisce informazioni su un altro evento, allora i due eventi sono indipendenti.

Due eventi sono **eventi indipendenti** se il verificarsi di uno dei due non influenza in alcun modo la probabilità che si verifichi anche l'altro. Quando lanciamo due volte consecutivamente lo stesso dado a 6 facce non truccato, la probabilità che il primo lancio dia 3 è 1/6, come abbiamo visto precedentemente:

$$\Pr[\textit{il risultato del primo lancio è } 3] = 1/6$$

Qual è la probabilità che sia 3 anche il risultato del lancio consecutivo? La probabilità di ottenere 3 anche nel secondo lancio è ancora 1/6, indipendentemente dal fatto che il risultato del primo lancio sia stato 3 o no. Poiché il risultato del primo lancio non influenza la probabilità di ottenere 3 nel secondo lancio, possiamo dire che i due eventi sono indipendenti (Figura 5.6-1).

◆ Due eventi sono **indipendenti** se il verificarsi di uno dei due non influenza la probabilità che si verifichi l'altro.

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

La probabilità di ottenere 3 nel secondo lancio è 1/6

La probabilità di ottenere 3 nel primo lancio è 1/6

Figura 5.6-1
Diagramma di Venn per tutti i risultati possibili del lancio di due dadi a 6 facce. Il primo numero di ciascuna coppia indica il risultato ottenuto con il primo dado e il secondo numero il risultato ottenuto con il secondo dado. Il risultato 3 nel primo lancio è indicato nel box blu. La probabilità di ottenere 3 nel secondo lancio è indicata nel box verde ed è la stessa (1/6), indipendentemente dal risultato del primo lancio.

Quando la probabilità di un evento dipende dal risultato di un altro evento, allora i due eventi sono **eventi dipendenti**.

La regola del prodotto

Quando due eventi sono indipendenti, la probabilità che si verifichino entrambi è data dal prodotto della probabilità del primo evento per la probabilità del secondo. Questa proprietà degli eventi indipendenti, detta **regola del prodotto** (un caso particolare del teorema delle probabilità composte), è essenziale per l'analisi dei dati perché permette di determinare se due o più variabili siano associate.

◆ **Regola del prodotto:** Se due eventi A e B sono indipendenti, allora

$$\Pr[A \text{ e } B] = \Pr[A] \times \Pr[B]$$

La base della regola del prodotto si può vedere nella Figura 5.6-1. L'area del diagramma di Venn che corrisponde al *risultato 3 sul primo dado* e al *risultato 3 sul secondo dado* è l'area della regione di sovrapposizione tra la regione blu e la regione verde. Poiché i due eventi sono indipendenti, l'area di questa regione di sovrapposizione è semplicemente la probabilità di essere nella regione blu per la probabilità di essere nella regione verde:

Pr[(*il risultato del primo lancio è 3*) e (*il risultato del secondo lancio è 3*)]
= Pr[*il risultato del primo lancio è 3*]
× Pr[*il risultato del secondo lancio è 3*]
= 1/6 × 1/6
= 1/36

La regola del prodotto riguarda combinazioni che implicano «e», cioè la probabilità che si verifichino *entrambi* gli eventi. Volendo conoscere la probabilità che si verifichino *questo* e *quello*, e se i due eventi sono indipendenti, possiamo moltiplicare le probabilità di ciascuno di essi per ottenere la probabilità che si verifichino entrambi. L'Esempio 5.6A applica la regola del prodotto a uno studio sul fumo e sull'ipertensione arteriosa.

Esempio 5.6A — *Fumo e ipertensione arteriosa*

Sia il fumo sia l'ipertensione arteriosa sono fattori di rischio di ictus e di altre patologie vascolari. Negli Stati Uniti, circa il 17% degli adulti fuma e circa il 22% soffre di ipertensione arteriosa. La ricerca al riguardo ha mostrato che l'ipertensione arteriosa non è associata al fumo; cioè, a quanto pare, i due fattori di rischio sono indipendenti l'uno dall'altro (Liang et al., 2001). Qual è la probabilità che uno statunitense adulto scelto a caso abbia entrambi questi fattori di rischio? ■

Poiché i due eventi sono indipendenti, la probabilità che un individuo sia *fumatore* e abbia anche l'*ipertensione arteriosa* è data dalla probabilità di fumare moltiplicata per la probabilità di essere affetto da ipertensione arteriosa:

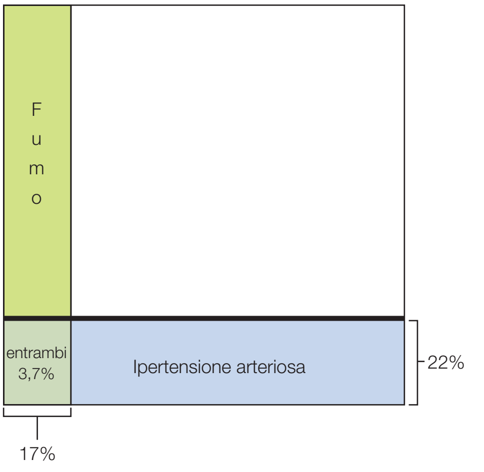


Figura 5.6-2
Diagramma di Venn per i due fattori di rischio indipendenti fumo e ipertensione arteriosa. La probabilità che si verifichino entrambi è direttamente proporzionale all'area del rettangolo in basso a sinistra.

Pr[*fumatore* e *ipertensione arteriosa*]
= Pr[*fumatore*] × Pr[*ipertensione arteriosa*]
= 0,17 × 0,22
= 0,037

Perciò, il 3,7% degli statunitensi adulti avrà entrambi i fattori di rischio di ictus.

Questo calcolo è visualizzato geometricamente nel diagramma di Venn nella Figura 5.6-2.

Enunciati che implicano «e» o «oppure»

Gli enunciati probabilistici che implicano «e» o «oppure» sono tanto comuni, e destano tanta confusione, da meritare di essere riassunti insieme:

- La probabilità di A oppure B implica la somma. Cioè, $\Pr[A \text{ oppure } B] = \Pr[A] + \Pr[B]$ se i due eventi A e B sono mutuamente esclusivi.
- La probabilità di A e B implica il prodotto. Cioè, $\Pr[A \text{ e } B] = \Pr[A] \times \Pr[B]$ se A e B sono indipendenti.

Ciò che può confondere è il fatto che l'enunciato che implica «e» richiede il prodotto, non la somma.

Indipendenza di più di due eventi

La regola del prodotto è valida anche per più di due eventi, come mostra l'Esempio 5.6B. Se più eventi sono tutti indipendenti, allora la probabilità che si verifichino tutti è il prodotto delle probabilità che si verifichi ciascuno di essi.

Esempio 5.6B

La slot machine si è mangiata tutti i soldi!

Le slot machine installate nei casinò e nelle sale giochi oggi sono in maggioranza elettroniche. Un certo tipo di slot machine dà una vincita il 9,8% delle volte che viene utilizzata, e ogni giocata è indipendente dal risultato della giocata precedente.[4] Qual è la probabilità che un giocatore che dispone di monete per otto giocate perda tutto? ■

Prima di tutto dobbiamo calcolare la probabilità di perdere in una singola giocata. La perdita e la vincita sono risultati mutuamente esclusivi di una singola giocata, quindi la probabilità di perdere in una giocata è semplicemente 1 meno la probabilità di vincere:

$$\Pr[\textit{perdita}] = 1 - \Pr[\textit{vincita}] = 1 - 0{,}098 = 0{,}902$$

La probabilità di perdere 8 volte consecutivamente è quindi data dalla probabilità di perdere una volta moltiplicata 8 volte per se stessa:

$$\begin{aligned}\Pr[\textit{perdere 8 volte}] &= \Pr[\textit{perdita nel primo gioco}]\\ &\quad\times \Pr[\textit{perdita nel secondo gioco}] \times \ldots\\ &= (\Pr[\textit{perdita}])^8 = (0{,}902)^8 = 0{,}438\end{aligned}$$

Esiste quindi una probabilità di circa il 44% di perdere tutti i soldi nelle 8 giocate.

L'Esempio 5.6C presenta un'applicazione biologica della regola dell'indipendenza nel caso di eventi multipli.

