

La valutazione di un esame diagnostico mediante curve ROC: alcune osservazioni

Marco B.L. Rocchi

Istituto di Biomatematica, Facoltà di Farmacia, Università di Urbino

ABSTRACT

Evaluation of a diagnostic test by means of ROC curves: some observations

In this paper the steps of the procedure of ROC curves drawing are analyzed; moreover, some remarks about methodological and statistical aspects are presented.

RIASSUNTO

Vengono analizzate le fasi della costruzione delle curve ROC e vengono fatte alcune considerazioni di carattere metodologico e statistico.

INTRODUZIONE

L'impiego delle curve ROC (Receiver Operating Characteristic curves) nella valutazione degli esami diagnostici ha conosciuto uno sviluppo esponenziale a partire dagli anni '70. A titolo di esempio, basterà osservare che una recente bibliografia ragionata sulle curve ROC in medicina, cita 25 lavori pubblicati negli anni '70 o in precedenza, 95 negli anni '80, 173 negli anni '90.

A fronte della sterminata letteratura sull'argomento, e probabilmente a causa del parallelo fiorire di software dedicati, sono pochi i lavori che si soffermano sulla costruzione delle curve come primo passo nella comprensione del loro significato e della loro potenzialità. Questo lavoro cerca di colmare tale lacuna.

CONCETTI PROPEDEUTICI: SENSIBILITÀ E SPECIFICITÀ

La validità di un esame diagnostico viene di norma affrontata mediante i consueti concetti di sensibilità, specificità, valore predittivo positivo e valore predittivo negativo; per una chiara panoramica sulla validità di un esame diagnostico si veda ad esempio: Carle et al. (1). Sebbene tali concetti siano per lo più familiari agli addetti ai lavori, sarà bene in questo contesto richiamarli alla memoria con la dovuta precisione, poiché è su di essi - ed in particolare sui primi due - che si basa la costruzione delle curve ROC. Consideriamo un ipotetico esame diagnostico a risposta quantale (cioè un esame che fornisca un risultato di tipo binario: positivo o negativo). La valutazione della validità di un esame del genere può dedursi da una tabella di contingenza 2 x 2 del tipo che segue:

| | | situazione vera | |
|-------------------------|----------|-----------------|------|
| | | malati | sani |
| risultato dell'esame | positivo | M+T+ | M-T+ |
| | negativo | M+T- | M-T- |

nella quale i simboli M+, M-, T+ e T- stanno a indicare rispettivamente i malati, i sani, i positivi all'esame e i negativi all'esame test, così che nella casella M+T+ saranno conteggiati i malati risultati positivi all'esame (Veri Positivi, VP), nella casella M+T- i malati negativi all'esame (Falsi Negativi, FN), nella casella M-T+ i sani positivi all'esame test (Falsi Positivi, FP) e nella casella M-T- i sani negativi all'esame (Veri Negativi, VN), come riassunto nella tabella seguente:

| | | situazione vera | |
|-------------------------|----------|-----------------|----------------|
| | | malati | sani |
| risultato dell'esame | positivo | Veri positivi | Falsi positivi |
| | negativo | Falsi negativi | Veri negativi |

dove appare evidente che le caselle della diagonale principale (VP e VN) rappresentano i casi in cui l'esame fornisce un risultato corretto e le caselle della diagonale secondaria rappresentano i casi in cui lo stesso esame dà luogo a risultati errati.

Una tabella siffatta si presta bene a formalizzare i concetti di sensibilità e di specificità, che vengono definiti come:

$$\text{Sensibilità} = \frac{VP}{VP + FN}$$

$$\text{Specificità} = \frac{VN}{FP + VN}$$

In altri termini, la sensibilità è la capacità di fornire risultati positivi nei soggetti malati (più brevemente: positività nella malattia) mentre la specificità è la capacità di fornire risultati negativi nei soggetti sani (negatività in assenza di malattia). Ancora, possiamo riassumere dicendo che la sensibilità è massima se l'esame è in grado di "catturare" (nel senso di identificare) *tutti* i malati, mentre la specificità è massima se l'esame "cattura" (cioè identifica) *solo* i malati.

Circa il valore predittivo positivo e negativo di un esame, basterà in questa sede ricordare che essi sono definiti come:

$$\text{Valore Predittivo Positivo} = \frac{VP}{VP + FP}$$

$$\text{Valore Predittivo Negativo} = \frac{VN}{VN + FN}$$

e che, mentre sensibilità e specificità sono proprietà intrinseche all' esame, i valori predittivi dipendono fortemente dalla prevalenza della malattia (ad esempio, possono fornire risultati fortemente diversi a seconda che siano calcolati nella popolazione generale o in una popolazione ospedalizzata).

