

28 Sensibilità, specificità, predittività, ROC

La questione fondamentale affrontata in questa Lezione è

$$P(\text{veramente malato} \mid \text{sapendo test positivo})$$

nel senso della probabilità frequentista, e argomenti affini.

- Specificità: indica quanto bene un test riconosce chi non ha una certa condizione, evitando i falsi positivi.
- Sensibilità: misura la capacità di un test di individuare correttamente chi ha davvero la condizione, riducendo i falsi negativi.
- Valore predittivo positivo: rappresenta la probabilità che una persona con un test positivo abbia davvero la condizione ricercata.

28.1 Sensibilità

“Specificità alta, il malato torna a casa informato”

Traiamo questo paragrafo da Wikipedia italiana, l’enciclopedia libera.

Con il termine sensibilità, in statistica, più precisamente nel campo della epidemiologia, si indica la capacità intrinseca di un test di screening di individuare in una popolazione di riferimento i soggetti malati. Essa è data dalla proporzione dei soggetti realmente malati e positivi al test (veri positivi) rispetto all’intera popolazione dei malati.

Un test sarà tanto più sensibile quanto più bassa risulterà la quota dei falsi negativi (cioè di soggetti malati erroneamente identificati dal test come sani). Un test molto sensibile, in definitiva, ci consente di limitare la possibilità che un soggetto malato risulti negativo al test.

Supponiamo che un test di screening dia come risultato solamente due opzioni: positivo al test e negativo. Essere positivi al test equivale ad essere ammalato secondo quel test, ma indagini diagnostiche successive possono rivelare l'effettiva malattia o meno.

Allora si otterranno 4 tipologie di osservati:

Sani Negativi (veri negativi)

Sani Positivi (falsi positivi) ← *risultano avere la malattia; ma non ce l'hanno...*

Malati Positivi (veri positivi)

Malati Negativi (falsi negativi) ← *non scoprono la loro malattia*

Rappresentabili così in tabella:

//////// MALATI - SANI
 POSITIVI Veri + Falsi +
 NEGATIVI Falsi - Veri -

La sensibilità del test verrà così calcolata:

$$S = \frac{V_+}{\text{totale MALATI}} = \frac{V_+}{V_+ + F_-}$$

ESEMPIO

25 2

4 55

Si troverà 86.2%.

Dalla wikipedia in inglese: *If the goal of the test is to identify everyone who has a condition, the number of false negatives should be low, which requires high sensitivity. That is, people who have the condition should be highly likely to be identified as such by the test. This is especially important when the consequence of failing to treat the condition are serious and/or the treatment is very effective and has minimal side effects.*

28.2 Specificità

“Specificità alta, il sano torna a casa contento”

Traiamo questo paragrafo da Wikipedia italiana, l’enciclopedia libera.

Con il termine specificità, in medicina, si indica la capacità di un test di dare un risultato normale ("negativo") nei soggetti sani:

$$Sp = \frac{V_-}{\text{totaleSANI}} = \frac{V_-}{V_- + F_+}$$

ESEMPIO

25 2

4 55

Si troverà 96.5%.

Dalla wikipedia in inglese: *If the goal of the test is to accurately identify people who do not have the condition, the number of false positives should be very low, which requires a high specificity. That is, people who do not have the condition should be highly likely to be excluded by the test. This is especially important when people who are identified as having a condition may be subjected to more testing, expense, stigma, anxiety, etc.*

28.3 Predittività

*“Valore predittivo alto: chi risulta positivo
“torna a casa quasi sicuro di avere la malattia”*

Traiamo questo paragrafo da Wikipedia italiana, l’enciclopedia libera.

Per predittività, in medicina, si intende la probabilità che un soggetto positivo ad un test di screening sia effettivamente malato. Il Valore

Predittivo Positivo, che esprime numericamente la predittività, si calcola come quota di soggetti veri positivi sul totale dei positivi (veri e falsi positivi).

$$VPP = \frac{V_+}{\text{totale POSITIVI}} = \frac{V_+}{V_+ + F_+}$$

ESEMPIO

25 2

4 55

Si troverà 92.6%.

