

Flavia Maria Mascioli

## Metodi bayesiani nella diagnosi medica

*Riprendendo le considerazioni esposte in un precedente articolo, l'A. illustra alcuni casi interessanti di applicazioni di metodi bayesiani nella diagnosi medica.*

»

1. — I problemi di inferenza statistica e quelli di decisione presentano una caratteristica comune, l'incertezza. Essa deriva in entrambi i casi, fondamentalmente, dalla incompletezza delle informazioni disponibili relativamente al problema, conoscitivo o decisionale, in esame.

In questi casi, un modo per ridurre l'incertezza è quello suggerito dall'approccio bayesiano che consiste nella revisione dell'opinione alla luce di nuove informazioni. Le opinioni sono espresse in termini di probabilità, le informazioni derivano dall'osservazione, e queste informazioni sono impiegate per rivedere l'opinione « iniziale » e ottenere una nuova opinione « finale ».

Il meccanismo formale alla base di questo procedimento è conosciuto come teorema di Bayes.

La teoria delle decisioni statistiche e i metodi dell'inferenza statistica bayesiana, sono stati ampiamente sviluppati negli ultimi vent'anni. Durante questo periodo è stata costruita una teoria unificata, e i concetti e i metodi hanno trovato applicazione in numerosi campi, in ingegneria e nelle comunicazioni, nel collaudo e nel controllo di qualità, in psicologia e nelle scienze sociali, nella diagnosi medica.

Proprio in quest'ultima, l'adozione di tecniche bayesiane, insieme con l'impiego di calcolatori elettronici, ha condotto di recente a risultati che sebbene ancora limitati, sembrano suscettibili di interessanti sviluppi.

Questo articolo si propone di illustrare questa possibilità, considerando due particolari problemi di diagnosi medica.

2. — La diagnosi medica, per la natura dei problemi che si trova ad affrontare è particolarmente adatta all'applicazione della teoria delle decisioni statistiche.

La corrispondenza fra malattie e sintomi, segni e dati clinici (nel seguito, per brevità, « sintomi ») è molto spesso di difficile individuazione. Malattie diverse possono presentare sintomi in parte identici o molto simili, mentre una medesima malattia si può presentare con sintomi di diversa intensità in pazienti diversi. Perciò in generale, il medico si trova a dover decidere diagnosi e terapie in situazioni di incertezza, basando i propri giudizi sulla conoscenza di probabilità. Ne è sufficiente la disponibilità di dati sempre più completi e affidabili: occorre, anche sviluppare metodi e tecniche adatti all'interpretazione dei dati disponibili.

3. — Un primo esempio di applicazione del teorema di Bayes in questo campo è quello proposto da Nugent (1) e citato da Lusted (2), e riguarda una ipotetica paziente quarantenne visitata in una clinica per malattie endocrine.

Ad un primo esame ella presenta alcune caratteristiche tipiche della sindrome di Cushing (tumore benigno dell'ipofisi): obesità generale, strie rosse, circa 105 di pressione diastolica, irsutismo, poca tolleranza al glucosio.

Al medico interessa la probabilità che la paziente sia effettivamente affetta da tale malattia, e per calcolarla il teorema di Bayes si rivela di grande utilità.

Il medico, basandosi sulla sua precedente esperienza e su dati statistici, può dare una valutazione della probabilità iniziale che la paziente sia malata:  $P_0(M) = 0,25$ , cui corrisponde, ovviamente, una probabilità che sia sana  $P_0(\bar{M}) = 0,75$ .

Per avere nuove informazioni, il medico esaminerà i dati

---

(1) C. A. Nugent ed al.: *Probability Theory in the Diagnosis of Cushing's Syndrome*. J. Clin. Endocr., 24: 621-627, 1964, Tavola 3.

(2) L. B. Lusted: *Introduction to Medical Decision Making*. Springfield, Thomas 1968.

Tabella 1 (\*)

Sintomi	Probabilità di un sintomo in un paziente colpito dalla sindrome di Cushing	Probabilità di un sintomo in un paziente non colpito dalla sindrome di Cushing
	$P \left( \frac{\text{sintomo/}}{\text{malato}} \right)$	$P \left( \frac{\text{sintomo/}}{\text{non malato}} \right)$
(S <sub>i</sub> )	$P(S_i/M)$	$P(S_i/\bar{M})$
1. Obesità centrale	0,90	0,29
2. Pletora	0,82	0,31
3. Oligomenorrea	0,72	0,51
4. Debolezza	0,65	0,07
5. Sesso femminile	0,65	0,77
6. Globuli bianchi 11.000 per mm <sup>3</sup> o più	0,58	0,30
7. Età 35 anni o meno	0,55	0,52
8. Ecchimosi	0,53	0,06
9. Acne	0,52	0,24
10. Irsutismo	0,50	0,29
11. Strie rosse	0,46	0,22
12. Mal di testa	0,41	0,37
13. Pressione diastolica del sangue 105 o più	0,39	0,17
14. Edema	0,38	0,17
15. Pressione venosa 49 o più	0,37	0,32
16. Osteoporosi	0,27	0,06
17. Potassio del siero 3,6 mEq/ litro o meno	0,25	0,04
18. Obesità generale	0,03	0,62

