Clinica Chimica Acta 425 (2013) 97–103

Contents lists available at ScienceDirect

Clinica Chimica Acta

journal homepage: www.elsevier.com/locate/clinchim

Beyond the 2 × 2 -contingency table: A primer on entropies and mutual information in various scenarios involving *m* diagnostic categories and *n* categories of diagnostic tests

Gilbert Reibnegger *

Institute of Physiological Chemistry, Center of Physiological Medicine, Medical University of Graz, A-8010 Graz, Austria

ARTICLE INFO

Article history:
Received 18 June 2013
Received in revised form 10 July 2013
Accepted 10 July 2013
Available online 22 July 2013

Keywords:
Diagnostic test evaluation
Information theory
Entropy
Mutual information

ABSTRACT

Background: Usual evaluation tools for diagnostic tests such as, sensitivity/specificity and ROC analyses, are designed for the discrimination between two diagnostic categories, using dichotomous test results. Information theoretical quantities such as mutual information allow in depth-analysis of more complex discrimination problems, including continuous test results, but are rarely used in clinical chemistry. This paper provides a primer on useful information theoretical concepts with a strong focus on typical diagnostic scenarios.

Methods and results: Information theoretical concepts are shortly explained. MATHEMATICA CDF documents are provided which compute entropies and mutual information as function of pretest probabilities and the distribution of test results among the categories, and allow interactive exploration of the behavior of these quantities in comparison with more conventional diagnostic measures. Using data from a previously published study, the application of information theory to practical diagnostic problems involving up to 4 × 4 -contingency tables is demonstrated.

Conclusions: Information theoretical concepts are particularly useful for diagnostic problems requiring more than the usual binary classification. Quantitative test results can be properly analyzed, and in contrast to popular concepts such as ROC analysis, the effects of variations of pre-test probabilities of the diagnostic categories can be explicitly taken into account.

1. Introduction

Evaluation of diagnostic tests is an important issue in medical disciplines. The overwhelming part of scientific literature in this field considers diagnostic test situations which can be conveniently represented by means of a 2 × 2 -contingency table: one dimension of such a table is defined by two diagnostic categories (e.g., "non-diseased" versus "diseased"), and the second dimension represents the dichotomous test result (e.g., "normal" versus "pathological"). A recent series of reviews provide an excellent overview of relevant methods and discuss the advantages as well as the limitations and problems of such analyses [1–4].

Information theoretical concepts for evaluation of diagnostic tests are not mentioned in these reviews. This is probably due to their somewhat higher complexity together with a lack of commercially available software and, hence, rather scarce presence in medical literature. However, these concepts have, on the one hand, a sound theoretical foundation [5,6] and, on the other hand, analysis of diagnostic situations with higher complexity than a 2 × 2 -table is straightforward within the framework of information theory [7–11]. Moreover, information theory makes explicit use of the pretest probabilities of the diagnostic categories investigated.

Here I intend to remind researchers of the potential of information theory in the field of test evaluation: by providing MATHEMATICA demonstration objects (CDF documents) I offer tools to interactively explore the properties of information theoretical quantities in comparison with traditional test evaluation methods. Using MATHEMATICA notebooks I demonstrate how information theory enables practical solutions for the treatment of various typical diagnostic scenarios involving two or more diagnostic categories and one or two quantitative diagnostic tests.

2. Methods

2.1. Demonstration data

For demonstration, I employ data from a previous study on the concentrations of neopterin, a marker for activated human and primate mononuclear cells, in serum and CSF samples from 218 children aged from 4 to 18 years [12]. Of these, 91 suffered from various diseases but had neither signs of central nervous nor of peripheral inflammation (diagnostic category 1), 43 had definite neuroborreliosis (category 2),

* Corresponding author at: Institute of Physiological Chemistry, Center of Physiological Medicine, Medical University of Graz, Harrachgasse 21/II, A-8010 Graz, Austria. Tel.: +43 664 92 20 577; fax: +43 316 380 9610.
E-mail address: gilbert.reibnegger@medunigraz.at.

