Gerd Gigerenzer

Quando i numeri ingannano

Imparare a vivere con l'incertezza

Raffaello Cortina Editore

LA COMPRENSIONE VISTA DALL'ESTERNO

Perché il fatto di presentare le informazioni in termini di frequenze naturali anziché di probabilità favorisce la comprensione? Le ragioni sono due: la semplicità computazionale (è la rappresentazione stessa a fare una parte del calcolo) e il primato sul piano dell'evoluzione e dello sviluppo (la nostra mente è adattata alle frequenze naturali).

LA RAPPRESENTAZIONE FA UNA PARTE DEL CALCOLO

La figura 4.2 illustra la differenza fra frequenze naturali e probabilità. A sinistra abbiamo delle frequenze naturali disposte ad albero; l'albero rappresenta il modo in cui una persona incontrerebbe delle informazioni statistiche attraverso la propria esperienza diretta. A destra le stesse informazioni vengono fornite come probabilità (ed è così che sono presentate agli studenti nella maggior parte dei libri di testo di medicina). I numeri sono ancora quelli incontrati nel problema del cancro al seno che il dottor Standing aveva trovato così difficile; i fumetti coi "pensierini" mostrano i calcoli necessari nei due casi per rispondere alla nostra domanda.

Entrambe le equazioni sono varianti della regola di Bayes,[7] che ha preso il nome dal pastore protestante (dissenziente) britannico al quale è attribuita, il reverendo Thomas Bayes (1702 (?)-1761).[8] Si vede subito che calcolare la probabilità di una malattia dato un esame positivo è più facile quando l'informazione è fornita in termini di frequenze naturali:

$$p\,(\text{malattia}|\text{positivo}) = \frac{a}{a+b}$$

La regola di Bayes con le frequenze naturali

Nella figura 4.2 a è il numero dei soggetti con un esame positivo *e* la malattia (7) e b quello dei soggetti con un esame positivo ma *senza* la malattia (70).[9] Con le probabilità, invece, il calcolo è più impegnativo:

$$p\,(\text{malattia}|\text{positivo}) = \frac{p(\text{malattia})p(\text{positivo}|\text{malattia})}{p(\text{malattia})p(\text{positivo}|\text{malattia}) + p(\text{non malattia})p(\text{positivo}|\text{non malattia})}$$

La regola di Bayes con le probabilità condizionali

Questa regola equivale a quella, più semplice, riportata sopra: entrambe ci dicono qual è la proporzione degli esiti positivi corretti (al numeratore) sulla totalità degli esiti positivi (al denominatore). La sola differenza è che nella seconda versione ogni frequenza naturale è stata sostituita dal prodotto di due probabilità. Quali? Ce lo spiega la tabella 4.1.

Un esame ha in generale quattro esiti possibili: quando una persona ha una malattia, il test può essere positivo (*positività*

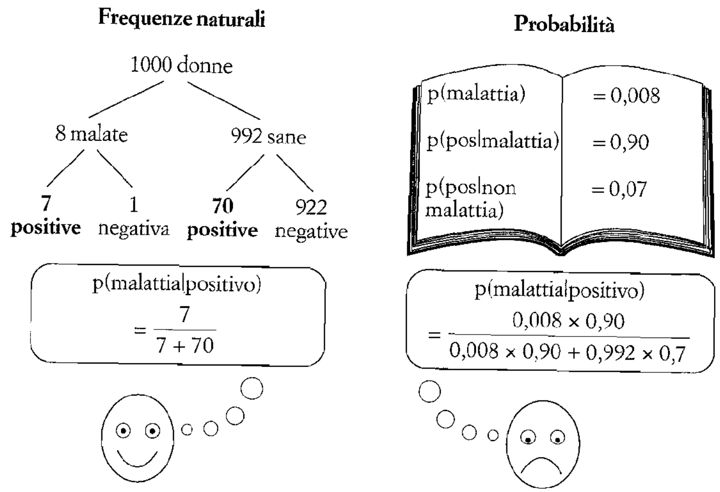

Figura 4.2 *Le frequenze naturali facilitano i computi bayesiani.* La persona dal viso felice ha avuto le informazioni pertinenti in frequenze naturali e può stimare la probabilità della malattia dato un test positivo (o un sintomo) con facilità; infatti, deve considerare due soli numeri, quello delle pazienti con un test positivo e la malattia (a = 7) e quello delle pazienti con un test positivo, ma senza la malattia (b = 70). La persona dal viso infelice ha avuto le stesse informazioni in linguaggio probabilistico, e per lei fare questa stima è difficile. La struttura della sua equazione è identica a quella usata dalla persona felice, $a/(a + b)$, ma le frequenze naturali a e b sono state trasformate in probabilità condizionali rendendo la formula molto più complessa.

