MATEMATICA, DEMOCRAZIA E VACCINI

di Marco Malvaldi

Recentemente, in seguito alle morti di alcune persone in concomitanza con la somministrazione del vaccino AstraZeneca, ho avuto modo di leggere alcuni arditi calcoli che tentavano di mostrare come tali decessi fossero fisiologici, ma giungendo a risultati paradossali. Il più semplicistico diceva questo:

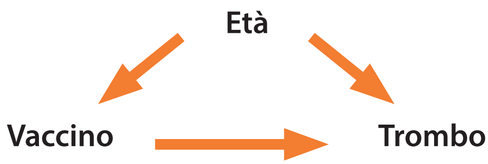
> *Dopo i titoloni allarmistici di certa stampa italiana, è normale che alcune persone si preoccupino perché si sono da poco sottoposte al vaccino di AstraZeneca (o perché lo devono ricevere a breve).*
> *Mettiamo quindi in chiaro un paio di cose:*
> *1) gli eventi di tromboemebolia sono stati circa 30 su quasi 5 milioni di persone vaccinate (0,0006%);*
> *2) non sappiamo se questi eventi siano collegati causalmente, ma solo che sono accaduti poco dopo la vaccinazione (ragione per cui l'Agenzia Europea del Farmaco ha cominciato a indagare SE le due cose fossero collegate);*
> *3) normalmente, 1 adulto su 1000 all'anno è colpito da tromboembolia, il che significa che su 5 milioni di persone mi aspetto di vedere circa 420 episodi al mese. Considerando che il vaccino AstraZeneca è stato approvato poco più di un mese fa, il fatto di aver osservato solo 30 casi di tromboembolia sembra indicare che il vaccino DIMINUISCE le possibilità di questo evento.*

Chi mi ha girato queste considerazioni, essendo un matematico accorto, mi ha sottolineato la stranezza dietro a questi risultati: un vaccino che diminuisce di un ordine di grandezza la probabilità che ti venga una trombosi, be', è piuttosto difficile da credere. Al tempo stesso, mi ha fatto notare che vaccino e stato di salute non sono variabili indipendenti: di solito, se stai già parecchio male per conto tuo non ti vai a far vaccinare. Ma, allora, come affrontare da un punto di vista matematico tale problema?

La semplice applicazione del teorema di Bayes, anche se è un ingrediente essenziale per capire se il vaccino causa il trombo, può portare fuori strada perché nel calcolo della probabilità non viene considerato nel modo adeguato il ruolo dell'età. Il teorema di Bayes, la sua conoscenza e la sua applicazione in modo adeguato, è necessaria per capire cosa succede, ma non è sufficiente – una distinzione che per un matematico è piuttosto clamorosa.

Negli ultimi due decenni, è stato sviluppato un metodo di calcolo che è in grado di mostrare in che modo poter distinguere tra cause e conseguenze tramite la semplice disponibilità di dati. Dati osservazionali, forniti dalla cara vecchia statistica, ma interpretati all'interno di un contesto, di una rete di relazioni: praticamente, un modello che tenta di simulare il meccanismo che, nel mondo reale, ha prodotto quei dati.

Se proviamo a costruire un modello causale per la domanda «se mi vaccino, mi verrà un trombo?» un primo tentativo potrebbe essere questo:



La freccia da X a Y significa, semplicemente, «X causa Y». Infatti, l'Età causa la somministrazione del Vaccino (al momento, se appartieni a una data fascia di età hai diritto al vaccino, altrimenti no ([1]), e inoltre determina anche a quale vaccino hai diritto), ma il Vaccino non causa l'Età (altrimenti nessuno se lo farebbe: se ti inietto questa fialetta non prendi il Covid ma ti ritrovi tutto d'un colpo ottantenne, sai che fregatura?).

L'Età è anche causa di Trombo (la probabilità cresce in maniera decisamente esponenziale con l'età, vedi ([2])). Ci chiediamo se il Vaccino causa il Trombo, e per questo tracciamo una freccia esplorativa tra le due variabili. Per testare se questo modello è valido o meno, dobbiamo confrontare tra loro due quantità:

- la probabilità di avere un trombo se sono stato vaccinato, che indico con $P(T|V)$;
- la probabilità di avere un trombo se vengo obbligato a essere vaccinato, che indico con $P(T|do[V])$.

Le due probabilità non sono uguali, se sono presenti delle correlazioni spurie che influenzano entrambe le variabili. Non sempre, purtroppo, è possibile calcolare la seconda quantità. Ma, se il grafo che abbiamo non contiene cicli, è sempre possibile poterla calcolare, sulla base della proprietà di indipendenza condizionale:

$$P(x_1,x_2,\ldots,x_n)=\prod_i P(x_i \vee pa[x_i])$$

dove $pa[x]$ indica i «genitori» della variabile x, ovvero le variabili che hanno una freccia che porta da loro a x. Se conosciamo tutte le variabili in gioco, e le relazioni tra esse obbediscono alla semplice regola appena detta, possiamo simulare l'effetto di un esperimento mentale, che risponde alla domanda «cosa accadrebbe (o cosa sarebbe accaduto) se obbligassi una certa variabile ad assumere il valore che voglio io?». Nel calcolo causale, indichiamo questa operazione, questo *Gedankenexperiment* – perdonate questo inutile sfoggio di tedesco, era una vita che aspettavo l'occasione di poter usare questa parola – con il simbolo $P\,(T \vee do\,[V])$.

