

LA STATISTICA BAYESIANA IN MEDICINA - PARTE II: LE PRINCIPALI APPLICAZIONI E L'INFERENZA¹

C. Montomoli, M. Nichelatti

Dipartimento di Scienze Sanitarie Applicate e Psicocomportamentali, Sezione di Statistica Medica ed Epidemiologia, Università degli Studi, Pavia

Bayesian statistics in medicine – Part II: main applications and inference

Bayesian statistics is not only used when one is dealing with 2-way tables, but it can be used for inferential purposes. Using the basic concepts presented in the first part, this paper aims to give a simple overview of Bayesian methods by introducing its foundation (Bayes' theorem) and then applying this rule to a very simple practical example; whenever possible, the elementary processes at the basis of analysis are compared to those of frequentist (classical) statistical analysis. The Bayesian reasoning is naturally connected to medical activity, since it appears to be quite similar to a diagnostic process. (G Ital Nefrol 2008; 25: 422-31)

Conflict of interest: None

KEY WORDS:

Renal
osteodistrophy,
Osteomalacia,
Osteoporosis,
Transplantation

PAROLE CHIAVE:

Osteodistrofia
uremica,
Osteomalacia,
Osteoporosi,
Trapianto

✉ Indirizzo degli Autori:

Prof.ssa Cristina Montomoli
Dipartimento di Scienze Sanitarie
Applicate e Psicocomportamentali
Università di Pavia
Via Agostino Bassi, 21
27100 Pavia
e-mail: cristina.montomoli@unipv.it

INTRODUZIONE

La statistica bayesiana è fondata sull'idea che i parametri (cioè le proprietà quantitative caratteristiche) di una popolazione, come la media dell'altezza o la media della pressione sistolica di tutti gli individui, abbiano delle distribuzioni di probabilità, e che queste riflettano la nostra conoscenza "a priori", cioè preesistente alla verifica dei dati reali. In altre parole, ciascuno può stimare con discreta approssimazione l'altezza media degli Italiani, ma questo non è tutto: se vogliamo, infatti, ciascun individuo già possiede a livello più o meno inconscio questa stima, e la utilizza per verificare delle ipotesi statistiche.

Facciamo un esempio. Quando, camminando per strada, si incrocia una ragazza alta 1 metro e 90 centimetri, probabilmente non ci si fa molto caso, perché - anche sapendo che l'altezza media di una donna in Italia è nettamente inferiore - si pensa semplicemente di avere visto un *outlier*, un caso "estremo", ma pur sempre compatibile con la nostra esperienza. Se però, dopo pochi passi, si incontra un'altra ragazza molto alta, e poi subito dopo un'altra ancora, allora viene

spontaneo chiedersi a quale fenomeno si stia assistendo, e si formulano alcune ipotesi. Se - infine - voltato l'angolo, si vede l'ingresso di una palestra da dove escono altre ragazze alte, allora si deduce che in quel posto si deve allenare una squadra femminile di pallacanestro o di pallavolo, ed il tutto rientra nella "normalità". Chiunque abbia vissuto un'esperienza simile, è stato costretto inconsapevolmente ad usare la statistica bayesiana: ha osservato un evento, ha formulato delle ipotesi soggettive in base alle sue conoscenze "a priori", attribuendo una probabilità a ciascuna ipotesi, ha accresciuto i dati disponibili mediante altre osservazioni, e quindi ha giudicato quale ipotesi fosse più probabile in base ai dati accumulati. La statistica bayesiana è tutta qui.

UN APPROCCIO AL CONFRONTO TRA STATISTICA FREQUENTISTA E BAYESIANA

La statistica frequentista assume che i parametri (ad esempio la media, la deviazione standard, o altro) di una popolazione abbiano un valore determinato (si badi, non casuale), incognito, e privo di una distribuzione di probabilità: pensiamo, ad esempio, all'altezza media della popolazione maschile adulta, od alla

¹ La prima parte è stata pubblicata sul G Ital Nefrol 2008; 25: 342-6.

pressione sistolica media delle donne al nono mese di gravidanza. Su questa base, lo statistico frequentista fa delle ipotesi sul valore del parametro, quindi valuta le ipotesi con degli appositi test, oppure ottiene l'intervallo di confidenza, ad esempio al 95%: in pratica, il frequentista analizza i dati per vedere quanto questi sono congrui al modello statistico da studiare.

