

DIXIEME COLLOQUE SUR LE TRAITEMENT DU SIGNAL ET SES APPLICATIONS

NICE du 20 au 24 MAI 1985

Les tests diagnostiques en médecine considérés comme des systèmes transmetteurs de signaux aléatoires binaires : espérance mathématique du coefficient de transmission et du gain informatif.

E. GREMILLET, A. CHAMPAILLER, T. HERRMANN, M. DECOUSUS,
J.C. HEALY.

Laboratoire de Biophysique et d'Informatique Médicale.
30 Rue Ferdinand Gamhon. 42023 SAINT-FETIENNE Cédex. FRANCE.

RESUME

L'évaluation de la performance d'un test diagnostique, c'est à dire son aptitude à affirmer ou à éliminer une hypothèse pathologique, repose sur l'étude d'un échantillon de sujets dont on détermine l'état malade ou non-malade et la réponse positive ou négative au test.

A partir de la matrice d'information ainsi constituée, l'approche classique consiste à calculer divers indices : sensibilité, spécificité, valeurs prédictives, etc...
Aucun de ces indices n'étant satisfaisant, deux approches nouvelles sont présentées en considérant le test diagnostique comme un système transmetteur de signaux aléatoires binaires, dont le rôle est de réduire l'incertitude.

La première, basée sur la théorie de l'information, consiste à calculer les entropies d'information aux différents points du système et le coefficient de transmission du test.

La deuxième utilise le théorème de Bayes pour définir la notion de gain informatif : somme des différences entre probabilités à postériori et à priori d'avoir ou de ne pas avoir la maladie.

Dans les deux cas la grandeur calculée est une fonction de la sensibilité, de la spécificité et de la prévalence. La prévalence étant variable d'une population à l'autre et souvent inconnue, nous proposons de calculer l'espérance mathématique du coefficient de transmission et du gain informatif sous l'hypothèse d'une prévalence aléatoire uniforme sur $[0,1]$. On caractérise ainsi de façon neutre le test diagnostique et on peut comparer facilement des tests concurrentiels. Pour le gain informatif, ce calcul est simple et de plus il est facile de calculer la prévalence optimale du test. Ces raisons nous font proposer l'étude du gain informatif d'un test pour son évaluation.

SUMMARY

The assessment of a diagnostic test, ie its ability to assert or to exclude a pathological hypothesis, is based on the study of a sample of patients whose healthy or ill state and positive or negative test answer is determined.

An information matrix is then built up, from which classical analysis provides some indices : sensitivity, specificity, predictive values, ...
Since none of the indices is quite satisfactory, two new approaches are presented, considering the diagnostic test as a system transmitting aleatory binary signals, whose aim is uncertainty reduction.

The first one, based on information theory, consists in computing information entropy in various points of the system and the transmission coefficient of the test.

The second one uses Bayes' theorem to define the information gain which is the sum of the differences between posterior and prior probabilities of having or not the disease.

In both cases, the calculated index is a function of sensitivity, specificity and prevalence. Since prevalence varies from one population to another it is often unknown, we propose to calculate the expectation of transmission coefficient and information gain, assuming that prevalence is uniformly aleatory on $[0,1]$

The diagnostic test is thus neutrally characterized and can be easily compared with competing tests. For the information gain, this computation is simple and in other respects computing the test's optimal prevalence is easy.

These considerations lead us to propose the study of the information gain for the assessment of a diagnostic test.



Les tests diagnostiques en médecine considérés comme des systèmes transmetteurs de signaux aléatoires binaires : espérance mathématique du coefficient de transmission et du gain informatif.

La mise au point d'une stratégie décisionnelle face à un problème médical, c'est à dire l'optimisation des rapports efficacité/coût, doit prendre en compte plusieurs éléments :

- prévalence de la pathologie-cible dans la population de référence ;
- performance des diverses procédures diagnostiques susceptibles d'affirmer ou d'éliminer l'hypothèse pathologique ;
- efficacité des thérapeutiques utilisables ;
- coût humain et financier des diverses étapes.

Le deuxième point est susceptible de bénéficier de concepts utilisés en traitement du signal.

Données initiales :

Elles sont fournies par l'étude d'un échantillon de n sujets dont le caractère malade (M) ou non-malade (\bar{M}) est établi avec certitude par un moyen non applicable en routine (biopsie, exploration chirurgicale, suivi à long terme...). Pour chaque sujet on effectue le test diagnostique à évaluer et on note sa réponse positive ($T+$) ou négative ($T-$). L'ensemble est colligé dans un tableau de contingence appelé matrice d'information (MI) qui regroupe vrais positifs (VP), faux positifs (FP), vrais négatifs (VN) et faux négatifs (FN) :

| | | | |
|------|-----|-----------|------|
| | M | \bar{M} | |
| $T+$ | VP | FP | $T+$ |
| $T-$ | FN | VN | $T-$ |
| | m | \bar{m} | n |

Le problème posé est alors de condenser l'information de cette MI pour obtenir un outil commode et fiable de comparaison de tests diagnostiques.