Esempio 5.6C

I piselli di Mendel

Come gli occhi azzurri negli esseri umani, anche i baccelli gialli di pisello sono un carattere recessivo. Cioè, i baccelli di pisello sono gialli soltanto se entrambe le copie del gene per il colore codificano il giallo. Una pianta di pisello che abbia una copia per il giallo e una copia per il verde del gene (una pianta «eterozigote») ha baccelli verdi come quelli delle piante che hanno due copie per il colore verde (una pianta «omozigote» verde). Gregor Mendel (1822-1884) ideò un metodo per determinare se una pianta verde fosse eterozigote oppure omozigote. Incrociò la pianta test con se stessa e valutò il colore dei baccelli di 10 discendenti scelti a caso. Se tutti i 10 discendenti erano verdi, Mendel deduceva che la pianta era omozigote; però, se anche un solo discendente era giallo, la pianta test veniva classificata come eterozigote. Qual è la probabilità che il metodo di Mendel fallisse? Se la pianta test è omozigote, ogni discendente è verde. D'altra parte, se la pianta test è eterozigote, la probabilità che un discendente sia verde è 3/4 e la probabilità che sia giallo è soltanto 1/4. Qual è la probabilità che tutti i 10 discendenti di una pianta test eterozigote siano verdi? ■

Mendel non eseguì questi calcoli, ma possiamo usare le regole della probabilità per determinare l'affidabilità del suo metodo. La probabilità che uno qualsiasi dei discendenti della pianta eterozigote sia verde è 3/4. Poiché il genotipo di ciascun discendente è indipendente dai genotipi degli altri discendenti, possiamo calcolare la probabilità che tutti i 10 discendenti siano verdi usando la regola del prodotto:

$$\begin{aligned}\Pr[\textit{tutti i 10 sono verdi}] &= \Pr[\textit{il primo è verde}]\\ &\quad\times \Pr[\textit{il secondo è verde}]\\ &\quad\times \Pr[\textit{il terzo è verde}] \times \ldots\\ &= 3/4 \times 3/4 \times 3/4 \times \ldots\\ &= (3/4)^{10} = 0{,}056\end{aligned}$$

È quindi verosimile che Mendel identificasse erroneamente circa il 5,6% degli individui eterozigoti. D'altra parte, il suo metodo identificava correttamente gli eterozigoti con probabilità (1 – 0,056) = 0,944.

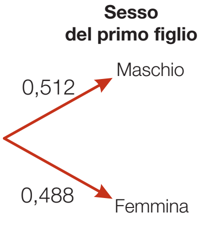
■ 5.7 Alberi di probabilità

Un **albero di probabilità** (o **albero probabilistico**) è un diagramma che può essere utilizzato per calcolare le probabilità di combinazioni di eventi che sono i risultati di più prove casuali. L'impiego degli alberi di probabilità è illustrato attraverso l'Esempio 5.7.

Esempio 5.7

Sesso e ordine di nascita

Molte coppie che pianificano di avere figli preferirebbero avere almeno un figlio di ciascun sesso. La probabilità che il primo figlio di una coppia sia maschio[5] è 0,512. In assenza di un intervento permesso dalla tecnologia, la probabilità che il secondo figlio della coppia sia maschio è indipendente dal sesso del primo figlio, e quindi rimane pari a 0,512. Immaginate di voler aiutare una nuova coppia nella sua pianificazione familiare. Se la coppia decide di avere soltanto due figli, qual è la probabilità di avere un figlio di ciascun sesso? ■



Questo problema richiede che si conoscano le probabilità di tutti i risultati mutuamente esclusivi di due variabili distinte. La prima variabile è «*il sesso del primo figlio*», la seconda è «*il sesso del secondo figlio*». Possiamo cominciare a costruire un albero di probabilità considerando le due variabili in successione, partendo dal sesso del primo figlio. Sono possibili due risultati incompatibili – cioè, «*maschio*» e «*femmina*» – che sono elencati verticalmente, uno sotto l'altro (vedi la figura a destra). Tracciamo poi delle frecce da un singolo punto a sinistra verso entrambi i due possibili risultati. Lungo ciascuna freccia scriviamo la probabilità

[4] Contrariamente a quanto si crede, una slot machine che non produce vincite da molte giocate *non* ha una maggiore probabilità di produrne nelle giocate successive rispetto a una slot machine che abbia appena fatto vincere un giocatore [Una spiegazione particolareggiata del perché si può trovare nel sito Web www.WizardofOdds.com/slots].

[5] Questa prevalenza di maschi è un pattern altamente ripetibile, misurato su decine di milioni di nati. Nascono molti più maschi che femmine. La frazione di maschi è ancora più alta al concepimento che alla nascita, perché durante la gravidanza i feti maschili muoiono a un tasso maggiore rispetto ai feti femminili. Si ritiene che gli spermatozoi che hanno il cromosoma Y siano capaci di nuotare più velocemente – e quindi di raggiungere prima la cellula uovo – rispetto agli spermatozoi con il cromosoma X, il che spiegherebbe l'eccesso di feti maschili.

© 978-88-08-06297-0

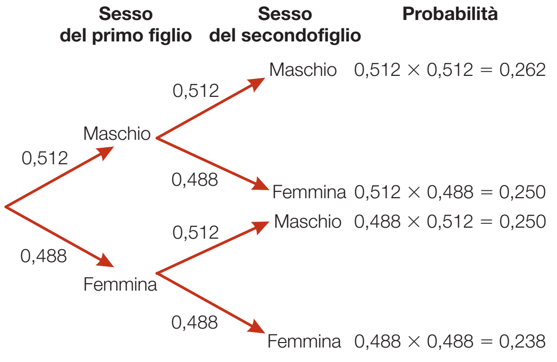


Figura 5.7-1
Un albero probabilistico per tutti i possibili risultati di una famiglia con due figli.

che si verifichi ciascun risultato (0,512 per «*maschio*» e 0,488 per «*femmina*»).

Elenchiamo poi tutti i possibili risultati per la seconda variabile, ma lo facciamo separatamente per ciascun possibile risultato della prima variabile. Per esempio, per il risultato «*maschio*» per il primo figlio, elenchiamo entrambi i risultati possibili (cioè, «*maschio*» e «*femmina*») per il sesso del secondo figlio. Quindi tracciamo frecce dal risultato «*maschio*» per la prima variabile verso i due possibili risultati per la seconda variabile, e scriviamo la probabilità di ogni risultato per la seconda variabile lungo ciascuna freccia. Ripetiamo questo procedimento nel caso in cui il risultato della prima variabile sia «*femmina*» . L'albero di probabilità che abbiamo ottenuto è mostrato nella Figura 5.7-1.

A questo punto è meglio controllare di avere scritto correttamente i risultati. Le probabilità lungo tutte le frecce che si originano da un singolo punto devono avere somma pari a 1 (a meno dell'errore di arrotondamento), perché rappresentano tutte le possibilità incompatibili. Se la loro somma è diversa da 1, abbiamo dimenticato di includere alcuni risultati oppure abbiamo scritto i valori in modo errato.

Con un albero di questo tipo siamo in grado di calcolare la probabilità di ogni possibile sequenza di risultati delle due variabili. Una sequenza di risultati è rappresentata da un cammino lungo le frecce dell'albero, il quale comincia dalla radice a sinistra e termina all'estremità di uno dei rami a destra. Per calcolare la probabilità di una data sequenza, si moltiplicano tutte le probabilità lungo il cammino percorso dalla radice all'estremità. Per esempio, la sequenza «*maschio e poi femmina*» nella Figura 5.7-1 ha una probabilità pari a $0{,}512 \times 0{,}488 = 0{,}250$. Sull'albero di probabilità di solito si elencano le probabilità di ciascuna sequenza di risultati in una colonna a destra delle estremità, come è mostrato nella Figura 5.7-1.

Ogni estremità dell'albero definisce una sequenza unica di eventi incompatibili. Esaminate la figura 5.7-1 (o qualsiasi altro albero di probabilità) per accertarvi che la somma delle probabilità di tutte le possibili sequenze sia 1. Se non è 1 (a meno dell'errore di arrotondamento), allora è stato commesso qualche errore nella costruzione dell'albero.

