

ROC Analysis for Multiple Markers with Tree-Based Classification

Mei-Cheng Wang and Shanshan Li

Abstract Multiple biomarkers are frequently observed or collected for detecting or understanding a disease. The research interest of this paper is to extend tools of ROC analysis from univariate marker setting to multivariate marker setting for evaluating predictive accuracy of biomarkers using a tree-based classification rule. Using an arbitrarily combined and-or classifier, an ROC function together with a weighted ROC function (WROC) and their conjugate counterparts are introduced for examining the performance of multivariate markers. Specific features of the ROC and WROC functions and other related statistics are discussed in comparison with those familiar properties for univariate marker. Nonparametric methods are developed for estimating the ROC and WROC functions, and area under curve (AUC) and concordance probability. With emphasis on population average performance of markers, the proposed procedures and inferential results are useful for evaluating marker predictability based on multivariate marker measurements with different choices of markers, and for evaluating different and-or combinations in classifiers.

Introduction

The Receiver Operating Characteristic (ROC) analysis has been widely used as tools for assessing the discriminant performance for biomarkers. Based on a univariate or combined-to-univariate marker, the ROC curve is known as a plot of the true positive rate versus the false positive rate for each possible cut point, for summarizing

The paper appeared in volume 19 (2013) of *Lifetime Data Analysis*.

M.-C. Wang (✉) • S. Li

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health,
615 N. Wolfe Street, Baltimore, MD 21205, USA
e-mail: mcwang@jhsph.edu; shli@jhsph.edu

sensitivity and specificity of a binary classifier system when marker measurements are continuous. In nonparametric, semiparametric or parametric models, the ROC curve and its associated measures such as area under curve (AUC) or partial area under curve (pAUC) have been used as useful indices for evaluating the predictive accuracy of markers or diagnostic tests [17]. In statistical literature, different measures have been developed to summarize and compare the predictive accuracy of biomarkers ([2, 8] among others).

This paper considers situations when multiple markers (M_1, M_2, \dots, M_k) are available for classification of disease state. The research interest is to establish criterion and tools for assessing predictive accuracy based on multivariate markers or multivariate test measurements, (M_1, M_2, \dots, M_k) , from observed data or a training data set. The proposed work includes at least two types of applications: (i) to quantify the result of dual or multiple readings from a single diagnostic test, or readings from multiple tests; (ii) to evaluate the predictability of combined multiple markers for a disease, where each marker characterizes a specific biological function for the disease. For the first type of applications, (i), multiple reading is employed for either reducing uncertainty of test classification or comparison of multiple diagnostic modalities [9, 15]. Applications of the second type, (ii), are important when multivariate markers are used as prognostic measurements for predicting or understanding the disease.

To analyze multiple marker data, several approaches have been developed to handle the correlation structure of marker measurements for different research goals. The most common approach is perhaps to combine multiple markers into a single composite score using logistic regression model, and evaluate the predictability of markers by the one-dimensional composite score [14]. For high-dimensional markers, or when markers come from different biological sources, it may not be analytically appropriate to combine the markers into a composite score and, in such situations, the tree-based regression model could serve as a good alternative for identifying a classification rule. The tree-based classification method is sometimes referred to as recursive partitioning, which is frequently used in data mining, machine learning and clinical practice as a predictive model [3, 22]. For example, Baker [1] and Etzioni et al. [6] considered discretized markers by keeping the marker values in multi-dimensional settings and proposed new definitions for ROC curves.

When markers are continuous, Jin and Lu [13] considered bivariate markers and proposed to use the area under the upper boundary of ROC region to evaluate diagnostic utilities. Jin and Lu's work can be viewed as an extension of Baker's approach [1] from discrete markers to continuous markers. Wang and Li [21] defined an ROC function for bivariate continuous markers via generalized inverse set of the quantile function FP , where the ROC function possesses a conditional expectation expression. In this paper, we generalize Wang and Li's results from bivariate marker to multivariate marker setting, and develop methods and inference for ROC analysis.

Assume a k -dimensional marker vector (M_1, M_2, \dots, M_k) is available and the disease state is determined by a sequence of arbitrarily combined and-or classifier with positivity specified in either direction of marker values; for example, $I((M_1 \geq$

m_1 or $M_2 < m_2$) and ($M_3 < m_3$ or $M_4 \geq m_4$). This extension links to potential applications related to classification tree with binary decision diagrams. The research interest is to establish criterion and tools for assessing predictive accuracy based on multivariate markers, (M_1, M_2, \dots, M_k) . Specifically, the ROC function is extended from univariate case to multivariate case, and a weighted ROC (WROC) function is introduced for examining the performance of predictive accuracy with arbitrarily combined and-or classifiers.

Let $(M_{l1}, M_{l2}, \dots, M_{lk})$, $l = 0, 1$, be the marker vector for a non-diseased or diseased subject. Let the arbitrarily combined and-or classifier be expressed as $I\{(M_{l1}, M_{l2}, \dots, M_{lk}) \in D(m_1, m_2, \dots, m_k)\}$ with $D(m_1, m_2, \dots, m_k) \subseteq R^k$ defined as the region for marker-based positivity. To simplify notation and formulation, hereafter we shall use bold face \mathbf{m} to represent the vector (m_1, m_2, \dots, m_k) , and let \mathbf{m}_l and \mathbf{M}_l , $l = 0, 1$, represent the vectors $(m_{l1}, m_{l2}, \dots, m_{lk})$ and $(M_{l1}, M_{l2}, \dots, M_{lk})$. Define the false and true positive rates respectively as

$$FP(\mathbf{m}) = P\{\mathbf{M}_0 \in D(\mathbf{m})\}, \quad TP(\mathbf{m}) = P\{\mathbf{M}_1 \in D(\mathbf{m})\}$$

The research interest is to extend rules and tools from univariate marker to multivariate marker setting for assessment of predictive accuracy of markers.

Using the US Alzheimer's Disease Neuroimaging Initiative (ADNI) data set as an example, the biomarkers of interest include measurements from different biological systems related to neuroimaging, genetics, CSF (Cerebrospinal fluid) and cognition. As the k markers are identified from different biological sources, it may not be appropriate to combine them using, say, a linear combination of the measurements. The and-or classifier also signifies the importance of interaction between markers. For example, using an Alzheimer's Disease study that the authors are currently involved (the BIOCARD study at Johns Hopkins School of Medicine), decreases in CSF Amyloid beta-42 and/or increases in total tau or phosphorylated-tau (p-tau) are hypothesized as strong predictors for AD or AD-related symptoms. It would be interesting to keep the k markers in multivariate setting and explore their respective roles and interaction nonparametrically.

The paper is organized as follows. Section "Univariate Marker Case" briefly reviews some of the fundamental definitions and properties for univariate ROC analysis, where emphasis is placed on those which will be extended to multivariate setting. In sections "Multivariate Markers: ROC, WROC and AUC" and "Other Types of ROC and WROC Functions", a set of ROC and ROC-related functions are introduced with discussion focused on contrasting features between univariate and multivariate cases. Section "Nonparametric Estimation" considers nonparametric estimators for ROC-related functions, AUC and concordance probabilities. Simulation and a real data analysis are presented in section "Simulation and Data Example" to illustrate the applicability of the proposed procedures. Section "Discussion" concludes the paper with a brief discussion.