(\*) Da C. A. Nugent et al.: *Probability Theory in the Diagnosis of Cushing's Syndrome*, cit.

della Tabella 1, dove sono rappresentate le probabilità che ogni sintomo  $S_i$  della sindrome di Cushing sia presente in un paziente, subordinatamente all'ipotesi che sia affetto o meno dalla malattia. Queste sono le probabilità subordinate o verosimiglianze che indicheremo con  $P(S_i / M)$ . Le  $P(S_i / M)$  nella Tabella 1 sono scritte in ordine decrescente. Si può ora applicare il teorema di Bayes e calcolare la probabilità finale che la paziente sia malata:

$$P(M/S_i) = \frac{P_0(M) P(S_i / M)}{P(S_i)} \quad (1)$$

Anziché considerare i diversi sintomi singolarmente, e procedere al calcolo della probabilità finale con un'analisi sequenziale, (dove cioè la distribuzione finale relativa al primo sintomo diviene iniziale per il secondo e così di seguito), è possibile considerare la serie di dati alla stregua di un unico dato ponendo:

$$P(\underline{S/M}) = P(S_1, S_2, S_3, \dots, S_{18} / M) \quad (2)$$

e supponendo i sintomi indipendenti subordinatamente alla malattia in esame, si ha:

$$P(\underline{S/M}) = \prod_{i=1}^{18} P(S_i / M) \quad (3)$$

Pertanto, scrivendo la (1) sotto forma di rapporti,

Rapporto finale = Rapporto iniziale  $\times$  Rapporto delle verosimiglianze degli  $S_i$

$$\frac{P(\underline{M/S})}{P(\underline{\tilde{M}/S})} = \frac{P_0(M)}{P_0(\tilde{M})} \times \prod_{i=1}^{18} \frac{P(S_i / M)}{P(S_i / \tilde{M})}$$

nel caso particolare in esame si avrà:

$$\frac{P(M/S)}{P(\tilde{M}/S)} = \frac{P_0(M)}{P_0(\tilde{M})} \times \frac{P(\tilde{S}_1/M)}{P(\tilde{S}_1/\tilde{M})} \times \dots \times \frac{P(S_{18}/M)}{P(S_{18}/\tilde{M})} \quad (4)$$

dove con  $P(\tilde{S}_i/M)$  si indica  $1 - P(S_i/M)$ , quando il sintomo è assente.

I rapporti nella (4) permettono una valutazione immediata di quanto la probabilità per M aumenti o diminuisca in rapporto alla probabilità per  $\tilde{M}$ .

Eseguito i calcoli si ottiene:

$$\frac{P(M/S)}{P(\tilde{M}/S)} = 0,001$$

Il rapporto finale ha un valore molto basso; rispetto al valore iniziale si ha una riduzione da 1/3 a 1/1000. Pertanto la probabilità che la paziente sia affetta dalla sindrome di Cushing diminuisce considerevolmente.

4. — Il secondo esempio riguarda la diagnosi dei tumori primari delle ossa, e, rispetto a quello esaminato in precedenza, si presenta più complesso ed interessante.

Per la diagnosi radiologica dei tumori ossei, risulta difficile al radiologo l'accumulazione di una esperienza sufficiente, in quanto avere un paziente con questo tipo di tumore può essere considerato un evento raro.

In questo caso, quindi, più ancora che in quello precedente, possono essere di grande aiuto il teorema di Bayes e l'uso del calcolatore.

Nella Tabella 2 che è dovuta a Lodwick (3) sono riportate le probabilità iniziali  $P_0(M)$  e le probabilità subordinate  $P(S_i/M)$

(3) G. S. Lodwick: *A Probabilistic Approach to the Diagnosis of Bone Tumors*, Radial Clin. N. Amer, 3: 487-497, 1963, Tavola II.

