

Valutazione dell' affidabilità dei test diagnostici legati al Covid 19 (Metodologia)

A cura di
Francesco Fabi



ce3s
CENTRO STUDI
STATISTICI
E SOCIALI

Basato su uno scritto di Fabrizio Fabi

Definizioni

- Sia “M ” l’evento “il paziente è malato” e “ #M ” l’evento complementare “il paziente è sano”.
- Sia poi con “ A ” l’evento “il test risulta positivo” e con “ #A ” l’evento complementare “il test risulta negativo”.
- Sia ancora $P(M)$ la probabilità di M valutata, per un certo paziente, prima di conoscere l’esito del test detta “probabilità a priori ” dell’evento M.
- Si indichi ancora con $P(A)$ la probabilità che il test dia risposta positiva, valutata prima di conoscere il risultato relativo allo specifico paziente. E’ anche essa una “ probabilità a priori ” e dell’evento A.
- E’ per definizione $P(\#M)=1-P(M)$ e $P(\#A)=1-P(A)$

Probabilità condizionate

- Per ognuno dei 4 eventi fondamentali vanno poi considerate le “probabilità condizionate”, ovvero la probabilità che si può attribuire ad uno di essi quando sia noto l’esito dell’altro; ad esempio la probabilità di A quando sia noto M, che si scrive $P(A|M)$.
- E’ privo di interesse considerare probabilità come $P(M|M)$ e $P(A|A)$, che ovviamente valgono 1; e sono banali anche $P(M|\#M)$ e $P(A|\#A)$, che ovviamente valgono 0.
- In definitiva, sono 8 le probabilità condizionate di interesse, ovvero:

$P(M|A)$, $P(M|\#A)$, $P(\#M|A)$, $P(\#M|\#A)$, $P(A|M)$, $P(A|\#M)$, $P(\#A|M)$, $P(\#A|\#M)$

Relazioni fra gli 8 valori delle probabilità condizionate

- Non sono indipendenti. Ognuno di essi è evidentemente legato ad un altro dalla complementarità a 1.
- Primo gruppo di relazioni: $P(M|A)=1-P(\#M|A)$ e $P(M|\#A)=1-P(\#M|\#A)$
 - *Le probabilità del primo gruppo, “probabilità a posteriori”, costituiscono le incognite del nostro problema.
- Secondo gruppo di relazioni: $P(A|M)=1-P(\#A|M)$ e $P(A|\#M)=1-P(\#A|\#M)$
 - *Le probabilità condizionate elencate nel secondo, vanno sotto il nome di “verosimiglianze” e possono essere considerate note.

Sensibilità e Specificità

- La verosimiglianza $P(A|M)$ è comunemente chiamata “sensibilità” del test;
- La verosimiglianza $P(\#A|\#M)$ è comunemente chiamata “specificità” del test.

*Si deve rimarcare, perché non è intuitivo e spesso è trascurato, che le probabilità a posteriori, ad esempio $P(M|A)$, non coincidono con le verosimiglianze.

Proprietà delle “Probabilità condizionate”

- $P(A) = P(A|M) P(M) + P(A|\#M) P(\#M)$ e $P(\#A) = P(\#A|M) P(M) + P(\#A|\#M) P(\#M)$

*Quindi, le probabilità a priori degli eventi A e #A discendono dalla sensibilità e specificità, che sono note (dalle sperimentazioni sui test che precedono la fase di applicazione sul campo).

- $P(M) = P(M|A)P(A) + P(M|\#A)P(\#A)$ e $P(\#M) = P(\#M|A)P(A) + P(\#M|\#A)P(\#A)$

*Tuttavia queste ultime formule sono di scarsa utilità, poiché, come detto, le “probabilità a posteriori” che si trovano nei secondi termini sono proprio le incognite del problema.

Considerazioni su $P(M)$ e $P(\#M)$

- Nel nostro caso, come per lo più avviene, le probabilità a priori $P(M)$ e $P(\#M)$ sono note con minor precisione delle verosimiglianze.
- In mancanza di informazioni particolari sul paziente che si sta per testare, la probabilità a priori di malattia $P(M)$ si può far coincidere con la frequenza ipotizzabile di casi complessivi nella popolazione in esame.
- Il range, o il valore puntuale per la $P(M)$, o potrebbe cambiare nel momento in cui si apprendesse che il paziente presenta sintomi tipici del Covid-19, oppure che il paziente ha avuto contatti con persone contagiate, o ha visitato una zona notoriamente infetta, ecc.
- Questo implica che sarà opportuno eseguire i calcoli bayesiani per un range di valori delle probabilità a priori e osservare le differenze nei risultati.

La formula di Bayes per $P(M|A)$

- La valutazione delle probabilità a posteriori, a partire da quelle a priori e delle verosimiglianze, è possibile facendo ricorso alla “formula di Bayes”, fondamentale nel calcolo delle probabilità.
- $P(M|A) = P(A|M) * P(M) / P(A) = P(A|M) * P(M) / ((P(A|M)P(M) + P(A|\#M)P(\#M)))$
- Come si vede, questa formula calcola la probabilità che il paziente sia malato quando si sia appreso che il test ha dato risposta positiva, utilizzando anche le probabilità a priori che sia malato. E se ne deduce subito anche la $P(\#M|A) = 1 - P(M|A)$

La formula di Bayes per $P(\#M | \#A)$

- $P(\#M | \#A) = P(\#A | \#M) * P(\#M) / P(\#A) = P(\#A | \#M) * P(\#M) / ((P(\#A | M)P(M) + P(\#A | \#M)P(\#M))$

*Questa formula dà la probabilità che il paziente sia sano quando si sia appreso che il test ha dato risposta negativa. Se ne deduce anche la $P(M | \#A) = 1 - P(\#M | \#A)$

*Le formule precedenti, come ben si vede, sono calcolabili perché tutti gli elementi a secondo membro sono noti.

* Se ci si riferisce a una situazione come quella media italiana dopo un anno di pandemia, un range di valori da 1% a 10% sarebbe già piuttosto ampio.