Journal of Statistical Planning and Inference 141 (2011) 3829–3838

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Nonparametric methodology for the time-dependent partial area under the ROC curve

Hung Hung, Chin-Tsang Chiang *

Department of Mathematics, National Taiwan University, Taipei 10617, Taiwan, ROC

ARTICLE INFO

Article history:
Received 25 August 2009
Received in revised form
9 March 2011
Accepted 24 June 2011
Available online 7 July 2011

Keywords:
AUC
Bandwidth
Censoring time
FPR
Gaussian process
Kaplan–Meier estimator
Marker-dependent censoring
Nonparametric estimator
pAUC
ROC
Survival time
TPR

ABSTRACT

To assess the classification accuracy of a continuous diagnostic result, the receiver operating characteristic (ROC) curve is commonly used in applications. The partial area under the ROC curve (pAUC) is one of the widely accepted summary measures due to its generality and ease of probability interpretation. In the field of life science, a direct extension of the pAUC into the time-to-event setting can be used to measure the usefulness of a biomarker for disease detection over time. Without using a trapezoidal rule, we propose nonparametric estimators, which are easily computed and have closed-form expressions, for the time-dependent pAUC. The asymptotic Gaussian processes of the estimators are established and the estimated variance–covariance functions are provided, which are essential in the construction of confidence intervals. The finite sample performance of the proposed inference procedures are investigated through a series of simulations. Our method is further applied to evaluate the classification ability of CD4 cell counts on patient's survival time in the AIDS Clinical Trials Group (ACTG) 175 study. In addition, the inferences can be generalized to compare the time-dependent pAUCs between patients received the prior antiretroviral therapy and those without it.

1. Introduction

Decision-making is an important issue in many fields such as signal detection, psychology, radiology, and medicine. For example, preoperative diagnostic tests are medically necessary and implemented in clinical preventive medicine to determine those patients for whom surgery is beneficial. For the sake of cost-saving or performance improvement, new diagnostic tests are often introduced and the classification accuracies of them are evaluated and compared with the existing ones. The ROC curve, a plot of the true positive rate (TPR) versus the false positive rate (FPR) for each possible cut point, has been widely used for this purpose when the considered diagnostic tests are continuous. One advantage of the ROC curve is that it describes the inherent classification capability of a biomarker without specifying a specific threshold. Moreover, the invariance characteristic of ROC curve in measurement scale provides a suitable base to compare different biomarkers. Generally, the more the curve moves toward the point (0,1), the better a biomarker performs.

In many applications, the area under the ROC curve (AUC), one of the most popular summary measures of the ROC curve, is used to evaluate the classification ability of a biomarker. It has the probability meaning that the considered biomarker of a randomly selected diseased case is greater than that of a non-diseased one. Generally, a perfect biomarker will have the AUC

* Corresponding author.
E-mail address: chiangct@ntu.edu.tw (C.-T. Chiang).

0378-3758/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2011.06.025

of one while a poor one takes a value close to 0.5. Since the AUC is the whole area under the ROC curve, relevant information might not be entirely captured in some cases. For example, two crossed ROC curves can have the same AUC but totally different performance. Furthermore, there might be limited or no data in the region of high FPR. In view of these drawbacks, it is more useful to see the pAUC within a certain range of TPR or FPR. To evaluate the performance of several biomarkers, McClish (1989) adopted the summary measure pAUC for the FPR over a practically relevant interval. On the other hand, Jiang et al. (1996) showed that women with false-negative findings at mammography cannot be benefited from timely treatment of the cancer. Thus, these authors suggested using the pAUC with a portion of the true positive range in their applied data. Although their perspectives are different, the main frame is the same: only the practically acceptable area under the ROC curve is assessed. As mentioned by Dwyer (1996), the pAUC is a regional analysis of the ROC curve intermediate between the AUC and individual points on the ROC curve. The pAUC becomes a good measure of classification accuracy because it is easier for a practitioner to determine a range of TPR or/and FPR that are relevant. Several estimation and inference procedures have been proposed by Emir et al. (2000), Zhang et al. (2002), Dodd and Pepe (2003), among others. A more thorough understanding of the ROC, AUC, and pAUC can be also found in Zhou et al. (2002).