Nella statistica bayesiana si parte da una filosofia diametralmente opposta: un parametro è una variabile *random*, cioè casuale, e possiede una sua distribuzione di probabilità. Il parametro non è noto, ma la sua distribuzione viene stimata a priori prima di iniziare l'analisi, sulla base delle impressioni soggettive e delle conoscenze pregresse di un fenomeno. Nel caso dell'altezza media dei maschi adulti, un bayesiano potrà stimare a priori una distribuzione di probabilità normale, centrata, ad esempio, a 175 cm, e poi si occuperà di raccogliere i dati, che sono il punto di inizio del suo lavoro. Lo statistico bayesiano parte, infatti, dai dati e non dai modelli, e decide quindi qual è il modello più congruo ai dati osservati. Il bayesiano, a questo punto, ha un vantaggio sul frequentista: infatti, il suo metodo permette di calcolare un "intervallo credibile", ad esempio al 95%, del parametro, ovvero, l'intervallo dei valori entro il quale è probabile al 95% che sia contenuto il parametro; si parte con la distribuzione a priori del parametro e si ottiene la distribuzione finale, che è condizionata dai dati. L'intervallo di confidenza che calcolano i frequentisti non ha invece lo stesso vantaggio: infatti (e malgrado una interpretazione che, pure se largamente utilizzata, non è completamente corretta) l'intervallo di confidenza non è l'intervallo che contiene il parametro (l'altezza media dei maschi adulti) nel 95% dei casi, perché il parametro dei frequentisti è incognito, non osservabile, e non è una variabile *random* (il frequentista non le assegna alcuna distribuzione di probabilità). L'altezza media dei maschi adulti - dopotutto - è una sola: la vera variabile *random* è l'altezza osservata durante ciascuna misura nel campione di soggetti.

Un frequentista può affermare che l'intervallo di confidenza è l'insieme dei possibili valori del parametro incognito per i quali la differenza tra il parametro stesso e la sua stima (ad esempio, la media campionaria è una stima della media della popolazione) non è significativamente diversa da zero (ad esempio, con $p < 0.05$), ma solo un bayesiano può dire che l'intervallo credibile del parametro lo contiene con il 95% di probabilità.

Non si può dire che i metodi bayesiani siano migliori di quelli frequentisti, né viceversa: si tratta di due approcci filosoficamente diversi allo stesso problema; in molti casi, può essere migliore l'approccio frequentista, in altri quello bayesiano. Sicuramente i metodi frequentisti sono più noti e più utilizzati nella letteratu-

ra, ma questo non è dovuto al fatto che la statistica frequentista sia più adatta alle discipline biomediche. La statistica bayesiana, infatti, trova molte applicazioni in medicina, dall'analisi dei processi diagnostici, alla valutazione degli studi clinici.

Nonostante questo, lo sviluppo dei metodi bayesiani in questo settore è ancora limitato, e la grande maggioranza degli articoli pubblicati sulle riviste mediche utilizza l'approccio statistico tradizionale. È tuttavia pensabile che la disponibilità di computer sempre più potenti e di *software* appositamente progettati renda sempre più frequente l'uso dell'approccio bayesiano, peraltro già molto utilizzato in parecchi altri settori della ricerca scientifica (ad esempio, in economia). Se al momento la metodologia dominante nel campo medico è la frequentista, questo avviene non solo (o non tanto) perché questo approccio sia più immediato o più abituale, ma anche perché i maggiori produttori di *software* statistico non hanno ancora messo sul mercato dei programmi capaci di lavorare con la logica bayesiana. In parte, dei *software* specifici sono già disponibili come *open source*, ma il loro uso è ancora limitato perché lo sviluppo è lasciato alla disponibilità di tempo dei ricercatori che se ne occupano.

IL TEOREMA DI BAYES

Il punto di partenza dell'analisi bayesiana è il teorema di Bayes, che riassume in un'unica formula molte delle proprietà del calcolo probabilistico. Consideriamo due possibili eventi, A_1 e A_2 , mutuamente esclusivi, che abbiano probabilità di verificarsi non nulla, e che possano determinare un dato effetto B .

Se si è certi che solo A_1 e A_2 possano causare B , allora la probabilità $P(A_1 | B)$ che si sia verificata la causa A_1 , dato che è stato osservato l'effetto B (che nella prima parte abbiamo chiamato probabilità *backward*, o *retrograda*, perché in un certo modo, guarda indietro nel tempo) è data dalla probabilità $P(A_1)$ dell'evento A_1 , moltiplicata per la probabilità $P(B | A_1)$ che si verifichi l'evento B , a condizione di avere prima osservato il verificarsi dell'evento A_1 , e divisa per la somma dei prodotti delle due probabilità per le loro probabilità condizionali; in altre parole, possiamo scrivere l'equazione:

$$P(A_1 | B) = \frac{P(A_1)P(B|A_1)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2)} \quad (1)$$

che ci dice quale sia la probabilità che - una volta osservato l'evento B - questo sia stato causato dal precedente evento A_1 (e solo da quello). L'equazione (1) rappresenta il *teorema di Bayes*, ed è facilmente generalizza-

bile ad un numero arbitrario n di possibili cause.

Il teorema di Bayes è il fondamento dell'approccio bayesiano all'inferenza statistica. Le alternative A_1 e A_2 rappresentano l'insieme di tutte le possibili cause di un certo effetto B , e - una volta fissate le probabilità a priori di ogni singola causa - la formula di Bayes, consente di assegnare a queste un valore a posteriori, dato il verificarsi dell'effetto. Le probabilità $P(A_1)$ e $P(A_2)$ diventano allora le conoscenze *a priori*, ossia quelle che si hanno prima di effettuare l'esperimento, ovvero prima di acquisire i dati. Le probabilità $P(B|A_1)$ e $P(B|A_2)$, invece, sono l'informazione su B che si acquisisce eseguendo l'esperimento, osservando cioè i dati, e sono dette anche *verosimiglianze*.