A titolo di esempio (dati fittizi), supponiamo che un esame eseguito su 1000 soggetti abbia fornito la tabella seguente:

| | | situazione vera | |
|----------------------|----------|-----------------|----------|
| | | malati | sani |
| risultato dell'esame | positivo | VP = 70 | FP = 150 |
| | negativo | FN = 30 | VN = 750 |

Essa determina i seguenti parametri:

$$\text{Sensibilità} = \frac{70}{70 + 30} = 0,70$$

$$\text{Specificità} = \frac{750}{150 + 750} = 0,83$$

$$\text{Valore Predittivo Positivo} = \frac{70}{70 + 150} = 0,32$$

$$\text{Valore Predittivo Negativo} = \frac{750}{750 + 30} = 0,96$$

Tutto quanto sin qui enunciato è perfettamente realizzabile per un esame a risposta quantale, ma cosa succede

quando l'esame misura una variabile quantitativa? In questo caso sensibilità e specificità variano al variare del valore di soglia (cut-point) fissato per l'esame, e la loro variazione mostra tendenze opposte, nel senso che un aumento di sensibilità (riduzione dei falsi negativi) "si paga" con una perdita di specificità (aumento dei falsi positivi) e viceversa, come è ben visibile dal grafico di Figura I.

Quindi un altro problema è: qual è il valore di soglia ottimale per un certo esame? A questa e ad altre domande le curve ROC sono in grado di fornire una risposta.

COSTRUZIONE DELLE CURVE ROC

La curva ROC è un diagramma in cui vengono riportati (per diversi possibili valori di soglia dell' esame) in ordinata la sensibilità e in ascissa il complemento a 1 della specificità (1- specificità, ossia il tasso dei Falsi Positivi).

Sensibilità e specificità vengono calcolate ex novo per ogni valore di soglia considerato, in base alla considerazione che ogni valore di soglia genera una nuova tabella di contingenza.

Per esemplificare quanto appena detto, prendiamo in considerazione dei dati (fittizi) concernenti la distribuzione della variabile oggetto del test nelle due popolazioni dei sani e dei malati (Tabella 1). I relativi istogrammi sovrapposti sono mostrati in Figura II.

Supponiamo ora di fissare un primo valore soglia per l' esame ad un livello molto basso, diciamo 1,5. E' evidente che una soglia così bassa corrisponderà ad un'ottima sensibilità del test (tutti i malati sono "catturati" = nessun Falso Negativo), ma ad una pessima specificità (accanto ai malati, anche moltissimi sani sono "catturati" erroneamente = tanti Falsi Positivi). A questa situazione, mostrata in Figura III, viene fatta corrispondere una tabella di con-

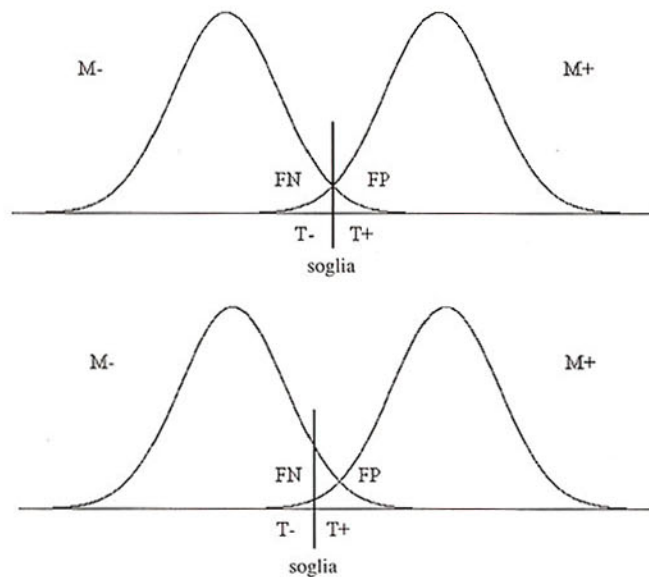


Figura I
Rappresentazione schematica della relazione fra livello di soglia (cut-point) del test, falsi positivi e falsi negativi.

Tabella 1

Ipotetica distribuzione di una popolazione rispetto ad una patologia (dati fittizi), nella suddivisione nelle due sottopopolazioni di sani e malati (cfr. Fig. II)

| Valore assunto dalla variabile Unità Arbitrarie (U. Arb.) | Frequenza (assoluta) nei | |
|--|--------------------------|------|
| | Malati | Sani |
| 1 | 0 | 100 |
| 2 | 0 | 300 |
| 3 | 0 | 500 |
| 4 | 50 | 350 |
| 5 | 100 | 250 |
| 6 | 150 | 150 |
| 7 | 200 | 50 |
| 8 | 100 | 0 |
| 9 | 50 | 0 |

tingenza, da cui i parametri di sensibilità e specificità (o meglio 1 - specificità) sono calcolati (vedi tabella 2). Questa coppia di parametri rappresenterà le coordinate del primo punto sulla curva ROC.