Approche classique :

Plusieurs indices ou couples d'indices ont été proposés pour décrire la MI (1,2,3) :

- Le couple sensibilité-spécificité (Se, Sp) :
 - . $Se = P(T+/M) = VP/m$: encore appelée "true positive fraction" (TPF)
 - . $Sp = P(T-/\bar{M}) = VN/\bar{m}$: encore appelée "true negative fraction" (TNF).
- Le couple dual "false negative fraction"- "false positive fraction" (FNF, FPF) :
 - . $FNF = 1 - Se$
 - . $FPF = 1 - Sp$
- Le couple des valeurs prédictives positive et négative (VPP, VPN) :
 - . $VPP = P(M/T+) = VP/t+$
 - . $VPN = P(\bar{M}/T-) = VN/t-$
- La précision, c'est à dire la fraction de bien-classés (notée A pour "accuracy") :
 - . $A = (VP+VN)/n$
- L'indice de Youden :
 - . $Y = Se/(1-Sp)$.
- Le rapport de vraisemblance (noté LR pour "likelihood ratio") :
 - . $LR = Se/(1-Sp)$

Aucun de ces index n'est exempt de critiques :

- l'information apparemment la plus intéressante est portée par le couple (VPP, VPN) mais :
 - . leurs valeurs dépendent de la prévalence $p(=P(M)=m/n)$ au sein de l'échantillon, et la représentativité de cet échantillon est non seulement rarement assurée mais encore pas forcément souhaitable dans le cas d'une pathologie rare où l'on serait amené à estimer des pourcentages à partir d'un effectif m faible.

. un couple de valeurs n'est pas un outil très commode de comparaison entre 2 ou plusieurs tests concurrentiels.

- Le couple (Se, Sp) (ou son homologue (FNF, FPF)) est soumis à la dernière critique mais est par contre indépendant de p :
- A dépend de p :
- Y et LR n'ont pas de signification probabilistique évidente, et le fait que pour un même Y on puisse avoir des LR très différents (et vice-versa) montre leurs limites.

D'autres approches paraissent donc nécessaires.

Coefficient de transmission :

La théorie de l'information permet de définir l'entropie d'information associée à un signal aléatoire pouvant prendre n valeurs assorties de probabilités P_i ($i=1..n$) :

$$H = \sum_{i=1}^n -P_i \cdot \log_2 P_i$$

avec $H = 1$ si $\forall i, P_i = \frac{1}{n}$ (incertitude maximale)

et $H \rightarrow 0$ si i tel que $P_i \rightarrow 0$ ou $P_i \rightarrow 1$ (incertitude minimale).

Le test diagnostique peut être considéré comme un système qui reçoit un signal binaire aléatoire M ou \bar{M} , et émet un signal binaire aléatoire $T+$ ou $T-$. On peut alors calculer (4) :

- l'entropie d'entrée :

$$H(E) = -P(M) \cdot \log_2 P(M) - P(\bar{M}) \cdot \log_2 P(\bar{M})$$
- l'entropie de sortie :

$$H(S) = -P(T+) \cdot \log_2 P(T+) - P(T-) \cdot \log_2 P(T-)$$
- les entropies conditionnelles :

$$H(E/S) = \sum_{E=T+} \sum_{S=T+} -P(Ts) \cdot P(Me/Ts) \cdot \log_2 P(Me/Ts)$$

$$H(S/E) = \sum_{E=T+} \sum_{S=T+} -P(Me) \cdot P(Ts/Me) \cdot \log_2 P(Ts/Me)$$
- l'entropie d'entrée-sortie :

$$H(E, S) = H(E) + H(S/E) = H(S) + H(E/S)$$
- et enfin le coefficient de transmission du test :

$$CT = \frac{H(E) + H(S) - H(E, S)}{\text{MIN}(H(E), H(S))}$$

Mais là encore CT dépend de $P(M)$, c'est à dire la prévalence p dans l'échantillon, et les mêmes restrictions que pour (VPP, VPN) et A peuvent s'appliquer.

Nous proposons donc de calculer l'espérance mathématique de CT en supposant p variable aléatoire uniforme sur $[0, 1]$. Cette hypothèse n'introduit aucun préjugé sur le type de patients auxquels le test va être appliqué. et permet donc une évaluation certes un peu ésotérique mais en tous cas parfaitement neutre du comportement du test en tant que système transmetteur de signaux aléatoires binaires, un peu de la même façon qu'on caractérise une "boîte noire" par sa réponse impulsionnelle, c'est à dire son comportement face à l'entrée élémentaire ou constitue le Dirac. Ici l'entrée élémentaire est une prévalence aléatoire uniforme sur $[0, 1]$.

Les tests diagnostiques en médecine considérés comme des systèmes transmetteurs de signaux aléatoires binaires : espérance mathématique du coefficient de transmission et du gain informatif.