Vediamo ora il caso in cui la prevalenza (frequenza) della malattia è decisamente minore, aumentando ad esempio di un fattore 100 le persone sane e lasciando inalterato il numero dei malati:

25 200

4 5500

Si troverà 11.1.

Esercizio. Relativamente all'ultima tabella, si calcolino prevalenza, sensibilità e specificità.

ESERCIZIO. (Tratto da Wikipedia, l'enciclopedia libera).

≈ % Per un test diagnostico in una determinata popolazione si abbia

	MALATI	SANI
POSITIVI	25	200
NEGATIVI	4	5500

Calcolare la predittività, ovvero il Valore Predittivo Positivo.

SVOLGIMENTO

Ricordando la definizione

$$\begin{aligned} \text{predittività} &= \text{Valore Predittivo Positivo} = VVP = \\ &= \frac{\text{veri positivi}}{\text{totale positivi}} = \frac{V_+}{V_+ + F_+} \end{aligned}$$

ora abbiamo

$$VVP = \frac{25}{25 + 200} = \frac{25}{225} = \frac{1}{9} \approx$$

$$\approx 0.111 = 11.1\%$$

(Notiamo che il valore è piuttosto basso nonostante la specificità sia alta:

$$\begin{aligned} Sp &= \frac{\text{veri negativi}}{\text{totale sani}} = \frac{V_-}{V_- + F_+} = \frac{5\,500}{5\,500 + 200} = \frac{5\,500}{5\,700} = \\ &= \frac{55}{57} \approx 0.965 = 96.5\% \end{aligned}$$

e questo è dovuto alla rarità della malattia nella popolazione considerata: circa una trentina di malati su poco più di 5 700 soggetti).

Nota di Wikipedia, l'enciclopedia libera:

Ossia la probabilità che un soggetto positivo al test sia effettivamente malato è pari all'11,1%, che equivale a dire che il soggetto ha una probabilità dell'88,9% di essere sano nonostante il test dica il contrario.

28.4 Esempio: i test rapidi per il covid-19.

In *Diagnostic accuracy of rapid antigen tests in asymptomatic and presymptomatic close contacts of individuals with confirmed SARS-CoV-2 infection: cross sectional study*, articolo scientifico del 2021 sul British Medical Journal su due test rapidi per il covid-19 in <https://www.bmj.com/content/374/bmj.n1676>, leggiamo i valori che gli Autori si prefiggono di misurare:

Main outcome measures Sensitivity, specificity, and positive and negative predictive values of Veritor System (Beckton Dickinson) and Biosensor (Roche Diagnostics) rapid antigen tests

28.5 Le curve ROC

Traiamo questo paragrafo da Wikipedia italiana, l'enciclopedia libera, alla voce *Receiver operating characteristic*.

Nella teoria delle decisioni, le curve ROC (Receiver Operating Characteristic, anche note come Relative Operating Characteristic[1]) sono degli schemi grafici per un classificatore binario. Lungo i due assi si possono rappresentare la sensibilità e (1-specificità), rispettivamente rappresentati da True Positive Rate (TPR, frazione di veri positivi) e False Positive Rate (FPR, frazione di falsi positivi). In altre parole, si studiano i rapporti fra allarmi veri (hit rate) e falsi allarmi.

La curva ROC viene creata tracciando il valore del True Positive Rate (TPR, frazione di veri positivi) rispetto al False Positive Rate (FPR, frazione di falsi positivi) a varie impostazioni di soglia. Il tasso di veri positivi è anche noto come sensibilità, richiamo o probabilità di rilevazione [2]. Il tasso di falsi positivi è anche noto come fall-out o probabilità di falsi allarmi [2] e può essere calcolato come (1 - specificità). Può anche essere pensato come un diagramma della potenza in funzione dell'errore di tipo I: quando la prestazione viene calcolata da un solo campione della popolazione, può essere considerata come una stima di queste quantità. La curva ROC è quindi il tasso dei veri positivi in funzione del tasso dei falsi positivi. In generale, se sono note le distribuzioni di sensibilità e 1-specificità, la curva ROC può essere generata tracciando la funzione di distribuzione cumulativa (area sotto la distribuzione di probabilità da $-\infty$ alla soglia di discriminazione) della probabilità di rilevamento nell'asse y rispetto alla funzione di distribuzione cumulativa della probabilità di falso allarme sull'asse x.