Che succede se decidiamo di spostare la soglia a 2,5? La sensibilità non viene ancora intaccata (anche in questo caso tutti i malati sono "catturati"), ma comincia ad aumentare, seppur lievemente, la specificità (rispetto al valore precedente di soglia, vi sarà un numero minore di Falsi Positivi).

Ancora una volta, a questa situazione viene fatta corrispondere una tabella di contingenza, da cui sono calcolate le coordinate (1 - specificità; sensibilità) per tracciare un secondo punto sulla curva ROC (tabella 2).

Immaginiamo di operare analogamente per il valore di soglia 3,5 che fornirà un nuovo punto sul grafico della

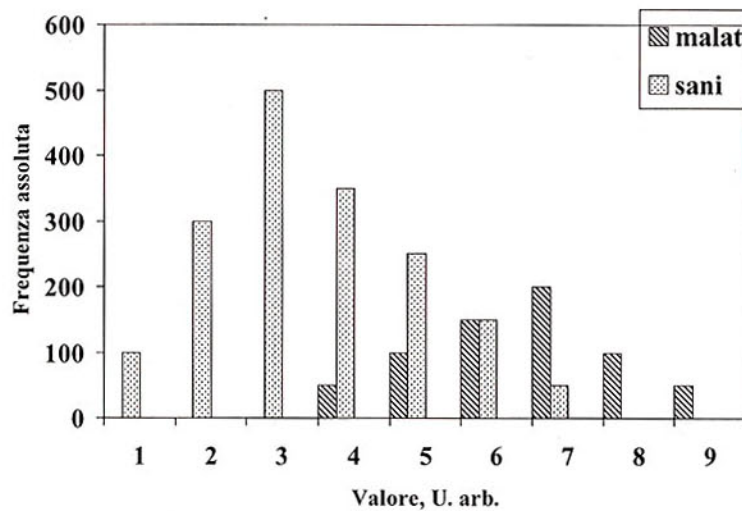


Figura II

Ipotetica distribuzione di una popolazione rispetto a una patologia (dati fittizi), nella suddivisione nelle due sottopopolazioni di sani e malati (cfr. Tab.1)

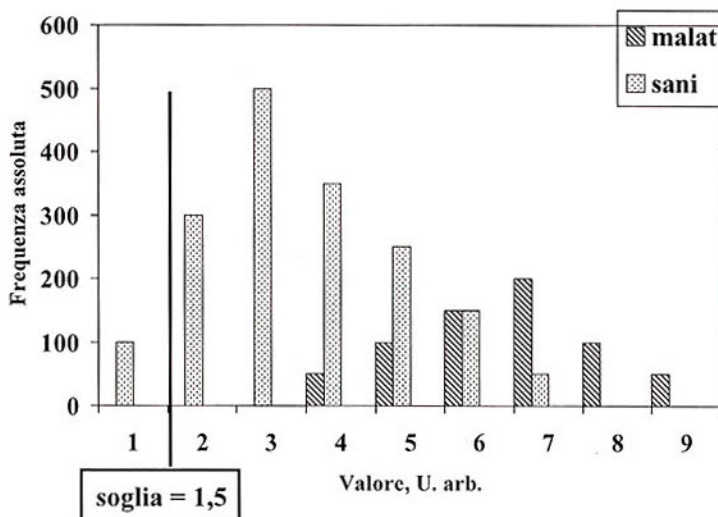


Figura III

Costruzione, a partire dalla distribuzione di Fig. II, della tabella di contingenza per il calcolo di sensibilità e specificità, per un valore di soglia = 1,5

Tabella 2

Variazione della specificità e della sensibilità di un esame al variare del valore soglia. Distribuzione ipotetica della tabella 1 e figura II

| Soglia U. Arb. | Positivi | | Negativi | | Sensibilità | Specificità | 1-Specif. | Totali |
|-------------------|----------|-------|----------|-------|-------------|-------------|-----------|--------|
| | Veri | Falsi | Veri | Falsi | | | | |
| 1,5 | 650 | 1600 | 100 | 0 | 1,00 | 0,06 | 0,94 | 2350 |
| 2,5 | 650 | 1300 | 400 | 0 | 1,00 | 0,24 | 0,76 | 2350 |
| 3,5 | 650 | 800 | 900 | 0 | 1,00 | 0,53 | 0,47 | 2350 |
| 4,5 | 600 | 450 | 1250 | 50 | 0,92 | 0,74 | 0,26 | 2350 |
| 5,5 | 500 | 200 | 1500 | 150 | 0,77 | 0,88 | 0,12 | 2350 |
| 6,5 | 350 | 50 | 1650 | 300 | 0,54 | 0,97 | 0,03 | 2350 |
| 7,5 | 150 | 0 | 1700 | 500 | 0,23 | 1,00 | 0,00 | 2350 |
| 8,5 | 50 | 0 | 1700 | 600 | 0,08 | 1,00 | 0,00 | 2350 |

curva ROC. Che succede al successivo valore di soglia (4,5)? Succede che per la prima volta l'esame, mentre continua a guadagnare in specificità, inizia a perdere in sensibilità; in altri termini, per la prima volta un certo numero di malati "sfugge" al test (e quindi la sensibilità non è più perfetta). Si veda ancora la tabella 2.

