UNA STORIA… PROBABILE

di Francesca D'Iapico

Si mostrano qui alcune delle tappe attraverso le quali si è compiuto il cammino che ha portato al calcolo delle probabilità come lo usiamo oggi. Un racconto pensato per lettori che abbiano già qualche familiarità con le proprietà di base della misura della probabilità di un evento e con alcuni dei risultati ad essa collegati.

La nascita della teoria della probabilità si può fissare nel 1654, anno in cui due grandi matematici del tempo Pierre de Fermat (1601-1665) e Blaise Pascal (1623-1662) iniziarono uno scambio epistolare che permise di risolvere il cosiddetto "problema dei punti", che può essere formulato in questo modo: si suppone che due giocatori scommettano due somme uguali su chi otterrà il punteggio migliore in una gara di testa o croce che preveda cinque lanci di una moneta (non truccata). La partita viene interrotta prima che uno dei due vinca: come verrà allora ripartita la posta?
La nascita del calcolo delle probabilità è così legata a un fatto di scarso interesse per la vita quotidiana, quale il gioco d'azzardo, ma questo problema (come altri analoghi) è stato l'inizio di una serie di ragionamenti che hanno permesso di fare previsioni non solo nell'ambito del gioco ma anche, come si sa bene, nella medicina, nella finanza, negli studi scientifici…
Naturalmente il problema dei punti era stato affrontato già precedentemente, per esempio da frate Luca Pacioli (1445-1514) nella "Summa" e da Girolamo Cardano (1501-1576) nel "De ludo alea", ma nessuno era ancora arrivato a una soluzione corretta.
Ecco come risolve il problema Pascal:

> "*Ecco approssimativamente come faccio a sapere il valore di ciascuna delle partite quando due giocatori giocano, per esempio, in tre partite e ciascuno ha messo in gioco 32 monete.*
> *Supponiamo che il primo abbia due punti e l'altro uno; ora giocano una partita la cui sorte è che, se il primo vince, prende tutto il denaro che è in gioco, cioè 64 monete ; se vince l'altro sono due punti contro due punti, di conseguenza, se vogliono ritirarsi è chiaro che ciascuno prende ciò che ha posto, cioè 32 monete; consideri signore che se vince il primo gli appartengono 64 monete; se perde gliene appartengono 32.*
> *Ora dunque, se non vogliono rischiare questa partita e ritirarsi senza giocarla, il primo deve dire: "Sono sicuro di avere 32 monete, perché la stessa perdita me le dà; però per le altre 32, forse le avrò io, forse le avrete voi; l'azzardo è lo stesso, dividiamo, dunque, queste 32 monete a metà e mi darete, oltre a queste, le 32 monete che mi vengono con sicurezza"'. Avrà dunque 48 monete e l'altro 16".*

In pratica, quello che Pascal considera sono i possibili esiti delle ultime due mani di una partita che viene sospesa sul punteggio di 2 a 1, dove il primo giocatore ha scommesso testa e il secondo croce. Pertanto al primo mancherà un solo punto per vincere mentre al secondo ne mancheranno due. Ciò che Pascal va a fare è elencare tutti gli esiti possibili di due lanci, segnando a lato quale giocatore avrebbe vinto se si fosse ottenuto quel risultato.

Lancio 1	Lancio 2	Punteggio
T	T	4-1
T	C	3-2
C	T	3-2
C	C	2-3

Pertanto Pascal si accorge che il primo giocatore vince in tre casi su quattro, mentre il secondo solo in un caso; quindi arriva alla conclusione che la posta vada data per tre quarti al primo e per un quarto al secondo giocatore.

In un linguaggio più formale, che ci può suggerire quali elementi sono generalizzabili e in quale modo, possiamo indicare lo spazio degli eventi con

$$\Omega = \{T,C\}X\{T,C\} = \{(T,T),(T,C),(C,T),(C,C)\}.$$

Poiché utilizziamo una moneta equilibrata e ogni lancio è indipendente dall'altro, la probabilità di ciascun esito, qui individuato da una coppia del prodotto cartesiano, è $\frac{1}{2} * \frac{1}{2} = \frac{1}{4}$.

L'insieme che descrive l'evento "vince il primo giocatore" è l'insieme A={(C,T)(T,C)}U{(T,T)} la cui probabilità è $\frac{1}{2} + \frac{1}{4} = \frac{3}{4}$, mentre la probabilità che vinca il secondo è $\frac{1}{4}$.

