## La Regola di Bayes

Introduciamo la *Regola di Bayes*, un argomento particolarmente importante (chiamato spesso anche *Formula di Bayes* o *Teorema di Bayes*).

Tale risultato rappresenta oggi il nocciolo centrale delle tematiche collegate all'inferenza statistica e permette di illustrare come le conoscenze personali su un un certo fenomeno possano essere aggiornate dopo ogni esperienza condotta su di esso.

L'importanza di questo risultato per la statistica è tale che la divisione tra le due scuole – statistica bayesiana e statistica frequentista (o tradizionale)– è praticamente nata dalla disputa sull'interpretazione che ad esso viene data.

*

* *

Il seguente è un risultato teorico (preliminare) di notevole importanza:

**Teorema 1** *: Per ogni coppia di eventi $A$ e $B$, di probabilità non nulla, si ha*

$$\mathbb{P}(A) = \mathbb{P}(A/B)\mathbb{P}(B) + \mathbb{P}(A/B^c)\mathbb{P}(B^c).$$

*(L'uguaglianza vale anche se scambiamo $A$ con $B$.)*

*Dimostrazione* Infatti, da $A = A \cap \Omega = A \cap (B \cup B^c)$ segue

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) = \mathbb{P}(A/B)\mathbb{P}(B) + \mathbb{P}(A/B^c)\mathbb{P}(B^c),$$

avendo usato direttamente le definizione di probabilità condizionata.

Iniziamo con un Esempio che apre la strada a quel che vogliamo fare.

**Esempio 1** *Un commerciante di televisori si serve principalmente di tre industrie produttrici diverse: Anvedi, Bellavista e Chiaroschermo (che indicheremo con $A$, $B$, $C$). La sua conoscenza delle ditte lo porta a sapere che $A$ produce metà dei televisori di $B$ ed il $20\%$ dei televisori prodotti è difettoso, $B$ il $10\%$ di difettosi e $C$, che produce il $50\%$ di televisori più di quanti ne produce $B$, ne produce il $5\%$ difettosi. Un giorno gli arriva al negozio un suo ordinativo di $600$ televisori.*

*1) qual è la probabilità che, scegliendone uno a caso, esso sia difettoso ?*

*2) qual è la probabilità che, sapendolo difettoso, esso sia stato prodotto da Anvedi ?*

**Svolgimento** Dai dati del problema si ha subito che, indicando con $D$ l'evento "televisore difettoso":

— $\mathbb{P}(A) = \dfrac{1}{6}; \quad \mathbb{P}(B) = \frac{1}{3}; \quad \mathbb{P}(C) = \frac{1}{2};$

— $\mathbb{P}(D/A) = \dfrac{1}{5}; \quad \mathbb{P}(D/B) = \dfrac{1}{10}; \quad \mathbb{P}(D/C) = \dfrac{1}{20}.$

Si ha allora:

1) $\mathbb{P}(D) = \mathbb{P}(D/A)\mathbb{P}(A) + \mathbb{P}(D/B)\mathbb{P}(B) + \mathbb{P}(D/C)\mathbb{P}(C) =$
$= \dfrac{1}{5}\dfrac{1}{6} + \dfrac{1}{3}\dfrac{1}{10} + \dfrac{1}{2}\dfrac{1}{20} = \dfrac{11}{120}.$

2) $\mathbb{P}(A/D) = \dfrac{\mathbb{P}(A \cap D)}{\mathbb{P}(D)} = \dfrac{\mathbb{P}(D/A)\mathbb{P}(A)}{\mathbb{P}(D)} = \dfrac{1}{5}\dfrac{\frac{1}{6}}{\frac{11}{120}} = \dfrac{4}{11}.$

Quanto abbiamo visto è solo un risultato utile alla comprensione del più importante:

**Teorema 2** (di Bayes) ([1]) *Dato un evento $A$ ed una classe completa di eventi incompatibili $H_j$, $j = 1...N$, si ha*

$$\mathbb{P}(H_j/A) = \frac{\mathbb{P}(A/H_j)\mathbb{P}(H_j)}{\sum_j \mathbb{P}(A/H_j)\mathbb{P}(H_j)}.$$

[1]Thomas Bayes (Londra, 1702 Tunbridge Wells, 1761) è stato un matematico e ministro presbiteriano britannico. È noto soprattutto per questo risultato, pubblicato postumo nel 1763. Tale risultato, che è una semplice conseguenza della definizione di probabilità condizionata, ha avuto sviluppi e contraccolpi (in filosofia, in statistica, di carattere culturale, ...) ancor oggi causa di controversie. Le sue considerazioni probabilistiche furono accettate da Laplace (nel 1781), poi da Condorcet (1743 – 1794) e da altri scienziati fino a quando George Boole (1815 – 1864) le ha (scientificamente) "contestate". Da allora è sorta la diatriba scientifico-culturale fra statistica tradizionale e statistica bayesiana.