Recent research in ROC methodology has extended the binary disease status to the time-dependent setting. Let T denote the time to a specific disease or death and Y represent the continuous diagnostic marker measured before or onset of the study with joint survivor function $S(t,y)=P(T>t,Y>y)$. For each fixed time point t, the disease status can be defined as a case if $T\le t$ and a control otherwise. To evaluate the ability of Y in classifying subjects who is diseased before time t or not, Heagerty et al. (2000) generalized the traditional TPR and FPR to the time-dependent TPR and FPR as $\mathcal{T}_t(y)=P(Y>y|T\le t)$ and $\mathcal{F}_t(y)=P(Y>y|T>t)$, which can be further derived to be $(S(0,y)-S(t,y))/(1-S_T(t))$ and $S(t,y)/S_T(t)$ with $S_T(t)=S(t,-\infty)$. For the time-dependent AUC, Chambless and Diao (2006), Chiang et al. (2009), and Chiang and Hung (2010) proposed different nonparametric estimators and developed the corresponding inference procedures. As for the time-dependent pAUC, there is still too little research on this topic so far. We propose nonparametric estimators, which are shown to converge weakly to Gaussian processes, and the estimators for the corresponding variance–covariance functions. The established properties facilitate us to make inference on the time-dependent pAUC and can be reasonably applied to the time-dependent AUC because it is a special case of this summary measure.

The rest of this paper is organized as follows. In Section 2, the nonparametric estimation and inference procedures are proposed for the time-dependent pAUC. The finite sample properties of the estimators and the performance of the constructed confidence bands are studied through Monte Carlo simulations in Section 3. Section 4 presents an application of our method to the ACTG 175 study. In this section, an extended inference procedure is further provided for the comparison of the time-dependent pAUCs. Some conclusions and future works are addressed in Section 5. Finally, the proof of main results is followed in the Appendix.

2. Estimation and inferences

In this section, we estimate the time-dependent pAUC and develop the corresponding inference procedures. Without loss of generality, the time-dependent pAUC is discussed for restricted $\mathcal{F}_t(y)$ because that the time-dependent pAUC for restricted $\mathcal{T}_t(y)$ can be derived in the same way by reversing the roles of case and control subjects.

2.1. Estimation

Let X be the minimum of T and censoring time C, $\delta=I(X=T)$ represent the censoring status, and $q_{\alpha t}=\mathcal{F}_t^{-1}(\alpha)=\inf\{y:\mathcal{F}_t(y)\le\alpha\}$, $\alpha\in(0,1]$, denote the $(1-\alpha)$th quantile of Y conditioning on $\{T>t\}$ at the fixed time point t. Following the expression $\{-\int\mathcal{T}_t(y)\,d_y\mathcal{F}_t(y)\}$ for the time-dependent AUC, the time-dependent pAUC $\theta_t(q_{\alpha t})$ with the $FPR_t(y)$ less than α is derived to be functional of $S(t,y)$:

$$\Theta_\alpha(S)=\frac{-\int(S(0,u)-S(t,u))I(u\ge q_{\alpha t})d_uS(t,u)}{S_T(t)(1-S_T(t))}\quad\text{for } t\in(0,\tau] \text{ with } P(X>\tau)>0. \tag{2.1}$$

Note that the value of $\theta_t(q_{\alpha t})$ for a perfect biomarker should be α while a useless one is $0.5\alpha^2$. Same with the interpretation of Cai and Dodd (2008), the rescaled time-dependent pAUC $\theta_t(q_{\alpha t})/\alpha$ can be explained as the probability that the test result of a case $\{T_i\le t\}$ is higher than that of a control $\{T_j>t\}$ with its value exceeding $q_{\alpha t}$ for $i\ne j$, i.e., $P(Y_i>Y_j|T_i\le t,T_j>t,Y_j>q_{\alpha t})$.

From the formulation in (2.1), an estimator of $\theta_t(q_{\alpha t})$ can be obtained if $S(t,y)$ is estimable. Under marker-dependent censoring (T and C are independent conditioning on Y), Akritas (1994) suggested estimating $S(t,y)$ by $\widehat{S}(t,y)=n^{-1}\sum_{i=1}^n\widehat{S}_T(t|Y_i)I(Y_i>y)$, where

$$\widehat{S}_T(t|y)=\prod_{\{i:X_i\le t,\delta_i=1\}}\left\{1-\frac{K_\lambda(\widehat{S}_Y(Y_i)-\widehat{S}_Y(y))}{n\widehat{S}_X(X_i|y)}\right\} \tag{2.2}$$

is an estimator of $S_T(t|y)=P(T>t|Y=y)$ with $\widehat{S}_Y(y)=n^{-1}\sum_{j=1}^n I(Y_j>y)$ and $\widehat{S}_X(t|y)=n^{-1}\sum_{j=1}^n I(X_j\ge t)K_\lambda(\widehat{S}_Y(Y_j)-\widehat{S}_Y(y))$ being estimators of $S_Y(y)=P(Y>y)$ and $S_X(t|y)=P(X>t|Y=y)$. Here, $K_\lambda(u)=(2\lambda)^{-1}I(|u|<\lambda)$ and λ is a nonnegative smoothing