Da queste considerazioni si deduce che la posta venga data per tre quarti a chi ha scommesso testa e per un quarto a chi ha scommesso croce.

Gestire il rischio

Nel 1657, a seguito della corrispondenza tra Pascal e Fermat, il matematico olandese Christiaan Huygens (1629-1695) scrisse il saggio "De ratiociniis in ludo aleae" in cui stabilì le regole basilari del calcolo delle probabilità e introdusse esplicitamente il concetto di aspettativa o guadagno atteso, ovvero la misura oggettiva del valore di una particolare scommessa per la persona che la fa. Per calcolare il guadagno atteso bisogna moltiplicare la probabilità di ciascun esito per la somma che con esso verrà vinta o persa (guadagno negativo) e sommare i risultati ottenuti.

Per esempio, si supponga di scommettere 100 € che esca rosso alla roulette; poiché la ruota ha trentasei caselle numerate da 1 a 36 alternativamente rosse e nere e 2 caselle con lo zero verdi, si ha che la probabilità che vinca rosso è pari a $\frac{18}{38} = \frac{9}{19}$. Pertanto si ha che l'aspettativa è pari a

$$\frac{9}{19} * 100\text{ €} + \frac{10}{19} * (-100\text{ €}) = -5.26\text{ €}.$$

Ciò significa che se si gioca ripetutamente, scommettendo ogni volta 100 € sul rosso, si perderanno in media 5.26 € a partita.

Oggi, il calcolo della quota da pagare per assicurarsi (detto altrimenti, il calcolo del premio di un'assicurazione) segue un processo simile a quello del gioco d'azzardo: le compagnie determinano il valore di tutti i possibili rischi che assicurano, moltiplicando la probabilità del rischio (che stimano a partire dai dati degli ultimi anni) per il costo medio dello stesso e sommando tutti i rischi possibili; a questa somma aggiungono il prezzo di gestione e il guadagno che vogliono ottenere. Il risultato di questo calcolo è il costo della polizza d'assicurazione stipulata.

La legge dei grandi numeri

Nel 1713 viene pubblicato, postumo, il libro di Jakob Bernoulli (1654-1705) "Ars conjectandi", in cui il matematico svizzero elabora la legge dei grandi numeri e introduce i concetti di probabilità a priori e a posteriori. Per "probabilità a priori" Bernoulli intendeva la probabilità calcolata prima del fatto, per esempio quella nei giochi di fortuna, mentre con l'espressione "probabilità a posteriori" egli si riferiva a quella calcolata dopo il fatto, per esempio le probabilità per eventi come le malattie o la morte. Allora Bernoulli si chiese in questo secondo caso con quale accuratezza la probabilità calcolata su un campione di una popolazione potesse essere correlato a ciò che accade per l'intera popolazione. Per risolvere questo problema scelse un esempio:

> *"Supponiamo che, senza che noi lo sappiamo, ci siano nascoste in un'urna 3000 biglie bianche e 2000 biglie nere; per cercare di determinare il numero di queste biglie le prendiamo una dietro l'altra (rimettendole di nuovo dentro) e osserviamo con quale frequenza prendiamo una biglia bianca e con quale frequenza una nera. Può succedere questo tanto spesso* […] *che il numero di biglie bianche e nere scelte sia nella ragione di 3:2 come le biglie nell'urna e non in una ragione differente?"*

Possiamo riproporre l'esperimento che Bernoulli ha effettuato nel modo che segue.

Si prenda un'urna con 5000 biglie rosse e blu delle quali non conosciamo quante siano blu e quante rosse. Si estragga per 50 volte una biglia osservandone il colore e reinserendola nell'urna. Se per

esempio ottenessimo che per 31 volte viene estratta una biglia rossa e per 19 una blu, ciò spingerebbe a supporre che ci siano circa 3000 biglie rosse e 2000 blu, così che si potrebbe affermare che la probabilità a posteriori di estrarre a caso una biglia rossa è di $\frac{3}{5}$.
Jakob Bernoulli dimostrò che, prendendo in considerazione un campione sufficientemente ampio, possiamo avere sempre più fiducia nella probabilità calcolata in questo modo. Questo risultato è noto come "legge dei grandi numeri" che in linguaggio formale si può tradurre come:

