Published in final edited form as:

Biometrics. 2012 December ; 68(4): 1207–1218. doi:10.1111/j.1541-0420.2012.01783.x.

Bivariate Marker Measurements and ROC Analysis

Mei-Cheng Wang* and Shanshan Li**

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Baltimore, MD 21205

Summary

This paper considers receiver operating characteristic (ROC) analysis for bivariate marker measurements. The research interest is to extend tools and rules from univariate marker to bivariate marker setting for evaluating predictive accuracy of markers using a tree-based classification rule. Using an and-or classifier, an ROC function together with a weighted ROC function (WROC) and their conjugate counterparts are proposed for examining the performance of bivariate markers. The proposed functions evaluate the performance of and-or classifiers among all possible combinations of marker values, and are ideal measures for understanding the predictability of biomarkers in target population. Specific features of 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 ROC-related functions, (partial) area under curve and concordance probability. With emphasis on average performance of markers, the proposed procedures and inferential results are useful for evaluating marker predictability based on a single or bivariate marker (or test) measurements with different choices of markers, and for evaluating different and-or combinations in classifiers. The inferential results developed in this paper also extend to multivariate markers with a sequence of arbitrarily combined and-or classifier.

Keywords

Concordance probability; Prediction accuracy; Tree-based classification; U-statistics

1. Introduction

Receiver Operating Characteristic (ROC) analysis has been widely used as tools for summarizing sensitivity and specificity of a binary classifier system in medicine and engineering (Metz, 1978). Based on a single or combined-to-univariate marker, the ROC curve and some associated measures such as area under curve (AUC), partial AUC and some other measures have been developed to summarize and compare the predictive accuracy of biomarkers (Pepe, 2003; Gu and Pepe, 2009; Baker et al., 2009; among others).

This paper considers situations when a pair of continuous marker variables (U , V) are available for classification of disease state with classifier $I(U > u \text{ and } V > v)$. The research interest is to extend rules and tools from univariate marker to bivariate marker setting for assessment of predictive accuracy of markers from observed data or a training data set. With intension to evaluate the overall performance of predictive accuracy, we extend the ROC

*mcwang@jhsph.edu

**shli@jhsph.edu

Supplementary Materials

Web Appendix, referenced in Section 6, together with computing software are available with this paper at the Biometrics website on Wiley Online Library.

function from univariate marker to bivariate marker case, and introduce a weighted ROC (WROC) function together with their conjugate counterparts. Using the US Alzheimers Disease Neuroimaging Initiative (ADNI) data set as an example, U and V can each be specified as a single or combined biomarker from different biological or functional systems such as cognition and brain. As the two markers are identified from different biological sources, it may not be appropriate to combine the two markers using, say, a linear combination of the two measurements. The and-or classifier also signifies the importance of interaction between markers. It would be interesting to keep the two markers in bivariate setting and explore their respective roles and interaction nonparametrically. In this type of applications, the classifier $I(U > u \text{ and } V > v)$ is used to emphasize the importance of both U and V for identifying the disease.

The proposed work includes at least two types of applications:

- I. A patient's diagnostic test result is examined twice by the same examiner, or examined by two different examiners, once by each. Or, a patient's health outcome is diagnosed by two different diagnostic tests (Metz and Shen, 1992).
- II. Bivariate markers each characterizing a specific biological function are collected to study their predictability for a disease.

For the first type of applications, multiple reading is employed for either reducing uncertainty of test classification or comparison of two diagnostic modalities (Metz et al., 1984, 1992; Hanley and McNeil, 1983). Applications of the second type are important when bivariate or multivariate markers are used as prognostic measurements for predicting or understanding the disease. To analyze bivariate or multivariate marker data, various 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 and evaluate the predictability of markers by the composite score (McIntosh and Pepe, 2002). Another direction of research is based on comparison of diagnostic markers through correlated ROC curves. DeLong et al. (1988) and Wieand et al. (1989) proposed different nonparametric approaches to analyze areas under curves and other associated measures from two correlated ROC functions. Zhou and Gatsonis (1996) extended the analysis and proposed methodology for analysis of incomplete data.