([1]) C'è una eccezione a questo, e riguarda le categorie lavorative come medici, insegnanti, forze dell'ordine ecc: è possibile trattarlo.

([2]) M. Silverstein – J. Heit – D. Mohr – T. Petterson – W. O'Fallon – L. Melton, *Trends in the incidence of deep vein thrombosis and pulmonary embolism: a 25-year population-based study*, «Arch Intern Med.», 158 (1998): 585-593.

Definiamo che V *causa* T se

$$P\,(T \mid do\,[V = \nu]) > P\,(T \vee V = \nu)$$

Per capire la differenza tra le due grandezze, immaginiamo di chiederci se è l'altitudine che causa la temperatura oppure il contrario. Sappiamo che le due grandezze sono correlate: in montagna fa più freddo che in pianura, per cui se fa più freddo del solito è più probabile che sia in montagna. Ma la causa può essere facilmente svelata cambiando, o forzando, una delle due variabili. Se mi porto a mille o duemila metri d'altezza, vedo che la temperatura cambia: ma se riempo di ghiaccio l'abitacolo della macchina, non è che mi sollevo…

Nel caso in cui siano note tutte le variabili presenti e il grafo causale con il quale descriviamo il fenomeno contenga solo frecce orientate singolarmente e non contenga cicli, possiamo ricavare sempre tale quantità, sulla base della proprietà di indipendenza condizionale: nel nostro caso, per esempio, possiamo scrivere

$$P(T \mid do[V]) = \sum_{E} P(T \mid V, E)P(E)$$

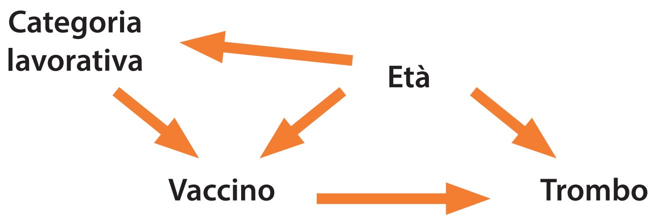
Scritta esplicitamente la ricetta, ci servono gli ingredienti.

Gli ingredienti che ci servono per il nostro calcolo sono delle tabelle di probabilità. In pratica, ci serve saper calcolare $P\,(T \mid V)$ per ogni fascia di età, $P\,(T \mid V, E)$, e $P\,(V)$ per ogni fascia di età, $P\,(V \mid E)$.

CASO (APPARENTEMENTE) PIÙ SPINOSO

Abbiamo finora fatto finta che l'Età sia l'unico fattore in gioco nella somministrazione del vaccino. In realtà, ce n'è un altro: alcune categorie (forze dell'ordine, insegnanti, medici ecc.) hanno ricevuto il vaccino, e tipologie diverse di vaccino a seconda della categoria di appartenenza.

Dato che l'età influenza l'appartenenza ad una data categoria lavorativa (se vai in pensione non lavori più) ma difficilmente appartenere alla categoria degli insegnanti ti modifica l'età, possiamo tracciare un nuovo diagramma:



In questo caso, abbiamo il colpo di fortuna che la stessa formula che abbiamo usato per il caso precedente resta valida. Infatti, Età blocca tutte le possibili correlazioni spurie tra Vaccino e Trombo, risultando in una specie di rubinetto che è in grado di separare efficacemente anche gli eventuali effetti della categoria lavorativa, secondo il cosiddetto principio *back-door* di Judea Pearl (3):

> *Data una coppia ordinata di variabili (X,Y) in un grafo diretto aciclico G, un set di variabili Z è in grado di bloccare le correlazioni spurie relativamente a (X,Y) se nessun nodo di Z è un discendente di X, e Z blocca ogni cammino tra X e Y che contiene una freccia diretta verso X.*

Avendo a questo punto tutti gli ingredienti necessari, e gli strumenti per lavorarli, possiamo mettere all'opera il nostro calcolo.

Nel giro di 37 giorni, sono stati riportati 30 casi di trombosi su 5×10^6 di vaccinati con vaccino AstraZeneca: una probabilità del 6×10^{-6}. A cose normali, ovvero su una popolazione dalla distribuzione pari a quella italiana, si sarebbero dovuti osservare – *nello stesso periodo di tempo (37 gg) e tra la popolazione vaccinata (5mln)* – circa 9 casi nella fascia 20-29 anni, 12 casi nella fascia di età 30-39 anni e circa 26 nella fascia 40-49, per un totale di 47 casi attesi. Riassumendo: 30 casi riportati rispetto a 47 casi attesi su fascia di età minore di 50 anni.

