Computational Statistics and Data Analysis 62 (2013) 39–51

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

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

Estimation of the volume under the ROC surface with three ordinal diagnostic categories

Le Kang[a], Lili Tian[b,*]

[a] US Food and Drug Administration, 10903 New Hampshire Ave, Silver Spring, MD 20993, United States
[b] Department of Biostatistics, University at Buffalo, 3435 Main St, Buffalo, NY 14214, United States

ARTICLE INFO

Article history:
Received 26 October 2011
Received in revised form 5 January 2013
Accepted 6 January 2013
Available online 20 January 2013

Keywords:
Box–Cox type transformation
Diagnostic accuracy
Kernel smoothing
Three ordinal diagnostic categories
Volume under the ROC surface

ABSTRACT

With three ordinal diagnostic categories, the most commonly used measure for the overall diagnostic accuracy is the volume under the ROC surface (VUS), which is the extension of the area under the ROC curve (AUC) for binary diagnostic outcomes. This article proposes two kernel smoothing based approaches for estimation of the VUS. In an extensive simulation study, the proposed estimators are compared with the existing parametric and nonparametric estimators in terms of bias and root mean square error. A real data example of 203 participants from a cohort study for the detection of Glycan biomarkers for liver cancer is discussed.

1. Introduction

Diagnostic testing is an extremely important aspect of medical care. Medical diagnosis usually involves the classification of patients into two or more categories. The evaluation of a diagnostic test procedure involves estimation of parameters that describe the accuracy of diagnostic test and it is therefore of paramount importance to correctly estimate the diagnostic accuracy to decide on the best test for certain disease. When subjects are categorized in a binary fashion, i.e., non-diseased and diseased, a receiver operating characteristic (ROC) curve, defined as a plot of sensitivity versus 1-specificity, is an important statistical tool for evaluating the accuracy of continuous diagnostic tests, and the area under the ROC curve (AUC) is one of the common indices used for overall diagnostic accuracy (e.g., Zhou et al., 2002; Pepe, 2003; Shapiro, 1999; Faraggi and Reiser, 2002).

In many situations the diagnostic decision is not limited to a binary choice. For example, a clinical assessment, NPZ-8, of the presence of HIV-related cognitive dysfunction (AIDS Dementia Complex—ADC) would discriminate between patients exhibiting clinical symptoms of ADC (combined stages 1–3), subjects exhibiting minor neurological symptoms (ADC stage 0.5) and neurologically unimpaired individuals (ADC stage 0) (Nakas and Yiannoutsos, 2004). Another example provided by Xiong et al. (2006) concerns mild cognitive impairment (MCI) or early stage Alzheimer's disease (AD) being a transitional stage between the cognitive changes from normal aging and the more severe problems caused by the AD. Thereafter, we refer the disease status between "non-diseased" and "diseased" as "intermediate", in other words, transitional status. For such disease processes with three stages, binary statistical tools such as ROC curve and AUC need to be extended. The ROC surface and the volume under the surface (VUS) have been proposed to assess the accuracy of tests with three ordinal diagnostic categories; e.g., Scurfield (1996) and Nakas and Yiannoutsos (2004) discussed ROC surface for multi-class

* Correspondence to: Department of Biostatistics, University at Buffalo, 706 Kimball Tower, 3435 Main Street, Buffalo, NY 14214-3000, United States. Tel.: +1 716 829 2715; fax: +1 716 829 2200.
E-mail address: ltian@buffalo.edu (L. Tian).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2013.01.004

diagnostic problems and developed the VUS to summarize the diagnostic accuracy with three ordinal diagnostic categories; Nakas and Yiannoutsos (2004) and He and Frey (2008) also discussed the nonparametric estimation of a single VUS; Xiong et al. (2006) proposed to estimate the VUS under the normality assumption and Li and Zhou (2009) studied the estimation of three-dimensional ROC surfaces with nonparametric and semi-parametric estimators.