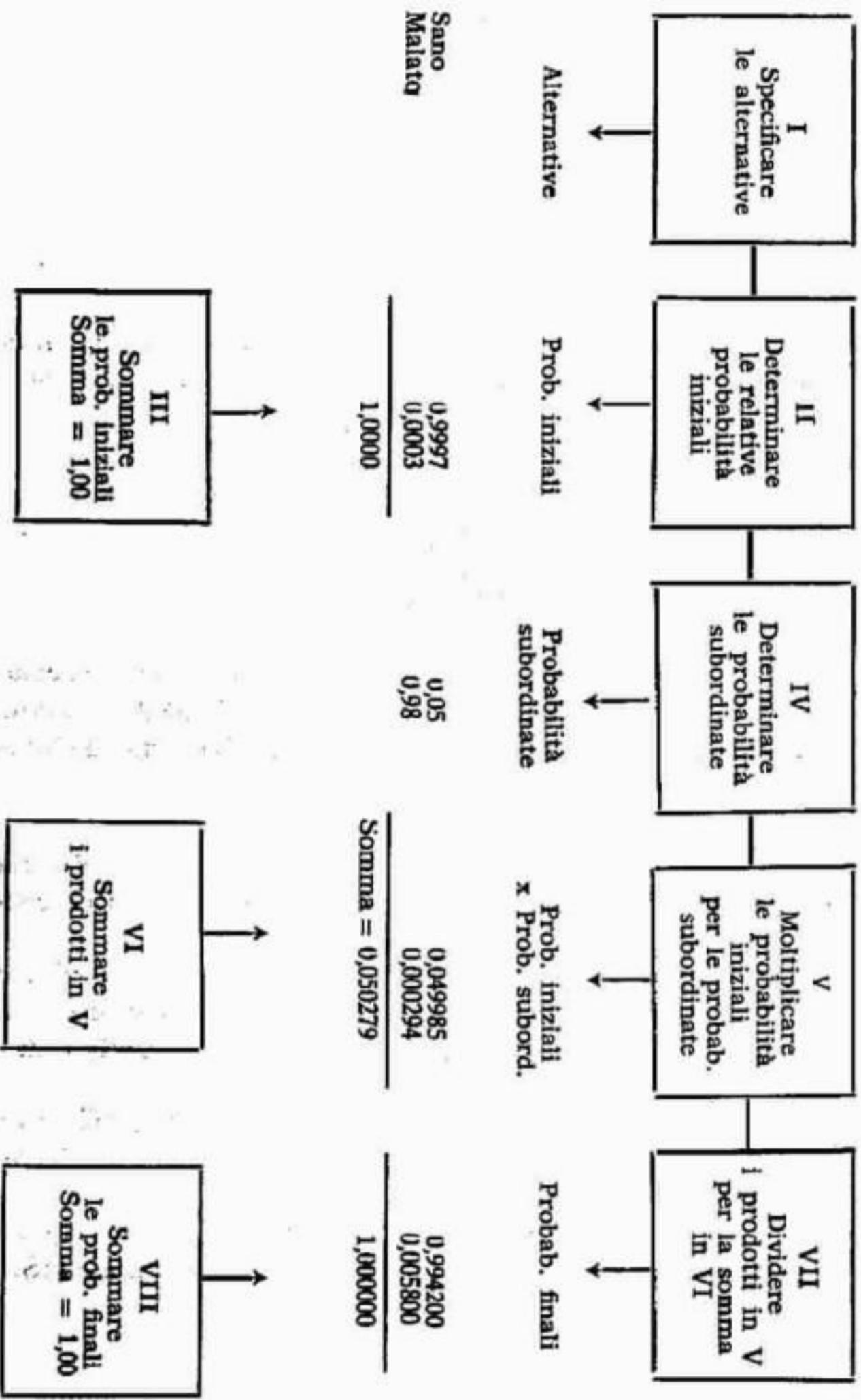


Fig. 1

per nove tipi di tumore osseo. Le probabilità subordinate sono ottenute attraverso l'esame radiologico, che permette di osservare le caratteristiche ed i sintomi di ciascun tumore.

Nella tabella i sintomi sono stati riuniti in gruppi. Il 1°, 2°, 4° gruppo sono costituiti da partizioni di sintomi e pertanto le relative probabilità subordinate danno somma 1. Il terzo e il quinto sono costituiti da sintomi che possono essere presenti contemporaneamente.

Nell'ultima sezione della tabella è rappresentata per ciascun tumore la distribuzione dei casi osservati fra vari stati di sviluppo, dallo stato iniziale (I) a quello più avanzato (V).

Utilizzando i dati della Tabella 2 è possibile l'applicazione del teorema di Bayes, riferito ad una partizione finita di ipotesi (tumori),

$$P(M_i | S_j) = \frac{P_0(M_i) P(S_j | M_i)}{P(S_j)} \quad \begin{array}{l} i = 1, 2, \dots, n \\ j = 1, 2, \dots, m \end{array} \quad (5)$$

dove  $P(S_j) = \sum_{i=1}^n P_0(M_i) P(S_j | M_i)$ .