Qual è la probabilità di avere un figlio di ciascun sesso in una famiglia con due figli? Secondo l'albero di probabilità, due delle quattro sequenze possibili determinano la nascita di 1 maschio e di 1 femmina. Nella prima sequenza, il maschio è procreato per primo, seguito dalla femmina, mentre nella seconda sequenza la femmina è procreata per prima, seguita dal maschio. Queste due differenti sequenze sono incompatibili e vogliamo conoscere la probabilità che si verifichi la prima *oppure* la seconda sequenza (entrambe producono un maschio e una femmina). Perciò, per la regola della somma, la probabilità di ottenere esattamente 1 maschio e 1 femmina quando si generano due figli è la somma delle probabilità delle due sequenze alternative che conducono a questo risultato: $0{,}250 + 0{,}250 = 0{,}500$.

Potremmo usare l'albero di probabilità presentato nella Figura 5.7-1 per calcolare le probabilità dei seguenti risultati:

- La probabilità che nasca almeno una femmina
- La probabilità che nasca almeno un maschio
- La probabilità che entrambi i figli siano dello stesso sesso.

Calcolate queste probabilità per verificare se avete compreso il concetto.[6]

Non è obbligatorio usare alberi di probabilità quando si calcolano le probabilità di sequenze di eventi, ma essi sono un utile strumento per accertarsi di avere tenuto conto di tutte le possibilità.

5.8 Eventi dipendenti

Gli eventi indipendenti sono semplici da trattare matematicamente, ma le cose diventano di solito più interessanti quando la probabilità che si verifichi un evento dipende da un altro evento. La maggior parte delle ricerche scientifiche studia variabili associate e non indipendenti.

La determinazione del sesso è sicuramente più affascinante in diversi specie di insetti che nella specie umana. In molti insetti, infatti, la madre è in grado di variare il numero relativo di figli maschi e femmine a seconda delle condizioni ambientali. In questo caso, il *sesso della prole* e l'*ambiente* sono eventi dipendenti, come illustrato dall'Esempio 5.8.

Esempio 5.8

Questo cibo è libero?

La vespa *Nasonia vitripennis* è un parassita e depone le proprie uova sul suo ospite, la pupa di una mosca. Le larve di Nasonia escono dalle uova all'interno del pupario (l'involucro protettivo della pupa), si alimentano dell'ospite ancora vivo e

[6] Risposte: Pr[*almeno 1 femmina*] = 0,738; Pr[*almeno 1 maschio*] = 0,762; Pr[*entrambi i figli dello stesso sesso*] = 0,500.

© 978-88-08-06297-0

crescono finché non emergono da adulti, mentre l'ospite è ormai morto. I maschi e le femmine, che possono essere fratelli e sorelle, si accoppiano sul posto.
Le femmine di *Nasonia* hanno una notevole capacità di manipolare il sesso delle uova che depongono.[7] Quando una femmina trova un ospite nuovo che non è già stato parassitato, depone principalmente uova femminili, producendo soltanto i pochi maschi necessari a fecondare tutte le sue figlie. Però, se l'ospite è già stato parassitato da un'altra femmina, allora depone soprattutto uova maschili.[8] Lo stato dell'ospite incontrato da una femmina e il sesso di un uovo deposto sono quindi variabili dipendenti (Werren, 1980). ■

Supponiamo che, quando una data femmina di *Nasonia* trova un ospite, ci sia una probabilità pari a 0,20 che questo contenga già uova deposte da una femmina precedente. Supponiamo, inoltre, che la femmina sia capace di individuare senza errori le precedenti infestazioni. Se l'ospite non è stato ancora parassitato, la femmina depone un uovo maschile con una probabilità pari a 0,05 e un uovo femminile con una probabilità pari a 0,95. Se l'ospite contiene già uova, allora la femmina depone un uovo maschile con probabilità 0,90 e un uovo femminile con probabilità 0,10. La Figura 5.8-1 mostra un diagramma di Venn per queste probabilità.

In base alla Figura 5.8-1, gli eventi «*ospite già parassitato*» e «*produzione di un uovo maschile*» sono dipendenti. La probabilità di deposizione di un uovo maschile cambia a seconda che l'ospite sia già stato parassitato oppure no. Supponiamo di volere conoscere la probabilità che un nuovo uovo, scelto a caso, sia maschile. Possiamo affrontare questo problema usando un albero di probabilità, come quello presentato nella Figura 5.8-2.

In base al diagramma, esistono esattamente due cammini che portano a un uovo maschile. Nel primo, l'ospite è già parassitato e la madre depone un uovo maschile. Questo cammino ha probabilità

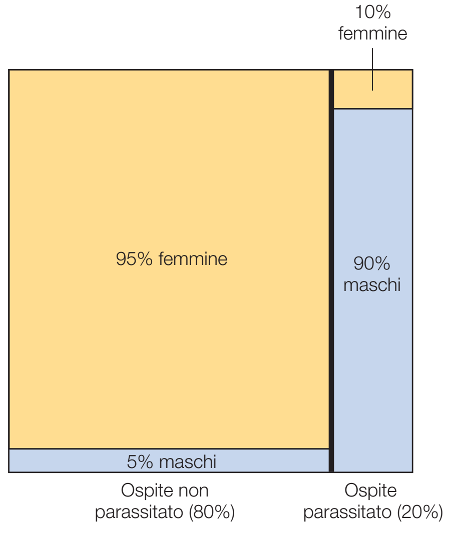


Figura 5.8-1
Diagramma di Venn che indica che il sesso delle uova deposte da *Nasonia* dipende dallo stato dell'ospite.

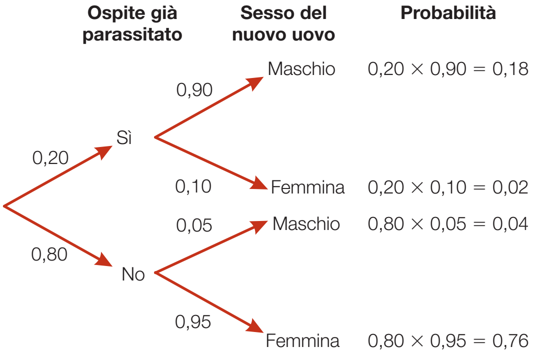


Figura 5.8-2
Un albero di probabilità per la determinazione del sesso delle uova deposte da *Nasonia* a seconda che l'ospite sia stato parassitato o meno.

$$\Pr[\textit{ospite già parassitato} \text{ e } \textit{nuovo uovo maschile}] = 0{,}20 \times 0{,}90 = 0{,}18$$

Nel secondo cammino, l'ospite non è già parassitato e la femmina depone un uovo maschile. Il secondo cammino ha probabilità

$$\Pr[\textit{ospite non già parassitato} \text{ e } \textit{nuovo uovo maschile}] = 0{,}80 \times 0{,}05 = 0{,}04$$

La probabilità che un uovo sia maschile è la somma delle probabilità di questi due cammini incompatibili:

$$\Pr[\textit{uovo maschile}] = 0{,}18 + 0{,}04 = 0{,}22$$

La probabilità che un uovo sia maschile in questa popolazione è 0,22.

L'albero probabilistico mostra che l'evento «*nuovo uovo maschile*» dipende dal fatto che l'ospite incontrato da una madre sia già stato parassitato. Ciò significa, quindi, che gli eventi «*ospite già parassitato*» e «*nuovo uovo maschile*» non sono indipendenti? Per confermarlo possiamo usare la regola del prodotto, valida soltanto per gli eventi indipendenti. La probabilità che *l'ospite sia già stato parassitato* e che *il sesso del nuovo uovo sia maschile* è 0,18. Ma questa probabilità *non* è quella che ci saremmo attesi assumendo l'indipendenza. Se moltiplichiamo la probabilità che il nuovo uovo sia maschile (0,22, come abbiamo calcolato poc'anzi) per la probabilità che un ospite sia già parassitato (0,20), otteniamo 0,22 × 0,20 = 0,044, che è diverso dalla probabilità effettiva di questi due eventi (0,18). Quindi, in base alla definizione di indipendenza, questi due eventi *non* sono indipendenti.