For analysis of biomarkers in multi-dimensional setting, Baker (2000) and Etzioni et al. (2003) proposed new definitions for ROC curves subject to tree-based (and-or) classifiers. They both divided the multi-dimensional marker space into small intervals, and defined the true and false positive rates based on subsets of the divided regions. The ROC curve was then defined as the optimal true positive rate for each given false positive rate. Baker (2000) viewed the true positive rate as benefit and the false positive rate as cost, and adopted an optimization algorithm based on a utility function. Etzioni et al. (2003) linked the evaluation of true and false positive rates to an objective function which was a weighted function of false positive rate and false negative rate. Besides and-or classifiers, other types of classifiers such as Swirls and Ripples were also proved useful for analyzing gene expression microarray data (Baker, 2010).

The paper is organized as follows. Section 2 briefly reviews some of the fundamental definitions and properties for univariate ROC analysis, where emphasis is placed on those which will be extended to bivariate setting. In Section 3, the ROC and WROC functions are introduced with discussion focused on contrasting features between univariate and bivariate cases. Section 4 considers nonparametric estimators for ROC-related functions, AUC and concordance probabilities. Section 5 introduces a set of ROC-related functions and provides further details on specific characteristics and features of bivariate ROC analysis. Simulation

and a real data analysis are presented in Section 6 to illustrate the applicability of the proposed procedures. The inferential results developed in this paper also extend to alternative classifier $I(U > u \text{ or } V > v)$ and to multivariate marker measurements with a sequence of arbitrarily combined and-or classifier with potential applications to tree-based classification with binary decision diagram. Section 7 describes potential extension of the proposed bivariate analysis to multivariate marker setting. Section 8 concludes the paper with a brief discussion.

2. Univariate Marker Case

Suppose the disease outcome D takes binary values 0 or 1, and U is a continuous marker variable. Let U_0 and U_1 respectively represent the marker variable from non-diseased ($D=0$) and diseased ($D=1$) group. Define $TP(u) = P(U_1 > u) = P(U > u | D=1)$ as the true positive rate (sensitivity), and $FP(u) = P(U_0 > u) = P(U > u | D=0)$ the false positive rate (1-specificity). Assume U_0 and U_1 are independent. Define $F_0(u) = 1 - FP(u)$ and $F_1(u) = 1 - TP(u)$ respectively as the cumulative distribution function of U_0 and U_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 $AUC(p) = \int I(0 < q < p) ROC(q) dq$. The area under ROC curve is defined as the total area with p ranging from 0 to 1, that is, $AUC(1)$. Define the partial concordance probability as $CON(p) = P(U_1 > U_0, FP(U_0) \leq p)$. For univariate marker model, the quantile variable $Q_0 = FP(U_0)$ is Uniform[0, 1] distributed and thus $CON(p)$ can be calculated using probability measure on (U_1, Q_0) and is simplified to

$$CON(p) = P(U_1 > FP^{-1}(Q_0), Q_0 \leq p) = \int_0^p \int I(u_1 > FP^{-1}(q)) dF_1(u_1) dq \\ = \int_0^p ROC(q) dq = AUC(p)$$

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

$$CON(p) = \int \int I(u_1 > u_0) I(FP(u_0) \leq p) dF_1(u_1) dF_0(u_0) \quad (1)$$

The equivalence between $CON(p)$ and $AUC(p)$ has led to development of nonparametric approaches for estimating $AUC(p)$ using the formula in (1). Dodd and Pepe (2003) 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

$$\int I(p_0^* \leq q < p_1) ROC(q) dq = P(U_1 > U_0, FP^{-1}(p_1) < U_0 \leq TP^{-1}(p_0)) \quad (2)$$

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 or equal to p_0 . As proposed by Dodd and Pepe (2003), by plugging the empirical distributions of U_0 and U_1 into (1) and (2), the partial area-under-curve can be estimated by nonparametric U-statistics.

