Journal of Statistical Planning and Inference 142 (2012) 1141–1150

Contents lists available at SciVerse ScienceDirect

Journal of Statistical Planning and Inference

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

Nonparametric Predictive Inference for diagnostic accuracy

Tahani Coolen-Maturi [a], Pauline Coolen-Schrijner [b,1], Frank P.A. Coolen [b,*]

[a] Kent Business School, University of Kent, Canterbury, Kent CT2 7PE UK
[b] Department of Mathematical Sciences, Durham University, Durham DH1 3LE, UK

ARTICLE INFO

Article history:
Received 1 September 2010
Accepted 29 November 2011
Available online 6 December 2011

Keywords:
Area under the ROC curve (AUC)
Diagnostic accuracy
Lower and upper probability
Nonparametric Predictive Inference (NPI)
Partial area under the ROC curve (pAUC)
Receiver Operating Characteristic (ROC)

ABSTRACT

Measuring the accuracy of diagnostic tests is crucial in many application areas including medicine and health care. Good methods for determining diagnostic accuracy provide useful guidance on selection of patient treatment, and the ability to compare different diagnostic tests has a direct impact on quality of care. In this paper Nonparametric Predictive Inference (NPI) methods for accuracy of diagnostic tests with continuous test results are presented and discussed. For such tests, Receiver Operating Characteristic (ROC) curves have become popular tools for describing the performance of diagnostic tests. We present the NPI approach to ROC curves, and some important summaries of these curves. As NPI does not aim at inference for an entire population but instead explicitly considers a future observation, this provides an attractive alternative to standard methods. We show how NPI can be used to compare two continuous diagnostic tests.

1. Introduction

The evaluation of the accuracy of diagnostic tests has particular importance in medicine and health care. Diagnostic test results may have only two values (binary test), or a value in a finite number of ordered categories (ordinal test), or real values (continuous test). There are several accuracy measures which vary depending on the type of diagnostic test results. For example, the Receiver Operating Characteristic (ROC) curve is a common statistical tool for describing the performance of certain medical tests. It is used to measure the accuracy of a diagnostic test that yields ordinal or continuous results. The ROC curve plays an important role in many areas such as signal detection, radiology, machine learning, data mining and credit scoring.

In this paper we introduce Nonparametric Predictive Inference (NPI) for diagnostic accuracy for continuous test results. The case of ordinal test results requires further development of NPI for ordinal data and is left for future research. For accuracy of binary tests we are investigating an approach using NPI for Bernoulli data (Coolen, 1998), we hope to report on this in the near future. NPI is a statistical method based on Hill's assumption $A_{(n)}$ (Hill, 1968), which gives direct probabilities for a future observable random quantity, given observed values of related random quantities (Augustin and Coolen, 2004; Coolen, 2006). During the last decade, NPI has been developed for different applications in statistics, operational research, and reliability and risk analysis.

Section 2 gives a brief overview of NPI, and Section 3 gives an introduction to the concepts of continuous diagnostic tests used in this paper. NPI for such tests is introduced in Section 4. This includes NPI for ROC curves, the area under the ROC curve, and the partial area under the ROC curve, which are commonly used summaries of the ROC curve. Comparison of diagnostic tests is important with regard to guidance on most useful test methods, the NPI approach to such comparison

* Corresponding author.
E-mail addresses: T.Coolen-Maturi@kent.ac.uk (T. Coolen-Maturi), frank.coolen@durham.ac.uk (F.P.A. Coolen).
[1] Pauline died on 23.04.2008, when the research reported in this paper was at an advanced stage.

0378-3758/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2011.11.015

of continuous diagnostic tests is discussed, and illustrated via an example, in Section 5. The paper is finished with some concluding remarks in Section 6.