Ciò si ricava da

$$\mathbb{P}(H_j/A) = \frac{\mathbb{P}(H_j \cap A)}{\mathbb{P}(A)}, \quad A = A \cap \Omega = A \cap (\cup_{j=1}^{N} H_j),$$

da cui

$$\mathbb{P}(H_j/A) = \frac{\mathbb{P}(A/H_j)\mathbb{P}(H_j)}{\sum_{k=1}^{N} \mathbb{P}(A/H_k)\mathbb{P}(H_k)}.$$

La Regola (o il Teorema) di Bayes cerca di risalire a quella che viene chimata la *probabilità delle cause*. L'esempio che stiamo per proporre cerca di rivelarne l'essenza.

**Esempio 2** *Un test farmacologico di un vaccino per malattie rare (una su* 100 000*) ha queste caratteristiche: per i malati il test è positivo nel* 95% *dei casi; se non si è malati il test è positivo allo* 0.005.
*Se Carlo esegue il test e risulta positivo, qual è la probabilità che Carlo sia effettivamente malato ?*

**Svolgimento** Sia $M$ l'evento che dice che Carlo è malato e $(T_+)$ l'evento che dice che il Test è positivo per Carlo. Allora cerchiamo

$$\mathbb{P}(M/(T_+) = \frac{\mathbb{P}(T_+/M)\mathbb{P}(M)}{\mathbb{P}(T_+/M)\mathbb{P}(M) + \mathbb{P}(T_+/M^c)\mathbb{P}(M^c)} =$$

$$\frac{0.95 \cdot 0.00001}{0.95 \cdot 0.00001 + 0.99999 \cdot 0.005} \approx 0.002.$$

Questo Test porta la probabilità iniziale, pari a 0.00001 ad un valore pari a 0.002, ovvero ci dice che la probabilità che Carlo sia malato è 200 volte superiore al dato di partenza.

Nelle condizioni in cui abbiamo svolto l'Esempio precedente, è chiaro che se, fra qualche tempo, Carlo ripetesse il Test, il suo punto di partenza non sarebbe più 1 su 100 000 ma 2 su 1000, cioè l'ultimo risultato disponibile (o ultima informazione per lui valida).

**Esempio 3** *In un palazzo vivono solo tre famiglie: Allegri, Beati, Contenti di* $4$ *componenti ciascuna. La famiglia Allegri è composta da* $4$ *maschi, la Beati da* $3$ *maschi e una femmina, la Contenti da* $4$ *femmine. Se dal portone del palazzo esce una persona di sesso maschile, quale è la probabilità che egli appartenga alla famiglia Beati ?.*

**Svolgimento** Dobbiamo preliminarmente decidere con quale probabilità esce un componente di una qualunque delle tre famiglie. In mancanza di altre informazioni supponiamo l'equiprobabilità: allora $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C)$.

Indicando con $M$ l'evento "l'individuo è maschio" dobbiamo calcolare

$$\mathbb{P}(B/M) = \frac{\mathbb{P}(M/B)\mathbb{P}(B)}{\mathbb{P}(M/A)\mathbb{P}(A) + \mathbb{P}(M/B)\mathbb{P}(B) + \mathbb{P}(M/C)\mathbb{P}(C)} = \frac{\frac{1}{4}}{\frac{1}{4} + \frac{1}{3}} = \frac{3}{7}.$$

**Esempio 4** *In un compito a quiz con cinque risposte multiple, uno studente su due si presenta all'esame senza aver studiato e quindi è "costretto" a rispondere a caso. Se lo studente Carlo risponde correttamente ad una domanda, qual è la probabilità che abbia studiato ?*

**Svolgimento** Indichiamo con $C = \{$ Carlo risponde correttamente $\}$ e con $S = \{$ Carlo ha studiato $\}$. Abbiamo allora:

$$\mathbb{P}(S/C) = \frac{\mathbb{P}(S)\mathbb{P}(C/S)}{\mathbb{P}(S)\mathbb{P}(C/S) + \mathbb{P}(S^c)\mathbb{P}(C/S^c)} = \frac{0.5 \cdot 1}{0.5 \cdot 1 + 0.5 \cdot 0.20} = \frac{5}{6}.$$

Presentiamo ancora due eventi atipici in cui la Regola di Bayes è assai evidenziata dalla sua peculiare applicazione.

**Esempio 5** *Qual è la probabilità che un imputato, condannato all'ergastolo, sia in effetti innocente ?*

**Nota:** La peculiarità di questo Esempio sta nel fatto che partiamo ..... senza dati. È importante dunque ricostruire il contesto numerico. Naturalmente le considerazioni che seguiranno potranno, a seconda dalla variabilità statistica, essere modificate.