The above properties will be extended to bivariate marker case for further analytical developments.

It is interesting to take an alternative approach by reversing the roles of true and false positive rates, and define the function

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

By property of composite function, it is seen that

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

Clearly, since the mapping $ROC(q)$ is one-to-one, the function $ROC^*(q)$ consists of 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, the sum of area under ROC curve and area under ROC^* curve equals 1. In Section 4, for bivariate marker model, a function parallel to $ROC^*(q)$ will be introduced and some interesting relationships similar to or different from those for univariate marker case will be explored.

3. Bivariate Markers: ROC, WROC and AUC

Suppose continuous bivariate measurements (U, V) are available for either classifying diagnostic test results or for studying the disease. Suppose (U_0, V_0) and (U_1, V_1) are independent pairs of bivariate markers respectively from non-diseased group ($D = 0$) and diseased group ($D = 1$). In some studies, the presence of disease is highly associated with large values of markers, and each of the two markers is essential for detecting the disease. In these cases, it would be scientifically relevant to classify the subject as a positive case when the values of U and V are both large.

Define $FP(u, v) = P(U_0 > u \text{ and } V_0 > v)$ and $TP(u, v) = P(U_1 > u \text{ and } V_1 > v)$. Let $F_0(u, v) = P(U_0 \leq u, V_0 \leq v) = P(U \leq u, V \leq v | D = 0)$ be the cumulative distribution function for non-diseased population, and $F_1(u, v) = P(U_1 \leq u, V_1 \leq v) = P(U \leq u, V \leq v | D = 1)$ the cumulative distribution function for diseased population. For bivariate markers, define the quantile variable $Q_0 = FP(U_0, V_0)$ and denote by H_0 the distribution function of Q_0 . As an important feature of bivariate 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 (U_0, V_0) , and therefore varies from bivariate-markers to bivariate-markers.

3.1 Definition of ROC Function

Now consider the question of how to define the ROC function for bivariate markers. When marker measurements are bivariate, the function $FP(u, v)$ 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 bivariate marker case. To keep the feature of controlling false positive rate in ROC function but avoiding the use of inverse function of FP , we consider the inverse set $B_0(q) = \{(u_0, v_0) : FP(u_0, v_0) = q\}$, $0 \leq q \leq 1$, and define the value of ROC function as the average of true positive rate over the inverse set.

For each fixed v satisfying $q = FP(-\infty, v)$, the q th-quantile on u -coordinate is defined as $FP_v^{-1}(q) = \inf\{u : FP(u, v) \leq q\}$, $0 \leq q \leq 1$, and an ROC function along the u -coordinate can be defined as $ROC_v(q) = TP(FP_v^{-1}(q), v)$. The function $ROC_v(q)$ behaves like a standard ROC function for univariate marker, except that the range of values of $q(ROC_v(q))$ is only a

subset of $(0, 1)$. Similar to a standard ROC function, $ROC_v(q)$ is non-decreasing in q , and therefore $ROC_v(q_1) \leq ROC_v(q_2)$ if $0 \leq q_1 \leq q_2 \leq FP(-\infty, v)$. Note that the inverse set $B_0(q)$ can also be expressed as $B_0(q) = \{(FP_v^{-1}(q), v) : v \in R, FP(-\infty, v) \geq q\}$. Let $\mathcal{V}_q$ represent the probability measure induced by F_0 on the subspace $B_0(q)$, the ROC function for bivariate markers is defined as

$$ROC(q) = \int ROC_v(q) d\mathcal{V}_q(v), \text{ for } 0 < q < 1 \quad (5)$$

or equivalently,

$$ROC(q) = E[TP(U_0, V_0) | FP(U_0, V_0) = q] \quad (6)$$

The value of $ROC(q)$ is bounded between 0 and 1, and can be interpreted 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. Note that $ROC_v(q)$ in (5) is non-decreasing in q for each fixed v , but such a monotonicity property may not hold for $ROC(q)$. Further, when $U_0 = V_0$, the support of (U_0, V_0) shrinks to the line $u = v$ on R^2 space and the ROC function reduces to the univariate ROC function.