Univariate Marker Case

In the section we consider the univariate marker case, $k = 1$. Suppose the disease outcome D takes binary values 0 or 1, and M is a continuous marker variable. Let M_0 and M_1 respectively represent the marker variable from non-diseased ($D = 0$) and diseased ($D = 1$) group. Define $TP(m) = P(M_1 > m) = P(M > m | D = 1)$ as the true positive rate (sensitivity), and $FP(m) = P(M_0 > m) = P(M > m | D = 0)$ the false positive rate ($1 - \text{specificity}$). Assume M_0 and M_1 are independent. Define $F_0(m) = 1 - FP(m)$ and $F_1(m) = 1 - TP(m)$ respectively as the cumulative distribution function of M_0 and M_1 .

There are multiple ways to define the ROC function for a univariate marker. A mathematically simple definition $ROC(q) = TP[FP^{-1}(q)]$, $q \in [0, 1]$, evaluates the magnitude of true positive rate at controlled false positive rate through inverse functional mapping between FP and TP . The comparison of two ROC functions from different markers should thus be interpreted as the comparison of TP values with the same FP rate. The partial area under ROC curve for false positive rate less than p , $0 \leq p \leq 1$, is defined as $AUC(p) = \int I(0 \leq q \leq p) ROC(q) dq$. The area under ROC curve is defined as the total area with the FP rate ranging from 0 to 1, that is, $AUC(1)$. Define the partial concordance probability as $CON(p) = P(M_1 > M_0, FP(M_0) \leq p)$. For univariate marker model, the quantile variable $Q_0 = FP(M_0)$ is Uniform $[0, 1]$ distributed and thus $CON(p)$ can be calculated using probability measure on (M_1, Q_0) and is simplified to

$$\begin{aligned} CON(p) &= P(M_1 > FP^{-1}(Q_0), Q_0 \leq p) = \int_0^p \int I(m_1 > FP^{-1}(q)) dF_1(m_1) dq \\ &= \int_0^p ROC(q) dq = AUC(p) \end{aligned}$$

Thus, an alternative way to define $ROC(p)$ is to obtain it as the derivative of the partial concordance probability with respect to p , namely $ROC(p) = CON'(p)$. By definition, $CON(p)$ can also be expressed as

$$CON(p) = \int \int I(m_1 > m_0) I(FP(m_0) \leq p) dF_1(m_1) dF_0(m_0) \tag{1}$$

The equivalence between $CON(p)$ and $AUC(p)$ has led to development of non-parametric approaches for estimating $AUC(p)$ using the formula in (1). Dodd and Pepe [4] showed that the partial area under curve possesses a concordance probability expression: Let $p_0^* = FP(TP^{-1}(p_0))$ and assume $p_0^* < p_1$, then

$$\int I(p_0^* \leq q < p_1) ROC(q) dq = P(M_1 > M_0, FP^{-1}(p_1) < M_0 \leq TP^{-1}(p_0)) \tag{2}$$

Thus, the partial concordance probability coincides with the partial AUC restricted to the interval that false positive rate less than p_1 and true positive rate greater than p_0 . As proposed by Dodd and Pepe [4], by plugging the empirical distributions

of M_0 and M_1 into (1) and (2), the partial area-under-curve can be estimated by nonparametric U-statistics. The above properties will be extended to multivariate marker case for further analytical developments.

An alternative approach can be adopted by reversing the roles of true and false positive rates to define a function similar to the ROC function:

$$ROC^*(q) = FP[TP^{-1}(q)], \quad q \in (0, 1) \tag{3}$$

By property of composite function, it is seen that

$$ROC^*(q) = ROC^{-1}(q) \tag{4}$$

Clearly, since the mapping $ROC(q)$ is one-to-one, the function $ROC^*(q)$ consists the same amount of information as that of $ROC(q)$. Graphically, $ROC(q)$ and $ROC^*(q)$ are symmetric with respect to the diagonal line which connects points $(0, 0)$ and $(1, 1)$. Thus, $ROC(q) + ROC^*(1 - q) = 1$ and the sum of area under ROC curve and area under ROC^* curve equals 1. In section “Other Types of ROC and WROC Functions”, for multivariate marker model, a function parallel to $ROC^*(q)$ will be introduced and some interesting relationships similar to or different from those of univariate maker case will be explored.

Multivariate Markers: ROC, WROC and AUC

Now consider continuous markers and classification rule in multivariate setting. Suppose \mathbf{M}_0 and \mathbf{M}_1 are independent k -dimensional marker vectors from non-diseased group ($D = 0$) and diseased group ($D = 1$) respectively. Define

$$FP(\mathbf{m}) = P\{\mathbf{M}_0 \in D(\mathbf{m})\} ,$$

$$TP(\mathbf{m}) = P\{\mathbf{M}_1 \in D(\mathbf{m})\} .$$

Let $F_0(\mathbf{m}) = P(M_{01} \leq m_1, M_{02} \leq m_2, \dots, M_{0k} \leq m_k)$ be the cumulative distribution function for non-diseased population, and $F_1(\mathbf{m}) = P(M_{11} \leq m_1, M_{12} \leq m_2, \dots, M_{1k} \leq m_k)$ the cumulative distribution function for diseased population. Define the quantile variable $Q_0 = FP(\mathbf{M}_0)$ and denote by H_0 the distribution function of Q_0 . As an important feature of multivariate markers, in general Q_0 is not uniformly distributed. The distribution of Q_0 depends on the classifier as well as the probability structure of \mathbf{M}_0 , and therefore varies from marker vector to marker vector.

Definition of ROC Function

When marker measurements are multivariate, the function $FP(\mathbf{M}_0)$ is not a one-to-one transformation, which implies that the ROC function for univariate marker, $TP(FP^{-1}(q))$, can not be used for multivariate marker case. Wang and Li [21] considered bivariate marker models and defined an ROC function via generalized inverse set of the quantile function FP , where the ROC function possesses a conditional expectation expression. For multivariate markers, instead of using the generalized inverse set to conceptualize the ROC function, the ROC function is defined as the average of the true positive rate conditioning on the set of marker values with false positive rate q , where the conditional average is calculated subject to the non-diseased population:

$$ROC(q) = E[TP(\mathbf{M}_0) \mid FP(\mathbf{M}_0) = q] \quad (5)$$

There are a few characteristics of $ROC(q)$ in (5), which may or may not be similar to characteristics of the ROC function for univariate marker:

- The value of the ROC function in (5) is bounded between 0 and 1.
- The function $ROC(q)$ may not be an increasing function in q , $0 \leq q \leq 1$.
- If the distributions of \mathbf{M}_0 and \mathbf{M}_1 are the same (i.e., the marker vector is non-predictive for disease), then for each Borel set $D(m_1, m_2, \dots, m_k)$, one has $TP(m_1, m_2, \dots, m_k) = FP(m_1, m_2, \dots, m_k)$. This implies $TP(\mathbf{M}_0) = FP(\mathbf{M}_0)$ with probability one and

$$E[TP(\mathbf{M}_0) \mid FP(\mathbf{M}_0) = q] = q.$$

Thus, if the markers are non-predictive for disease, the ROC function coincides with the diagonal line which connects points (0,0) and (1,1), which is similar to the ROC function for univariate marker.