Tabella 4.1 *Risultati degli esami.* Un esame può dare quattro esiti possibili: (*a*) positivo, essendoci la malattia (o qualche altra condizione ignota), (*b*) positivo senza malattia, (*c*) negativo essendoci la malattia, (*d*) negativo senza malattia. Le percentuali dei casi in cui si verificano questi quattro esiti sono dette, rispettivamente, (*a*) sensibilità (o proporzione delle positività vere), (*b*) proporzione delle positività false, (*c*) proporzione delle negatività false e (*d*) specificità (proporzione delle negatività vere). Le due aree ombreggiate indicano i due tipi di errore possibili. La frequenza delle positività vere e di quelle false sono, rispettivamente, la *a* e la *b* della regola di Bayes.

	Malattia	
Risultato dell'esame	**Sì**	**No**
Positivo	(*a*) sensibilità	(*b*) tasso di false positività
Negativo	(*c*) tasso di false negatività	(*d*) specificità

vera) o negativo (*negatività falsa*). La probabilità *p*(positivo|malattia) è la *sensibilità* dell'esame; la sensibilità della mammografia è la proporzione delle donne con un risultato positivo fra quelle col cancro al seno (di solito varia fra l'80% e il 95%, con valori più bassi fra quelle più giovani). La somma delle proporzioni di questi due esiti (sensibilità e tasso di false negatività) è uguale a 1.

Quando una persona non ha la malattia, l'esito può essere positivo (*falsa positività*) o negativo (*negatività vera*). Anche qui la somma delle rispettive proporzioni – tasso di false positività e specificità – è uguale a 1. La probabilità *p*(positivo|non malattia) è il tasso di false positività di un esame; il tasso di false positività della mammografia è la percentuale delle donne con un esito positivo fra quelle che non hanno il cancro al seno; varia fra il 5 e il 10% e ha valori più elevati fra le più giovani.

Fra i quattro esiti possibili, due (quelli ombreggiati nella tabella 4.1) sono errati. Le proporzioni di questi due errori sono interdipendenti: se il tasso di false positività di un esame diminuisce, quello di false negatività aumenta, e viceversa. Le quattro probabilità della tabella 4.1 sono dette *condizionali*

perché esprimono la probabilità di un certo evento (per esempio, un esito positivo) nell'ipotesi che se ne verifichi un certo altro (per esempio, la malattia) – *dato* cioè questo altro evento. La probabilità non condizionale *p*(malattia) è il *tasso di base* della malattia stessa; in altre parole, il tasso di base è la proporzione delle persone con una determinata malattia su una certa popolazione e in un momento specifico – ed è ben noto che le probabilità condizionali ci fanno confondere, il tasso di base no.

Ma adesso capiamo finalmente la ragione esatta per cui le cose stanno così. Quando trasformiamo le frequenze naturali in probabilità condizionali, togliamo di mezzo il tasso di base (è la cosiddetta *normalizzazione*), cosa che ha il vantaggio di far cadere tutti i valori risultanti in uno stesso intervallo, quello fra 0 e 1, ma anche lo svantaggio che quando facciamo inferenze a partire dalle probabilità anziché dalle frequenze naturali dobbiamo ri-moltiplicare queste probabilità condizionali per i rispettivi tassi di base.[10]

Ricapitolando, le frequenze naturali facilitano le inferenze effettuate sulla base di informazioni numeriche. È la rappresentazione a eseguire una parte del ragionamento fornendoci direttamente il risultato di certe moltiplicazioni che dovremmo fare noi, se ci venissero fornite delle probabilità. È in questo senso che la *comprensione* può venire alla mente dall'*esterno*.*

* Gioco di parole intraducibile. *Insight* significa "comprensione" ma anche, letteralmente, "visione interna" (*in* = "dentro", *sight* = "visione"). [NdT]