In the case that the bivariate markers are independent of the disease outcome, i.e., the distributions of (U_0, V_0) and (U_1, V_1) are the same, then $ROC_v(q) = q$ for each v and therefore $ROC(q) = q$ for $0 \leq q \leq 1$. The plot for ROC function in this extreme case coincides with the diagonal line which connects points $(0, 0)$ and $(1, 1)$, a result similar to the univariate marker case. When bivariate markers are 'predictive' for disease outcome, note that $ROC_v(q) \geq q$ along the u -coordinate for each v satisfying $FP(-\infty, v) \geq q$, and therefore $ROC(q) \geq q$.

In Section 3.2, we shall further propose a WROC function to characterize the role of the probability structure of the bivariate classifier to conceptualize the area under curve in bivariate marker setting.

3.2 WROC and AUC

In use of the ROC function, a question of interest is whether the function in (5) can be directly used for comparisons of predictive accuracy of markers. To address the question, we recall that for univariate marker the area under the ROC curve is calculated with uniform distribution on q -axis (i.e., FP -axis). For bivariate markers, the ROC function defined in (6) can be used to compare the performance of true positive rate locally by conditioning on $FP(U_0, V_0) = q$. To evaluate bivariate 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, $AUC = \int ROC(q) \cdot h_0(q) dq$, 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. 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

$$AUC(p) = \int_0^p WROC(q) dq$$

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

Next we prove the equivalence between partial concordance probability and partial area under *WROC* curve as an extension of a property for univariate marker (Dodd and Pepe, 2003). The partial concordance probability with restriction on false positive rate is

$$\begin{aligned} CON(p) &= P(U_1 > U_0, V_1 > V_0, FP(U_0, V_0) \leq p) \\ &= \int \int I(u_1 > u_0, v_1 > v_0) \cdot I(FP(u_0, v_0) \leq p) dF_1(u_1, v_1) dF_0(u_0, v_0) \quad (7) \\ &= \int TP(u_0, v_0) \cdot I(FP(u_0, v_0) \leq p) dF_0(u_0, v_0) \end{aligned}$$

where the full concordance probability corresponds to the special case $p = 1$, $CON(1) = P(U_1 > U_0, V_1 > V_0)$. Using probability measure of (Q_0, V_0) instead of F_0 in (7) to calculate $CON(p)$, we obtain

$$\begin{aligned} CON(p) &= \int \int TP(FP_v^{-1}(q), v) I(q \leq p) d\mathcal{H}_q(v) dH_0(q) \\ &= \int ROC(q) I(q \leq p) dH_0(q) = \int_0^p WROC(q) dq = AUC(p) \quad (8) \end{aligned}$$

The equivalence between $CON(p)$ and $AUC(p)$ is essentially an extension of the result from univariate marker to bivariate markers. The property that $WROC(p)$ can be obtained as the derivative of $AUC(p)$ implies that it is also the derivative of $CON(p)$. This property implies that estimation of $WROC(p)$ can be obtained through estimation of the derivative of $CON(p)$, which will be used in nonparametric estimation in Section 4. 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 (7) can be extended to

$$\begin{aligned} CON(p, q) &= P(U_1 > U_0, V_1 > V_0, FP(U_0, V_0) \leq p, TP(U_1, V_1) > q) \\ &= \int \int I(u_1 > u_0, v_1 > v_0) I(FP(u_0, v_0) \leq p, TP(u_1, v_1) > q) dF_1(u_1, v_1) dF_0(u_0, v_0), \end{aligned}$$