L'equazione (1) è quindi un meccanismo di apprendimento dall'esperienza, che associa $P(B|A_1)$ e $P(B|A_2)$, cioè le conoscenze *a priori* (oppure soggettive) sugli eventi A_1 e A_2 alle informazioni $P(B|A_1)$ e $P(B|A_2)$, per cui $P(A_1|\hat{B})$ è la quantificazione del miglioramento delle nostre conoscenze sul fenomeno, miglioramento che otteniamo dopo avere effettuato un esperimento.

TEST DI VERIFICA

1) Come si può definire la statistica bayesiana?

- La statistica bayesiana è un tipo di analisi dei dati basata sulla probabilità soggettiva di un evento
- La statistica bayesiana è una metodologia di analisi alternativa che si utilizza nei casi in cui la statistica frequentista non sia adatta
- La statistica bayesiana è un tipo di analisi dei dati che combina le conoscenze pregresse (anche soggettive) su un dato evento con quelle ottenute dai dati disponibili
- La statistica bayesiana associa le conoscenze soggettive relative ad un determinato fenomeno con le conoscenze derivate dall'analisi frequentista per ottenere la significatività dei dati.

2) Qual è il fondamento dell'analisi bayesiana?

- L'analisi bayesiana è basata sul teorema di Bayes, che consente di calcolare la probabilità di una certa causa, dato un determinato effetto, per mezzo della probabilità a priori e della verosimiglianza
- L'analisi bayesiana si fonda sul teorema di Bayes, che calcola la significatività statistica (*p-value*) di un evento relativamente alla significatività di tutti gli altri eventi possibili
- La statistica bayesiana è basata sulla probabilità soggettiva (*a priori*) di un evento, per cui la significatività è influenzata dalla scelta del

ricercatore: per alcuni di essi l'evento sarà significativo, mentre per altri no

- La statistica bayesiana si fonda sull'associazione dei dati ottenuti in una certa indagine con i valori della significatività ottenuta in tutte le indagini precedenti.

3) Quali possono essere alcune delle principali differenze tra analisi frequentista (tradizionale) e bayesiana?

- La statistica frequentista calcola la probabilità di un evento valutandola come la frequenza con cui questo si è verificato, mentre la statistica bayesiana calcola la probabilità sulla base delle conoscenze del ricercatore circa il fenomeno studiato; in entrambi i casi si ottengono gli intervalli di confidenza entro cui sono contenuti i valori del parametro cercato
- La statistica frequentista calcola la probabilità dei dati una volta fissato il modello statistico, ottenendo la significatività (*p-value*) come la probabilità di avere dati più estremi di quelli effettivamente osservati, mentre la statistica bayesiana calcola la probabilità dei vari modelli una volta ottenuti i dati sperimentali; i dati più estremi non hanno alcuna rilevanza sui modelli
- La statistica frequentista si occupa di ottenere la significatività di un fenomeno statistico per mezzo dell'inferenza, ottenendo l'intervallo di confidenza, che contiene il valore vero del parametro cercato, mentre la statistica bayesiana è utilizzata l'inferenza per calcolare il valore vero di un parametro, una volta fissato l'intervallo di confidenza
- La statistica frequentista analizza i dati sperimentali basandosi sui modelli statistici, ed ottiene la significatività rispetto ai modelli più estremi, mentre la statistica bayesiana analizza in vari modelli a priori ed ottiene la significatività statistica rispetto al modello fissato a priori.

La risposta corretta alle domande sarà disponibile sul sito internet www.sin-italy.org/gin e in questo numero del giornale cartaceo dopo il Notiziario SIN

COME FUNZIONA IL METODO BAYESIANO: IL CONFRONTO TRA MODELLI

Chiamiamo modello l'insieme delle informazioni utili per descrivere un fenomeno non noto. Esempi possibili di modello sono la proporzione di cittadini votanti durante le elezioni, lo stipendio medio mensile in una popolazione lavorativa, oppure la quantità di pazienti che guariscono da una certa malattia dopo l'assunzione di un dato farmaco. In generale, il modello migliore viene scelto dopo l'osservazione dei dati. Per quanto riguarda lo stato di salute rispetto ad una data malattia, i modelli possibili sono solo due: "il soggetto è malato" oppure "il soggetto non è malato".

La statistica bayesiana è molto utile per effettuare un confronto tra modelli. Per fare un esempio, ipotizziamo che in una popolazione si manifesti una malattia stagionale che in media contagia il 40% della popolazione; assumiamo inoltre che, tra i soggetti a rischio, i malati siano indistinguibili dai sani, e che l'unico sistema per rilevare la malattia sia un test diagnostico che nei soggetti malati risulta positivo nel 60% dei casi (quindi la *sensibilità* del test è del 60%), mentre nei soggetti sani risulta negativo nel 90% dei casi (quindi la *specificità* del test è del 90%). La domanda che ci si deve porre è: *se un paziente è positivo al test diagnostico, qual è la probabilità che sia realmente malato?*

La situazione iniziale è riassunta dalle Tabelle I e II.