Presumendo che gli eventi che sono stati denunciati siano quelli inattesi, ovvero per una età minore di 50 anni, vediamo che siamo pienamente nel campo di quanto era lecito attendersi sulla base della statistica di questa patologia. Non so, però, se i casi riportati sono stati quelli per cui c'è una stretta correlazione temporale tra la somministrazione del vaccino e il manifestarsi del trombo. Se, per esempio, i casi fossero solo quelli riportati lo stesso giorno della vaccinazione, la statistica cambierebbe brutalmente.

Ci sono varie cose che impediscono un calcolo causale completo. In primo luogo, non siamo in grado di assegnare una distribuzione di fasce di età al solo vaccino AstraZeneca: le fasce di età sono riportate (almeno sui dati a cui ho accesso io) solo per la vaccinazione globale.

In secondo luogo, non sono in grado di capire l'intervallo di tempo entro il quale è stato riportato l'evento trombotico. Un conto è se i casi riportati sono solo quelli avvenuti lo stesso giorno del vaccino, un conto è se sono nell'intervallo di un mese.

In terzo luogo, finora si è parlato genericamente di trombosi: ma se un evento trombotico inizia a presentarsi in condizioni particolari con frequenza maggiore del solito (per esempio al cervello, come è successo nei casi studiati in Germania e in Danimarca) allora il calcolo sopra riportato non è più valido.

In pratica, ho capito e definito meglio il problema, ma non ho i dati sufficienti per poter dare una risposta. Certo, ho evitato di cadere nella trappola di cui si parlava all'inizio: dare una risposta assurda giustificandola con un calcolo applicato in maniera balorda. Mostrando calcoli del genere, rendiamo un cattivo servizio sia alla matematica che alla società. Ma al tempo stesso ho mostrato un problema non secondario: non abbiamo abbastanza dati. Lasciate che lo ripeta, perché è importante: nell'era dei big data, non siamo in grado molto spesso di raccogliere e classificare dati in modo che siano utili alla società (quella con la s minuscola, che ci riguarda tutti).

(3) J. Pearl, *Causality*, Cambridge University Press 2009.

Per quale motivo ho scelto di parlare del calcolo causale, in un articolo su matematica e democrazia? La risposta credo sia chiara: potrebbe essere un ottimo strumento di democrazia.

Innanzitutto, perché è un metodo per prendere decisioni. Il calcolo causale si basa sulla possibilità di poter intervenire e di valutare gli effetti di un intervento: non è una mera testimonianza, una statistica immota e imbelle di fronte al futuro, ma un modo per appiccicare un numero a una o più possibili azioni, e quindi poter confrontare i risultati.

Una delle capacità fondamentali della matematica è la possibilità di assegnare un rango e di confrontare: siccome siamo sicuri che cinque è maggiore di tre, possiamo sempre mettere a confronto due risultati e dire con certezza quale è il minore e quale il maggiore. È una delle poche operazioni dell'intelletto umano che non è ambigua, che non teme complotti o fraintendimenti.

Al tempo stesso, però, il numero e il rango che stiamo guardando sono il frutto di un calcolo, che contiene ipotesi, ed è valido solo finché sono valide quelle ipotesi stesse. Se il rango è inequivocabile, ma il procedimento è sbagliato, o traballante, i nostri numeri sono solo cifre prive di significato, e sono inutili.

Ma c'è un altro modo in cui dei risultati possono risultare inutili, ed è se la procedura per ottenerli è incomprensibile. Se chi legge, chi vede i risultati non capisce in che modo sono stati ottenuti, il nostro calcolo è destinato a rimanere lettera morta per queste persone – ed è qui la seconda possibile applicazione del calcolo causale. Che non riguarda i risultati, ma la possibilità di farli accettare, di renderli più comprensibili anche solo per il principio che seguono: quello di causa ed effetto. Lo stesso principio che seguiamo quando raccontiamo storie, quando trasmettiamo la nostra conoscenza attraverso le imprese degli eroi, che si chiamino Ulisse o Luke Skywalker non importa. Le storie interessanti sono sempre quelle che seguono una concatenazione di cause ed effetti: questo succede perché prima è successo quello, da cui ne consegue che eccetera eccetera.

L'essere umano è abituato a organizzare il mondo in cause ed effetti, in un ingranaggio di meccanismi complicati ma ineluttabili, in cui alcune cause non si vedono ma i loro effetti sono palesi, o almeno così crediamo. Se il nostro risultato è giusto, ma nessuno lo capisce, potremo usarlo per costruire un razzo, o sviluppare una nuova tecnologia – funzionerà, la natura è molto più fiduciosa degli esseri umani nella matematica. Ma non potremo usarlo per convincere altri uomini a prendere decisioni ponderate. Non giuste, ma ponderate: cioè rispondenti alle loro intenzioni, quali esse siano, e magari potrebbero essere anche diametralmente opposte alle nostre. Se la maggioranza delle persone decide che è preferibile così, in democrazia dobbiamo seguire la maggioranza. Conviene a tutti che più persone possibile siano in grado di valutare gli effetti delle loro intenzioni – che sono cause, e che potrebbero avere delle conseguenze assolutamente opposte a quelle che credono.

Marco Malvaldi
marcoampelio@hotmail.com