- When the markers are predictive subject to the classifier $D(m_1, m_2, \dots, m_k)$, it means that $TP(m_1, m_2, \dots, m_k) \geq FP(m_1, m_2, \dots, m_k)$ for each $(m_1, m_2, \dots, m_k) \in R^k$, and this implies $TP(\mathbf{M}_0) \geq FP(\mathbf{M}_0)$ with probability one and

$$ROC(q) = E[TP(\mathbf{M}_0) \mid FP(\mathbf{M}_0) = q] \geq E[FP(\mathbf{M}_0) \mid FP(\mathbf{M}_0) = q] = q,$$

for $0 \leq q \leq 1$. Thus, the ROC function is above the diagonal line if the markers are predictive for disease.

WROC and AUC

In use of the ROC function, a question of interest is whether the function in (5) can be used for comparisons of markers' predictive accuracy at population level. To address the question, we recall that for univariate marker the area under ROC curve is calculated with uniform distribution on q -axis (i.e., FP-axis).

For multivariate markers, the ROC function defined in (5) can be used to compare the performance of true positive rate locally by conditioning on $FP(\mathbf{M}_0) = q$. To evaluate multivariate markers' predictability unconditionally, the evaluation should take into account the distribution of Q_0 besides the use of the conditionally defined ROC function.

Using the probability distribution of Q_0 , the AUC can be naturally defined as the area under ROC curve subject to Lebesgue integration with measure H_0 on q -axis, namely $AUC = \int ROC(q)dH_0(q)$, or equivalently,

$$AUC = \int_0^1 ROC(q) \cdot h_0(q) dq \tag{6}$$

where $h_0(q)$ is the derivative of $H_0(q)$, which is assumed to exist. Define

$$WROC(q) = ROC(q) \cdot h_0(q)$$

as the weighted ROC (*WROC*) function. Note that $WROC(q)$ is the unconditional average of the true positive rate with fixed false positive rate q :

$$WROC(q) = E[TP(\mathbf{M}_0)I(FP(\mathbf{M}_0) = q)] . \tag{7}$$

It is seen that AUC is interpreted as area under *WROC* curve with uniform measure over the unit interval $[0, 1]$. Subsequently, the partial area under *WROC* curve can be defined as

$$AUC(p) = \int_0^p WROC(q)dq , \tag{8}$$

which can be used for comparison of markers in terms of their population-average predictability.

The concordance probability is naturally defined as $CON = P(\mathbf{M}_1 \in D(\mathbf{M}_0))$. Next we prove the equivalence between the concordance probability and the area under *WROC* curve, which is an extension of a property for univariate marker [4]:

$$\begin{aligned} CON &= P(\mathbf{M}_1 \in D(\mathbf{M}_0)) = \int \int I(\mathbf{m}_1 \in D(\mathbf{m}_0)) dF_1(\mathbf{m}_1)dF_0(\mathbf{m}_0) \\ &= \int TP(\mathbf{m}_0) dF_0(\mathbf{m}_0) = \int_0^1 E[TP(\mathbf{M}_0) | Q_0 = q] \cdot h_0(q)dq \\ &= \int_0^1 WROC(q)dq = AUC \end{aligned} \tag{9}$$

With an additional constraint on the false positive rate p , $0 \leq p \leq 1$, the partial concordance probability can be expressed as

$$CON(p) = P(\mathbf{M}_1 \in D(\mathbf{M}_0), FP(\mathbf{M}_0) \leq p) ,$$

where the full concordance probability corresponds to the special case $p = 1$. The partial concordance probability is

$$\begin{aligned}
 CON(p) &= P(\mathbf{M}_1 \in D(\mathbf{M}_0), FP(\mathbf{M}_0) \leq p) \\
 &= \int \int I(\mathbf{m}_1 \in D(\mathbf{m}_0))I(FP(\mathbf{m}_0) \leq p) dF_1(\mathbf{m}_1)dF_0(\mathbf{m}_0) \\
 &= \int TP(\mathbf{m}_0)I(FP(\mathbf{m}_0) \leq p) dF_0(\mathbf{m}_0) = \int_0^p E[TP(\mathbf{M}_0) | Q_0 = q] \cdot h_0(q) dq \\
 &= \int_0^p WROC(q) dq = AUC(p)
 \end{aligned} \tag{10}$$

The equivalence between $CON(p)$ and $AUC(p)$ is again an extension of the result from univariate marker model to multivariate marker model. Further, with the restrictions that the false positive rate is less than or equal to p and that the true positive rate is greater than q , the formula in (10) can be extended to

$$\begin{aligned}
 CON(p, q) &= P(\mathbf{M}_1 \in D(\mathbf{M}_0), FP(\mathbf{M}_0) \leq p, TP(\mathbf{M}_1) > q) \\
 &= \int \int I(\mathbf{m}_1 \in D(\mathbf{m}_0))I(FP(\mathbf{m}_0) \leq p, TP(\mathbf{m}_1) > q) dF_1(\mathbf{m}_1)dF_0(\mathbf{m}_0),
 \end{aligned}$$

which is a useful formula for constructing a U-statistic in estimation of the concordance probability with two-sided constraints. It is also clear that $CON(p, 0) = AUC(p)$.

Nonparametric Estimation

Suppose the observations include independent samples of iid copies of \mathbf{M}_0 and iid copies of \mathbf{M}_1 , where marker vectors are represented by $\{\mathbf{M}_{i,0} : i = 1, \dots, n_0\}$ and $\{\mathbf{M}_{j,1} : j = 1, \dots, n_1\}$, and realization values by $\{\mathbf{m}_{i,0} : i = 1, \dots, n_0\}$ and $\{\mathbf{m}_{j,1} : j = 1, \dots, n_1\}$, respectively from non-diseased and diseased populations. In this section we consider nonparametric approaches for estimation of ROC, WROC, AUC and CON. Denote by \widehat{TP} , \widehat{FP} , \widehat{F}_1 and \widehat{F}_0 respectively the empirical distribution of the corresponding function. For those p with $FP(\mathbf{m}_{i,0}) = p$, initially one can use a crude empirical estimate $TP(\mathbf{m}_{i,0})$ to estimate $ROC(p)$. Or, alternatively, we can consider the ROC function in its form as a conditional expectation in (5), $ROC(q) = E[TP(\mathbf{M}_0) | FP(\mathbf{M}_0) = q]$, and construct a kernel average estimate, which can be thought of as a smoothed version of the crude empirical estimate, to estimate $ROC(q)$:

$$\widehat{ROC}(p) = \frac{\int \widehat{TP}(\mathbf{m}_0) \cdot k\left(\frac{p - \widehat{FP}(\mathbf{m}_0)}{b}\right) d\widehat{F}_0(\mathbf{m}_0)}{\int k\left(\frac{p - \widehat{FP}(\mathbf{m}_0)}{b}\right) d\widehat{F}_0(\mathbf{m}_0)} = \frac{\sum_{i=1}^{n_0} \widehat{TP}(\mathbf{m}_{i,0}) \cdot k\left(\frac{p - \widehat{FP}(\mathbf{m}_{i,0})}{b}\right)}{\sum_{i=1}^{n_0} k\left(\frac{p - \widehat{FP}(\mathbf{m}_{i,0})}{b}\right)},$$

where the kernel $k(\cdot)$ is a mean zero density function and b is a bandwidth [7].