Dato che i soggetti sani non si distinguono dai malati, scegliendo a caso un soggetto a rischio, la probabilità che questo sia realmente malato è del 40%. Il calcolo può essere effettuato con la formula della probabilità condizionale (vista nella prima parte), ovvero

che nel nostro caso si può scrivere

dove il termine $Pr(\text{malato e positivo})$ indica la probabilità che un soggetto sia simultaneamente malato e positivo al test, mentre $Pr(\text{malato} \mid \text{positivo})$ indica la probabilità di avere un paziente malato, una volta accertato che tale paziente sia risultato positivo al test.

Si osservi che la probabilità condizionale dà la sensazione di una sequenza temporale, per cui la positività al test sembra venire prima della diagnosi; in realtà, non è proprio questo ciò che succede: la probabili-

TABELLA I - SITUAZIONE INIZIALE ("A PRIORI") RELATIVA ALLA PRESENZA DELLA MALATTIA NELLA POPOLAZIONE E AGLI ESITI DEL TEST

Popolazione		Risultato del test	
Malati	40%	positivi	60%
		negativi	40%
		<i>totale</i>	<i>100%</i>
Sani	60%	positivi	10%
		negativi	90%
		<i>totale</i>	<i>100%</i>

tà condizionata si limita ad analizzare gli eventi che si verificano solo nei soggetti dove la condizione è verificata.

Facendo i calcoli con i valori in nostro possesso otteniamo la probabilità che il soggetto positivo sia effettivamente malato:

che è quindi pari all'80%.

TABELLA II - TAVOLA DI CONTINGENZA CHE MOSTRA I RISULTATI ATTESI SULLA BASE DELLE CONDIZIONI A PRIORI ILLUSTRATE NELLA TABELLA I

	Risultato del test		Totale
	Positivi	Negativi	
Malati	24%	16%	40%
Sani	6%	54%	60%
<i>Totale</i>	<i>30%</i>	<i>70%</i>	<i>100%</i>

Abbiamo ora modificato in modo sensibile la nostra conoscenza del sistema; in pratica, la nostra conoscenza iniziale è stata migliorata dall'aumento delle informazioni disponibili. Ripetendo il test una seconda volta sullo stesso soggetto, non partiamo più dalla probabilità iniziale (che avevamo *a priori*), ma da quanto illustrato nelle Tabelle III e IV.

TABELLA III - SITUAZIONE DI PARTENZA PRIMA DI FARE IL SECONDO TEST DIAGNOSTICO

Positivi al primo test		Risultato del secondo test	
Malati	80%	positivi	60%
		negativi	40%
Sani	20%	positivi	10%
		negativi	90%

Dai calcoli si ottiene

per cui la probabilità che il soggetto sia effettivamente malato dopo due test positivi è del 96%: la nostra conoscenza del sistema è ancora migliorata rispetto alla situazione iniziale. Il valore ottenuto viene interpretato come una probabilità; in pratica, chiamando F la probabilità che le indagini diagnostiche siano positive, otteniamo due modelli, che potremmo chiamare $F_{0,6}$ e $F_{0,1}$, dai valori della frequenza dei test positivi nei malati e nei sani.

Questi modelli sono assimilabili alle ipotesi che vengono studiate con i test di significatività della statistica frequentista. Dopo le due indagini diagnostiche, possiamo concludere che c'è una probabilità del 96% che il modello $F_{0,6}$ sia quello corretto.

Lo svolgimento simultaneo o in tempi differenti dei due test diagnostici non altera la probabilità ottenuta. Infatti, la *verosimiglianza* di due test diagnostici positivi, assumendo rispettivamente che sia vero il modello $F_{0,6}$ o che sia vero il modello $F_{0,1}$ è data da:

$$(2) \qquad (2)$$

dove si sono utilizzate le proprietà della distribuzione binomiale (quella che interessa il sistema che stiamo studiando).

Per ciascuno dei modelli possiamo associare le probabilità a priori alla verosimiglianza calcolata per ciascun *outcome*: i dati ottenuti sono forniti nella Tabella V.

Con i dati possiamo ora calcolare la probabilità

TABELLA IV - TAVOLA DI CONTINGENZA CON I RISULTATI DEL SECONDO TEST DIAGNOSTICO EFFETTUATO SOLAMENTE TRA I SOGGETTI POSITIVI AL PRIMO TEST

	Risultato del test		Totale
	Positivi	Negativi	
Malati	48%	32%	80%
Sani	2%	18%	20%
Totale	50%	50%	100%

di essere effettivamente ammalati dopo due test positivi. Per farlo, occorre ottenere una cosa che chiamiamo "verosimiglianza relativa" di ciascun modello (Tab. V) e dividerla per la somma delle verosimiglianze relative di tutti i modelli. Nel nostro caso abbiamo:

il cui risultato è identico a quello ottenuto quando i due test sono stati effettuati in sequenza, e conferma che la probabilità che il modello $F_{0,6}$ sia valido è del 96%.