Calcul de E (CT) :

- on a $CT = g(p)$ avec $p \sim U[0,1]$ de densité $f(p)$ et de répartition $F(p)$

donc $f(p) = 1$ si $p \in [0,1]$
 $f(p) = 0$ sinon

$$E(CT) = E(g(p)) = \int_0^1 g(p) dF(p) = \int_0^1 g(p) \cdot f(p) \cdot dp$$

$$E(CT) = \int_0^1 g(p) \cdot f(p) dp = \int_0^1 g(p) dp$$

- d'où $E(CT) = \int_0^1 CT \cdot dp$

- CT peut être mis sous la forme d'une fonction de (Se, Sp, p) ; Se et Sp sont tirés de la MI et E (CT) est calculé par la méthode des trapèzes.

Gain informatif :

Si on adopte une approche probabilistique bayésienne on peut développer une autre méthode : un test est utile si les probabilités à postériori $P(M/T+)$ et $P(\bar{M}/T-)$ sont respectivement supérieures aux probabilités a priori $P(M)$ et $P(\bar{M})$. C'est une autre façon de dire que le test doit diminuer l'incertitude.

Nous proposons de chiffrer ce "gain informatif" par :

$$GI = [P(M/T+) - P(M)] + [P(\bar{M}/T-) - P(\bar{M})]$$

au'on peut réduire à :

$$GI = P(M/T+) - P(M/T-)$$

Là encore GI est une fonction de p et nous proposons la même démarche que pour CT : calcul de E (GI) si $p \sim U[0,1]$.

On trouve une formule similaire :

$$E(GI) = \int_0^1 GI \cdot dp$$

GI peut être exprimé en fonction de (Se, Sp, p) et une résolution analytique de l'intégrale fournit :
 $E(GI) =$

$$\frac{(Se+Sp-1) - Se(1-Sp)\ln(Se/1-Sp) - Sp(1-Se)\ln(Sp/1-Se)}{(Se+Sp-1)^2}$$

Il est même possible de spécifier la population la plus à même de tirer bénéfice du test en calculant la prévalence pour GI maximum : $\frac{dGI}{dp} = 0 \Rightarrow$

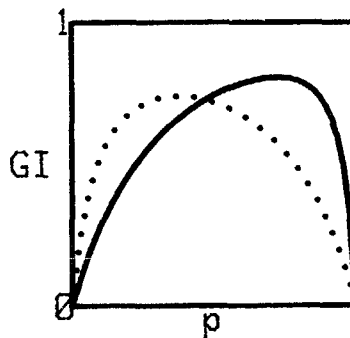
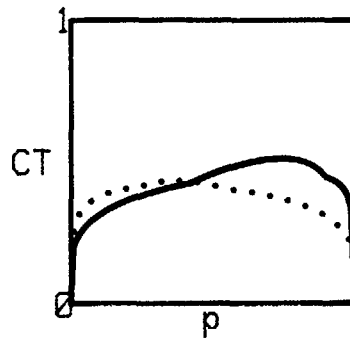
$$P_m = \frac{Sp(1-Sp) \pm \sqrt{Sp(1-Sp)Se(1-Se)}}{(Se+Sp-1)(Se-sp)}$$

(une seule des 2 racines est comprise entre 0 et 1).

Comparaison de tests par CT et GI :

Si on trace CT (ou GI) en fonction de p on obtient une courbe hyperboloïde. La surface sous la courbe est d'autant plus grande que Sp et Se sont élevés et vaut E(CT) (ou E(GI)). La courbe est d'autant plus dissymétrique que Se et Sp sont différents, mais ce phénomène est plus marqué pour GI que pour CT (voir figures). Ce dernier point, le calcul plus facile de E(GI) et surtout de p_m nous font préférer GI à CT.

Nous proposons donc d'évaluer un test diagnostique, système transmetteur de signaux aléatoires binaires, par les caractéristiques E (GI) et p_m de la courbe GI en fonction de P.



| | A | B... |
|-------|------|------|
| Se | 0,97 | 0,77 |
| Sp | 0,73 | 0,93 |
| Y | 0,7 | 0,7 |
| LR | 3,6 | 11 |
| E(CT) | 0,42 | 0,37 |
| E(GI) | 0,6 | 0,55 |
| p_m | 0,72 | 0,38 |

Exemple : comparaison de 2 tests

- L'analyse classique concluerait que B est meilleur que A ;
- CT et GI montrent tous deux qu'en fait A est globalement meilleur mais surtout que A est adapté aux populations à haut risque alors que B convient à celles à bas risque.

- 1 : McNEIL B.J., ADELSTEIN S.J. : Determining the value of diagnostic and screening tests. J. Nucl. Med., 1976, 17, 6, 439-448.
- 2 : PATTON D.D. : Introduction to clinical decision making. Semin. Nucl. Med., 1978, 8, 4, 273-282.
- 3 : ALPEROVITCH A. : Assessment of diagnostic tests. Bull. Cancer, 1980, 67, 4, 376-383.
- 4 : VITULO F., PARISI E., LAVAGNINI L., ASCENTI E. : La théorie de l'information. Standardisation d'une méthode d'analyse statistique pour l'évaluation des examens instrumentaux alternatifs. Ann. Radiol. 1983, 26, 5, 381-385.