0009-8981/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cca.2013.07.011

51 suffered from other central nervous system infections (category 3) and 33 had peripheral infections (category 4).

Medians and ranges of neopterin (nmol L^{-1}) in CSF in the four diagnostic categories 1 to 4 were, respectively: 3.6 (0.2–9.3), 24.0 (5.9–45.0), 37.8 (14.4–84.1) and 6.9 (2.5–13.4); in serum: 6.3 (2.7–10.7), 7.4 (3.4–14.5), 10.2 (4.9–19.1), and 21.6 (13.8–41.9). Briefly, in category 1 both serum and CSF neopterin concentrations were low. CSF neopterin was raised in category 2 and, even stronger, in category 3. Serum neopterin was raised only in peripheral infections (category 4).

2.2. Basics of information theoretical evaluation of diagnostic tests

Prior to performing a diagnostic test, there exists some a priori probabilities for a patient to suffer from a suspected disease or not (pre-test probabilities). There exists uncertainty about the true status of the patient. After applying a test with known diagnostic sensitivity and specificity, we can estimate the a posteriori probabilities (post-test probabilities) of disease or non-disease using Bayes' theorem. The test outcome normally should lead to reduction of the prior uncertainty. In the terminology of information theory, the patient is regarded to be a "sender" in the sense of being a signal source (clinical symptoms, results of diagnostic tests, etc.), and the physician functions as a "receiver" trying to decipher the signals sent by the patient. The communication channel between sender and receiver is rarely free from disturbance, and even after receiving a signal (such as a test result), some uncertainty will probably remain at the site of the receiver about the sender's true status because, in conventional terminology, the diagnostic test used may lead with some probability to false positive or false negative decisions, due to imperfect specificity and sensitivity.

The information theoretical concept of uncertainty is closely related to probabilities: Shannon [5,6] chose a logarithmic measure to define uncertainty, or "entropy" as it is usually called, based on the respective probabilities of all possible events:

$$H = -\sum_{i=1}^{n} P_i \cdot ld(P_i).$$

Here, H is the entropy and the P_i are the probabilities with which each of n possible, mutually exclusive, outcomes or events may occur (note that $\sum_{i=1}^{n} P_i = 1$; the probabilities of all possible events sum up to unity). *ld* (*logarithmus dualis*) denotes the logarithm to base 2; this base is usually employed in information theory, and the information theoretical quantities then carry the unity bit (short for binary digit).

To give the simplest example, suppose we throw a coin. The two outcomes (head and tail) for an ideal (fair) coin both have probabilities $P_{Head} = P_{Tail} = \frac{1}{2}$, and thus, $H = -\frac{1}{2} ld(\frac{1}{2}) - \frac{1}{2} ld(\frac{1}{2}) = 1$ bit.

However, if an unfair coin had been used with, say, $P_{Head} = \frac{3}{5}$ and $P_{Tail} = \frac{2}{5}$, $H = -\frac{3}{5} \mathrm{ld}\left(\frac{3}{5}\right) - \frac{2}{5} \mathrm{ld}\left(\frac{2}{5}\right) = 0.97$ bit: the entropy is smaller because we know in advance that there is a slightly higher probability for outcome "head".

Similarly, if we had a patient for whom we estimate a pre-test probability for having a certain disease of 50%, our prior uncertainty was 1 bit. Could we apply a perfect diagnostic test to this patient, and this test would yield a positive result (i.e., the patient with certainty suffers from the disease), the test would have reduced our prior uncertainty of 1 bit to zero; the information gain or "mutual information" is 1 bit. As the example with the unfair coin demonstrates, the expected information gain through a perfect diagnostic test would decline for pre-test probabilities of disease either smaller or larger than 50%.

