4. Sapendo che...

Gli assiomi di Kolmogorov e la teoria della probabilità che se ne ricava non ci dicono nulla su come ricavare e assegnare le probabilità stesse, problema che abbiamo affrontato nel Capitolo 3, né su cosa significhi dire che un evento ha una determinata probabilità. Un possibile punto di vista, come accennato, è interpretare la probabilità di un evento come misura della nostra incertezza o ignoranza riguardo ad esso: non siamo sicuri che avverrà, ma in qualche modo ci sentiamo di dire, ad esempio, che accadrà due volte su tre, se potessimo ripetere questo "esperimento" più volte, oppure che siamo disposti a scommettere che accada, purché in tal caso riceviamo almeno una volta e mezzo quello che abbiamo puntato.
In questo modo la probabilità diventa in un certo senso una misura di informazione, ed è quindi naturale pensare che, se accumuliamo nuovi dati riguardo a un evento, ne possiamo e dobbiamo aggiornare la probabilità. Questo punto di vista è legato molto strettamente al metodo scientifico: non possiamo mai avere certezza di nulla, ma una serie di risultati sperimentali favorevoli a una nostra teoria farà aumentare la nostra confidenza nel fatto che tale teoria possa dare una buona spiegazione.

4.1 Condizionamenti

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

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

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

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

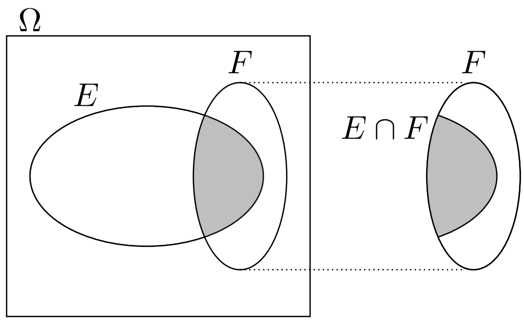

Figura 4.1: Nel condizionamento F è il "nuovo" universo.

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

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

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

La probabilità che non sia uscito 1 è:

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

Esempio 4.1.2. In una sperduta isola dell'arcipelago matematico, i traghetti passano in modo abbastanza casuale. Gli abitanti sono interessati a studiare le probabilità del tempo (in ore) passato ad aspettare un traghetto. Lo spazio Ω sarà quello dei reali positivi o, meglio, quello di tutti i reali, facendo poi attenzione che la probabilità nei negativi sia nulla. Nella Sezione 3.2 abbiamo visto che la tribù che scegliamo in questo caso è quella dei Boreliani, $\mathcal{B}$. Come probabilità P scegliamo quella indotta dalla funzione F descritta nell'Esempio 3.2.3, per cui $P((-\infty, a]) = 1 - e^{-a}$, per ogni $a \geqslant 0$ (ed è 0 per $a < 0$). Qual è la probabilità di aspettare almeno 2 ore? E quella di aspettare almeno 4 ore, se dopo due ore il traghetto non è ancora passato?

Soluzione. Qual è l'evento che ci interessa, nel primo caso? Che il tempo sia maggiore di 2, quindi nella semiretta $(2, +\infty)$, che è il complementare di $(-\infty, 2]$. Quindi

$$P((2, +\infty)) = 1 - P((-\infty, 2]) = 1 - (1 - e^{-2}) = e^{-2} \approx 13.5\%.$$

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

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

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

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

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

in cui abbiamo usato la distributività dell'intersezione rispetto all'unione.
Di conseguenza anche per P_F valgono tutte le proprietà viste per una qualunque misura di probabilità P. Tutto questo non ci dovrebbe sorprendere più di tanto, dal momento che abbiamo dato la definizione di probabilità condizionata proprio con l'idea di avere alla fine una misura di probabilità.

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

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

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

quindi $P(M^c \mid S) = 1 - P(M \mid S) = 30\%$.
Nel dare la definizione di probabilità condizionata, avevamo come scopo il quantificare l'effetto di un evento su un altro, in termini di probabilità. Tuttavia, il bello delle identità è che possiamo rigirarle un po' per mettere in evidenza altri aspetti. In particolare, dalla definizione di probabilità condizionata possiamo ricavare un modo (anzi, due) per scrivere la probabilità dell'intersezione tra due eventi:

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

La doppia identità (4.1) prende anche il nome di teorema (o regola) del prodotto. Osserviamo che siamo stati un po' imprecisi: non abbiamo specificato che $P(E) \neq 0 \neq P(F)$. Tuttavia in entrambi questi casi abbiamo che $P(E \cap F) = 0$, perché evento incluso in un evento di probabilità nulla e, qualunque valore (finito) vogliamo assegnare a $P(E \mid F)$ (o $P(F \mid E)$), lo annulleremo moltiplicandolo per 0.