Quando si esegue un'inferenza statistica frequentista, i test di significatività si fanno assumendo che una certa ipotesi (l'ipotesi nulla) sia vera; da questa viene calcolata la probabilità di osservare dei valori più estremi di quelli effettivamente trovati (cioè si calcola il *p-value*). La statistica frequentista, di conseguenza, studia la probabilità dei dati osservati, una volta stabilito il modello di riferimento (l'ipotesi nulla): quindi, la probabilità frequentista è identificabile con la formula $\Pr(\text{dati} \mid \text{modello})$. La statistica bayesiana, invece, studia la probabilità che un determinato modello sia valido, una volta ottenuti i dati: la probabilità bayesiana è allora identificata con la formula $\Pr(\text{modello} \mid \text{dati})$.

Riassumiamo il percorso logico effettuato sinora. Siamo partiti da una probabilità iniziale del 40% (la probabilità di scegliere un malato estraendolo a caso dai soggetti a rischio): questa è già stata definita *probabilità a priori*, perché è stata ricavata da un'informazione pregressa e perché è la probabilità che il nostro modello sia valido *prima di osservare i dati*. Questa probabilità iniziale è stata sostituita successivamente con una *probabilità a posteriori* (pari al 96%), ottenuta dopo avere osservato i dati ed avere fatto gli opportuni calcoli.

Il calcolo della probabilità a posteriori a partire da quella a priori è stato fatto per mezzo della *verosimiglianza*, che connette la probabilità a priori con il

TABELLA V - PERCENTUALI DEGLI *OUTCOME* DEI DUE TEST DIAGNOSTICI, BASATE SULLA PROBABILITÀ A PRIORI STABILITE PRIMA DELLA RACCOLTA DEI DATI

Probabilità a priori		Verosimiglianza del risultato dei due test		Verosimiglianza "relativa"	
Malati	40%	2 test positivi	36%	14.4%	
		primo positivo, secondo negativo	24%	9.6%	
		primo negativo, secondo positivo	24%	9.6%	
		2 test negativi	16%	6.4%	
		<i>Totale nei soggetti malati</i>	<i>100%</i>	<i>40%</i>	
Sani	60%	2 test positivi	1%	0.6%	
		primo positivo, secondo negativo	9%	5.4%	
		primo negativo, secondo positivo	9%	5.4%	
		2 test negativi	81%	48.6%	
		<i>Totale nei soggetti sani</i>	<i>100%</i>	<i>60%</i>	
<i>Totale</i>	<i>100%</i>			<i>100%</i>	

Nella penultima colonna a destra sono fornite le verosimiglianze di ciascun outcome: si osservi come i valori riportati nel caso dei due test positivi (nei malati e nei sani) siano identici a quelli ricavati nelle equazioni (2) utilizzando il calcolo combinatorio. L'ultima colonna contiene dei valori che chiamiamo verosimiglianze relative, ottenute moltiplicando la verosimiglianza per la probabilità a priori

modello ed i dati per mezzo del teorema di Bayes espresso nell'equazione (1), cioè, individuati i modelli $F_{0,6}$ e $F_{0,1}$:

Le verosimiglianze $V(F_{0,6})$ e $V(F_{0,1})$ erano state calcolate nelle equazioni (2) per entrambi i modelli considerati. I risultati finali sono riassunti nella Tabella VI.

È da rimarcare che i calcoli effettuati secondo la logica bayesiana coinvolgono solo i valori osservati,

TABELLA VI - TAVOLA SINOTTICA DELLE OPERAZIONI NECESSARIE PER CALCOLARE LE PROBABILITÀ A POSTERIORI DEI MODELLI ANALIZZATI

	Modello	
	$F_{0,6}$	$F_{0,1}$
Probabilità a priori	0.4	0.6
Verosimiglianza	0.36	0.01
Probabilità a priori × verosimiglianza	$0.4 \times 0.36 = 0.144$	$0.6 \times 0.01 = 0.006$
Probabilità a posteriori	$\frac{0.144}{0.144 + 0.006} = 0.96$	$\frac{0.006}{0.144 + 0.006} = 0.04$

I modelli da confrontare in questo caso sono solamente due, ma il processo è generalizzabile ad un numero arbitrario di modelli. Si nota che le probabilità a posteriori sono molto differenti da quelle a priori, per cui, se i dati disponibili sono in numero ragionevolmente sufficiente, il rischio che la probabilità a priori influenzi in modo determinante il risultato finale è molto limitato.

mentre le informazioni sugli eventi non osservati non sono utilizzate (ad esempio, nella statistica bayesiana i valori "più estremi" di quelli osservati non hanno importanza). Nella statistica bayesiana la probabilità che un modello sia vero è determinata solo da quanto visto sperimentalmente, mentre tutte le altre informazioni, non hanno alcun rilievo. La struttura del ragionamento bayesiano, inoltre, è tale che l'accumularsi delle osservazioni sperimentali riduce in modo progressivo l'effetto esercitato dalla probabilità a priori sull'inferenza statistica, che quindi è sempre meno influenzata dai convincimenti soggettivi. In pratica, due differenti ricercatori potranno attribuire differenti probabilità a priori allo stesso modello, ma se l'analisi verrà condotta con un numero di dati sufficientemente grande, allora le probabilità a posteriori dei due ricercatori dovranno necessariamente convergere.

