Teorema di Bayes

Andrea Centomo

Liceo “F. Corradini” di Thiene

Anno scolastico 2010/2011

Negli anni sessanta, negli Stati Uniti, era piuttosto popolare un quiz televisivo chiamato Let’s Make a Deal. Durante il programma Monty Hall, il conduttore, proponeva sempre il seguente gioco.

Ci sono tre porte, diciamo A, B e C, e dietro a una di queste c’è un ricco premio mentre dietro alle rimanenti due nulla. Monty Hall chiede al giocatore di scegliere una delle tre porte: supponiamo A. Quindi invece di aprire A, per vedere se il giocatore a o meno vinto, apre una delle due rimanenti porte, diciamo C, dietro a cui non si trova nulla. A questo punto Monty Hall chiede al concorrente se preferisce tenere la sua prima scelta A o se preferisce cambiare A con B.

Problema di Monty Hall. Conviene cambiare porta o è indifferente?

Non è difficile osservare che la probabilità di indovinare la porta con il premio all’inizio del gioco è

$$P(A) = \frac{1}{3}.$$

La parte interessante viene dopo. Nel momento in cui viene scelto se cambiare o meno la porta molte persone sono portate alle seguenti conclusioni:

1. la probabilità che il premio si trovi in A o in B è sempre $1/3$ e quindi cambiare o meno è indifferente;

2. la probabilità che il premio si trovi in A o in B è $1/2$ e quindi cambiare o meno è indifferente.

mentre la conclusione corretta è che

$$P(B) = \frac{2}{3}.$$

Il motivo, anche se non intuitivo, si può spiegare facilmente.

Immaginate che vi venga chiesto di calcolare la probabilità di indovinare la porta con il premio aprendo non una ma due porte, è chiaro che questa probabilità sarà $2/3$. Sapendo che C non nasconde un premio tutti i $2/3$ di probabilità vengono scaricati su B la cui scelta è più favorevole.

Il risultato precedente si può ottenere utilizzando il Teorema di Bayes. Supponiamo di avere due eventi E e F, sappiamo che per le probabilità condizionate valgono le formule

$$P(F|E) = \frac{P(E \cap F)}{P(E)} \qquad P(E|F) = \frac{P(E \cap F)}{P(F)}$$

da cui

$$P(F|E) = \frac{P(E|F)\,P(F)}{P(E)}.$$

Sia ora F l’evento “il premio si trova in B” ed E l’evento “il premio non è in C”. Allora

$$P(B|E) = \frac{P(E|B)\,P(B)}{P(E)} = \frac{1 \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

dove è chiaro che $P(E) = 1/2$ in quanto la probabilità a priori di scelta di una delle due porte rimaste è $1/2$.

Se invece F è l'evento "il premio si trova in A" allora

$$P(A|E) = \frac{P(E|A)P(A)}{P(E)} = \frac{\frac{1}{2}\cdot\frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}.$$

Un altro problema che si può risolvere con la formula di Bayes.

> **Problema**. *Si consideri una scuola che ha il 60% di studenti maschi e il 40% di studentesse femmine. Le studentesse indossano in egual numero gonne o pantaloni; gli studenti indossano tutti quanti i pantaloni. Un osservatore, da lontano, nota un generico studente coi pantaloni. Qual è la probabilità che quello studente sia una femmina?*

Ora M e F siano gli eventi "essere maschio" e "essere femmina":

$$P(M) = \frac{60}{100} = \frac{3}{5} \qquad P(F) = \frac{2}{5}.$$

Indichiamo con B l'evento "avere i pantaloni". Allora

$$P(B|M) = 1 \qquad P(B|F) = \frac{1}{2}.$$

$$P(F|B) = \frac{P(B|F)P(F)}{P(B)} = \frac{\frac{1}{2}\frac{2}{5}}{P(B)} = \frac{1}{5P(B)}.$$

Per calcolare $P(B)$ utilizziamo il Teorema della Probabilità Totale:

$$P(B) = P(B|M)P(M) + P(B|F)P(F) - P(M \cap F) = 1 \cdot \frac{3}{5} + \frac{1}{2}\cdot\frac{2}{5} - 0 = \frac{4}{5}$$

da cui

$$P(F|B) = \frac{1}{5P(B)} = \frac{1}{4}.$$

C'è pertanto $1/4$ di probabilità che lo studente sia femmina.

> **Problema del test positivo**. Si consideri un test destinato a rivelare la presenza nel siero di anticorpi anti-HIV. Si assuma che questo test abbia una *sensibilità* del 100% (il test, quindi, è positivo nel 100% dei malati). Si assuma che questo test abbia una *specificità* del 99.7% (il test, quindi, è negativo nel 99.7% dei soggetti sani). Si sa che la prevalenza dell'infezione da HIV è del 3 per mille (nella popolazione, su 1000 soggetti presi a caso, 3 sono infetti da virus). Quale è il valore predittivo del test positivo?

Indichiamo con N l'evento "il test è negativo" e con $\bar{N}$ il suo complementare. Indichiamo con M l'evento "essere ammalato" e con $\bar{M}$ il suo complementare. Dai dati del problema abbiamo

$$P(\bar{N}|M) = 1 \qquad P(N|\bar{M}) = \frac{997}{1000} \qquad P(M) = \frac{3}{1000}.$$

$$P(M|\bar{N}) = \frac{P(\bar{N}|M)P(M)}{P(\bar{N})} = \frac{P(\bar{N}|M)P(M)}{P(\bar{N}|M)P(M) + P(\bar{N}|\bar{M})P(\bar{M})} = \frac{\frac{3}{1000}}{\frac{3}{1000} + \left(1 - \frac{997}{1000}\right)\left(1 - \frac{3}{1000}\right)}$$

da cui

$$P(M|\bar{N}) = \frac{1}{1 + \left(1 - \frac{3}{1000}\right)} = \frac{1000}{1997} \approx 50\%$$

Nel caso degli anticorpi anti-HIV, la differenza tra la probabilità di essere malati a posteriori (dopo avere effettuato il test, pari al 50%) e la probabilità di essere malati a priori (prima di avere effettuato il test, pari al 3 per mille) rappresenta appunto il valore aggiunto che il test è in grado di fornire, in termini di informazione, alla diagnosi clinica.