Una cosa interessante potrebbe essere caratterizzare quegli eventi che non interagiscono tra loro, quelli che intuitivamente chiameremmo eventi indipendenti. Vogliamo dare una definizione matematica di indipendenza tra eventi, per poi vedere come essa si sposi con l'idea intuitiva di eventi indipendenti.

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

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

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

cioè, supponendo $P(E) \neq 0$ e $P(F) \neq 0$ (ci torniamo, tra un attimo),

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

cioè sapere che è accaduto F non cambia quello che sappiamo della probabilità di E e, viceversa, sapere che è accaduto E non cambia la probabilità di F. Osserviamo anche che, avendo due catene di uguaglianze, possiamo invertire il ragionamento, quindi sapere che E non ci dà informazioni su F ed F non ci dà informazioni su E implica che E ed F sono indipendenti, secondo la definizione di indipendenza data sopra. Quindi la definizione data caratterizza esattamente quello che ci aspettavamo e il termine usato è giustificato.

Se siamo invece interessati alla probabilità dell'intersezione di due eventi, sappiamo che essa è uguale al prodotto delle probabilità dei due eventi se questi ultimi sono tra loro indipendenti. Se non abbiamo questa informazione, dobbiamo usare la regola del prodotto (4.1) vista sopra, oppure l'identità incontrata nel Capitolo 2:

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

Esempio 4.1.4. Gaia sa che le sue professoresse di Storia e Agiografia[1] interrogano in ognuna delle due materie a sorteggio tra coloro che ancora non hanno un voto. Sapendo che della classe di 25, tra studentesse e studenti, nessuno è ancora stato interrogato in Storia e 6 persone (ma non Gaia) hanno un voto in Agiografia, con che probabilità Gaia verrà interrogata domani?

Soluzione. Gaia ha una probabilità di essere interrogata in Storia uguale a $P(S) = 1/25$, come tutte le sue compagne e i suoi compagni di classe, e $P(A) = 1/19$ di essere interrogata in Agiografia. Le due estrazioni vengono fatte da liste (o contenitori) diversi, quindi non si influenzano l'una con l'altra: possiamo allora considerare le due interrogazioni come indipendenti. La probabilità di interrogazione di Gaia è allora,

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

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

Esempio 4.1.5. Nicolò possiede un'auto sportiva gialla. Un giorno, in un parcheggio, vede che l'auto in sosta accanto alla sua è anch'essa una sportiva gialla. Mentre rientra verso casa, si chiede quanto sia probabile che un'auto sia una sportiva gialla. Da una rapida ricerca online scopre che solamente 1 auto ogni 100 è un'auto sportiva e che solo 1 auto su 200 è gialla. Ne conclude quindi che la probabilità che un'auto sia una sportiva gialla è 1/20000.
In realtà questo ragionamento non è corretto, perché nulla garantisce che i due eventi "auto sportiva" e "auto gialla" siano indipendenti. In effetti, con un po' di attenzione, Nicolò scopre poco dopo che tra le auto gialle, 1 su 3 è un'auto sportiva, quindi la probabilità che un'auto a caso sia gialla e sportiva è

$$P(S \cap G) = P(S \mid G) \cdot P(G) = \frac{1}{3} \cdot \frac{1}{200} = \frac{1}{600} \neq \frac{1}{20000} = P(S) \cdot P(G).$$

In questo contesto l'errore non ha grosse conseguenze, ma un errore simile è stato uno degli ingredienti che ha portato alla condanna, poi annullata, di Malcom Ricardo Collins[2].

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

[1] Lo studio delle vite dei santi.

[2] È un caso giudiziario realmente accaduto negli anni Sessanta in California. Per saperne di più, basta cercare il nome su un qualunque motore di ricerca. In alternativa, una discussione più dettagliata di questo e altri esempi reali analoghi di errori matematici in ambito processuale si trova nel libro *Math on Trial* [8].