La tendenza ad un aumento di specificità, a scapito della sensibilità dell'esame, continua a realizzarsi anche per i valori di soglia 5,5 e 6,5. Al seguente valore soglia (7,5) si registra una novità: a fianco di una sensibilità ormai fortemente ridotta, per la prima volta l'esame mostra una specificità perfetta; infatti (tabella 2 e figura IV) nessun soggetto sano risulta positivo al test (azzeramento dei Falsi Positivi).

E' da osservare che il prezzo pagato per l'elevata specificità è il crollo della sensibilità, chiaramente al di sotto di valori accettabili. Situazione analoga, e addirittura peggiorativa, si registra per il successivo valore di soglia (8,5).

La costruzione della curva ROC mostra probabilmente meglio di ogni altro procedimento il compromesso fra sensibilità e specificità di un esame diagnostico. Natural-

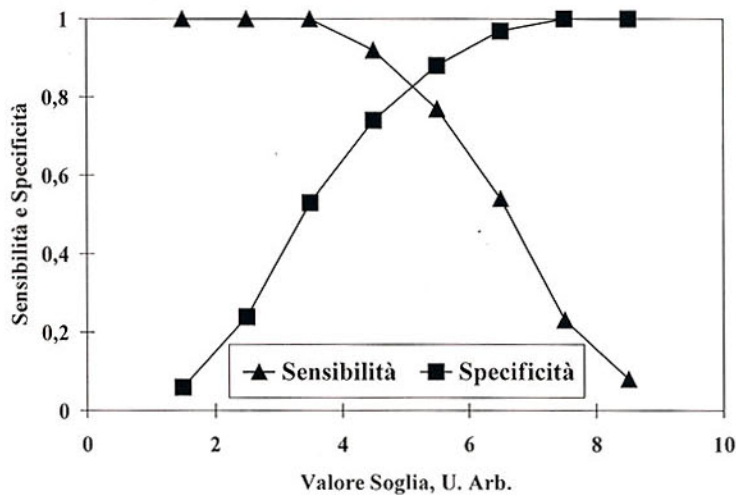
mente, per completezza di analisi, va osservato che, qualora le distribuzioni dei sani e dei malati non siano affatto sovrapposte, la soglia ideale è un qualunque valore compreso tra le due distribuzioni e che in questo fortunato (ma purtroppo assai raro) caso nessun compromesso è necessario tra sensibilità e specificità.

A questo punto, i valori su cui tracciare la curva ROC sono tutti definiti, e la curva può essere tracciata (figura V).

Sottolineiamo che nel nostro caso la curva è più precisamente una spezzata; ciò è dovuto al fatto che sono stati considerati pochi valori di soglia, compatibili con le distribuzioni - fittizie - della variabile oggetto di studio, la quale è stata scelta, per comodità espositiva, su scala discreta. Ovviamente, qualora la variabile sia realmente continua, e scegliendo numerosi valori soglia per l'esame, la spezzata tende a una curva.

ALCUNE OSSERVAZIONI SULLE CURVE ROC

Ora che la curva ROC è tracciata, quali informazioni se ne possono dedurre? Quello che segue è un breve

**Figura IV**

Variazione della specificità e della sensibilità di un esame al variare del valore soglia (cfr. Tab. 2)

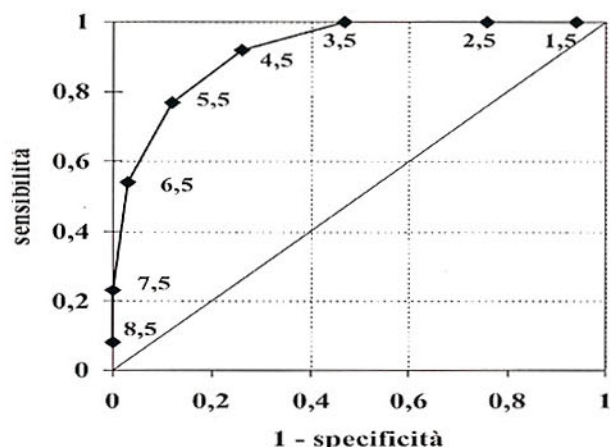


Figura V

Costruzione, a partire dalla distribuzione di Fig. II, della curva ROC corrispondente. I valori riportati lungo la spezzata indicano i valori di soglia su cui si è valutato l'esame

elenco, necessariamente non esaustivo, del ruolo che le curve ROC possono avere in clinimetria.