2. Nonparametric Predictive Inference (NPI)

Nonparametric Predictive Inference (NPI) is a statistical method based on Hill's assumption $A_{(n)}$ (Hill, 1968), which gives direct probabilities for a future observable random quantity, given observed values of related random quantities (Augustin and Coolen, 2004; Coolen, 2006). To introduce $A_{(n)}$ we use the following notation. Suppose that $X_1,\ldots,X_n$, X_{n+1} are real-valued absolutely continuous and exchangeable random quantities. Let the ordered observed values of $X_1,\ldots,X_n$ be denoted by $x_1<x_2<\cdots<x_n$, and let $x_0=-\infty$ and $x_{n+1}=\infty$ for ease of notation. For ease of presentation, throughout this paper we assume that no ties occur, it is easy to generalize the method to allow ties by breaking the ties in all possible ways and deriving the overall NPI lower and upper probabilities as the minimum and maximum, respectively, of the lower and upper probabilities corresponding to each way of breaking the ties (Maturi, 2010). For X_{n+1}, representing a future observation, based on n observations, $A_{(n)}$ (Hill, 1968) is

$$P(X_{n+1}\in(x_j,x_{j+1}))=\frac{1}{n+1},\quad j=0,1,\ldots,n \tag{1}$$

$A_{(n)}$ does not assume anything else, and can be considered to be a post-data assumption related to finite exchangeability (De Finetti, 1974). Hill (1988) discusses $A_{(n)}$ in detail, including its justification from several foundational perspectives. Inferences based on $A_{(n)}$ are predictive and nonparametric, and can be considered suitable if there is hardly any knowledge about the random quantity of interest, other than the n observations, or if one does not want to use such information, e.g. to study effects of additional assumptions underlying other statistical methods. We consider NPI as a frequentist statistical framework, which has the important advantages of being exactly calibrated (Lawless and Fredette, 2005) and not requiring the use of counterfactuals nor relying on asymptotic justifications as is the use for many commonly used frequentist statistics procedures. $A_{(n)}$ is not sufficient to derive precise probabilities for many events of interest, but it provides bounds for probabilities via the 'fundamental theorem of probability' (De Finetti, 1974), which are lower and upper probabilities in interval probability theory (Walley, 1991; Weichselberger, 2001). For more details on NPI, including references to applications in statistics and related topics, see www.npi-statistics.com.

3. Continuous diagnostic tests

In this section, some common measures of diagnostic accuracy are briefly reviewed, following Pepe (2003) and Dodd and Pepe (2003). Let D be a binary variable describing the disease status, i.e. $D=1$ for disease and $D=0$ for non-disease. Suppose that Y is a continuous random quantity of a diagnostic test result, and that large values of Y are considered more indicative of disease. Using a threshold c, the test result is called positive if $Y>c$, so if it indicates the disease, and negative if $Y\le c$, where $c\in(-\infty,\infty)$.

The sensitivity (SN) of a test is the probability of a positive test result for an individual with the condition (disease), and is also known as True Positive Fraction (TPF). The specificity (SP) is the probability of a negative test result for an individual without the condition (non-disease). Clearly, an accurate diagnostic test will have sensitivity and specificity both close to one. The False Positive Fraction (FPF) is the probability of a positive test result for an individual without the condition, hence $\mathrm{SP}=1-\mathrm{FPF}$.

Throughout this paper, Y^1 and Y^0 are used to refer to test results for the disease and non-disease groups, respectively, and n_1 and n_0 are the corresponding numbers of individuals in the disease and the non-disease groups. With threshold $c\in(-\infty,\infty)$, $\mathrm{FPF}(c)$ and $\mathrm{TFP}(c)$, are the main probabilities considered in this paper for continuous diagnostic tests, with

$$\mathrm{FPF}(c)=P(Y^0>c|D=0)=S_0(c) \tag{2}$$

$$\mathrm{TPF}(c)=P(Y^1>c|D=1)=S_1(c) \tag{3}$$

where $S_0(c)$ and $S_1(c)$ are the survival functions for the random quantities Y^0 and Y^1 for the diagnostic test results for the non-disease and disease groups, respectively.

3.1. Receiver Operating Characteristic (ROC) curve

The Receiver Operating Characteristic (ROC) curve is a common statistical tool for describing the performance of diagnostic tests which yield ordinal or continuous results. For continuous test results, the ROC curve is defined as the combination of FPF and TPF over all values of the threshold c, i.e.

$$\mathrm{ROC}=\{(\mathrm{FPF}(c),\mathrm{TPF}(c)),c\in(-\infty,\infty)\} \tag{4}$$