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

3.3 Comparing Different ROC Functions

Jin and Lu (2009), abbreviated as JL hereafter, considered bivariate continuous markers and proposed to use the area under the upper boundary of ROC region to evaluate diagnostic utilities. The value of upper boundary of ROC region corresponds to $TP(u^*, v^*)$, where $(u^*, v^*) = \arg \max \{ TP(u, v) : FP(u, v) = q \}$. JL's work can be viewed as an extension of Baker's approach (Baker, 2000) from discrete markers to continuous markers, and the work placed emphasis on optimization of marker values for clinical decisions. In contrast with the upper boundary function, our ROC function is defined as the conditional expectation of the true positive rate given a fixed false positive rate and, by definition, it measures the average

performance of marker predictability based on the set $\{TP(u, v): FP(u, v) = q\}$. While JL's upper boundary ROC function identifies the optimal classification rule for each specific false positive rate, the area under their function does not possess a probabilistic interpretation. Essentially their AUC can be viewed only as a subjectively defined summary measure for evaluation of marker predictability. In contrast, our ROC function in (6) has the feature as the average true positive rate with controlled false positive rate, and after accounting for the role of the distribution of quantile variable, the AUC under the WROC curve has a concordance probability interpretation, which is a nice extension of the well known property from univariate marker case.

To apply our method, as will be seen from next section, either the crude empirical estimators or nonparametric smoothing estimators can be constructed for the estimation of ROC and WROC functions, where threshold is determined empirically. And, by taking advantage of the equivalence between AUC and CON, nonparametric U-statistics can be constructed for the estimation of AUC and partial AUC. The statistical methods and inferential results are much more structural, accessible and workable when compared with statistical properties of JL's boundary function and AUC. Although our approach is not designed to achieve optimality as a decision rule, it enjoys the advantage of preserving the distributional structures of markers, and the associated summary measure 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 focus largely on the understanding of predictability of biomarkers in target population, and less on optimization of clinical decision rules.

Similar to constraints from use of any statistical approach, the proposed methods have limitations since they can be applied only to and-or classifier for bivariate markers (and arbitrarily combined and-or classifier for multivariate markers, as discussed in Section 7). The and-or classifiers are generally considered an important class of classifiers, but there are situations where they may not reflect how multiple markers should be used in practice. Note that, however, the ROC function in (6) is formulated as a conditional expectation which calculates the average of true positive rate conditioning on a fixed false positive rate from non-diseased population. The definition is very general and conceptually extendable to other types of classifiers. For broader applications, it will be interesting to extend the proposed approach to alternative classifiers, for bivariate or multivariate markers, when the conventional ROC analysis does not apply.

4. Nonparametric Estimation

Suppose the observations include independent samples of iid copies of (U_0, V_0) and iid copies of (U_1, V_1) , with observations represented by $\{(U_{0i}, V_{0i}): i = 1, \dots, m\}$ and $\{(U_{1j}, V_{1j}): j = 1, \dots, n\}$, respectively from non-diseased and diseased populations. In this section we consider nonparametric approaches for estimation of ROC, WROC, AUC and CON. Denote by $\widehat{FP}$, $\widehat{F}_1$ and $\widehat{F}_0$ respectively the empirical distribution of the corresponding function. For those p with $FP(u_{0p}, v_{0p}) = p$, initially one can use a crude empirical estimate $TP(u_{0p}, v_{0p})$ to estimate $ROC(p)$. Or, alternatively, we can consider the ROC function in its form as a conditional expectation in (6), $ROC(q) = E[TP(U_0, V_0) | FP(U_0, V_0) = q]$, and construct a kernel average estimate, which can be thought of as a smoothed version of the crude empirical estimate, to estimate $ROC(q)$:

$$\widehat{ROC}(p) = \frac{\int \widehat{TP}(u_0, v_0) \cdot k\left(\frac{p - \widehat{FP}(u_0, v_0)}{b}\right) d\widehat{F}_0(u_0, v_0)}{\int k\left(\frac{p - \widehat{FP}(u_0, v_0)}{b}\right) d\widehat{F}_0(u_0, v_0)} = \frac{\sum_{i=1}^m \widehat{TP}(u_{0i}, v_{0i}) \cdot k\left(\frac{p - \widehat{FP}(u_{0i}, v_{0i})}{b}\right)}{\sum_{i=1}^m k\left(\frac{p - \widehat{FP}(u_{0i}, v_{0i})}{b}\right)},$$

where the kernel $k(\cdot)$ is a mean zero density function and b is a bandwidth (Green and Silverman, 1994).