[7] Le vespe, come le formiche e le api, hanno un meccanismo di determinazione del sesso molto diverso dal nostro. Tutto ciò che una femmina deve fare per determinare il sesso di un uovo al momento della deposizione è controllare che questo venga fecondato o no con gli spermatozoi che ha immagazzinato. Se l'uovo è fecondato diventa una femmina, altrimenti diventa un maschio.
[8] Il numero dei figli maschi aumenta nel secondo caso perché ci sono molte femmine non imparentate con cui accoppiarsi.

5.9 La probabilità condizionata e il teorema di Bayes

Se vogliamo conoscere la probabilità di un evento, dobbiamo tenere conto di tutte le informazioni esistenti che potrebbero influenzare il suo risultato. Se volessimo conoscere la probabilità di vedere un elefante durante una passeggiata pomeridiana, per esempio, otterremmo una risposta diversa a seconda del luogo in cui ci troviamo, come il parco del Serengeti in Tanzania oppure il centro di Milano. L'algebra della probabilità condizionata permette di sviluppare e precisare i nostri ragionamenti riguardo alle probabilità di eventi casuali in presenza di informazioni addizionali.

La probabilità condizionata

La **probabilità condizionata** (o **subordinata**) è la probabilità di un evento condizionata (subordinata) al verificarsi di un altro evento.

- La **probabilità condizionata** di un evento è la probabilità che l'evento si verifichi *assumendo che* si sia verificata una certa condizione.

Nell'Esempio 5.8, la probabilità condizionata che una vespa femmina deponga un uovo maschile è pari 0,90, *assumendo che* l'ospite su cui depone l'uovo contenga già uova di un'altra femmina (cioè, sia già stato parassitato). Possiamo confermarlo esaminando di nuovo la Figura 5.8-2. La probabilità condizionata si indica nel modo seguente:

Pr[*il nuovo uovo è maschile* | *l'ospite è già parassitato*] = 0,90

Più in generale, Pr[*evento* | *condizione*] rappresenta la probabilità che si verifichi l'evento quando la condizione è soddisfatta. La barra verticale al centro di questa espressione è un simbolo che significa «dato che» o «quando la condizione seguente è soddisfatta». (Attenzione a non confonderla con il segno di divisione.)

Il diagramma di Venn nella Figura 5.8-1 illustra il significato di questa probabilità condizionata. Il 90% dell'area corrispondente a «*ospite parassitato*» rappresenta i casi in cui la prole è maschile, mentre il restante 10% si riferisce alla prole femminile. La probabilità che un uovo sia maschile è diversa per la condizione «*ospite parassitato*» (0,9) rispetto alla condizione «*ospite non parassitato*» (0.05).

La probabilità condizionata trova molte applicazioni importanti. Volendo conoscere la probabilità complessiva di un particolare evento, dobbiamo sommare tutte le sue probabilità per ogni possibile condizione, pesate sulla probabilità di quella condizione. Questa regola è nota come **teorema della probabilità totale**.

- Secondo il **teorema della probabilità totale**, la probabilità di un evento X è

$$\Pr[X] = \sum_{\text{tutti i valori di } Y} \Pr[Y]\Pr[X \mid Y]$$

dove Y rappresenta tutti i possibili valori mutuamente esclusivi (incompatibili) della condizione.

Un modo per interpretare questa formula è considerarla come la probabilità media dell'evento X, pesata sulla base delle probabilità con cui le diverse condizioni Y in cui si osserva X si possono verificare.

Il diagramma di Venn nella Figura 5.8-1 permette anche di visualizzare questo teorema. La probabilità che l'uovo sia maschile si ottiene sommando le aree dei due rettangoli blu, una per la condizione in cui *l'ospite è già parassitato* e l'altra per la condizione in cui *l'ospite non è già parassitato*. La larghezza di questi rettangoli è direttamente proporzionale alla probabilità della condizione, mentre l'altezza è direttamente proporzionale a Pr[*maschio* | *condizione dell'ospite*]. Moltiplicando la larghezza per l'altezza di ciascun rettangolo otteniamo la sua area (la sua probabilità), e sommando tra loro queste aree otteniamo la probabilità totale.

Per calcolare la probabilità che un nuovo uovo sia maschile, dobbiamo considerare due condizioni possibili: (1) l'ospite è già parassitato e (2) l'ospite non è già parassitato. Perciò, avremo in questo caso due termini nel secondo membro dell'equazione:

$$\begin{aligned}\Pr[\textit{uovo maschile}] = {} & \Pr[\textit{ospite già parassitato}]\\ & \Pr[\textit{uovo maschile} \mid \textit{ospite già parassitato}]\\ & + \Pr[\textit{ospite non parassitato}]\ \Pr[\textit{uovo maschile} \mid \textit{ospite non parassitato}]\\ = {} & (0{,}20 \times 0{,}90) + (0{,}80 \times 0{,}05) = 0{,}22\end{aligned}$$

È la stessa risposta che abbiamo ottenuto considerando l'albero di probabilità, ma ora siamo in grado di derivarla in termini di probabilità condizionate.

La regola del prodotto generalizzata (teorema della probabilità composta)

Le probabilità condizionate ci permettono di calcolare la probabilità di una combinazione di due eventi anche se essi non sono indipendenti. In questa situazione, possiamo valutare la probabilità che si verifichino entrambi moltiplicando la probabilità di un evento per la probabilità condizionata dell'altro evento, ovvero per la probabilità dell'altro evento condizionata al verificarsi del primo evento. Si tratta della **regola del prodotto generalizzata**, detta anche **teorema della probabilità composta**.

- La **regola del prodotto generalizzata** permette di trovare la probabilità che si verifichino due eventi anche quando sono dipendenti:

$$\Pr[A \text{ e } B] = \Pr[A]\ \Pr[B \mid A]$$

Riflettendo attentamente, possiamo renderci conto di come questa regola sia ragionevole. Affinché due eventi (A e B) si verifichino, deve verificarsi l'evento A. Per definizione, ciò avviene con probabilità Pr[A]. Sapendo che A si è verificato, la probabilità che si verifichi anche B è Pr[B | A]. Moltiplicando tra loro queste probabilità otteniamo la probabilità che si verifichino sia A sia B.

Non importa quale evento si denoti con A e quale con B. Vale infatti anche l'inverso, ovvero

$$\Pr[A \text{ e } B] = \Pr[B]\ \Pr[A \mid B]$$

Nel caso delle vespe parassite, per esempio, se volessimo conoscere la probabilità che *un ospite sia già stato parassitato* e che *la vespa madre abbia deposto un uovo maschile*, moltiplicheremmo la probabilità che l'ospite sia già stato parassitato (0,2) per la probabilità che venga deposto un uovo maschile *dato che* l'uovo è già stato parassitato (0,9), ottenendo 0,18.

Se A *e* B sono indipendenti, allora le informazioni disponibili su A non forniscono informazioni su B, e quindi $\Pr[B \mid A] = \Pr[B]$. Cioè, la regola del prodotto generalizzata si riduce alla regola del prodotto quando gli eventi sono indipendenti.

Il teorema di Bayes

Una potente relazione matematica sulla probabilità condizionata è data dal **teorema di Bayes**:[9]

$$\Pr[A|B] = \frac{\Pr[B|A]\Pr[A]}{\Pr[B]}$$

Il teorema di Bayes può apparire piuttosto complicato, ma può essere dedotto dalla regola del prodotto generalizzata. Poiché

$$\Pr[A \text{ e } B] = \Pr[B]\ \Pr[A \mid B]$$

e

$$\Pr[A \text{ e } B] = \Pr[A]\ \Pr[B \mid A]$$

è anche vero che

$$\Pr[B]\ \Pr[A \mid B] = \Pr[A]\ \Pr[B \mid A]$$

Dividendo entrambi i membri per $\Pr[B]$, otteniamo il teorema di Bayes.

L'Esempio 5.9 applica il teorema di Bayes alla diagnosi della sindrome di Down.