Note that the ROC function in (5) is defined as the average of true positive rate given a fixed value of the false positive rate, where the calculation of the conditional expectation is through the two one-dimensional variables $TP(\mathbf{M}_0)$ and $FP(\mathbf{M}_0)$. Thus, the ‘curse of dimensionality’ does not occur when the ROC function is estimated nonparametrically. A nonparametric estimator of $WROC(p)$ can be constructed by estimating the derivative of $CON(p)$ in (10) using kernel estimation technique:

$$\widehat{WROC}(p) = \frac{1}{b} \int \widehat{TP}(\mathbf{m}_0) \cdot k\left(\frac{p - \widehat{FP}(\mathbf{m}_0)}{b}\right) d\widehat{F}_0(\mathbf{m}_0) = \frac{1}{n_0 b} \sum_{i=1}^{n_0} \widehat{TP}(\mathbf{m}_{i,0}) \cdot k\left(\frac{p - \widehat{FP}(\mathbf{m}_{i,0})}{b}\right)$$

which is seen to be the same as the product of $\widehat{ROC}(p)$ and the kernel estimate of $h(p)$,

$$\frac{1}{b} \int k\left(\frac{p - \widehat{FP}(\mathbf{m}_0)}{b}\right) d\widehat{F}_0(\mathbf{m}_0) .$$

Based on the equivalence between $AUC(p)$ and $CON(p)$, a nonparametric estimator of $AUC(p)$ can be obtained:

$$\widehat{AUC}(p) = \int \int I(\mathbf{m}_1 \in D(\mathbf{m}_0)) I(\widehat{FP}(\mathbf{m}_0) \leq p) d\widehat{F}_1(\mathbf{m}_1) d\widehat{F}_0(\mathbf{m}_0) \quad (11)$$

With the restriction that the false positive rate is less than or equal to p and the true positive rate greater than q , the formula in (11) can be extended to

$$\begin{aligned} \widehat{CON}(p, q) &= \int I(\mathbf{m}_1 \in D(\mathbf{m}_0)) \cdot I(\widehat{FP}(\mathbf{m}_0) \leq p, \widehat{TP}(\mathbf{m}_1) > q) d\widehat{F}_1(\mathbf{m}_1) d\widehat{F}_0(\mathbf{m}_0) \\ &= \frac{1}{n_0 n_1} \sum_{i=1}^{n_0} \sum_{j=1}^{n_1} I(\mathbf{m}_{j,1} \in D(\mathbf{m}_{i,0})) \cdot I(\widehat{FP}(\mathbf{m}_{i,0}) \leq p, \widehat{TP}(\mathbf{m}_{j,1}) > q) , \end{aligned}$$

where the estimator has the form of a U-statistic [12].

Theorem 1. Let $N = n_0 + n_1$. Assume $0 < \lim_{N \rightarrow \infty} n_0/N = \lambda < 1$. Then, for $p, q \in [0, 1]$, (i) $\widehat{CON}(p, q)$ converges to $CON(p, q)$ in probability as $N \rightarrow \infty$, and (ii) $\sqrt{N}\{\widehat{CON}(p, q) - CON(p, q)\} \xrightarrow{d} \text{Normal}(0, \sigma^2)$, where σ^2 is specified in the Appendix.

The asymptotic results require that N be large and $0 < n_0/N = \lambda < 1$. This condition is generally satisfied with random sampling while disease status D could be either random or fixed, which is respectively relevant in prospective and retrospective (case-control) study. In the case D is random, N corresponds to the total sample size and n_0/N converges to $P(D = 0) = \lambda, 0 < \lambda < 1$, with probability 1 and the asymptotic normality holds with the usual interpretation.

Other Types of ROC and WROC Functions

Similar to considerations of using (3) in univariate marker case, for multivariate markers we may want to consider a function with the roles of true and false positive rates reversed. Define $Q_1 = TP(\mathbf{M}_1)$, and let H_1 and h_1 respectively be the distribution function and density function of Q_1 . Then, similar to the structure of $ROC(q)$, where $ROC(q) = E[TP(\mathbf{M}_0) \mid FP(\mathbf{M}_0) = q]$, for multivariate markers we may define

$$ROC^*(q) = E[FP(\mathbf{M}_1) \mid TP(\mathbf{M}_1) = q] .$$

In general, as a part of the main features which distinguish the univariate and multivariate ROC inferences, the functional transformation $ROC^*(q)$ is not one-to-one and therefore does not have the inverse functional relationship with $ROC(q)$. Further define

$$\overline{ROC}(q) = E[FN(\mathbf{M}_0) \mid TN(\mathbf{M}_0) = q] \quad \text{and} \quad \overline{ROC}^*(q) = E[TN(\mathbf{M}_1) \mid FN(\mathbf{M}_1) = q]$$

where $FN(\mathbf{m}) = P(\mathbf{M}_1 \notin D(\mathbf{m}))$ is the false negative rate and $TN(\mathbf{m}) = P(\mathbf{M}_0 \notin D(\mathbf{m}))$ is the true negative rate. The weighted functions corresponding to ROC^* , $\overline{ROC}(q)$ and $\overline{ROC}^*(q)$ can be defined in such ways similar to the $WROC$ function: for $0 < q < 1$,

$$\begin{aligned} WROC(q) &= ROC(q) \cdot h_0(q); \quad WROC^*(q) = ROC^*(q) \cdot h_1(q) \\ \overline{WROC}(q) &= \overline{ROC}(q) \cdot h_0(1 - q); \quad \overline{WROC}^*(q) = \overline{ROC}^*(q) \cdot h_1(1 - q) \end{aligned}$$

These weighted ROC functions serve to study the performance of predictive accuracy for multivariate markers from different perspectives. For example, $WROC^*(p)$ serves to study the performance of false positive rate with true positive rate controlled at value p . It is shown in the appendix that

$$\begin{aligned} ROC(q) + \overline{ROC}(1 - q) &= 1; \quad ROC^*(q) + \overline{ROC}^*(1 - q) = 1 \\ WROC(q) + \overline{WROC}(1 - q) &= h_0(q); \quad WROC^*(q) + \overline{WROC}^*(1 - q) = h_1(q) \end{aligned}$$

Thus, the function ROC provides the same amount of information as \overline{ROC} , and similarly ROC^* is as informative as \overline{ROC}^* . Also, with knowledge of $h(q)$, $WROC(q)$ provides the same amount of information as \overline{WROC} for predictive accuracy, and similar argument applies to the relationship between $WROC^*$ and \overline{WROC}^* . Essentially, the pair-wise relationship can be thought of as the conjugate partnership.

For evaluation based on partial area under curve, subject to either smaller FP ($FP \leq p$) or larger TP ($TP > q$), choices of these weighted ROC functions should be $WROC$ and \overline{WROC}^* so that maximization of area under curve would make sense. These two weighted ROC functions together with their corresponding ROC functions are used in our simulation to study the performance of the proposed criteria and methods for multivariate markers. Note that the partial concordance probability for true negativity is $\overline{CON}^*(p) = P(\mathbf{M}_0 \notin D(\mathbf{M}_1), FN(\mathbf{M}_1) \leq p)$. By similar technique employed in section “WROC and AUC”, it can be proved that this concordance probability coincides with the area under $\overline{WROC}^*(p)$ function, $\overline{CON}^*(p) = \overline{AUC}^*(p)$, and therefore a U-statistic $\widehat{\overline{CON}^*}(p)$ can be constructed to estimate $\overline{CON}^*(p)$.

In case of requiring both $FP \leq p$ and $TP > q$, these ROC or WROC functions cannot be used for evaluation, but $CON(p, q)$ can be used and estimated by the technique described in section “Multivariate Markers: ROC, WROC and AUC”. For estimation of \overline{ROC}^* , \overline{WROC}^* and $\overline{CON}^*(p, q)$, nonparametric estimates can be constructed using methods similar to those for the functions ROC , $WROC$ and $CON(p, q)$. Also, a property similar to Theorem 1 can be established for $\widehat{\overline{CON}^*}(p)$ by the same technique.