Come è ovvio i valori molto bassi delle probabilità subordinate eliminano certi tipi di tumore. Per esempio, un tumore con grado di sviluppo pari al IV, non può essere un tumore a cellule giganti, un condroblastoma o un condroma.

Analogamente, un tumore che abbia colpito le ossa piccole e piatte non può essere un sarcoma periostale.

Può essere interessante osservare le probabilità finali per ogni tipo di tumore in un paziente che presenti determinati sintomi, determinandone una scala dal più probabile al meno probabile.

Per analisi di questo tipo con molti sintomi e una serie di pazienti il calcolatore può essere di grande aiuto.

Per concludere occorre ancora osservare, che le probabilità subordinate  $P(S_i/M)$ , possono essere ottenute con l'analisi quantitativa o invece direttamente col giudizio umano. In alcuni casi la quasi totalità dei medici si trova d'accordo sui valori delle probabilità subordinate: come ad esempio sulla probabilità che un paziente colpito da tubercolosi polmonare abbia l'emottisi.

Altre volte, invece, si hanno valutazioni anche molto di-

Tab. 2

Tipo di tumore	Prob. iniziali	Età			Sede del tumore				Dimensione		Matrice	Grado di sviluppo							
		Fino a 21 anni	21 - 30	31 o più	Ossa piccole e piatte		Ossa lunghe		P (S <sub>i</sub> /M)	P (S <sub>i</sub> /M)									
1) Tumore a cellule giganti	0,15	0,20	0,35	0,45	0,20	0,80	0,99	0,01	1,00	0,20	0,40	0,60	0,01	0,00	0,15	0,35	0,50	0,00	0,00
2) Condrolablastoma	0,05	0,75	0,20	0,05	0,20	0,80	1,00	0,50	0,75	0,00	0,90	0,10	0,00	0,30	0,50	0,35	0,15	0,00	0,00
3) Condroma	0,03	0,50	0,35	0,15	0,30	0,70	0,30	0,20	1,00	0,25	0,85	0,15	0,00	0,02	0,85	0,15	0,01	0,00	0,00
4) Condrosarcoma	0,17	0,25	0,25	0,50	0,35	0,65	0,40	0,01	0,85	0,65	0,20	0,80	0,05	0,65	0,15	0,20	0,25	0,25	0,15
5) Fibrosarcoma	0,10	0,20	0,20	0,60	0,05	0,95	0,55	0,00	0,90	0,65	0,20	0,80	0,25	0,02	0,00	0,10	0,40	0,30	0,20
6) Osteosarcoma	0,25	0,65	0,25	0,10	0,10	0,90	0,30	0,05	0,95	0,75	0,15	0,85	0,98	0,05	0,00	0,00	0,10	0,30	0,60
7) Sarcoma periostale	0,05	0,20	0,35	0,45	0,00	1,00	0,30	0,01	1,00	0,50	0,25	0,75	1,00	0,05	0,15	0,25	0,55	0,05	0,00
8) Tumore di Ewing	0,15	0,70	0,25	0,05	0,35	0,65	0,20	0,05	0,85	0,90	0,15	0,85	0,00	0,00	0,00	0,05	0,10	0,20	0,65
9) Reticolo sarcoma	0,05	0,10	0,25	0,65	0,20	0,80	0,50	0,01	0,85	0,80	0,15	0,85	0,00	0,00	0,00	0,00	0,20	0,30	0,50

verse delle  $P(S_i/M)$ .

Non si può quindi dare una regola generale di valutazione; è stato tuttavia dimostrato da Edwards et al. (4) che, in un contesto non statistico, come quello della diagnosi medica, è possibile pervenire alle  $P(S_i/M)$  basandosi esclusivamente sul giudizio di medici con notevole esperienza clinica.

### Appendice

— Nell'applicare il teorema di Bayes può essere utile effettuare i successivi passaggi logici secondo lo schema tabulare riprodotto in Figura 1. I dati che vi compaiono, illustrano il calcolo delle probabilità finali relative ad un ipotetico caso di diagnosi medica tratto dal precedente articolo dello scrivente (5).

---

(4) W. Edwards, H. Lindman, L. J. Savage: *Bayesian Statistical Inference for Psychological Research*, *Psychol. Rev.*, 70: 193-242 Maggio 1963.

(5) F.M. Mascioli: *Come funziona il ragionamento bayesiano*, *Periodico di Matematiche*, 5: 57-62, 1973.