Esempio 5.9 — *La diagnosi della sindrome di Down*

La sindrome di Down (DS, Down Syndrome) è una condizione cromosomica patologica [presenza nel nucleo cellulare di 3 copie del cromosoma 21 (trisomia 21)] che interessa circa un feto su 1000. Il più accurato test per la sindrome di Down richiede l'amniocentesi, un intervento che, purtroppo, implica un piccolo rischio di aborto (circa un caso su 200). Sarebbe quindi auspicabile avere un test accurato della sindrome di Down privo di rischi: un test del genere esiste ed è noto come triplo test. Questo esame è diventato recentemente di uso comune e si basa sulla concentrazione di tre ormoni nel sangue materno a circa 16 settimane di gravidanza.

Il triplo test, però, non è perfetto: non sempre identifica correttamente un feto con sindrome di Down (un errore detto «falso negativo») e talvolta identifica erroneamente come affetto da sindrome di Down un feto con un corredo cromosomico normale (un errore detto «falso positivo»). In condizioni normali, il tasso di identificazione del triplo test (cioè, la probabilità che un feto con sindrome di Down venga identificato correttamente come affetto dalla sindrome) è 0,60. Il tasso di falsi positivi (la probabilità che un test indichi erroneamente che un feto normale è affetto da sindrome di Down) è 0,05 (Newberger, 2000). ■

Alla maggior parte delle persone questi valori numerici sembrano intuitivamente accettabili: in base alle probabilità date, il triplo test sembrerebbe corretto la maggior parte delle volte in cui viene applicato. Però, se il test su un feto scelto a caso dà un risultato positivo (cioè indica che il feto è affetto da sindrome di Down), qual è la probabilità che il feto sia realmente affetto dalla sindrome? Provate a indovinare la risposta prima di risolvere il problema.

Per affrontare questo problema utilizziamo il teorema di Bayes. Vogliamo conoscere una probabilità condizionata: la probabilità che un feto sia affetto da sindrome di Down quando il triplo test dà un risultato positivo. In altre parole, vogliamo conoscere $\Pr[DS \mid \textit{risultato positivo}]$. Usando il teorema di Bayes, possiamo scrivere

$$\Pr[DS \mid \textit{risultato positivo}] = \frac{(\Pr[\textit{risultato positivo} \mid DS]\ \Pr[DS])}{(\Pr[\textit{risultato positivo}])}$$

Conosciamo $\Pr[\textit{risultato positivo} \mid DS]$ e $\Pr[DS]$, i due fattori che compaiono al numeratore, ma non $\Pr[\textit{risultato positivo}]$, il termine al denominatore. Possiamo però determinare la probabilità di un risultato positivo usando il teorema della probabilità totale introdotta precedentemente. Cioè, possiamo sommare tutte le possibilità per trovare la probabilità di un risultato positivo.

$$\begin{aligned}\Pr[\textit{risultato positivo}] &= (\Pr[\textit{risultato positivo} \mid DS]\ \Pr[DS]) \\ &\quad + (\Pr[\textit{risultato positivo} \mid \textit{no } DS] \\ &\quad \Pr[\textit{no } DS]) \\ &= (0{,}60 \times 0{,}001) + [0{,}05 \times (1 - 0{,}001)] \\ &= 0{,}05055\end{aligned}$$

La probabilità che un evento *non* si verifichi è uguale a 1 meno la probabilità che si verifichi. Quindi, la probabilità che un feto scelto a caso *non* sia affetto da sindrome di Down è uguale a 1 meno la probabilità che sia affetto dalla sindrome. Sulla base dell'Esempio 5.9, $\Pr[DS] = 0{,}001$, quindi $\Pr[\textit{no } DS] = 1 - 0{,}001 = 0{,}999$ nella precedente equazione.

Tornando al teorema di Bayes, siamo ora in grado di rispondere alla domanda:

$$\Pr[DS \mid \textit{risultato positivo}] = \frac{(0{,}60 \times 0{,}001)}{0{,}050\,55} = 0{,}012$$

Esiste una bassissima probabilità (cioè, l'1,2%) che un

[9] Il teorema di Bayes prende il nome dal matematico e ministro presbiteriano britannico Thomas Bayes (1702-1761), che lo formulò. Il teorema fu pubblicato postumo nel 1763.

 © 978-88-08-06297-0

feto risultato positivo al triplo test sia effettivamente affetto da sindrome di Down!

Per molti è più intuitivo ragionare in termini di numeri anziché di probabilità in questo tipo di calcoli. Per ogni milione di feti sottoposti a test, 1000 saranno affetti da sindrome di Down, e 999 000 non lo saranno. Di quei 1000, il 60%, ossia 600, risulteranno positivi al test. Dei 999 000, il 5%, ossia 49 950, risulteranno falsi positivi al test. Perciò, su 1 milione di test, vi sono 600 + 49 950 = 50 550 risultati positivi, dei quali soltanto 600 sono veri positivi. I 600 veri positivi divisi per i 50 550 positivi totali sono pari all'1,2%, la stessa risposta che abbiamo ottenuto prima. I bambini affetti da sindrome di Down hanno un'elevata probabilità di essere identificati, ma essi rappresentano una piccolissima frazione di tutti i bambini. Perciò, i risultati veri positivi sono nettamente inferiori rispetto ai falsi positivi.

Questa frazione elevata di falsi positivi non è insolita. Molti strumenti diagnostici generano alte percentuali di falsi positivi rispetto al numero totale di test positivi. In questo caso è meglio eccedere in cautela, perché quando il triplo test dà un risultato positivo, può essere controllato con l'amniocentesi.

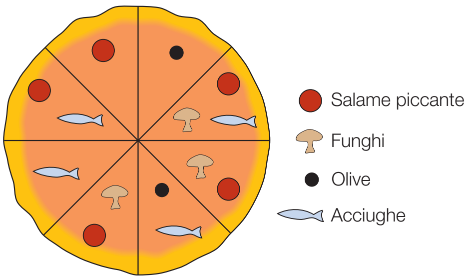
Pensavate forse che, con un risultato positivo al test, la probabilità di avere la sindrome di Down fosse più elevata? Non siete i soli. Un'indagine effettuata fra medici ha trovato che la loro comprensione della probabilità condizionata con falsi positivi era estremamente scarsa (Elstein, 1988). In una domanda sui tassi di falsi positivi, in cui la risposta corretta era che il 7,5% delle pazienti risultate positive al test erano affette da cancro mammario, il 95% dei sanitari pensava che la risposta corretta fosse il 75%. Se questi medici avessero conosciuto meglio la teoria della probabilità, avrebbero probabilmente evitato di spaventare le loro pazienti riguardo i rischi di una grave malattia, provocando loro inutili stress.

■ 5.10 Riepilogo

- Una prova casuale è un processo o un esperimento che ha due o più risultati (o esiti) possibili, il cui verificarsi non può essere previsto.
- La probabilità di un evento è la frazione di volte che l'evento si verificherebbe se si ripetesse una prova casuale moltissime volte nelle stesse condizioni.
- Una distribuzione di probabilità descrive le probabilità di tutti i possibili risultati di una prova casuale.
- Due eventi (A e B) sono incompatibili (o mutuamente esclusivi) se non possono verificarsi entrambi simultaneamente (cioè, $\Pr[A \text{ e } B] = 0$). Se A e B sono incompatibili, allora la probabilità che si verifichi A oppure B è la somma della probabilità che si verifichi A e della probabilità che si verifichi B (cioè $\Pr[A \text{ o } B] = \Pr[A] + \Pr[B]$). Si tratta della regola della somma.
- La regola della somma generalizzata fornisce la probabilità che si verifichi l'uno o l'altro evento quando essi non sono incompatibili: $\Pr[A \text{ o } B] = \Pr[A] + \Pr[B] - \Pr[A \text{ e } B]$. La regola della somma generalizzata si riduce alla regola della somma quando A e B sono incompatibili, perché in questo caso $\Pr[A \text{ e } B] = 0$.
- Due eventi sono indipendenti se la conoscenza di uno dei due risultati non fornisce informazioni sull'altro risultato. Più formalmente, A e B sono indipendenti se $\Pr[A \text{ e } B] = \Pr[A]\Pr[B]$. Si tratta della regola del prodotto.
- Gli alberi di probabilità sono utili strumenti per calcolare le probabilità di serie complicate di eventi.
- Se gli eventi non sono indipendenti, si dice che sono dipendenti. La probabilità che due eventi dipendenti si verifichino simultaneamente è data dalla regola del prodotto generalizzata: $\Pr[A \text{ e } B] = \Pr[A]\Pr[B \mid A]$.
- La probabilità condizionata di un evento è la probabilità che si verifichi quell'evento dato il verificarsi di una certa condizione.
- Gli alberi di probabilità e il teorema di Bayes sono importanti strumenti per i calcoli che implicano probabilità condizionate.
  La legge della probabilità totale, $\Pr[X] =$
  $$\sum_{\text{tutti i valori di } Y} \Pr[Y]\,\Pr[X \mid Y],$$
  permette di calcolare la probabilità di un evento (X) in base a tutte le probabilità condizionate di quell'evento. Nella legge si somma, per tutte le possibili condizioni (Y), la probabilità di quella condizione ($\Pr[Y]$) moltiplicata per la probabilità condizionata dell'evento data quella condizione ($\Pr[X \mid Y]$).