Real diagnostic tests tend to be more or less imperfect; their sensitivities and specificities will be smaller than 100%. Therefore, information theoretical treatment of real diagnostic tests is somewhat more involved. Briefly, there is the input entropy $H(D)$ depending solely on the pre-test probabilities of the diagnostic categories. The output entropy $H(T)$ depends on the probabilities of the test outcomes. If the diagnostic categories and the diagnostic test possessed some mutual association, $H(D)$ and $H(T)$ would overlap to some degree, and the joint entropy $H(D,T)$ would be smaller than the sum $H(D) + H(T)$. The mutual information (transinformation) is defined as the extent of this overlap:

$$I(D;T) = H(D) + H(T) - H(T,D).$$

The difference $H(D)$–$I(D;T)$ is called equivocation entropy $H(D|T)$; it denotes the uncertainty about the disease status of the patient after having performed the test. The difference $H(T)$–$I(D;T)$ is denoted ambiguity entropy $H(T|D)$; it is the uncertainty about test outcome when the disease status is given. For practical applications, the mutual information is the most important quantity because it measures the reduction of uncertainty by applying the diagnostic test.

2.3. Notation of relevant quantities

In the following sections, I shall adopt the following notation (upper part of Table 1): diagnostic categories are denoted $D1$, $D2$,..., and test outcomes by $T1$, $T2$, ... Pre-test probabilities for diagnostic category i are given by $P(Di)$.

In a test evaluation experiment we obtain a contingency table reflecting the absolute frequencies of all patients according to their classification by disease category and test result. The probabilities $P(Di,Tj)$ are obtained by dividing the respective cell frequency by the total number of subjects investigated.

Table 1
Relevant probabilities and entropies for 3 × 3-contingency tables (for 2 × 2-tables, omit all terms including "3").

Probabilities	Test result T1	Test result T2	Test result T3	Row sums
Diagnostic category D1	$P(D1,T1) = P(T1\|D1) \cdot P(D1)$	$P(D1,T2) = P(T2\|D1) \cdot P(D1)$	$P(D1,T3) = P(T3\|D1) \cdot P(D1)$	$P(D1)$
Diagnostic category D2	$P(D2,T1) = P(T1\|D2) \cdot P(D2)$	$P(D2,T2) = P(T2\|D2) \cdot P(D2)$	$P(D2,T3) = P(T3\|D2) \cdot P(D2)$	$P(D2)$
Diagnostic category D3	$P(D3,T1) = P(T1\|D3) \cdot P(D3)$	$P(D3,T2) = P(T2\|D3) \cdot P(D3)$	$P(D3,T3) = P(T3\|D3) \cdot P(D3)$	$P(T3)$
Column sums	$P(T1)$	$P(T2)$	$P(T3)$	
Input entropy	$-(P(D1) \cdot ld(P(D1)) + P(D2) \cdot ld(P(D2)) + P(D3) \cdot ld(P(D3)))$			
Output entropy	$-(P(T1) \cdot ld(P(T1)) + P(T2) \cdot ld(P(T2)) + P(T3) \cdot ld(P(T3)))$			
Joint entropy	$-\begin{pmatrix} P(D1,T1) \cdot ld(P(D1,T1) + P(D1,T2) \cdot ld(P(D1,T2) + P(D1,T3) \cdot ld(P(D1,T3) + \\ P(D2,T1) \cdot ld(P(D2,T1) + P(D2,T2) \cdot ld(P(D2,T2) + P(D2,T3) \cdot ld(P(D2,T3) + \\ P(D3,T1) \cdot ld(P(D3,T1) + P(D3,T2) \cdot ld(P(D3,T2) + P(D3,T3) \cdot ld(P(D3,T3) \end{pmatrix}$			
Equivocation	Joint entropy − output entropy			
Ambiguity	Joint entropy − input entropy			
Mutual information	Input entropy + output entropy − joint entropy = joint entropy − equivocation − ambiguity			