IL PROCESSO DIAGNOSTICO È BAYESIANO

Il ragionamento bayesiano è del tutto simile ad un processo diagnostico. Il medico che visita un paziente con una malattia ignota stabilisce una serie di probabilità a priori per ogni possibile diagnosi; la probabilità a priori sarà calcolata sulla base di informazioni proprie ed ambientali, così che il medico potrà costruire un insieme di modelli mutuamente esclusivi (con probabilità complessiva del 100%, ovviamente), ciascuno

dei quali avrà una certa probabilità di essere vero; la somma delle probabilità di tutti i modelli dovrà essere ovviamente pari a 1 (il 100%). La probabilità a priori dei vari modelli sarà sia soggettiva (due medici potranno attribuire dei pesi differenti allo stesso sintomo), sia obiettiva, perché terrà in debito conto l'anamnesi, ed escluderà i modelli non congrui al quadro patologico. Il medico ha quindi selezionato un insieme di modelli più o meno probabili a priori, tra i quali si trova anche il modello vero.

La fase successiva del processo è la valutazione della verosimiglianza: il medico prescrive o esegue una serie di manovre o di accertamenti strumentali o di laboratorio, atti a generare i *dati* con i quali si potranno preferire alcuni modelli ed escluderne altri. In buona sostanza, la positività o la negatività del referto di ogni esame (può essere una semplice sfigmomanometria come una tomografia ad emissione di positroni, non importa) serve a generare la verosimiglianza del modello cui l'esame fa riferimento. Con le varie verosimiglianze, il medico effettua (inconsapevolmente) dei calcoli che gli consentono di ottenere le probabilità a posteriori (oggettive e suffragate dai dati) di ciascun modello ipotizzato, e quindi il modello più "verosimilmente" vero.

Come visto precedentemente, potrebbe sussistere il problema del differente peso (probabilità a priori) dato da due medici allo stesso sintomo: è quindi possibile che le diagnosi iniziali possano essere differenti, ma nella sostanza, se gli accertamenti sono completi e sufficientemente accurati (se le informazioni disponibili sul paziente aumentano), ambedue i medici dovranno necessariamente convergere verso la stessa diagnosi, indipendentemente dalla loro opinione iniziale. L'approccio bayesiano è quindi particolarmente importante nelle diagnosi differenziali, in special modo per le malattie rare: ogni medico - anche se non lo sa - fa della statistica bayesiana.

TEST DI VERIFICA

4) Quale di queste affermazioni è vera?

- Nella statistica frequentista i parametri di una popolazione hanno valori noti, mentre non è nota la loro distribuzione di probabilità, mentre nella statistica bayesiana sono conosciuti sia i valori dei parametri, sia la loro distribuzione
- Nella statistica frequentista i parametri di una popolazione hanno dei valori determinati ma incogniti, mentre nella statistica bayesiana i parametri sono delle variabili random

- Nella statistica frequentista i parametri della popolazione sono in genere caratterizzati da distribuzione normale e media e varianza note, mentre nella statistica bayesiana la distribuzione dei parametri è determinata solo dai valori degli stessi parametri calcolati nei campioni studiati precedentemente
- Sia la statistica frequentista, sia la statistica bayesiana calcolano i parametri basandosi sull'ipotesi che questi abbiano valori determinati e conosciuti; però l'analisi frequentista assume che le distribuzioni siano note, mentre in quella bayesiana le distribuzioni non hanno importanza.

5) I risultati di un'analisi statistica ottenuti con metodo frequentista e con metodo bayesiano possono coincidere?

- I risultati possono essere praticamente identici se la stima del parametro fatta col metodo frequentista ha dato un valore simile a quella ottenuta con il metodo bayesiano
- I risultati possono essere praticamente identici se i dati analizzati sono in numero sufficientemente grande
- I risultati possono essere simili se, e solo se, i valori della probabilità a priori utilizzati nell'analisi bayesiana sono simili alle probabilità calcolate nell'analisi frequentista
- I risultati possono essere praticamente identici se la probabilità a priori dell'analisi bayesiana ha un valore molto vicino a quello della probabilità a posteriori, ma questo si verifica solamente se la verosimiglianza della probabilità a priori è molto elevata.

6) Ci sono motivi per preferire l'analisi frequentista a quella bayesiana o viceversa?