Problemi di approfondimento

1. La pizza mostrata nella figura seguente, ordinata alla Pizzeria Venn in Via Bayes, è divisa in 8 fette:



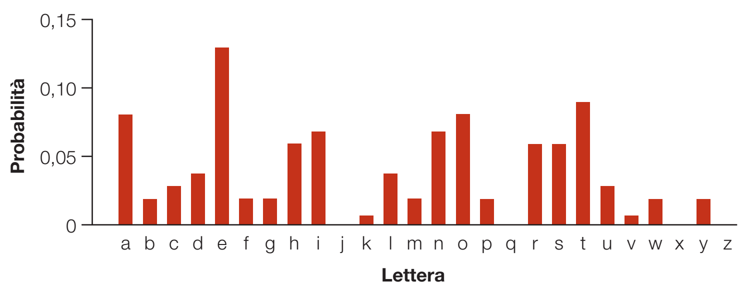
Sulle fette possono esserci salame piccante, funghi, olive e/o acciughe. Immaginate di prendere una fetta di pizza totalmente a caso (cioè, esiste una probabilità pari a 1/8 di prendere una qualsiasi delle 8 fette). Rispondete alle seguenti domande basandovi sul disegno.

a. Qual è la probabilità che la vostra fetta contenga salame piccante?
b. Qual è la probabilità che la vostra fetta contenga sia salame piccante sia acciughe?
c. Qual è la probabilità che la vostra fetta contenga salame piccante o acciughe?
d. Il salame piccante e le acciughe sono incompatibili sulle fette di pizza?
e. Le olive e le acciughe sono incompatibili sulle fette di pizza?
f. Ottenere funghi e ottenere acciughe sono indipendenti quando si scelgono fette di pizza?
g. Se qualcun altro prende una fetta da questa pizza e vi dice che contiene olive, qual è la probabilità che contenga anche acciughe?
h. Se qualcun altro prende una fetta da questa pizza e vi dice che contiene acciughe, qual è la probabilità che contenga anche olive?
i. Sette vostri amici prendono una fetta di pizza a caso e la mangiano senza dirvi come fosse farcita. Qual è la probabilità che l'ultima fetta che resta contenga olive?
j. Prendete due fette a caso da questa pizza. Qual è la probabilità che entrambe contengano olive? (Attenzione: dopo aver preso la prima fetta, la probabilità di prendere una delle fette restanti cambia.)
k. Qual è la probabilità che una fetta presa a caso *non* contenga salame piccante?
l. Disegnate una pizza dove funghi, olive, acciughe e salame piccante siano incompatibili.

2. In una data escursione di caccia, la probabilità che un leone trovi una preda è 0,80. Se il leone trova la preda, la probabilità che riesca a catturarla è 0,10. Qual è la probabilità che un leone catturi una preda ogni volta che va a caccia?

3. Le cavità negli alberi sono importanti siti di nidificazione per un'ampia varietà di animali selvatici, compreso il picchio muratore pettobianco (*Sitta carolinensis*) mostrato nella prima pagina di questo capitolo. Le cavità negli alberi sono molto più comuni nelle foreste composte prevalentemente da alberi vecchi che in quelle con una prevalenza di alberi giovani. Un recente rilevamento nel Missouri (USA) ha trovato che 45 di 273 alberi in una zona di foresta di alberi vecchi avevano cavità, mentre gli alberi restanti ne erano privi (Fan et al., 2005). Qual è la probabilità che un albero scelto a caso abbia una cavità?

4. Il seguente diagramma a barre mostra la frequenza relativa di lettere in testi in lingua inglese. Tali diagrammi sono utili per la decifrazione di codici semplici.[10]
a. Se una lettera è stata scelta a caso da un libro scritto in inglese, qual è la probabilità che sia una vocale (cioè, A, E, I, O o U)?
b. Qual è la probabilità che cinque lettere scelte indipendentemente e casualmente formino la sequenza «S-T-A-T-S»?
c. Qual è la probabilità che due lettere scelte a caso siano entrambe E?



5. In una grande popolazione di panda gigante, sono presenti cinque alleli per un certo gene (contrassegnati da A_1, A_2, A_3, A_4 e A_5). Le frequenze relative di questi alleli sono 0,1, 0,15, 0,6, 0,05 e 0,1, rispettivamente. In questa popolazione, assumendo che tutti gli individui abbiano le stesse probabilità di riprodursi, i due alleli presenti in un qualsiasi individuo rappresentano un campione casuale estratto dall'intera popolazione.
a. Qual è la probabilità che un singolo allele scelto a caso da questa popolazione sia A_1 oppure A_4?
b. Qual è la probabilità che un individuo abbia due alleli A_1?
c. Qual è la probabilità che un individuo *non* sia A_1A_1?
d. Se si estraessero due individui a caso da questa popolazione, qual è la probabilità che nessuno dei due abbia un genotipo A_1A_1?
e. Se si estraessero due individui a caso da questa popolazione, qual è la probabilità che almeno uno dei due abbia un genotipo A_1A_1?
f. Qual è la probabilità che tre individui scelti a caso *non* abbiano l'allele A_2 o l'allele A_1? (Tenete presente che ogni individuo ha due alleli, uno sul cromosoma di origine materna e uno su quello di origine paterna.)

6. Costruite un albero di probabilità che mostri la probabilità di ottenere una somma pari a 7 lanciando due dadi a 6 facce non truccati (vedi Figura 5.4-2).

7. Dopo esservi laureati in biologia, partecipate a una selezione per un posto di lavoro (ben pagato) come addestra-

[10] Vedi: www.simonsingh.net/The_Black_Chamber/frequencyanalysis.html.

Problemi di approfondimento

tore di serpenti in un circo. Nel corso dello spettacolo, dovete prendere due serpenti da una fossa. La fossa contiene 8 serpenti, 3 dei quali sono innocui, mentre gli altri 5 sono pericolosi. I tagli di bilancio avevano causato la cancellazione dell'insegnamento di erpetologia dal vostro corso di laurea, quindi non siete in grado di distinguere gli animali innocui da quelli velenosi. Prendete un serpente con la mano sinistra e un altro con la mano destra.
   a. Qual è la probabilità che *non* abbiate preso serpenti pericolosi?
   b. Supponete che ogni serpente pericoloso abbia un'uguale probabilità di mordervi. Questa probabilità è 0,8. I serpenti innocui non mordono. Qual è la probabilità che, quando prendete i due serpenti, veniate morsi almeno una volta?
   c. Supponendo ancora che i serpenti innoqui non mordano e che i serpenti pericolosi abbiano una probabilità dello 0,8 di mordere, ponete il caso di avere preso soltanto un serpente e che esso non vi abbia morso. Qual è la probabilità che questo serpente sia innoquo?

8. Cinque ricercatori estraggono indipendentemente un campione casuale dalla stessa popolazione e calcolano un intervallo di confidenza al 95% per lo stesso parametro.
   a. Qual è la probabilità che i cinque ricercatori abbiano calcolato un intervallo che include il valore vero del parametro?
   b. Qual è la probabilità che almeno uno non includa il valore vero del parametro?

9. Il «gatto di Schrödinger»[11] vive sotto la minaccia costante di morte a causa del rilascio casuale di un veleno mortale. La probabilità di rilascio del veleno è pari all'1% ogni giorno e il rilascio è indipendente nei giorni successivi.
   a. Qual è la probabilità che il gatto sopravviva 7 giorni?
   b. Qual è la probabilità che il gatto sopravviva un anno (365 giorni)?
   c. Qual è la probabilità che il gatto muoia entro un anno?