Remark. By setting $M_{l1} = M_{l2} = \dots = M_{lk}$, $l = 0, 1$, univariate marker model can be viewed as a degenerated case of multivariate markers. For this degenerated case, the quantile variable $Q_0 = FP(\mathbf{M}_0)$ and $Q_1 = TP(\mathbf{M}_1)$ both follow Uniform[0, 1] distribution, and $\overline{ROC}(q) = FN(TN^{-1}(q))$ and $\overline{ROC}^*(q) = TN(FN^{-1}(q))$. In this case, each of the WROC functions coincides with their counterpart of ROC functions. Further, besides the relationship $ROC(q) + \overline{ROC}(1 - q) = 1$ and $ROC^*(q) + \overline{ROC}^*(1 - q) = 1$, it is seen that $ROC^*(q) = ROC^{-1}(q)$, which implies that each of the four ROC functions provides the same amount of information as the other three functions for predictive accuracy of the marker.

Simulation and Data Example

Simulation

To show the performance of predictive accuracy for multivariate markers, we conduct simulation studies under different scenarios. We compare ROC and WROC curves for multivariate markers under each scenario, along with the weight function $h_0(q)$. We also compare univariate and multivariate marker cases to evaluate the gain and loss by using multiple markers.

Since this paper is a generalization of the bivariate ROC analysis of Wang and Li [21], we take $k \geq 3$ markers for evaluation. For simplicity, we take $k = 3$. Consider the simulation model where (M_{01}, M_{02}, M_{03}) and (M_{11}, M_{12}, M_{13}) follow a multivariate normal distribution. By convention we assume higher marker value

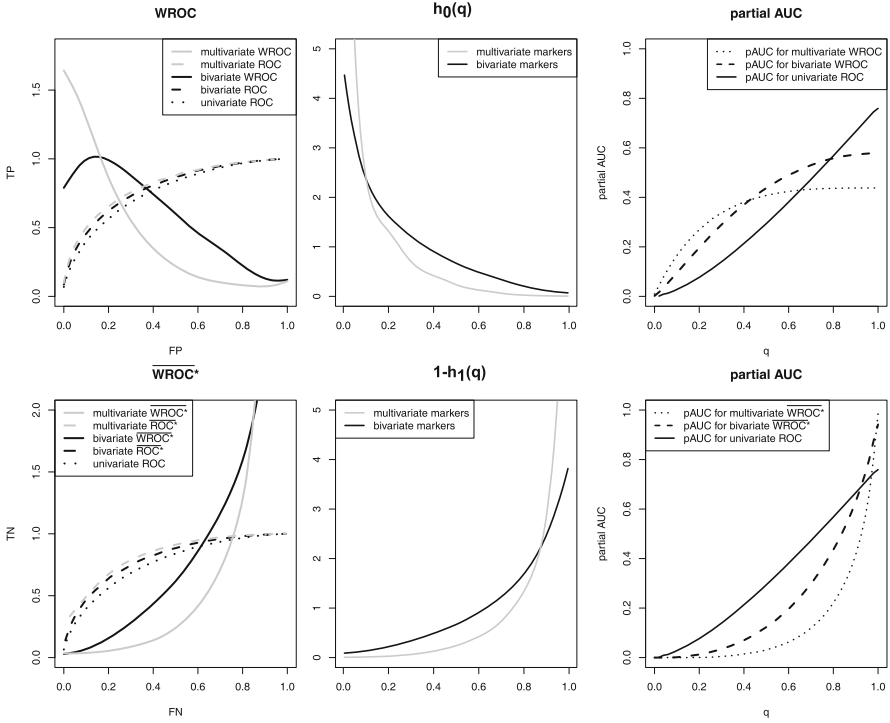


Fig. 1 Simulation for classifier $I(M_1 > m_1, M_2 > m_2, M_3 > m_3)$ with $\rho_0 = \rho_1 = \mathbf{0}$

indicates presence of disease. Let $N_1 = 200$ be the number of diseased individuals and $N_2 = 200$ be the number of non-diseased individuals. We generate data so that (M_{01}, M_{02}, M_{03}) have mean $(0, 0, 0)$ and unit deviations. We generate data so that (M_{11}, M_{12}, M_{13}) have mean $(1, 1, 1)$ and unit deviations. Let $\rho_l = (\rho_{l12}, \rho_{l23}, \rho_{l13})$, $l = 0, 1$, where ρ_{lij} denote the correlation between M_{li} and M_{lj} . We consider different scenarios according to different correlations ρ_l . The ROC analysis for univariate marker is based on data generated from the distributions of M_{11} , bivariate ROC analysis is based on data generated from the distribution of (M_{11}, M_{12}) , and multivariate ROC analysis is based on data generated from the distribution of (M_{11}, M_{12}, M_{13}) .

Figures 1–3 exhibit simulation results when $\rho_0 = \rho_1 = \mathbf{0}, \mathbf{0.5}$ and $\mathbf{1}$ respectively. As discussed in section “Other Types of ROC and WROC Functions”, $WROC$ is the conjugate partner of \overline{WROC} and $WROC^*$ is the conjugate partner of \overline{WROC}^* , and with the knowledge of $h_0(q)$ and $h_1(q)$, each of paired-partners provides the same amount of information for prediction as its partner. Choices of these weighted ROC functions should include only $WROC$ and \overline{WROC}^* so that maximization of area under curve makes sense.

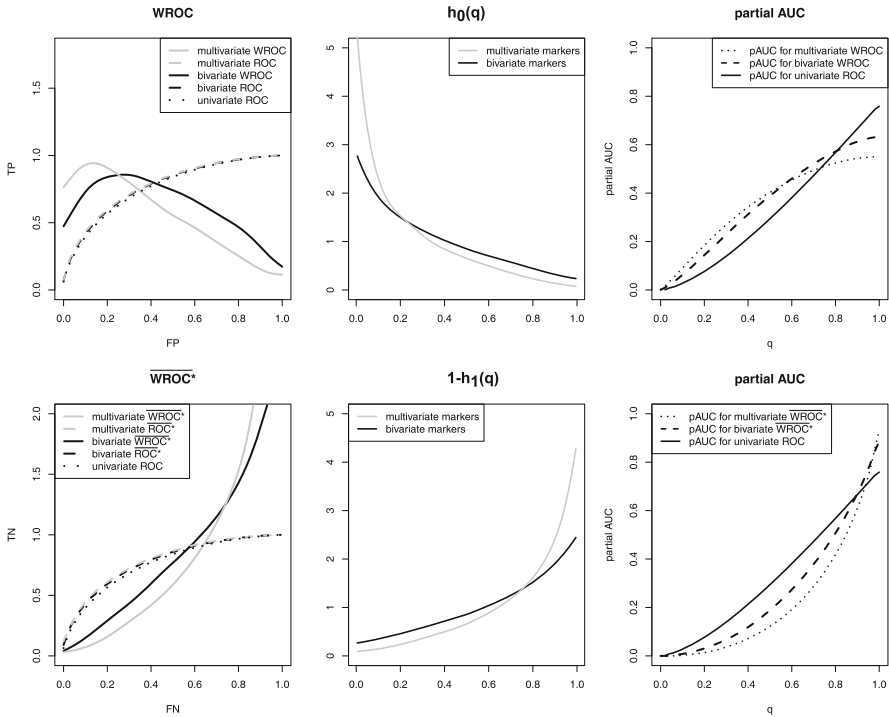


Fig. 2 Simulation for classifier $I(M_1 > m_1, M_2 > m_2, M_3 > m_3)$ with $\rho_0 = \rho_1 = 0.5$