- La statistica bayesiana è solo uno strumento in più per analizzare un fenomeno, e non vi sono vantaggi decisivi che facciano preferire, in generale, l'una o l'altra tecnica; al momento, infatti, l'unica differenza pratica è la limitata disponibilità o la scarsa flessibilità dei *software* "bayesiani", mentre quelli frequentisti sono disponibili sul mercato già da molto tempo: di fatto, la statistica bayesiana ha enormi potenzialità di sviluppo nella medicina
- La statistica frequentista è più obiettiva perché parte dalle osservazioni dei fatti concreti (frequenze degli eventi analizzati), mentre la statistica bayesiana è fondata sulle conoscenze personali o sulle esperienze o le informazioni del ricercatore, per cui è bene che usino i metodi bayesiani solo coloro che hanno una

conoscenza diretta del fenomeno da studiare; in tutti gli altri casi, i metodi frequentisti sono largamente preferibili, in quanto gestibili anche da un ricercatore che non li conosce in modo diretto

- c. La statistica frequentista è sempre preferibile quando il campione da analizzare è molto grande, per cui l'intervallo di confidenza sul parametro si riduce, ed il suo valore vero viene ottenuto facilmente: in tal modo non sussistono più i motivi per utilizzare la probabilità a priori tipica dei metodi bayesiani
- d. Malgrado le due metodiche siano ugualmente giustificabili dal punto di vista scientifico, la statistica frequentista è talvolta preferibile nella pratica, perché fornisce la probabilità dei dati, che sono quanto "concretamente" raccolto nell'indagine sperimentale.

RIASSUNTO

L'analisi statistica dei dati biomedici è in grandissima parte effettuata utilizzando le tecniche frequentiste della statistica "classica". I metodi bayesiani sono usati in misura molto minore, anche se hanno mostrato una grandissima utilità in molte discipline scientifiche, inclusa la medicina. La differente utilizzazione dei modelli è dovuta in parte all'abitudine, in parte alla disponibilità ancora limitata di pacchetti informatici capaci di gestire l'inferenza statistica con i modelli bayesiani; poco alla volta, questo divario sta colmandosi, ma occorreranno alcuni anni prima che la statistica bayesiana possa mostrare tutte le sue potenzialità.

Non esistono motivi particolari per ritenere che un approccio sia migliore di un altro; è però vero che in alcuni aspetti della ricerca medica il processo analitico bayesiano sembra molto più "naturale" di quello frequentista.

I frequentisti interpretano la probabilità del verificarsi di un evento come la sua frequenza, per cui "vedono" la probabilità nel mondo esterno e calcolano l'inferenza sulla probabilità dei dati, una volta noto il modello. I bayesiani, invece, interpretano la probabilità come la misura dell'incertezza di un evento in base alle loro conoscenze, informazioni ed esperienze, e calcolano l'inferenza come la probabilità del modello, una volta noti i dati.

Queste differenze concettuali si riverberano sulla misura della significatività: i frequentisti calcolano il p-value sulla base dell'ipotesi nulla, come la probabilità di avere dei dati più estremi di quelli osservati e con un algoritmo di calcolo opportuno ottengono un intervallo di confidenza al 95%, che è una realizzazione di un processo che ha il 95% di probabilità di generare un insieme che con-

tiene il parametro; i bayesiani, invece calcolano l'intervallo credibile, cioè l'insieme dei valori che effettivamente contiene il parametro con il 95% di probabilità.

Fermo restando che la probabilità bayesiana è una opportunità in più per qualsiasi ricercatore. I confronti effettuati hanno voluto mostrare che la statistica bayesiana non è estranea al ragionamento diagnostico, ma - anzi - ne è un'appendice naturale, e che la sua dipendenza dalle conoscenze soggettive ha un effetto molto limitato sui risultati finali quando i dati sono ragionevolmente numerosi. Va aggiunto che le probabilità a priori vanno esplicitate: un bayesiano deve dichiarare i valori utilizzati, e deve anche documentare i motivi alla base della loro scelta; nell'esempio di analisi bayesiana svolto in precedenza, si sono utilizzate come probabilità a priori le frequenze della malattia, quindi si è usata la probabilità frequentista per fare un'inferenza bayesiana.

APPENDICE: ALCUNE RIFLESSIONI SUL P-VALUE

Nella statistica frequentista, il valore di p (p -value) che si ottiene dopo un test statistico (come un t di Student o un χ^2), informa sulla probabilità che l'evento osservato sia dovuto al caso, ovvero sulla probabilità di osservare valori più estremi di quelli effettivamente osservati. In modo del tutto arbitrario, ma universalmente accettato, si è stabilita da molti anni la convenzione che $p < 0.05$ caratterizzi un test significativo. In altre parole, se, ad esempio, l'escrezione tubulare di bicarbonato è significativamente minore in un gruppo di pazienti rispetto ad un altro, allora non è verosimile che il caso abbia selezionato in un gruppo i pazienti con escrezione minore e nell'altro gruppo quelli ad escrezione normale; è invece molto più probabile (anche se non certo in modo assoluto) che in uno dei due gruppi l'escrezione sia minore a causa di un fattore specifico, ad esempio per un trattamento con acetazolamide. Il valore di p è dunque lo strumento che consente di riconoscere la presenza o l'assenza di un effetto dovuto al fattore rispetto al quale è stata fatta l'ipotesi.