$$\lim_{n\to\infty} P\left(\left|\frac{m}{n}-p\right|>\varepsilon\right)=0\,,$$

dove *m* è la frequenza con cui si osserva il successo (nell'esempio sopra il numero di volte in cui esce una biglia rossa), mentre *n* è il numero di esperimenti effettuati e *p* è la probabilità effettiva del successo .
In questo modo Jakob Bernoulli descrisse come determinare il numero di osservazioni necessarie per avere una stima, il più accurata possibile, della probabilità di un evento. Partendo da qui, la domanda che si pose fu se, dato uno specifico numero di osservazioni, fosse possibile calcolare la probabilità che esse rientrino in un dato intervallo.
La risposta definitiva a tale quesito fu data dal matematico francese Abraham De Moivre (1667-1754) che la pubblicò nel 1733 nel suo libro "Doctrine des chances". Egli mostrò come un numero di osservazioni casuali si distribuiscano attorno al valore medio; oggi questa distribuzione viene chiamata "distribuzione normale". Inoltre De Moivre scoprì che i dati raccolti si distribuivano secondo quella curva che siamo abituati a chiamare "a campana".
Ciò non rimase senza seguito. Otto anni più tardi, infatti, il matematico tedesco Karl Friedrich Gauss (1777-1855) si accorse che ogni volta che riportava in un grafico un gran numero di misurazioni la curva risultante aveva sempre una forma a campana. Gauss allora comprese che avrebbe potuto sfruttare la distribuzione normale per stimare il valore dei dati e per assegnare agli eventi probabilità numeriche approssimative.

Il teorema di Bayes

Nel 1764 comparve per la prima volta nella rivista "Philosophical Transactions" la formula di Bayes (1702-1761), ministro presbiteriano deceduto nel 1761, che aveva lasciato in eredità i suoi scritti matematici all'amico Richard Price (1723-1791), il quale si accorse dell'importanza di alcuni suoi studi e provvide a farli pubblicare nella rivista appena citata.
L'importanza di questa formula sta nel fatto che essa permette di rivedere la stima di una probabilità alla luce di nuove informazioni. Sia *I* un certo evento che chiamiamo ipotesi; sia *P*(*I*) la probabilità che l'ipotesi *I* risulti corretta, probabilità a priori. Sia *E* un evento tale che *P*(*I*/*E*) sia la probabilità che *I* risulti corretta dato l'evento *E*; questa è la stima riveduta della probabilità dell'ipotesi ed è questa la probabilità che oggi chiamiamo "condizionata". Per calcolare questa nuova probabilità c'è bisogno di conoscere la probabilità che accada *E* sapendo che *I* è corretta (*P*(*E*|*I*)) ,la probabilità che *E* accada sapendo che *I* è falsa ($P(E|I_{falsa})$ e la probabilità che *I* sia falsa data da $P\left(I_{falsa}\right)=1-P(I)$. La formula di Bayes è quindi :

$$P(I|E)=\frac{P(I)*P(E|I)}{P(I)*P(E|I)+P\left(I_{falsa}\right)*P(E|I_{falsa})}\quad .$$

Riformulato in termini moderni, il teorema di Bayes diventa:

Sia $\{E_i\}_{i=1\ldots n}$ una partizione finita di uno spazio degli eventi Ω e sia A⊆ Ω tale che P(A)>0. Allora per un generico elemento della partizione E_k (con $k\in\{1,\ldots,n\}$ si ha