When $\rho_0 = \rho_1 = 0$, the three markers are mutually independent, so the use of all three markers is expected to be more informative than one marker or two markers alone. Figure 1 shows a clear pattern of gain and loss as the number of markers increases. The gain in $WROC(q)$ for small values of q , when compared to univariate ROC curve, is substantial for multivariate ROC curve but only moderate for bivariate ROC curve. Similarly, the loss in $WROC(q)$ for large values of q is substantial for bivariate ROC curve but only moderate for bivariate ROC curve. This phenomenon can partly be explained by the right skewness of the weight function $h_0(q)$: the distribution of FP is uniform in univariate case, but it distributes more probability toward smaller values for bivariate marker case, and the inclusion of the third marker makes the weight function more skewed. By the equivalence between partial concordance probability and partial area under $WROC$ curve, we find that multivariate markers outperform univariate marker and bivariate marker for the region with small FP . The function \bar{WROC}^* for multivariate markers shows the opposite direction of gain and loss, compared to univariate or bivariate marker case. There is loss in $\bar{WROC}^*(q)$ for small values of q (FP) and gain for large values of q , which is due to the left skewness of the weight function $h_1(1 - q)$.

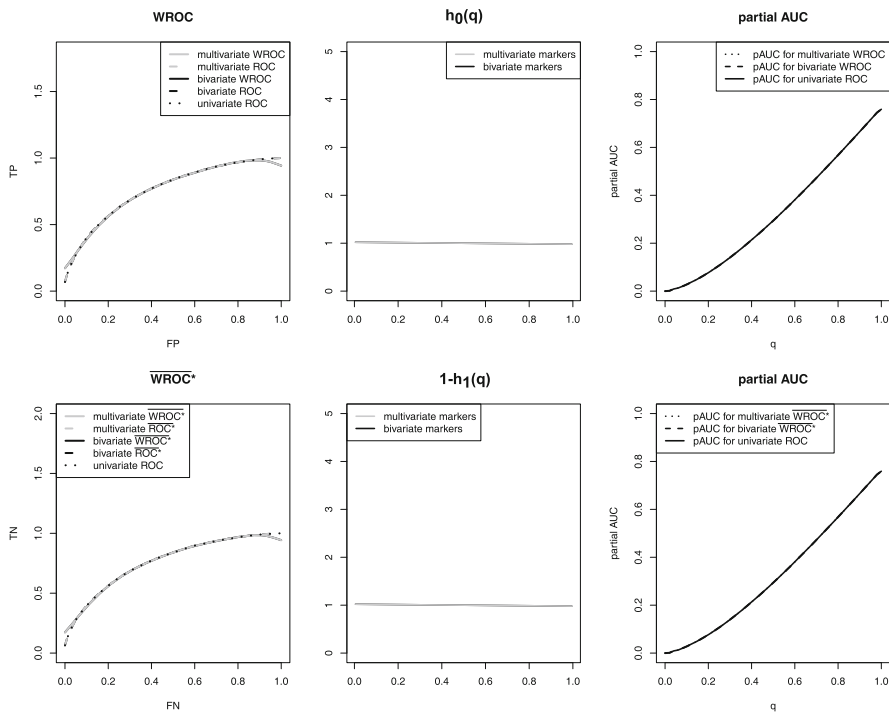


Fig. 3 Simulation for classifier $I(M_1 > m_1, M_2 > m_2, M_3 > m_3)$ with $\rho_0 = \rho_1 = 1$

When $\rho_0 = \rho_1 = 0.5$, the three markers are moderately correlated, similar to the case $\rho_0 = \rho_1 = 0$, the distribution of Q_0 and Q_1 still distribute more probability to small values, so we can observe the same pattern of tradeoff between gain at small FP and loss at large FP .

When $\rho_0 = \rho_1 = 1$, the three markers are identical and they provide the same information as one marker case (or two marker case). The ROC ($WROC$) functions for multivariate case coincides with the ROC function for univariate case (Fig. 3). The univariate case can thus be viewed as a degenerated case of multivariate markers.

A Data Example

We apply the proposed methods to the Alzheimer’s Disease Neuroimaging Initiative (ADNI) data for multivariate ROC analysis. The ADNI study is a research project with research focus on

changes of cognition, function, brain structure and function, and biomarkers in elderly controls, subjects with mild cognitive impairment, and subjects with Alzheimer’s disease

(quoted from <http://adni.loni.ucla.edu/>). The study is supported by the NIH, private pharmaceutical companies, and nonprofit organizations. Enrollment target was 800 participants – 200 normal controls, 400 patients with amnesic MCI, and 200 patients with mild AD – at 58 sites in the United States and Canada. Participants were enrolled on a rolling basis, and evaluated every six months. One of the major goals of the ADNI study is to identify biomarkers that are associated with progression from MCI to AD, and determine which biomarker measures (alone or in combination) are the best predictors of disease progression. Sensitivity and specificity for both cross-sectional and longitudinal diagnostic classification were considered important statistical techniques for assessing biomarkers in disease progression [18].

Investigations of the risk of progressing from MCI to AD dementia have largely focused on measures from the following categories: demographics, cognition, apolipoprotein E (APOE), magnetic resonance imaging (MRI), and cerebrospinal fluid (CSF) data. Demographic variables include age, education and gender. Cognitive measures represent five domains respectively: memory, language, executive function, spatial ability, and attention. Neuroimaging measures include brain volume, ventricular volume, and bilateral hippocampal volumes. The CSF variables include T-tau, $A\beta_{42}$, p-tau181, the ratio of the first two variables, and the ratio of the last two variables.

For this section, we selected three markers, hippocampus volume, memory score and executive function for illustration. To account for censoring, we used a reduced sample data set to create time-independent binary disease outcomes ($D = 0, 1$). We chose the 24th month as the cut-off time to define disease state. Of the 274 subjects who had complete data for the three markers, 49 subjects were loss to follow up before 24 months, so we focused on the 225 subjects who have had follow-up time longer than 24 months: there were 89 failures ($D = 1$) and 136 survivors ($D = 0$) at the 24th month. Let M_1 be hippocampus volume, M_2 be executive function score, and M_3 be memory score. Figure 4 compares the diagnostic performance of three markers (M_1, M_2, M_3), bivariate markers (M_1, M_2), and univariate marker M_1 . If the classifier is $I(M_1 > m_1, M_2 > m_2, M_3 > m_3)$, there is gain for small values of FP and loss for large values of FP . The partial AUC plot indicates that multivariate markers produce higher partial concordance summary than univariate marker when $q < 0.6$, and multivariate markers produce higher partial concordance summary than bivariate marker when $q < 0.3$. In diagnostic testing, it is crucial to maintain the false positive rate to be low to avoid unnecessary monetary costs. Thus, if the prognostic capacity is evaluated in terms of partial AUC, the multivariate marker hippocampus volume, executive function and memory score together would be considered performing much better than hippocampus volume alone.