10. Alcuni semi cadono su un suolo costituito al 70% da rocce nude. Il resto del suolo è un habitat idoneo per la germinazione. Immaginate che la posizione del punto in cui cade un seme su questo suolo sia casuale rispetto all'idoneità del sito per la germinazione.
   a. Qual è la probabilità che un singolo seme cada su un sito idoneo per la germinazione?
   b. Se due semi cadono indipendentemente, qual è la probabilità che entrambi cadano su habitat idonei?
   c. Immaginate che tre semi siano sparsi casualmente e indipendentemente. Usando un albero di probabilità, trovate la probabilità che 2 di questi 3 semi cadano su un habitat idoneo.

11. Il primo test usato nello screening per l'infezione da HIV è stato l'esame ELISA, che misurava le concentrazioni di anticorpi anti-HIV in un campione di sangue. In una particolare popolazione di 5000 individui sottoposti al test, 20 erano infettati dall'HIV ma erano negativi al test (falsi negativi), 980 erano infettati ed erano positivi al test, 8 non erano infettati ma erano positivi al test (falsi positivi), e 3992 non erano infettati ed erano negativi al test.
   a. Qual era il tasso di falsi positivi?
   b. Qual era il tasso di falsi negativi?
   c. Se un individuo campionato casualmente da questa popolazione è positivo al test, qual è la probabilità che abbia l'HIV?

12. In una particolare slot machine, la probabilità di vincere in una giocata è pari al 9% e le vincite sono indipendenti da giocata a giocata. Se giocate 6 volte consecutivamente, qual è la probabilità di vincere (V) e di perdere (P) nei seguenti ordini:
   a. VVPVVV
   b. VVVVVP
   c. PVVVVV
   d. VPVPVP
   e. VVVPPP
   f. VVVVVV

13. In una competizione come la finale di tennis femminile a Wimbledon, la vincitrice è la tennista che si aggiudica per prima due set. Ogni set prosegue finché non viene vinto da una delle due; non vi sono pareggi. Se una tennista vince i primi due set, il match termina in due set. Il numero massimo possibile di set in un match è tre.
   a. Supponete che due tenniste siano tecnicamente equivalenti, cosicché la probabilità di ciascuna di vincere ogni singolo set è pari al 50%. Usando un albero di probabilità, determinate la probabilità che un match duri due set. Quali sono le probabilità che il match duri tre set?
   b. Supponete che una tennista sia più forte dell'altra, e che la probabilità della sua vittoria in ogni set sia pari al 55%. Usando un albero di probabilità, determinate la probabilità che la tennista più debole vinca il match.

14. Uno studio ha stimato che la probabilità che un uomo si lavi le mani dopo avere usato il bagno è pari a 0,74, mentre questa probabilità è pari a 0,83 per una donna. (Questi dati sono reali, e provengono da un comunicato stampa della American Society for Microbiology diffuso il 15 settembre 2003.) Una grande stanza contiene 40 uomini e 60 donne. Supponete che uomini e donne abbiano un'uguale probabilità di usare il gabinetto. Qual è la probabilità che il prossimo individuo – uomo o donna – che utilizza il bagno si lavi le mani?

15. L'istogramma nella Figura 5.4-2 mostra la probabilità di ogni risultato che si può ottenere nel lancio di due dadi a 6 facce. Usando un albero di probabilità, determinate la probabilità di ottenere una somma pari a 5 nel lancio.

16. Se avete mai scattato una foto di famiglia, dovreste sapere che è molto difficile ottenere uno scatto in cui nessuno viene immortalato con gli occhi chiusi (ossia nel momento in cui batte le palpebre). La probabilità che un individuo chiuda gli occhi durante una fotografia è circa 0,04 (Svenson, 2006).
   a. Se scattate una fotografia di una singola persona, qual è la probabilità che *non* chiuda le palpebre?
   b. Se scattate una fotografia di 10 persone, qual è la probabilità che almeno una persona chiuda le palpebre durante lo scatto?

[11] Il «gatto di Schrödinger» è un esperimento mentale ideato dal fisico e matematico austriaco Erwin Schrödinger (1887-1961) nell'ambito della meccanica quantistica. [*N.d.T.*]

Ulteriori problemi

17. Un insieme di 1600 piante di pisello in uno degli esperimenti di Mendel conteneva 900 piante alte con baccelli verdi, 300 piante alte con baccelli gialli e 300 piante basse con baccelli verdi.

a. «Piante alte» e «baccelli verdi» sono incompatibili per questo insieme di piante?

b. «Piante alte» e «baccelli verdi» sono indipendenti per questo insieme di piante?

18. Un mazzo di carte ordinario contiene 52 carte, con 13 carte per ogni seme (picche, cuori, quadri e fiori). Le carte di cuori e di quadri sono rosse, mentre quelle di picche e di fiori sono nere. Ogni seme ha un asso, 9 carte numerate da 2 a 10, e 3 «figure». Le figure sono un jack, una regina e un re. Rispondete alle seguenti domande per una singola carta estratta a caso da un mazzo di carte ben mescolato.

a. Qual è la probabilità di estrarre un re di qualsiasi seme?

b. Qual è la probabilità di estrarre una figura che sia anche una carta di picche?

c. Qual è la probabilità di estrarre una carta senza numero?

d. Qual è la probabilità di estrarre una carta rossa? Qual è la probabilità di estrarre un asso? Qual è la probabilità di estrarre un asso rosso? Questi eventi («asso» e «rosso») sono incompatibili? Sono indipendenti?

e. Elencate due eventi che siano incompatibili per una singola estrazione da un mazzo di carte.

f. Qual è la probabilità di estrarre un re rosso? Qual è la probabilità di estrarre una figura di cuori? Questi due eventi sono incompatibili? Sono indipendenti?

19. Il genoma umano è costituito da DNA i cui nucleotidi contengono le quattro basi azotate A, T, G e C. Alcune regioni del genoma umano sono estremamente ricche di G-C (cioè, in queste regioni un'elevata frazione di nucleotidi del DNA contiene guanina e citosina). Altre regioni sono relativamente ricche di A-T (cioè, in queste regioni un'elevata frazione di nucleotidi del DNA contiene adenina e timina). Immaginate di volere confrontare le sequenze nucleotidiche di due regioni del genoma. Il 60% dei nucleotidi nella prima regione contiene G-C (30% di guanina e 30% di citosina) e il 40% contiene A-T (20% di adenina e 20% di timina). La seconda regione ha il 25% di ciascuna delle quattro basi.

a. Se si sceglie a caso una singola base da ciascuna regione, qual è la probabilità che le due basi siano identiche?

b. Supponete che le basi siano presenti indipendentemente all'interno delle diverse regioni e che si campioni casualmente una sequenza di 3 basi da ciascuna regione. Qual è la probabilità che queste due triplette di basi siano identiche?

20. A Vancouver (British Columbia), la probabilità di pioggia in un giorno invernale è 0,58, in un giorno primaverile è 0,38, in un giorno estivo è 0,25 e in un giorno autunnale è 0,53. Ciascuna di queste stagioni dura 3 mesi.

a. Qual è la probabilità di pioggia in un giorno scelto a caso a Vancouver?

b. Se vi dicessero che un certo giorno a Vancouver pioveva, potreste calcolare la probabilità che quel giorno fosse invernale?