Definiamo gli eventi $A$ = { *l'imputato è giudicato colpevole* } e con $B$ = { *l'imputato è innocente* }.

Dobbiamo calcolare $\mathbb{P}(B/A)$. Per applicare la Regola di Bayes ($\mathbb{P}(B/A) = \dfrac{\mathbb{P}(A/B)\mathbb{P}(B)}{\mathbb{P}(A)}$), dobbiamo fare alcune ipotesi di lavoro:

- il tribunale ha un archivio statistico che dice che il 90% degli imputati è veramente colpevole. Poniamo dunque $\mathbb{P}(B^c) = 0.9$;
- il giudice è considerato uno molto attento a non commettere ingiustizie; a maggior ragione è pensabile che nel caso di ergastolo, l'errore sia minimo. Poniamo dunque $\mathbb{P}(A/B) = 0.001$;
- l'opinione publica giudica un colpevole non condannato *meno grave* che un innocente in carcere. Poniamo tale fatto come $\mathbb{P}(A^c/B^c) = 0.05$.

Abbiamo allora

$$\mathbb{P}(A) = \frac{1}{1000}\frac{1}{10} + \frac{95}{100}\frac{9}{10} \approx 0.8551.$$

Pertanto $\mathbb{P}(B/A) = \dfrac{0.0001}{0.8551} = 0.00012$. In queste ipotesi quindi una persona (su 10 000) rischia di essere condannata ingiustamente.

**Nota**: La (eventuale) difficoltà di questo Esempio non sta ovviamente nella parte aritmetica (molto semplice) ma nell'impostazione del problema; la trasformazione di un dato (di archivio, di cronaca, ..., ) in un dato numerico è non semplice, non automatico, forse da qualcuno ritenuto addirittura opinabile: ma, a pensarci bene, la presunta opinabilità non è altro che un aspetto della definizione soggettiva. È questo infatti uno dei casi in cui l'assegnazione "frequentista" di probabilità deve coincidere con la definizione (di probabilità da adottare). Ed è naturale ritenere che tale procedura è assolutamente necessaria per avere indicazioni (numeriche) che abbiano un qualche significato scientifico.

**Esempio 6** ( Tale Esempio ricalca il "paradosso del carceriere", ben presente in molti libri di Calcolo delle Probabilità.)

*Al master di Matematica Sublime vengono ammessi* 5 *studenti ogni anno: nel 2006 si sono presentati* 9 *candidati, fra i quali il giovane Carlo. Due giorni dopo la prova Carlo incontra il Direttore della Scuola e gli chiede come è andata: il Direttore risponde che non può dirgli niente che riguardi Carlo stesso, che peraltro la commissione è ancora al lavoro e che ha quindi solo informazioni parziali: gli può dire comunque che il candidato Adelmo ha superato la prova.*

*Ci chiediamo allora se sia possibile determinare:*

*a) la* $\mathbb{P}(C) = \mathbb{P}\{$ *Carlo venga ammesso* $\}$.

*b) la* $\mathbb{P}(C/A) = \mathbb{P}\{$ *Carlo venga ammesso sapendo che Adelmo è stato ammesso* $\}$.

*c) la* $\mathbb{P}(C/D) = \mathbb{P}\{$ *Carlo venga ammesso dopo aver ricevuto dal Direttore l'informazione che un altro candidato ha superato la prova* $\}$.

**Svolgimento**

a) L'unica ipotesi che possiamo fare è che i 9 candidati siano "equibravi" e dunque

$$\mathbb{P}(C) = \mathbb{P}\{ \text{ Carlo venga ammesso } \} = \frac{5}{9}.$$

**Nota**: Lo stesso risultato si ottiene osservando che i casi favorevoli sono $\binom{8}{4}$, mentre i casi possibili sono $\binom{9}{5}$.

b)

$$\mathbb{P}(C/A) = \mathbb{P}\{ \text{ Carlo venga ammesso sapendo che Adelmo viene ammesso } \} =$$

$$\frac{\binom{7}{3}}{\binom{8}{4}} = \frac{1}{2}.$$

c) Indichiamo con $D$ l'evento { il Direttore rivela a Carlo che un altro dei candidati è stato ammesso. }

Notiamo anzitutto che $D \subset A$ in quanto Adelmo può essere assunto anche se il Direttore avesse fatto un altro nome. Da questa osservazione segue

$$\mathbb{P}(C/D) = \frac{\mathbb{P}(C \cap D)}{\mathbb{P}(D)} = \frac{\mathbb{P}(C \cap D \cap A)}{\mathbb{P}(D \cap A)} =$$

$$\frac{\mathbb{P}(A)\mathbb{P}(C/A)\mathbb{P}(D/A \cap C)}{\mathbb{P}(A)\mathbb{P}(C/A)\mathbb{P}(D/A \cap C) + \mathbb{P}(A)\mathbb{P}(C^c/A)\mathbb{P}(D/A \cap C^c)}.$$