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

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

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

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

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

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

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

A questo punto possiamo osservare che i due eventi $E \cap F$ ed $E \cap F^c$ sono disgiunti e la loro unione è E, quindi, siccome la probabilità dell'unione disgiunta è la somma delle probabilità, abbiamo $P(E \cap F) = P(E)P(F)$, ossia l'indipendenza.
Torniamo allora a guardare il testo iniziale e proviamo a rileggere quello che c'è scritto. La condizione $P(E \mid F) = P(E \mid F^c)$ ci dice che sapere che F sia accaduto o meno non ci dà alcuna informazione su E, infatti non cambia la sua probabilità.

4.2 Formula di fattorizzazione

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

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

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

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

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

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

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

Esempio 4.2.1. Durante un allenamento per la Gara a Squadre di Matematica, ci sono solamente tre giocatori della squadra: il Capitano, il Consegnatore e un'altra studentessa del primo anno. Il Capitano è un esperto di problemi di combinatoria e ne risolve sei su sette, gli altri due preferiscono entrambi geometria e risolvono quelli di combinatoria solo una volta su quattro. Oggi lavorano indipendentemente su tre problemi, assegnati a caso, di cui solo uno di combinatoria. Con che probabilità lo avranno risolto, come squadra, alla fine dell'allenamento?

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

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

Grazie alla formula di fattorizzazione possiamo riscrivere la probabilità cercata come

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

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

Soluzione. Indicando con M l'essere un matematico e con S, N ed E l'essere abitante sulle tre isole Idilos, Iremun ed Erettel rispettivamente, abbiamo

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

In pratica quello che stiamo facendo è prendere la media pesata delle probabilità dell'evento che ci interessa nei vari casi, con i pesi dati dalle probabilità dei casi stessi.

4.3 Intermezzo - Il valore atteso

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

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

La risposta è meno ovvia di quanto non sembri: una possibilità è infatti fare la media aritmetica degli esiti. Nel caso del lancio di un dado avremmo allora $(1 + 2 + 3 + 4 + 5 + 6)/6 = 3.5$. Ma se sapessimo che il dado è truccato, come possiamo tenerne conto? E poi, con tutte le medie che ci sono, perché proprio quella aritmetica e non, per esempio, quella geometrica? Non abbiamo una buona giustificazione da dare, la scelta che abbiamo fatto non usa nulla in particolare del contesto in cui siamo.

Un modo per includere le informazioni che abbiamo (cioè la probabilità) è il seguente: scegliamo come valore centrale la media dei possibili valori (cioè degli esiti) pesati con le loro[3] rispettive probabilità. Come strategia sembra promettente: i pesi hanno somma 1 (essendo gli esiti una partizione dello spazio Ω ed essendo la probabilità di Ω uguale a 1). Non solo, visto che stiamo considerando gli esiti del nostro esperimento aleatorio e le loro probabilità, stiamo usando informazioni sul nostro problema, perché gli eventi più probabili hanno un peso maggiore.
Questa quantità appena introdotta prende il nome di *valore atteso*, *media* o *speranza matematica*. Ha un ruolo importantissimo in probabilità, perché permette di riassumere in un numero alcune caratteristiche di un esperimento.

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

Soluzione. La moneta è bilanciata, quindi con probabilità $1/2$ paghiamo 1 euro per giocare, vinciamo e incassiamo 2 euro, per un guadagno di 1 euro. Con probabilità $1/2$ paghiamo 1 euro per giocare, perdiamo e incassiamo 0 euro, per un guadagno di -1 euro. Il guadagno medio è allora $1/2 \cdot 1 + 1/2 \cdot (-1) = 0$.
Avere valor medio nullo può sembrare noioso (in realtà nel caso di una scommessa si dice che è *equa*), vediamo allora un altro esempio in cui la speranza ha un valore diverso.

Esempio 4.3.2. Quando Marianna gioca a carte contro suo nonno ha una probabilità di vincere uguale a 0.25. Se in una settimana giocano a carte 12 volte, quante partite vincerà in media Marianna, supponendo che i risultati delle partite siano tra loro indipendenti?

Soluzione. Cerchiamo di risolvere il problema più in generale, chiamando p la probabilità che Marianna vinca una partita ed n il numero di partite giocate. Cominciamo col chiederci quale sia la probabilità che Marianna non ne vinca nemmeno una: vuol dire che in tutte le n partite giocate ha vinto il nonno, cosa che ha probabilità $(1-p)^n$. Qual è invece la probabilità che Marianna ne vinca esattamente una? Dobbiamo scegliere qual è la partita che vince, tra le n giocate, cosa che possiamo fare in $\binom{n}{1} = n$ modi. Ciascuno di questi casi ha una probabilità uguale a $p(1-p)^{n-1}$.