Without restriction on the false positive rate, the AUC under the multivariate WROC curve is 0.358 (SE: 0.022) and the AUC under the multivariate \overline{WROC}^* is 0.964 (SE: 0.024); the AUC under the bivariate WROC curve is 0.437 (SE: 0.030)

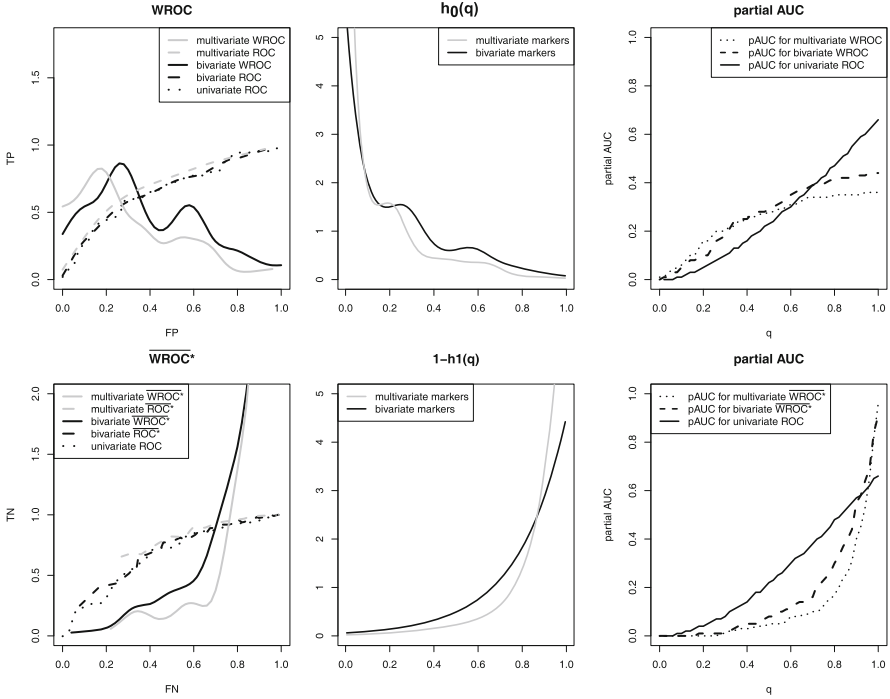


Fig. 4 $(M_1, M_2, M_3) = (\text{hippocampus, executive function, memory})$, with classifier $I(M_1 > m_1, M_2 > m_2, M_3 > m_3)$

and the AUC under the bivariate \overline{WROC}^* is 0.906 (SE: 0.030); the AUC under the univariate ROC curve is 0.658 (SE: 0.040). The bootstrap method was adopted to calculate the standard errors for estimation of AUC.

Discussion

Existing ROC methods to incorporate multiple markers typically consider a composite score based on combined markers by modeling the relationship between the marker vector \mathbf{M} and the binary outcome D [14], where $P(Y = 1|\mathbf{M}) = p(\mathbf{M})$ is used as the optimal score to identify the combination of multiple markers for classifying the disease outcome. In general, by the Neyman-Pearson lemma, the optimality of $p(\mathbf{M})$ is a very general property which holds without dimensionality constraint on \mathbf{M} . In the case that the linear logistic regression model assumption holds, the optimal classification rule, $p(\mathbf{M})$, becomes equivalent to the regression function $\beta\mathbf{M}$ under the logit link. Thus, the optimality property of a one-dimensional

classification score heavily relies on the assumption of logistic regression model. In this paper, we extend tools from univariate marker to multivariate markers for evaluating predictive accuracy of markers under a nonparametric setting based on tree-based classification rules.

The proposed ROC and WROC functions together with the AUC are intended to measure the average performance of and-or classifier among all possible combinations of true positive rate for a given false positive rate for evaluating predictability of markers and comparing curves, and they may not reflect the optimized use of markers for clinical decisions. Although the proposed approach is not designed to achieve optimality as a decision rule such as the one proposed by Jin and Lu [13], our methods and inferential results are much more structural, accessible and workable. The proposed ROC and WROC functions enjoy the advantage of preserving the distributional structures of markers, and the associated summary measures such as AUC or partial AUC serve as very appropriate summary measures to evaluate the performance of and-or classifier among all possible combinations of marker values – this is a feature similar to the univariate marker case. These summary measures are useful in applications, since many biomarker studies (such as the ADNI study and two other Alzheimer’s Disease studies that the authors are currently involved) have research emphasis largely focused on the understanding of predictability of biomarkers in target population, and less emphasis toward optimization of clinical decision rules.

The evaluation takes into account the distributions of quantile variables Q_0 and Q_1 in the diseased and non-diseased populations, which leads to the result of equivalence between AUC and CON, a property similar to the case of univariate marker. We also provide estimation procedures using nonparametric smoothing estimators for the ROC and WROC function, and U-statistic for the AUC. For applications of the proposed analysis, as the ‘curse of dimensionality’ is not a concern for nonparametric estimation of ROC, WROC and other related properties, the usual random split into training sample (for model fitting) and test sample (for creating ROC curve and calculating AUC) would be as proper as it is for univariate marker case, and therefore is advisable.

For future and further research, similar to the considerations for univariate ROC analysis [16, 20], it would be interesting to consider methodology to adjust for covariates such as age, sex or other demographical factors for bivariate or multivariate markers.

Also, given that the disease outcomes typically change with time, it would be interesting to extend the ROC analysis for high-dimensional markers to accommodate time-to-disease information using the ‘survival-tree methodology’ [22], along the lines of extending ROC techniques from binary disease outcome model to right-censored survival data model in univariate marker settings [5, 10, 11, 19].

Appendix

Proof of Theorem 1. Define the kernel function of the U-statistic [12] as

$$h(\mathbf{M}_{0i}, \mathbf{M}_{1j}; FP, TP) = I(\mathbf{M}_{1j} \in D(\mathbf{M}_{0i})) \cdot I(FP(\mathbf{M}_{0i}) \leq p, TP(\mathbf{M}_{1j}) > q).$$

Note that

$$\begin{aligned} \widehat{AUC}(p, q) &= \frac{1}{n_0 n_1} \sum_{i=1}^{n_0} \sum_{j=1}^{n_1} h(\mathbf{M}_{0i}, \mathbf{M}_{1j}; \widehat{FP}, \widehat{TP}) = \frac{1}{n_0 n_1} \sum_{i=1}^{n_0} \sum_{j=1}^{n_1} h(\mathbf{M}_{0i}, \mathbf{M}_{1j}; FP, TP) \\ &\quad + \frac{1}{n_0 n_1} \sum_{i=1}^{n_0} \sum_{j=1}^{n_1} \{h(\mathbf{M}_{0i}, \mathbf{M}_{1j}; \widehat{FP}, \widehat{TP}) - h(\mathbf{M}_{0i}, \mathbf{M}_{1j}; FP, TP)\} \\ &= I + II \end{aligned}$$

The kernel function in Term I satisfies $E[h^2] < \infty$ and by two-sample U-statistics theory, I converges to $AUC(p, q)$ in probability. Term II can be expressed as

$$II = \frac{1}{n_0 n_1} \sum_{i=1}^{n_0} \sum_{j=1}^{n_1} I(\mathbf{M}_{1j} \in D(\mathbf{M}_{0i})) \{I(\widehat{FP}(\mathbf{M}_{0i}) \leq p, \widehat{TP}(\mathbf{M}_{1j}) > q) - I(FP(\mathbf{M}_{0i}) \leq p, TP(\mathbf{M}_{1j}) > q)\}$$