Livello di validità di un esame

È chiaro che rappresentando la combinazione di sensibilità e specificità, cioè dei due principali parametri di validità di un esame diagnostico, la curva ROC è utile a valutare, anche quantitativamente, quanto un esame sia globalmente valido (dove il termine globalmente indica che la validità dell'esame è valutata indipendentemente dalla scelta del valore di soglia). È intuitivo che, quanto più la curva ROC si avvicina ai "lati" sinistro e superiore del diagramma, tanto maggiore è la validità del test. Si ricordi infatti che, essendo rappresentate in ordinata la sensibilità e in ascissa il complemento a 1 della specificità, l'angolo "in alto a sinistra" del diagramma rappresenta l'optimum per un esame diagnostico, indicando la combinazione di massima sensibilità e massima specificità.

Viceversa, una curva ROC che si avvicini alla diagonale del diagramma indicherà un esame la cui validità è nulla (un esame di questo tipo si realizza quando le due distribuzioni di sani e malati siano completamente sovrapposte e la sua validità diagnostica è paragonabile a quella che si otterrebbe col lancio di una moneta).

Da tutto ciò si deduce anche che la curva ROC può permettere la quantificazione della validità di un esame, mediante il calcolo dell'area sottesa dalla curva stessa; tale area avrà un intervallo di variazione compreso tra 0,5 (test assolutamente non valido) e 1 (test ideale). Si noti anche che alcuni autori preferiscono considerare l'area compresa tra la curva stessa e la diagonale; l'intervallo di variazione sarà in questo caso compreso tra 0 e 0,5. L'area può facilmente essere calcolata come somma di trapezi, ma non mancano attualmente numerosi software che sono in grado, mediante algoritmi numerici, di calcolare l'area anche quando questa è sottesa da una curva ottenuta con tecniche di smoothing (per approfondimenti circa gli aspetti quantitativi del calcolo dell'area sotto la curva, si consiglia Metz (2)).

Ai valori d'area, comunque ottenuti, si può associare una interpretazione qualitativa del livello di validità dell'esame, mediante l'applicazione della scala, un po' empirica ma efficace, descritta di seguito:

0,9-1 = eccellente

0,8-0,9 = buono

0,7-0,8 = discreto

0,6-0,7 = scarso

0,5-0,6 = nullo

Sceita del valore soglia ottimale per un esame

Da quanto detto al paragrafo precedente, risulta che l'angolo "in alto a sinistra" del diagramma rappresenta la situazione ideale per un esame diagnostico, in quanto combinazione di massima sensibilità e massima specificità. Appare quindi ovvio che la "teorica" soglia ottimale per un esame è quella in corrispondenza alla quale la curva ROC è più vicina a tale angolo (ad esempio in figura V il valore soglia ideale è 5,5).

Abbiamo parlato di soglia ottimale "teorica" per sottolineare il fatto che tale valore va interpretato alla luce di considerazioni globali rispetto alla patologia indagata. Con ciò si vuole dire che, ad esempio, di fronte a patologie fortemente invalidanti e d'altra parte curabili se diagnosticate tempestivamente (si pensi ad esempio allo screening per Ipotiroidismo Congenito), la scelta del valore soglia cadrà non sul valore "teoricamente" ottimale, ma si preferirà tenere il più alto possibile il livello di sensibilità dell'esame perché nessun malato sfugga alla diagnosi (nessun Falso Negativo), accettando di pagare lo scotto di una ridotta specificità (un certo numero di Falsi Positivi).

Tutto ciò, dal punto di vista formale, significa che nella scelta della soglia ideale occorrerà tenere conto del "costo" (clinico, etico, sociale, economico) associato ad ogni possibile valore di soglia dell'esame (3).

Tale costo sarà dato dalla somma di due componenti, la prima legata alla imperfetta sensibilità (costo dei Falsi Negativi), la seconda alla imperfetta specificità (costo dei Falsi Positivi).

Denotando con:

cFN = costo di un Falso Negativo;

cFP = costo di un Falso Positivo;

Prev = prevalenza della malattia nella popolazione;

FN = tasso di Falsi Negativi;

FP = tasso di Falsi Positivi;

CFN = costo globale dei Falsi Negativi;

CFP = costo globale dei Falsi Positivi;

CT = costo dell'esame;

il costo globale dei falsi negativi e quello dei falsi positivi sono definiti rispettivamente da:

$$CFN = Prev \cdot FN \cdot cFN$$

$$CFP = (1 - Prev) \cdot FP \cdot cFP$$

Dunque, il costo del test

$$CT = CFN + CFP = Prev \cdot FN \cdot cFN + (1 - Prev) \cdot FP \cdot cFP$$

potrà essere calcolato, e graficato, per ogni valore di soglia. Un grafico di questo tipo consentirà di stabilire la soglia ideale, cioè quella alla quale corrisponda il minor costo possibile. Si veda, ad esempio, il grafico di figura VI,