Per valutare questa espressione, ricordiamo che

$$\mathbb{P}(C^c/D) = 1 - \mathbb{P}(C/D) = \frac{1}{2}$$

e che

$$\mathbb{P}(A) = \mathbb{P}(C) = \frac{5}{9} \text{ (per l'“equibravura”) }.$$

Indichiamo poi con

- $P_A = \mathbb{P}(D/A \cap C)$ ovvero la probabilità che il Direttore riveli l'ammissione del candidato Adelmo, sapendo che sia Carlo che Adelmo vengono ammessi;
- $P_{A^c} = \mathbb{P}(D/A \cap C^c)$ ovvero la probabilità che il Direttore riveli l'ammissione del candidato Adelmo, sapendo che che Adelmo è stato ammesso ma Carlo No.

Abbiamo allora

$$\mathbb{P}(C/D) = \frac{\frac{5}{9}\frac{1}{2}P_A}{\frac{5}{9}\frac{1}{2}P_A + \frac{5}{9}\frac{1}{2}P_{A^c}}$$

Cosa possiamo dire sui valori $P_A$ e $P_{A^c}$ ? Collegata a questa domanda è la questione dell'indipendenza degli eventi $D$ e $C$. In genere *non possiamo dire* che essi sono indipendenti in quanto tutto dipenderà da situazioni di scelta della correzione da parte della commissione, ecc...

Se, ad es., comunque la commissione avesse iniziato a caso a correggere gli elaborati, avremmo

$$P_A = \frac{1}{4}, \quad P_{A^c} = \frac{1}{5}$$

(la prima nell'ipotesi che il Direttore sappia che anche Carlo è stato ammesso, la seconda nel caso in cui sappia che Carlo non è stato ammesso).

Avremo allora

$$\mathbb{P}(C/D) = \frac{\frac{5}{9}\frac{1}{2}\frac{1}{4}}{\frac{5}{9}\frac{1}{2}\frac{1}{4} + \frac{5}{9}\frac{1}{2}\frac{1}{5}} = \frac{5}{9}.$$

Vediamo, come forse ci si poteva aspettare, che in questo caso gli eventi $C$ e $D$ vengono ad essere eventi indipendenti: ciò è dipeso dall'aver postulato l'"equibravura" dei candidati e dal fatto che il Direttore conosca solo la situazione di Carlo ed Adelmo; peraltro se i candidati non fossero "equibravi" oppure se il Direttore conoscesse un altro nome di candidato ammesso, le cose cambierebbero; ad es., sempre in ipotesi di equibravura ma con questa ultima "conoscenza" avremmo $P_A = \frac{1}{3}$, $P_{A^c} = \frac{1}{4}$ e non avremmo più indipendenza; se poi avessimo informazioni sulla "bravura" dei concorrenti le cose ancora cambierebbero (e l'indipendenza pure).

Ribadiamo pertanto il fatto che il concetto di indipendenza fra eventi, che a priori sembrerebbe "naturale" o quantomeno frutto del senso comune, è invece – e speriamo che con questi Esempi ciò sia stato compreso – un concetto molto delicato, non "assoluto" ma "dipendente" dalle ipotesi in campo e, conseguentemente, dalla concezione stessa di probabilità.

**Nota**: Se $A$ rappresenta un'ipotesi di cui si voglia determinare la probabilità, e $B$ indica il risultato di un esperimento, la Regola di Bayes permette di rideterminare la probabilità stimata soggettivamente, tenendo conto dei risultati effettivamente ottenuti nell'esperimento.

Da qui è nato il nome, per alcuni scienziati infelice, di *assegnazione soggettiva* di probabilità. La denominazione richiama però il fatto che la probabilità calcolata in questo modo dipende fondamentalmente dalle informazioni in possesso del "soggetto" che valuta la probabilità: naturalmente il **Principio di coerenza** di Bruno de Finetti fà sì che la valutazione *personale* del soggetto sia scientifica e razionale e non un opinione arbitraria.

Abbiamo già detto, parlando di probabilità condizionata, che uno stesso giudizio, esposto di fronte ad una folta assemblea, viene accettato a diversi livelli, secondo le profondità delle conoscenze degli ascoltatori.

Se colui che lo espone ne è intimamente persuaso e se, per la sua posizione e per la sua autorevolezza ispira una grande fiducia, la sua esposizione, per quanto straordinaria, avrà, per gli ascoltatori privi di altre fonti, lo stesso grado di verosimiglianza di un fatto comune.

La stessa situazione – fiducia, autorevolezza, . . ., spesso "definisce" i buoni insegnanti.