Note that

$$\begin{aligned} |II| &\leq \frac{1}{n_0 n_1} \sum_{i=1}^{n_0} \sum_{j=1}^{n_1} \left| I(\widehat{FP}(\mathbf{M}_{0i}) \leq p, \widehat{TP}(\mathbf{M}_{1j}) > q) - I(FP(\mathbf{M}_{0i}) \leq p, TP(\mathbf{M}_{1j}) > q) \right| \\ &\leq \frac{1}{n_0 n_1} \sum_{i=1}^{n_0} \sum_{j=1}^{n_1} \left| I(\widehat{FP}(\mathbf{M}_{0i}) \leq p) - I(FP(\mathbf{M}_{0i}) \leq p) \right| + \frac{1}{n_0 n_1} \sum_{i=1}^{n_0} \sum_{j=1}^{n_1} \left| I(\widehat{TP}(\mathbf{M}_{1j}) > q) - I(TP(\mathbf{M}_{1j}) > q) \right| \\ &= \frac{1}{n_0} \sum_{i=1}^{n_0} \left| I(\widehat{FP}(\mathbf{M}_{0i}) \leq p) - I(FP(\mathbf{M}_{0i}) \leq p) \right| + \frac{1}{n_1} \sum_{j=1}^{n_1} \left| I(\widehat{TP}(\mathbf{M}_{1j}) > q) - I(TP(\mathbf{M}_{1j}) > q) \right| \\ &= o_p(n_0^{-1/2}) + o_p(n_1^{-1/2}) = o_p(N^{-1/2}) \end{aligned}$$

The consistency result, (i), in Theorem 1 follows by viewing the fact that term II converges to 0 in probability. To prove (ii), first note that Term I converges in distribution to a normal distribution by U-statistics theory: $\sqrt{N}\{I - AUC(p, q)\} \xrightarrow{d} \text{Normal}(0, \sigma^2)$, where $\sigma^2 = \lambda^{-1} \tau_{1,0} + (1 - \lambda)^{-1} \tau_{0,1}$ with

$$\tau_{1,0} = \text{COV}[h(\mathbf{M}_{01}, \mathbf{M}_{11}), h(\mathbf{M}_{01}, \mathbf{M}_{12})]$$

and

$$\tau_{0,1} = \text{COV}[h(\mathbf{M}_{01}, \mathbf{M}_{11}), h(\mathbf{M}_{02}, \mathbf{M}_{11})].$$

Also,

$$\begin{aligned} \sqrt{N}\{\widehat{AUC}(p, q) - AUC(p, q)\} &= \sqrt{N}\{I - AUC(p, q)\} + \sqrt{N} \cdot II \\ &= \sqrt{N}\{I - AUC(p, q)\} + o_p(1) \xrightarrow{d} \text{Normal}(0, \sigma^2) \end{aligned}$$

Property in section “Other Types of ROC and WROC Functions”.

- (i) $ROC(q) + \overline{ROC}(1 - q) = 1$, and $WROC(q) + \overline{WROC}(1 - q) = h_0(q)$
- (ii) $ROC^*(q) + \overline{ROC}^*(1 - q) = 1$, and $WROC^*(q) + \overline{WROC}^*(1 - q) = h_1(q)$

Proof. Note that

$$\begin{aligned} ROC(q) + \overline{ROC}(1 - q) &= E[TP(\mathbf{M}_0)|FP(\mathbf{M}_0) = q] + E[FN(\mathbf{M}_0)|FP(\mathbf{M}_0) = q] \\ &= E[TP(\mathbf{M}_0) + FN(\mathbf{M}_0)|FP(\mathbf{M}_0) = q] \\ &= E[1 |FP(\mathbf{M}_0) = q] = 1, \end{aligned}$$

and it follows $WROC(q) + \overline{WROC}(1 - q) = ROC(q) \cdot h_0(q) + \overline{ROC}(1 - q)h_0(q) = h_0(q)$, which proved (i). Similar argument can be used to prove (ii).

References

1. Baker, S.G.: Identifying combinations of cancer markers for further study as triggers of early intervention. *Biometrics* **56**, 1082–1087 (2000)
2. Baker, S.G., Cook, N.R., Vickers, A., Kramer, B.S.: Using relative utility curves to evaluate risk prediction. *J. R. Stat. Soc. A* **172**, 729–748 (2009)
3. Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J.: *Classification and Regression Trees*. Wadsworth and Brooks/Cole Advanced Books and Software, Monterey (1984)
4. Dodd, L., Pepe, M.S.: Partial AUC estimation and regression. *Biometrics* **59**, 614–623 (2003)
5. Etzioni, R., Pepe, M.S., Longton, G., Hu, C., Goodman, G.: Incorporating the time dimension in receiver operating characteristic curves: a prostate cancer case study. *Med. Decis. Making* **19**, 242–251 (1999)
6. Etzioni, R., Kooperberg, C., Pepe, M.S., Smith, R., GANN, P.H.: Combining biomarkers to detect disease with application to prostate cancer. *Biostatistics* **4**, 523–538 (2003)
7. Green, P.J., Silverman, B.W.: *Nonparametric Regression and Generalized Linear Models: A Robust Penalty Approach*. Chapman and Hall, London (1994)
8. Gu, W., Pepe, M.S.: Measures to summarize and compare the predictive capacity of markers. *Int. J. Biostat.* **5**(1), (2009)
9. Hanley, J.A., McNeil, B.J.: A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* **148**, 839–843 (1983)
10. Heagerty, P.J., Zheng, Y.: Survival model predictive accuracy and ROC curves. *Biometrics* **61**, 92–105 (2005)
11. Heagerty, P.J., Lumley, T., Pepe, M.S.: Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* **56**, 337–344 (2000)
12. Hoeffding, W.: A class of statistics with asymptotically normal distributions. *Ann. Stat.* **19**, 293–325 (1948)
13. Jin, H., Lu, Y.: ROC region of a regression tree. *Stat. Probab. Lett.* **79**, 936–942 (2009)

14. McIntosh, M.W., Pepe, M.S.: Combining several screening tests: optimality of the risk score. *Biometrics* **58**, 657–664 (2002)
15. Metz, C.E., Shen, J.H.: Gains in accuracy from replicated readings of diagnostic images: prediction and assessment in terms of ROC analysis. *Med. Decis. Making* **12**, 60–75 (1992)
16. Pepe, M.S.: Three approaches to regression analysis of receiver operating characteristic curves for continuous test results. *Biometrics* **54**, 124–135 (1998)
17. Pepe, M.S.: *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford, New York (2003)
18. Risacher, S.L., Saykin, A.J., West, J.D., Shen, L., Firpi, H.A., McDonald, B.C., Alzheimer's Disease Neuroimaging Initiative (ADNI): baseline MRI predictors of conversion from MCI to probable AD in the ADNI cohort. *Curr. Alzheimer Res.* **6**, 347–361 (2009)
19. Slate, E.H., Turnbull, B.W.: Models for longitudinal biomarkers of disease onset. *Stat. Med.* **19**, 617–637 (2000)
20. Tosteson, A.N., Begg, C.B.: A general regression methodology for ROC curve estimation. *Med. Decis. Making* **8**, 205–215 (1988)
21. Wang M.-C., Li, S.: Bivariate marker measurements and ROC analysis. *Biometrics* **68**, 1207–1218 (2012)
22. Zhang, H., Crowley, J., Sox, H.C., Olshen, R.A.: Tree-structured statistical methods. *Encycl. Biostat.* **6**, 4561–4573 (1998)