Il valore di p ha però dei limiti: il primo è la sua arbitrarietà: in molti casi un effetto potrebbe essere clinicamente significativo senza esserlo statisticamente, oppure potrebbe accadere il contrario, e si potrebbe avere un effetto che è significativo statisticamente, ma non clinicamente. In altre parole, ci sono situazioni in cui anche un $p < 0.10$ potrebbe essere percepito come significativo: pensiamo al caso in cui una nuova tecnica terapeutica riduca con tale significatività la mortalità dei pazienti.

Un secondo possibile limite di p può essere verificato con un semplice esempio. Ipotizziamo di effettuare un esperimento binario, cioè un esperimento con due

soli possibili *outcome*, come il lancio di una moneta, le cui uscite siano *testa* (*T*) o *croce* (*C*). Dopo un certo numero di lanci, il numero di volte n_T in cui è uscito *testa* dovrà essere più o meno simile al numero n_C di volte in cui è si è ottenuta *croce*; molto difficilmente, però, i numeri saranno uguali: ma allora come sarà possibile riconoscere se la moneta è truccata? In altre parole, se consideriamo *T* il risultato favorevole e *C* quello sfavorevole, fissato un numero *N* di lanci complessivi della moneta, qual è il valore soglia di n_T al di sotto del quale (o quello di n_C al di sopra del quale) potremo dire che la moneta è truccata? La domanda può essere posta anche in un altro modo: dato che in una moneta non truccata la probabilità di uscita di *testa* e quella di *croce* sono uguali (50%, quindi 0.5), quali valori di n_T e n_C , fissato il numero di lanci, danno valori significativamente differenti da 0.5? La risposta viene dalla statistica. La probabilità di ottenere n_T volte *testa* e n_C volte *croce* in *N* lanci di moneta (in cui $n_C = N - n_T$) è data da

dove *P* è la probabilità di ottenere *testa* e $1 - P$ quella di ottenere *croce*: se la moneta non è truccata, allora $P = 1 - P = 0.5$. Ad esempio, la probabilità di otte-

nere 3 volte *testa* e 7 volte *croce* in 10 lanci è:

quindi di poco inferiore al 12%. Se invece volessimo calcolare la probabilità, in *N* lanci, di ottenere *testa* da 0 a *R* volte (con $R \leq N$), allora dovremmo usare la formula

Utilizzando quest'ultima equazione, possiamo calcolare quali sono i valori critici delle uscite dell'evento *testa*, dato il numero di lanci *N*, al di sotto dei quali potremo affermare (con significatività $p < 0.05$) che la moneta è truccata (Tab. VII).

Dalla Tabella vediamo che la significatività al 5% dipende dalla grandezza del campione (dal numero di lanci), e sembra più idonea a individuare la falsità

TABELLA VII - RISULTATI DELLA SIMULAZIONE DEL LANCIO DI UNA MONETA, RIPETUTO DA 10 FINO AD UN MILIARDO DI VOLTE (T = TESTA, C = CROCE).

Numero di lanci della moneta	La moneta è truccata se n_T è minore di:	Percentuale minima di lanci in cui si deve ottenere <i>croce</i> per ritenere la moneta truccata
10	2	90%
100	42	59%
1,000	474	52.7%
10,000	4,918	50.83%
100,000	49,740	50.261%
1,000,000	499,178	50.082%
10,000,000	4,997,399	50.026%
100,000,000	49,991,776	50.008%
1,000,000,000	499,973,995	50.002%

Per ciascun esperimento (serie di lanci), la tabella riporta il "valore soglia" del numero di volte in cui deve uscire *testa* (il risultato che consideriamo favorevole) per poter ritenere truccata la moneta. L'ultima colonna (facilmente ottenibile a partire dalla seconda) mostra invece quale sia la percentuale minima di uscite del risultato *croce* (quello considerato sfavorevole) perché si possa significativamente ritenere ($p < 0.05$) che la moneta sia truccata. Si osserva che, all'aumentare del numero di lanci, il valore della percentuale critica decresce molto velocemente, fino ad arrivare molto vicino al 50%

dell'ipotesi nulla con campioni piccoli.

Infatti, almeno dal punto di vista "psicologico", appare difficile credere che una moneta sia truccata quando (dopo un miliardo di lanci) il risultato fosse sfavorevole solo nel 50.002% dei casi, mentre, dopo soli 10 lanci, sarebbe molto più immediato ritenere il

gioco truccato osservando un risultato sfavorevole nel 90% dei casi.

DICHIARAZIONE DI CONFLITTO DI INTERESSI

Gli Autori dichiarano di non avere conflitto di interessi.

BIBLIOGRAFIA

1. Albert JH, Rossman AJ. Workshop statistics: discovery with data, a Bayesian approach. Key College Publishing, Emeryville, USA: 2001.
2. Berry DA. Statistics: a Bayesian perspective. Duxbury Press - Wadsworth Publishing, Belmont, USA: 1996.