A nonparametric estimator of $WROC(p)$ can be constructed by estimating the derivative of $CON(p)$ in (7) using kernel estimation technique:

$$\widehat{WROC}(p) = \frac{1}{b} \int \widehat{TP}(u_0, v_0) \cdot k\left(\frac{p - \widehat{FP}(u_0, v_0)}{b}\right) d\widehat{F}_0(u_0, v_0) = \frac{1}{mb} \sum_{i=1}^m \widehat{TP}(u_{0i}, v_{0i}) \cdot k\left(\frac{p - \widehat{FP}(u_{0i}, v_{0i})}{b}\right)$$

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

$$\frac{1}{b} \int k\left(\frac{p - \widehat{FP}(u_0, v_0)}{b}\right) d\widehat{F}_0(u_0, v_0).$$

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

$$\widehat{AUC}(p) \int \int I(u_1 > u_0, v_1 > v_0) \cdot I(\widehat{FP}(u_0, v_0) \leq p) d\widehat{F}_1(u_1, v_1) d\widehat{F}_0(u_0, v_0) \quad (9)$$

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 (9) can be extended to

$$\begin{aligned} \widehat{CON}(p, q) &= \int \int I(u_1 > u_0, v_1 > v_0) \cdot I(\widehat{FP}(u_0, v_0) \leq p, \widehat{TP}(u_1, v_1) > q) d\widehat{F}_1(u_1, v_1) d\widehat{F}_0(u_0, v_0) \\ &= \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n I(U_{1j} > U_{0i}, V_{1j} > V_{0i}) \cdot I(\widehat{FP}(U_{0i}, V_{0i}) \leq p, \widehat{TP}(U_{1j}, V_{1j}) > q), \end{aligned}$$

where the estimator has the form of a U-statistic (Hoeffding, 1948).

Theorem 4.1

Let $N = m + n$. Assume $0 < \lim_{N \rightarrow \infty} m/N = \lambda < 1$. 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)
- $$\sqrt{N}\{\widehat{CON}(p, q) - CON(p, q)\} \xrightarrow{d} \text{Normal}(0, \sigma^2), \text{ where } \sigma^2 \text{ is specified in the Appendix.}$$

The asymptotic results require that N be large and $0 < m/N = \lambda < 1$. 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 m/N converges to $P(D=0) = \lambda$, $0 < \lambda < 1$, with probability 1 and the asymptotic normality holds with the usual interpretation.

5. Other Types of ROC and WROC Functions

Similar to considerations of using (3) in univariate marker case, for bivariate markers we may want to consider an ROC-type function with the roles of true and false positive rates reversed, or with true and false positive rates replaced by true and false negative rates. Define $Q_1 = TP(U_1, V_1)$, and let H_1 and h_1 respectively be the distribution function and density function of Q_1 . Similar to the construction of $ROC(q)$, we may define

$$\begin{aligned} ROC^*(q) &= E[FP(U_1, V_1) | TP(U_1, V_1) = q], \text{ for } 0 < q < 1 \\ \overline{ROC}(q) &= E[FN(U_0, V_0) | TN(U_0, V_0) = q], \text{ for } 0 < q < 1 \\ \overline{ROC}^*(q) &= E[TN(U_1, V_1) | FN(U_1, V_1) = q], \text{ for } 0 < q < 1 \end{aligned}$$

where $FN(u, v) = P(U_1 \leq u \text{ or } V_1 \leq v)$ is the false negative rate and $TN(u, v) = P(U_0 \leq u \text{ or } V_0 \leq v)$ is the true negative rate. The weighted functions corresponding to ROC^* , $\overline{ROC}(q)$ and $\overline{ROC}^*(q)$ can be defined: for $0 < q < 1$, $WROC^*(q) = ROC^*(q) \cdot h_1(q)$, $\overline{WROC}(q) = \overline{ROC}(q) \cdot h_0(1-q)$, and $\overline{WROC}^*(q) = \overline{ROC}^*(q) \cdot h_1(1-q)$. These weighted ROC functions serve to study the performance of predictive accuracy for bivariate 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