$$P(E_k|A)=\frac{P(A|E_k)P(E_k)}{\sum_{i=1}^{n}P(A|E_i)P(E_i)}$$

Dimostrazione

Per definizione di probabilità condizionata si ha

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

Ma per il teorema delle probabilità totali si ha

$$\text{P(A)}=\sum_{i=1}^{n} P(A|E_i)P(E_i)$$

Quindi

$$P(E_k|A) = \frac{P(A|E_k)P(E_k)}{\sum_{i=1}^{n} P(A|E_i)P(E_i)}$$

■

Vediamo un esempio: supponiamo di effettuare un test medico per diagnosticare un certo tipo di allergia che abbia un'incidenza dell'1% nella popolazione. Ampie sperimentazioni cliniche hanno mostrato che l'affidabilità del test è del 79%. Più precisamente, anche se il test non fallisce nell'identificare un individuo allergico, esso dà un risultato positivo nel 21% dei casi in cui il soggetto non è allergico (questi casi vengono detti "falsi positivi"). Quindi, se ci si sottopone al test e si ottiene un esito positivo, qual è la probabilità che si sia effettivamente allergici?
Proviamo, prima di utilizzare la pratica formula di Bayes, ad applicare il ragionamento che porta alla nascita del teorema di Bayes. Assumiamo che la popolazione complessiva ammonti a 10000 persone e che le varie probabilità si riflettano esattamente nei numeri effettivi degli individui: pertanto, su una popolazione complessiva di 10000 persone, ad essere allergici saranno precisamente 100. In assenza del test ciò che si può dire è che la probabilità di essere allergici è dell'1% (questa è la probabilità anteriore). Se un individuo effettua il test con esito positivo bisogna rivedere la probabilità che sia allergico. Si sa che, tra la popolazione, ci sono 100 individui affetti dall'allergia, e per tutti il test darà correttamente esito positivo, identificandoli come allergici. Passando agli altri 9900 soggetti sani, per il 21% di questi il test darà un falso positivo, ovvero identificherà $9900 * 0.21 = 2179$ di loro come affetti dall'allergia. Pertanto, complessivamente, il test identificherà 2179 soggetti affetti dall'allergia. Ora la domanda è: se si appartiene a questo gruppo di persone, qual è la probabilità di appartenere effettivamente al sottogruppo di persone allergiche? Essa è pari a 100/2179=0.0459.
Infatti utilizzando la formula di Bayes si ha:

$P(I) = 0.01$	(l'allergia ha un'incidenza dell'1% tra la popolazione)
$P(E|I) = 1$	(il test dà sempre risultato positivo se si è allergici)
$P(I_{falsa}) = 0.99$	(il 99% della popolazione non è allergica)
$P(E|I_{falsa}) = 0.21$	(il test dà un falso positivo nel 21% dei casi)

$$P(I|E) = \frac{0.01*1}{(0.01*1)+(0.99*0.21)} = \frac{0.01}{0.2179} = 0.0459.$$

Una domanda per il lettore: come cambierebbe questa probabilità se nel 92% dei casi il test desse risultato positivo se l'individuo è allergico?

Il paradosso dei compleanni

La formula di Bayes venne ripresa in seguito da Laplace (1749-1827) che nel 1773 nella memoria intitolata "Sulla probabilità delle cause dei successi" enuncia e dimostra il teorema di Bayes. In questo scritto egli formula anche espressamente la "regola" per determinare la probabilità che può essere così formulata:

La probabilità di un successo è uguale al numero dei casi favorevoli diviso per il numero di tutti i casi possibili.

Un esempio dell'utilizzo di questa regola di Laplace è il cosiddetto "problema dei compleanni": in una riunione ci sono N persone che si sono incontrate in modo casuale. Per quale valore di N la probabilità che ci siano almeno due persone che festeggino il compleanno lo stesso giorno è superiore al 50%?
L'intuito spingerebbe a dire 183, ossia il numero immediatamente superiore alla metà dei giorni dell'anno, ma non è così.
Si supponga che l'anno abbia 365 giorni. Se determiniamo la probabilità che non ci siano coincidenze, allora la probabilità cercata è la complementare di questa. Si scelga una a caso fra le N persone che possa compiere gli anni in uno qualsiasi dei 365 giorni, poi se ne scelga una seconda, una terza fino a completare le N persone. Si ha allora che il numero dei casi possibili è

$$CP = 365 * 365 * 365 * \dots * 365 = 365^N \ .$$

Contiamo i casi in cui non si ripete una data di compleanno: ci sono 365 modi di scegliere la data della prima persona, 364 per la seconda, 363 per la terza… fino alla N-sima che potrà compiere gli anni in 365-(N-1) giorni. Pertanto il numero di casi favorevoli è

$$CF = 365 * 364 * 363 * \ \dots * (365 - N + 1) = \frac{365!}{(365-N)!} \ .$$

Pertanto la probabilità che non ci sia una coppia di compleanni coincidenti è:

$$\frac{CF}{CP} = \frac{\frac{365!}{(365-N)!}}{365^N} = \frac{365!}{365^N * (365-N)!} .$$

Da ciò la probabilità che ci siano almeno due persone i cui compleanni coincidano è

$$p = 1 - \frac{365!}{365^N * (365-N)!} \ .$$

Per N=23 si ha *p* = 0.507.