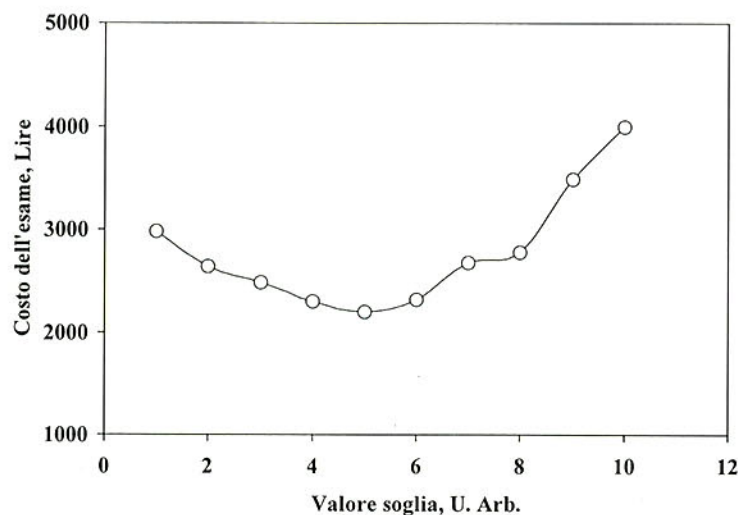


Figura VI

Grafico "costo dell'esame vs valore di soglia", per la scelta del livello di soglia ottimale (dati fittizi). Il valore di soglia= 5 minimizza il costo dell'esame, e pertanto rappresenta il valore di soglia ideale

tracciato su dati fittizi.

Altri metodi, basati sul concetto di "utilità" dell' esame, complementare a quello di "costo", sono stati analizzati e messi a confronto da Krieg et al. (4), ai quali si rimanda.

Curve ROC e rapporto di verosimiglianza

Da alcune semplici nozioni di analisi matematica (la cui trattazione esula dagli scopi del presente lavoro) si deduce facilmente che la pendenza (coefficiente angolare) della retta tangente alla curva in un certo punto (corrispondente a un certo valore soglia) fornisce il rapporto di verosimiglianza (Likelihood Ratio, LR) di un risultato positivo per quel valore soglia, ovvero il rapporto tra la probabilità di avere un esame positivo in un soggetto malato e la probabilità di avere un esame positivo in un soggetto sano. Formalmente ciò si esprime mediante la seguente notazione:

$$LR + = \frac{\text{Pr (T+| M+)}}{\text{Pr (T+| M-)}} = \frac{\text{sensibilità}}{1 - \text{specificità}}$$

E' ovvio che il valore soglia per cui LR+ è massimo non va interpretato come il valore ottimale di soglia per l'esame (aspetto già discusso sopra), ma in un certo senso esprime il punto in cui è più elevata la "velocità" (matematicamente: la derivata prima) con cui il parametro "sensibilità" cresce rispetto al parametro "1 - specificità".

Confronti mediante curve ROC

Una comparazione mediante curve ROC può riferirsi a casi differenti, quali un confronto fra esami diversi per la diagnosi della stessa patologia, un confronto di validità di uno stesso esame in situazioni patologiche differenti (diverse patologie, diversi stadi di una stessa patologia), o infine un confronto in cui lo stesso esame è applicato in differenti sottogruppi di pazienti aventi la stessa patologia, ma che differiscono per altre caratteristiche (età, sesso, etc.).

Ovviamente, un solido confronto tra esami diagnostici non può limitarsi a valutare quale curva ROC passi più

vicino ai lati sinistro e superiore del diagramma (ovvero quale sottenda la maggiore area), se questa comparazione non è sostenuta da un test statistico (figura VII).

Un test idoneo è stato approntato da Hanley e McNeil (5,6). Sebbene tale test non sia il solo approccio proposto in letteratura, tuttavia esso ha goduto e gode di grande fortuna, probabilmente perché, a dispetto dell'apparente complessità per il calcolo della statistica test, un semplice foglio elettronico può essere programmato con poche istruzioni per eseguirlo.

Prima di descrivere le fasi di calcolo di questo test, occorre premettere alcune notazioni: anzitutto occorre considerare separatamente le due curve ROC che si vogliono confrontare, e di ognuna si devono calcolare l'area sottesa dalla curva e la sua deviazione standard. Per fare ciò, tutta una serie di calcoli va eseguita per vari intervalli della variabile; infatti, se noi consideriamo un certo numero di valori soglia per l'esame, è chiaro che l'intervallo della variabile oggetto del test può essere suddivisa in un numero di intervalli pari al numero dei valori soglia considerati, maggiorato di uno (si può visivamente pensare a questi intervalli come alle colonne dell'istogramma relativo). Nelle formule che seguono, useremo le seguenti notazioni:

x = gli intervalli della variabile, come definiti dai cut-point;
 M^{x-} = numero di soggetti sani la cui variabile diagnostica assume valore compreso nell'intervallo x;
 M^{x+} = numero di soggetti malati la cui variabile diagnostica assume valore compreso nell'intervallo x;
 $M^{<x}$ = numero di soggetti sani la cui variabile diagnostica assume valore inferiore all'intervallo x;
 $M^{>x}$ = numero di soggetti malati la cui variabile diagnostica assume valore superiore all'intervallo x;
 $M^{\text{TOT-}}$ = totale dei soggetti sani;
 $M^{\text{TOT+}}$ = totale dei soggetti malati;
 AUC = area sottesa dalla curva;
 SD_{AUC} = deviazione standard di AUC.