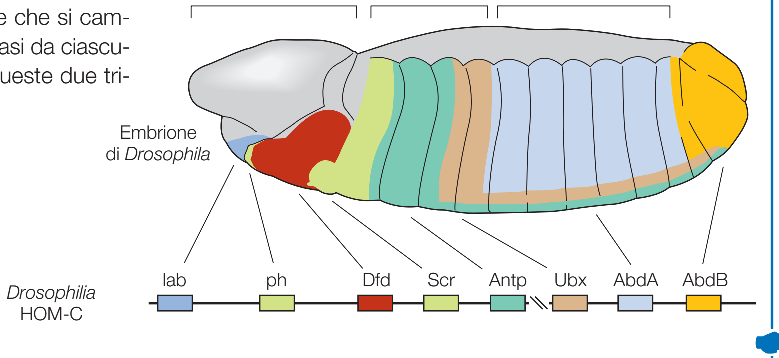
21. Quando si pongono domande su argomenti potenzialmente imbarazzanti, i ricercatori possono preoccuparsi che le risposte date non siano sincere. Per esempio, se vi chiedessero se abbiate mai rubato in un negozio, potreste dare una risposta falsa temendo che la vostra risposta vi possa mettere in cattiva luce. Una tecnica per promuovere la sincerità nelle indagini implica la probabilità. Ogni intervistato lancia due volte una moneta. Se il risultato del primo lancio è testa, allora egli risponde «sì» nel caso in cui il lancio della seconda moneta darà testa e «no» se il lancio della seconda moneta darà croce. Però, se il lancio della prima moneta dà croce, l'intervistato lancia la moneta una seconda volta ma ignora il risultato dei lanci e risponde sinceramente alla domanda imbarazzante «sì o no» riguardo al taccheggio. In questo modo chi vede i risultati dei singoli intervistati non potrà mai sapere se un particolare individuo abbia rubato in un negozio, anche se la risposta data era «sì». Supponete che il 20% delle persone di un dato gruppo abbia rubato in un negozio e che a tutti i membri di questo gruppo si chieda di svolgere la procedura precedente.

a. Costruite un albero probabilistico che descriva i possibili risultati di tale indagine.

b. Qual è la probabilità totale che una persona di questo gruppo abbia risposto «sì»?

22. Immaginate che un lungo tratto di DNA contenga adenina, timina, citosina e guanina in uguali proporzioni. Qual è la probabilità di avere 10 adenine consecutive in un campione di 10 nucleotidi estratti a caso?

23. I biologi dello sviluppo hanno scoperto che una serie di geni, detti geni Hox, è responsabile della determinazione dell'identità cellulare lungo l'asse anteroposteriore delle regioni dell'embrione in sviluppo. I geni vengono «accesi» (espressi) in differenti regioni del corpo. Un fatto sorprendente riguardo ai geni Hox è che essi tipicamente sono presenti in fila sullo stesso cromosoma e nello stesso ordine delle parti del corpo che controllano. Per esempio, la drosofila, o moscerino della frutta (*Drosophila melanogaster*), ha otto geni Hox che sono presenti sul cromosoma nello stesso ordine in cui sono espressi nel corpo, dalla testa verso l'addome (Lewis et al., 1978; Negre et al., 2005).



 © 978-88-08-06297-0

Ulteriori problemi

Se gli otto geni venissero distribuiti casualmente sullo stesso cromosoma, quale sarebbe la probabilità che essi si allineino nello stesso ordine dei segmenti in cui sono espressi?

24. Il coleottero della farina (*Tribolium* sp.) ha 10 cromosomi, di dimensioni circa uguali, e ha anche 8 geni Hox (vedi il Problema 23) (Brown et al., 2002). Se gli 8 geni Hox fossero distribuiti casualmente in tutto il genoma del coleottero, quale sarebbe la probabilità di trovarli sullo stesso cromosoma?

25. Un seme viene trasportato dal vento in un habitat complesso. Può atterrare su uno qualsiasi di tre differenti tipi di suolo: un suolo di alta qualità che dà una probabilità di sopravvivenza del seme pari a 0,8, un suolo di media qualità con una probabilità di sopravvivenza di 0,3 e un suolo di bassa qualità con una probabilità di sopravvivenza di 0,1. Questi tre tipi di suolo (alta, media e bassa qualità) sono presenti nell'habitat nelle proporzioni 30:20:50 rispettivamente. La probabilità che un seme atterri su un particolare tipo di suolo è direttamente proporzionale alla frequenza di quel tipo di suolo nell'habitat.

a. Costruite un albero probabilistico per determinare le probabilità di sopravvivenza considerando tutte le possibili circostanze.

b. Qual è la probabilità di sopravvivenza di un seme, assumendo che atterri ancora vitale al suolo?

c. Supponete che il seme abbia una probabilità pari a 0,2 di morire prima di atterrare in un habitat. Qual è la sua probabilità totale di sopravvivenza?

26. Un sacchetto contiene 5 sassi: 3 neri e 2 bianchi. Rispondete a ciascuna delle seguenti domande, assumendo che il sacchetto ritorni allo stato iniziale dopo ogni domanda.

a. Si estrae un sasso a caso dal sacchetto. Qual è la probabilità che sia bianco?

b. Si estrae un sasso dal sacchetto e *non lo si reintroduce nel sacchetto*. Questo primo sasso è nero. Qual è la probabilità che un secondo sasso estratto a caso sia nero?

c. Si estrae un sasso a caso dal sacchetto, si registra il suo colore e lo si reintroduce nel sacchetto. Si ripete questo processo altre due volte (per un totale di tre estrazioni). Qual è la probabilità che tutti e tre i sassi siano bianchi?

d. Si estrae un sasso a caso dal sacchetto, si registra il suo colore e poi *non lo si reintroduce nel sacchetto*. Si estrae un secondo sasso a caso, e non lo si reintroduce nel sacchetto. Infine, si estrae un terzo sasso dal sacchetto. Qual è la probabilità che tutti e tre i sassi estratti siano bianchi?

e. Perché le risposte alle parti (**c**) e (**d**) sono diverse?

27. Il blackjack viene giocato con un mazzo ordinario di 52 carte. (Vedi il Problema 18 per una descrizione di questo tipo di mazzo.) «Fare blackjack» significa ottenere 21 con due sole carte, ovvero una di esse è un asso (che vale 11 punti) e l'altra è un 10, un jack, una regina o un re (totale = 21). Se un giocatore riceve due carte estratte a caso dallo stesso mazzo, qual è la probabilità che faccia un blackjack? (Tenete presente che, quando una carta viene data a un giocatore, essa viene tolta dal mazzo.)

28. Trascurando gli anni bisestili, un anno ha 365 giorni.

a. Se le persone nascono con uguale probabilità in ciascuno dei 365 giorni, qual è la probabilità che tre persone scelte a caso compiano gli anni in giorni diversi?

b. Se le persone nascono con uguale probabilità in ciascuno dei 365 giorni, qual è la probabilità che 10 persone scelte a caso compiano gli anni in giorni diversi?

c. Se, come in effetti risulta essere, i tassi di natalità sono più elevati durante alcuni periodi dell'anno, aumenta o diminuisce la probabilità che 10 persone scelte a caso compiano gli anni in giorni diversi, rispetto alla risposta della parte (**b**)?

29. Durante il Progetto Manhattan (che portò alla realizzazione delle prime bombe nucleari negli Stati Uniti durante la Seconda guerra mondiale), il fisico Enrico Fermi chiese al generale Leslie R. Groves, capo militare del progetto, «In base a quale criterio definite "grande" un generale» Il generale Groves rispose: «Ogni generale che vinca cinque battaglie di fila è grande.» Aggiunse che soltanto circa il 3% dei generali sono «grandi». Se le battaglie venissero vinte completamente a caso con una probabilità del 50% per ogni parte belligerante, quale frazione di generali sarebbe «grande» secondo questa definizione? Confrontate questa percentuale con quella indicata dal generale Groves.

30. Qual è la probabilità di ricevere una scala reale all'asso (cioè, un 10, un fante, una regina, un re e un asso dello stesso seme) in una mano di cinque carte estratte da un mazzo ordinario? Tenete presente (Problema 18) che un mazzo ordinario contiene quattro semi, ciascuno con 13 carte distinte, comprendenti le cinque necessarie per la scala reale all'asso. L'ordine in cui le carte vengono date è ininfluente, e ogni carta viene trattenuta dal giocatore a cui viene data (cioè, ciascuna carta *non* viene reintrodotta nel mazzo prima che venga data la carta successiva).

31. «Dopo essere stata sottoposta a 10 mammografie, una paziente ha una probabilità del 50% di avere ricevuto almeno un falso allarme.» Date queste informazioni (tratte da Elmore et al., 2005), e assumendo che i falsi allarmi siano indipendenti l'uno dall'altro, qual è la probabilità di un falso allarme in seguito a una singola mammografia?