In this article, we focus on estimation of the VUS. The purpose of this paper is two-fold: (1) We propose two kernel smoothing based approaches (K1 and K2) for estimation of the VUS, and examine the use of Box–Cox type transformation before applying the kernel smoothing based approaches (K1T and K2T) and (2) We perform an extensive simulation study to compare bias and root mean square error (RMSE) of the proposed approaches with those of the existing parametric and nonparametric methods. One of the existing nonparametric VUS estimators is based on the Mann–Whitney U statistic (MW) and the other one is from empirical distribution plug-in (EP), and the existing parametric estimators are either based on the normality of the data (N) or the normality of transformed data using Box–Cox type transformation (NT). The details of the existing estimators of the VUS are reviewed in Section 2. The proposed kernel smoothing based approaches will be presented in Section 3. An extensive simulation study comparing different VUS estimation procedures in terms of bias and RMSE is conducted in Section 4. In Section 5, we apply all the methods to a real data set of 203 participants from a cohort study for the detection of Glycan biomarkers for liver cancer to estimate the diagnostic accuracy of a particular protein segment in discriminating different stages of liver cancer. Section 6 provides a concluding discussion.

2. Existing approaches for estimation of the VUS

This section presents the preliminaries of VUS (Section 2.1) and reviews the existing nonparametric and parametric VUS estimators (Sections 2.2 and 2.3).

2.1. Preliminaries

ROC surface, analogous to ROC curve, has been proposed to assess the accuracy of tests with three ordinal diagnostic categories. Let Y_1, Y_2 and Y_3 denote the scores resulting from a diagnostic test and let F_1, F_2 and F_3 be the corresponding cumulative distribution functions for non-diseased, intermediate and diseased subjects, respectively. Assume the results of a diagnostic test are measured on continuous scale and higher values indicate greater severity of the disease. Given a pair of threshold values c_1 and c_3 ($c_1 < c_3$), let $\delta_1 = F_1(c_1), \delta_3 = 1 - F_3(c_3)$ be the true classification rates for non-diseased and diseased category, respectively. Then the probability that a randomly selected subject from intermediate category has a score between c_1 and c_3 is

$$\delta_2 = F_2(c_3) - F_2(c_1) = F_2\left[F_3^{-1}(1-\delta_3)\right] - F_2\left[F_1^{-1}(\delta_1)\right]. \tag{1}$$

The triplet $(\delta_1, \delta_2, \delta_3)$, where $\delta_2 = \delta_2(\delta_1, \delta_3)$ is a function of (δ_1, δ_3), would produce an ROC surface in the three-dimensional space for all possible $(c_1, c_3) \in \mathbb{R}^2$. As the ROC curve for a binary diagnosis represents the trade-off between sensitivity and specificity for the two categories (non-diseased and diseased), the ROC surface represents the three-way trade-off among the correct classification probabilities for the three categories.

In order to summarize the overall diagnostic accuracy for the diagnostic test, the volume under the ROC surface (VUS) has been considered. It is defined as

$$\text{VUS} = \int_0^1 \int_0^{1-F_3\left[F_1^{-1}(\delta_1)\right]} F_2\left[F_3^{-1}(1-\delta_3)\right] - F_2\left[F_1^{-1}(\delta_1)\right] d\delta_3 d\delta_1. \tag{2}$$

This is a generalization of the AUC for an ROC curve under a binary classification. One could show that VUS is mathematically equivalent to the probability $P(Y_1 < Y_2 < Y_3)$ when Y_1, Y_2 and Y_3 are randomly selected from each diagnostic category, respectively. For a useless test (when Y_1, Y_2 and Y_3 have identical distributions), VUS is 1/6. For more details, see Xiong et al. (2006).

2.2. Nonparametric approaches (MW and EP)

Assume the sample sizes for non-diseased, intermediate and diseased subjects are n_1, n_2 and n_3, respectively. The unbiased nonparametric Mann–Whitney U statistic of the probability $P(Y_1 < Y_2 < Y_3)$, i.e., the VUS, is given by

$$\text{U} = \frac{1}{n_1 n_2 n_3} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} I(Y_{1i} < Y_{2j} < Y_{3k}), \tag{3}$$

where $I(\cdot)$ stands for the indicator function (Nakas and Yiannoutsos, 2004). We denote this estimator by MW. Li and Zhou (2009) proposed the nonparametric estimator of ROC surface by replacing all the cumulative distribution functions in (1) with their empirical counterparts. Thus, the estimated VUS is given by

$$\widehat{\text{VUS}} = \int_0^1 \int_0^{1-\hat{F}_3\left[\hat{F}_1^{-1}(\delta_1)\right]} \hat{F}_2\left[\hat{F}_3^{-1}(1-\delta_3)\right] - \hat{F}_2\left[\hat{F}_1^{-1}(\delta_1)\right] d\delta_3 d\delta_1, \tag{4}$$