Il ROC è anche noto come curva Receiver Operating Characteristic, poiché è un confronto tra due caratteristiche operative (TPR e FPR) al cambiare del criterio.[3]

Applicazioni

Le curve ROC furono utilizzate per la prima volta durante la sec-

Distributions of the Observed signal strength

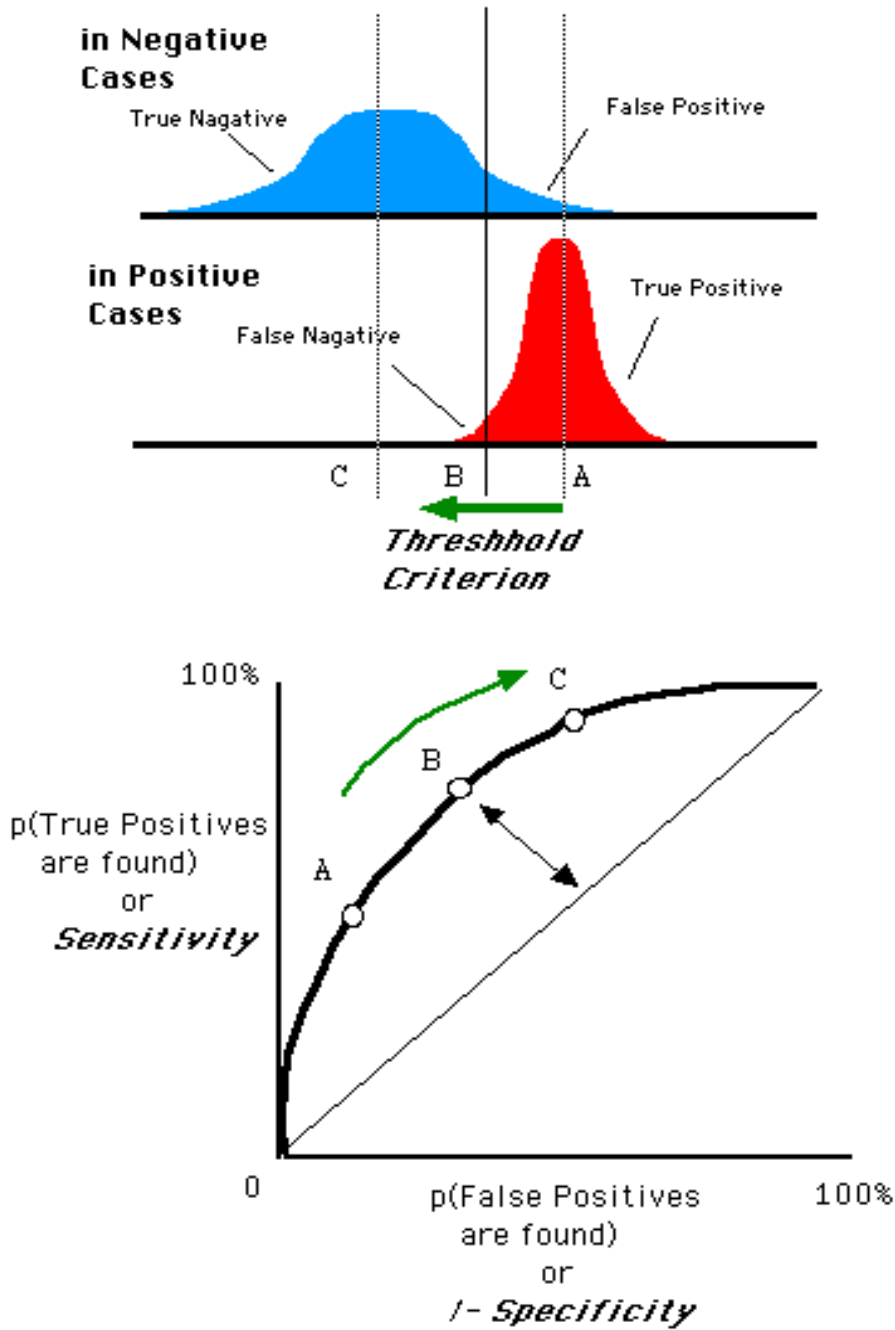


Figure 42: From <https://commons.wikimedia.org/wiki/File:ROCfig.PNG>. Author (assumed) NekoJaNekoJa~commonswiki

onda guerra mondiale, da alcuni ingegneri elettrotecnici che volevano individuare i nemici utilizzando il radar durante le battaglie aeree. Recentemente le curve ROC sono utilizzate in medicina,[4][5] radiologia,[6] psicologia,meteorologia [7], veterinaria[8], fisica e altri ambiti, come il machine learning ed il data mining.

Concetto basilare

Se si considera un problema di predizione a 2 classi (classificatore binario come da figura: distribuzione rossa e azzurra), scelto un valore di soglia (threshold o cut-off), rispetto a cui decidere il risultato, ovvero se appartenente alla classe positiva (p) o negativa (n), dato che le due curve di distribuzione di probabilità risultano in parte sovrapposte, sono possibili quattro risultati a seconda della posizione del valore di cut-off:

- se il risultato della predizione è positivo p e il valore vero è anche positivo p, viene chiamato vero positivo (true positive - TP);
- se invece il valore vero è negativo, il risultato viene chiamato falso positivo (false positive - FP);
- al contrario, si ha un vero negativo (true negative - TN) quando entrambi, il risultato e il valore vero, sono negativi;
- un falso negativo (false negative - FN) invece si ha quando il risultato è negativo e il valore vero è positivo.