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

A questo punto non è difficile passare al caso più generale di un numero k (compreso tra 0 e n) di partite vinte da Marianna: la probabilità sarà $q_k = \binom{n}{k} p^k (1-p)^{n-k}$. Osserviamo che se sommiamo questa quantità al variare di tutti i valori di k possibili (cioè per k da 0 ad n) otteniamo 1. Non può essere altrimenti, perché sono tutti i casi possibili, disgiunti tra loro a due a due, quindi la loro unione dà l'intero spazio e la somma delle loro probabilità dà 1.
Il problema, però, riguarda la media, che chiameremo M. Andiamo quindi a pesare i possibili esiti, ossia il numero k di partite vinte da Marianna, ciascuno con la propria probabilità q_k. Abbiamo

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

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

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

dove abbiamo usato il fatto che la somma nella penultima riga è uguale a 1 (per lo stesso motivo per il quale lo era quella dei q_k prima[4]). Nel caso particolare che ci interessa, il numero delle partite è $n = 12$ e la probabilità di vittoria è $p = 0.25$, quindi il numero medio di partite che Marianna vince è $m = 12 \cdot 0.25 = 3$.
Come abbiamo detto, la definizione di media richiede che gli esiti siano numeri, altrimenti non sapremmo farne la media. In realtà si può mostrare, mediante il concetto di variabile aleatoria, che questa non è una grande restrizione: bisogna solo trovare come associare a ogni evento un numero, in un modo compatibile con gli assiomi della probabilità, dal momento che dovremo trasferire la probabilità

[4] Questo è in una veste diversa lo stesso modello binomiale visto nell'Esercizio 18, di cui adesso sappiamo quindi calcolare la media.

dai nostri eventi alle loro immagini mediante questa trasformazione. Queste trasformazioni, che sono a tutti gli effetti delle particolari funzioni dallo spazio di probabilità considerato alla retta dei numeri reali, prendono il nome di *variabili aleatorie*. Usando questo formalismo è possibile, tra le altre cose, lasciar cadere nella definizione di media la richiesta che gli esiti siano al più un numero numerabile, al costo di qualche accortezza aggiuntiva, tra cui sostituire integrali a somme per fare la media pesata. Questo piccolo assaggio, però, è tutto quello che vedremo delle variabili aleatorie in questo libro. Per chi fosse stato stuzzicato, però, è sufficiente prendere un qualunque testo universitario per un corso di probabilità o, meglio ancora, seguire un corso di probabilità.
Tornando alla speranza matematica, possiamo osservare che per definizione essa è determinata dalla misura di probabilità. Quindi, dal momento che la probabilità condizionata ad un evento F è essa stessa una probabilità (come abbiamo visto nell'Osservazione 4.1.2), indurrà un particolare valore atteso, che prende il nome di *speranza condizionata a* F. Questa altro non è che la speranza matematica nel mondo in cui l'evento F si è verificato.

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

Soluzione. In generale, la media di un dado a 12 facce è $M = 13/2 = 6.5$. Ma abbiamo un'informazione in più: sappiamo che non sono usciti alcuni valori, cosa che rappresentiamo con l'evento F. Vogliamo allora calcolarci il valore medio condizionato ad F. Per farlo andiamo a fare la media pesata di tutti i possibili esiti del dado, con i pesi dati dalla probabilità condizionata ad F:

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

4.4 Variazioni sull'indipendenza

Facciamo un passo indietro e torniamo all'indipendenza: l'abbiamo definita a partire dalla probabilità dell'intersezione siamo poi passati al legame con la probabilità condizionata, che ci ha dato una caratterizzazione molto più intuitiva dell'indipendenza stessa. Perché allora non abbiamo usato direttamente la probabilità condizionata per dare la definizione? Torniamo un momento alla questione, lasciata in sospeso, del caso in cui abbiamo un evento di probabilità nulla. Cosa succede? Supponiamo che sia $P(E) = 0$. Allora, per monotonia, $P(E \cap F) \leqslant P(E) = 0$, cioè $P(E \cap F) = 0 = P(E) \cdot P(F)$, ossia un evento di probabilità nulla è indipendente