A proposito di DNA

Il fenomeno del paradosso dei compleanni ha oggi dimostrato di avere un'enorme importanza nei tribunali in relazione all'identificazione del DNA, quando un individuo viene identificato confrontando un suo campione di DNA con un altro preso da un elemento di riferimento ritrovato sulla scena del crimine.
Molte persone credono che il confronto avvenga tra le intere sequenze di DNA, ma di fatto non è così. Ad essere confrontati sono soltanto i cosiddetti profili del DNA. Nella profilazione del DNA, un certo numero di specifiche sequenze di geni (attualmente 13) sono scelte come punti di confronto, e per ciascuna di esse viene calcolato un marcatore numerico basato sulla sua struttura molecolare. La risultante sequenza di numeri costituisce il profilo del DNA dell'individuo. Sono i profili del DNA che vengono comparati per l'identificazione dei campioni di DNA. Quando gli scienziati che eseguono il confronto dichiarano che c'è una corrispondenza, essi intendono dire che i due campioni di DNA hanno lo stesso profilo del DNA.
Ognuno dei tredici punti di confronto è scelto in modo tale che la probabilità che due individui non imparentati scelti a caso abbiano il medesimo marcatore numerico in corrispondenza di quel punto sia circa 1/10. Poiché si ritiene che i diversi punti siano indipendenti l'uno dall'altro, la probabilità che due

persone scelte a caso abbiano tredici punti in comune è $1/10^{13}$, valore che permette di dire che questo metodo ha un alto grado di affidabilità.
Tuttavia esistono dei rischi nell'utilizzo della profilazione del DNA come prova. Innanzitutto, se due persone sono imparentate è molto più probabile che ci siano corrispondenze tra alcuni punti dei loro profili; inoltre il paradosso dei compleanni implica che in una grande città con diversi milioni di abitanti ci sia una concreta possibilità che due persone abbiano lo stesso profilo. Infatti, anche in una piccola città di 65000 abitanti la probabilità che due persone scelte a caso abbiano dei profili di DNA che combacino in 9 punti è del 5%. Pertanto i tribunali nei casi più delicati non accettano come prova un confronto tra profili di DNA.
Comunque la regola di Laplace suppone che tutti i casi, favorevoli e possibili, siano tra loro ugualmente probabili. Nel caso in cui qualcuno di essi abbia una probabilità di verificarsi maggiore o minore del resto, la probabilità del successo si deve determinare con l'aiuto della regola della somma, sempre di Laplace:

> *Se un successo si può presentare in due o più forme incompatibili tra loro, la probabilità del successo è uguale alla somma di ognuna delle probabilità di ogni caso favorevole.*

che scritta in termini formali diventa:

siano A e B due eventi disgiunti, allora $P(A \cup B) = P(A) + P(B)$.

Inoltre può accadere che il successo cercato sia composto, pertanto si deve ricorrere alla regola del prodotto, trovata ancora una volta da Laplace:

> *Se affinché si verifichi un successo si devono verificare a loro volta due successi che dipendono l'uno dall'altro, allora la probabilità del successo di partenza è uguale al prodotto delle probabilità del primo successo per la probabilità del secondo successo, supponendo che il primo sia avvenuto".*

Questa regola è quella che in termini attuali viene chiamata formula della "probabilità condizionata":

> *siano* B *l'evento di cui si vuole calcolare la probabilità e* A *il primo evento successo; allora, supponendo* $B \subseteq A$, si ha

$$\mathrm{P(B) = P(B|A) * P(A)}\ .$$

La storia continua
Nel 1812 Laplace pubblica la "Teoria analitica delle probabilità" in cui cerca di avvicinare due discipline: l'analisi matematica e il calcolo delle probabilità. In particolare discusse il teorema del limite centrale, sviluppando le teorie di Gauss ottenute dalle osservazioni sugli errori di misurazione.
Ed è questa teoria delle probabilità basata sull'analisi che dominò fino al 1933, quando il matematico sovietico Andréi Kolmogórov (1903-1987) appoggiò il calcolo delle probabilità sulla teoria della misura, proponendo una serie di assiomi che rispettassero le intuizioni descritte dalla regola di Laplace e dalla definizione frequentista di probabilità.