È inoltre possibile rappresentare questo tipo di situazione utilizzando una tabella di contingenza di tipo 2×2 , dove le colonne rappresentano la distinzione tra soggetti sani e malati; le righe invece rappresentano il risultato del test sui pazienti. Un risultato qualitativo del test potrebbe essere quello di andare a valutare il numero di falsi positivi e negativi; meno ve ne saranno e maggiormente il test sarà valido.

Una curva ROC è il grafico dell'insieme delle coppie (FP, TP) al variare di un parametro del classificatore. Per esempio, in un classificatore a soglia, si calcola la frazione di veri positivi e quella di falsi positivi per ogni possibile valore della soglia; tutti i punti così

ottenuti nello spazio FP-TP descrivono la curva ROC.

Attraverso l'analisi delle curve ROC si valuta la capacità del classificatore di discernere, ad esempio, tra un insieme di popolazione sana e malata, calcolando l'area sottesa alla curva ROC (Area Under Curve, AUC). Il valore di AUC, compreso tra 0 e 1, equivale infatti alla probabilità che il risultato del classificatore applicato ad un individuo estratto a caso dal gruppo dei malati sia superiore a quello ottenuto applicandolo ad un individuo estratto a caso dal gruppo dei sani.[9]

Le curve ROC passano per i punti $(0,0)$ e $(1,1)$, avendo inoltre due condizioni che rappresentano due curve limite:

- una che taglia il grafico a 45° , passando per l'origine. Questa retta rappresenta il caso del classificatore casuale (linea di "nessun beneficio"), e l'area sottesa AUC è pari a 0,5.
- la seconda curva è rappresentata dal segmento che dall'origine sale al punto $(0,1)$ e da quello che congiunge il punto $(0,1)$ a $(1,1)$, avendo un'area sottesa di valore pari a 1, ovvero rappresenta il classificatore perfetto.

VIII – Variabili aleatorie discrete

BOZZA - DRAFT