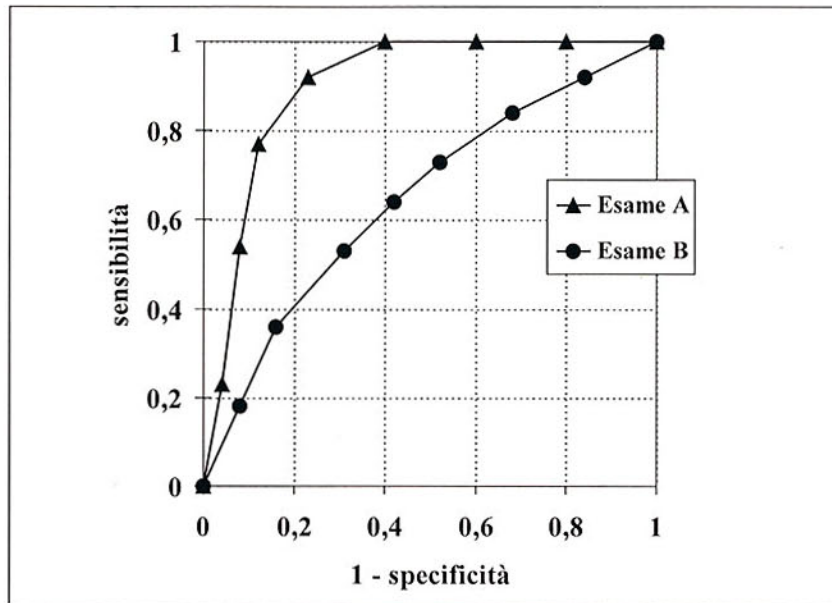


Figura VII
 Grafico di confronto fra due curve ROC, tracciate per due ipotetici esami diagnostici. L'esame A risulta più valido dell'esame B, ma per stabilire che tale differenza è significativa occorrerà eseguire un test statistico

A questo punto, i calcoli per il metodo di Hanley e McNeil possono essere sintetizzati nei seguenti step:

step 1: si considera la prima delle due curve ROC
 step 2: si calcola

$$R = \sum_x (M_+^x \cdot M_-^x + \frac{1}{2} M_-^x M_+^x)$$

step 3: si calcola

$$AUC = \frac{R}{M_+^{TOT} M_-^{TOT}}$$

step 4: si calcolano

$$N1 = \sum_x M_-^x \left[(M_+^x)^2 + (M_+^x M_-^x) + \frac{1}{3} (M_+^x)^2 \right]$$

$$N2 = \sum_x M_+^x \left[(M_-^x)^2 + (M_-^x M_+^x) + \frac{1}{3} (M_-^x)^2 \right]$$

step 5: si calcolano

$$Q1 = \frac{N1}{M_-^{TOT} (M_+^{TOT})^2}$$

$$Q2 = \frac{N2}{M_+^{TOT} (M_-^{TOT})^2}$$

step 6: si calcola

$$SD_{AUC} = \sqrt{\frac{AUC(1-AUC) + (M_+^{TOT} - 1)(Q1 - AUC^2) + (M_-^{TOT} - 1)(Q2 - AUC^2)}{M_+^{TOT} M_-^{TOT}}}$$

step 7: assegnate all'area sottesa sotto la prima curva ROC e alla sua deviazione standard le notazioni AUC1, SD_{AUC1}, si procede agli stessi calcoli (step 1-6) per la seconda curva ROC, ottenendo così AUC2, SD_{AUC2}

step 8: si calcola la statistica test:

$$z = \frac{AUC1 - AUC2}{\sqrt{SD_{AUC1}^2 + SD_{AUC2}^2}}$$

che si distribuisce, sotto l'ipotesi nulla di una identità fra le due curve ROC, come una normale gaussiana standar-

dizzata

step 9: si applica la seguente regola di decisione del test (per $\alpha = 0,05$): se $z < -1,96$ o $z > 1,96$, si respinge l'ipotesi nulla e si conclude per una differenza significativa in termini di validità fra i due esami diagnostici considerati.

Come detto, il metodo di Hanley e McNeil non esaurisce le modalità di confronto fra curve ROC; ad esempio, altri metodi sono stati proposti da DeLong et al. (7) e da Ma e Hall (8), e di recente è stato proposto un test applicabile qualora le curve siano state costruite mediante tecniche di smoothing basate sul metodo kernel (9).

Altri campi applicativi delle curve ROC in ambito medico

Solo per completezza di informazione, sarà bene ricordare che la tecnica delle curve ROC trova numerose altre applicazioni in campo medico, anche in branche lontane dalla medicina di laboratorio. Di seguito vengono forniti alcuni esempi, senza che essi abbiano naturalmente la minima pretesa di completezza, di applicazioni delle curve ROC:

(a) confronto fra differenti metodi di valutazione del segmento ST in un ECG sotto sforzo nella diagnosi di coronaropatie (10);

(b) valutazione di una scala per la diagnosi di faringiti streptococciche, basata su quattro segni (essudato tonsillare, febbre, adenopatia, assenza di tosse) (11);

(c) confronto fra indagine mammografica e palpazione per lo screening del tumore della mammella (12);

(d) confronto fra tre differenti scale psicologiche per la valutazione del rischio di comportamenti molesti a sfondo sessuale (13).

UNA DIGRESSIONE STORICA: L'ORIGINE DELLE CURVE ROC

Le curve ROC (acronimo di Receiver Operating Characteristic; cioè Caratteristiche Operative del Ricevente) hanno una origine militare.

Durante la Seconda Guerra Mondiale, i tecnici radar erano chiamati a distinguere sul monitor i segnali corrispondenti a un aereo nemico dai normali rumori di fondo (i quali essendo di natura stocasticamente poissoniana, danno di tanto in tanto origine a picchi anomali). E' per valutare la capacità dei singoli tecnici radar che le curve ROC sono state inventate.

Immaginiamo due addetti che rappresentino altrettanti casi limite di comportamento.

Consideriamo dapprima un tecnico - chiamiamolo "ansioso" - che ad ogni piccolo segnale radar fa scattare l'allarme; il risultato è ovvio: nessun aereo nemico gli sfugge, ma fa scattare moltissimi allarmi a vuoto (il parallelo medico è chiaro: se gli aerei nemici sono i "malati" e i rumori di fondo sono i "sani", questo addetto radar rappresenta un test molto sensibile, che identifica tutti i malati, ma poco specifico, cioè che insieme ai malati segnala anche molti sani).

Prendiamo ora in esame un tecnico - chiamiamolo "imperturbabile" - che non si lascia impressionare da piccoli segnali, ed è molto più cauto nel fare scattare l'allarme; in questo caso il risultato è che difficilmente l'allarme scatterà a vuoto (aumento di specificità), ma potrà sfuggirgli qualche aereo nemico (perdita di sensibilità).

Solo con un trentennio di ritardo - a partire dagli anni '70 - le curve ROC sono state utilizzate con successo nell'interpretazione di esami clinici; non a caso i primi ad impiegarle furono i radiologi, il cui problema, analogamente a quello tecnici radar, poteva essere ricondotto a una questione di analisi del segnale (signal detection).

BIBLIOGRAFIA

1. Carle F, Mattei A, Necozone S, Di Orio F. Caratteristiche del test diagnostico. In: Di Orio, editor. Elementi di metodologia epidemiologica clinica. Padova: Piccin Editore 1994;19-76.
2. Metz CE. Basic principles of ROC analysis. *Sem Nuc Med* 1978;8:283-298.
3. Burlina A. Medicina di laboratorio. Fondamenti di diagnostica. Torino: C.G. Edizioni Medico Scientifiche 1992;551-558.
4. Krieg AF, Abendroth TW, Bongiovanni MB. When is a diagnostic test result positive? Decision tree models based on net utility and threshold. *Arch Pathol Lab Med* 1986;110:787-791.
5. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29-36.
6. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983;148:839-843.
7. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves. A nonparametric approach. *Biometrics* 1988;44:837-845.
8. Ma G, Hall WJ. Confidence bands for receiver operating characteristic curves. *Med Decis Mak* 1993;13:191-197.
9. Zou KH, Hall WJ, Shapiro D. Smooth non-parametric ROC curves for continuous diagnostic tests. *Stat Med* 1997;16:2143-2156.
10. Detrano R, Salcedo EE, Go TR. Stress test options in suspected coronary disease. *Postgrad Med* 1984;75:153-164.
11. Wigton RS, Connor JL, Centor RM. Transportability of a decision rule for the diagnosis of streptococcal pharyngitis. *Arch Intern Med* 1986;146:81-83.
12. Gohagan JK, Spitznagel EL, McCrate MM, Frank TB. ROC analysis of mammography and palpation for breast screening. *Invest Radiol* 1984;19:587-592.
13. Hanson RK, Thornton D. Static 99: improving actuarial risk assessment for sex offenders. Ottawa: Department of the Solicitor General of Canada 1999.