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

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 bivariate markers. Note that the partial concordance probability for true negativity is

$\overline{CON}^*(p) = P(U_0 \leq U_1 \text{ or } V_0 \leq V_1, FN(U_1, V_1) \leq p)$. By similar technique employed in Section 3.2, 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 $TP > q$, $CON(p, q)$ can be used and estimated by the technique described in Section 3. 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 4.1 can be established for $\widehat{\overline{CON}^*}(p)$ by the same technique.

By setting $U_0 = V_0$ and $U_1 = V_1$, univariate marker model can be viewed as a degenerated case. For this degenerated case, the quantile variable $Q_0 = FP(U_0, V_0)$ and $Q_1 = TP(U_1, V_1)$ 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 $ROC(WROC)$ functions in bivariate setting coincides with the ROC function in univariate setting. 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.

In case that U_k and V_k are weakly or moderately correlated, the distribution of Q_0 and Q_1 distribute more probability to small values. There would be an expected trade-off between gain in $\overline{WROC}^*(q)$ for small values of $q(FP)$ and loss in $WROC^*(q)$ for small values of $q(FN)$. The proposed plots based on $(ROC, WROC)$ and $(\overline{ROC}^*, \overline{WROC}^*)$ and (provide clear contrast, with nice interpretation, to evaluate the amount of gain and loss in selection of different classifiers or markers. The simulations in Section 6 clearly illustrate some of these features.

The total area under $WROC$ curve, $AUC(1)$, for bivariate markers (U, V) is never larger than $AUC(1)$ for univariate marker U . This is clear in view of the fact that, by (8), $AUC(1) = CON(1) = P(U_1 > U_0, V_1 > V_0) - P(U_1 > U_0)$. In contrast, the total area under $\overline{WROC}^*$ curve for (U, V) is never less than the total area under ROC curve for U . In general, the predictability of bivariate markers should not be judged by the total area under $WROC$ or $\overline{WROC}^*$ curve, but by partial areas under these two curves subject to constraints on the false positive rate and the false negative rate.

6. Simulation and Data Example

6.1 Simulation

To show the performance of predictive accuracy for bivariate markers, we conduct simulation studies under different scenarios. We compare ROC and WROC curves to evaluate the gain and loss by using univariate or bivariate markers.

Consider the simulation model where (U_0, V_0) and (U_1, V_1) follow bivariate normal distributions. Let $m = n = 200$ be the number of diseased (non-diseased) individuals. We generate 500 replications of bivariate normal data at different levels of correlation so that (U_0, V_0) has mean $(0, 0)$ and unit deviations, and (U_1, V_1) has mean $(1, 1)$ and unit deviations. Let $\rho_k, k = 0, 1$, denote the correlation between U_k and V_k , and we consider different scenarios according to different correlations ρ_k . The ROC analysis for univariate marker is based on data generated from the distributions of U_k , and bivariate ROC analysis is based on data generated from the distribution of $(U_k, V_k), k = 0, 1$.

Figures 1~3 exhibit simulation results for three cases: $\rho_0 = \rho_1 = 0, \rho_0 = \rho_1 = 0.5$ and $\rho_0 = \rho_1 = 1$. The ROC, WROC, density curves h_0 and h_1 , and partial AUC are calculated as the average of the proposed nonparametric estimates based on 500 replications of bivariate normal data. When $\rho_0 = \rho_1 = 0$, the two markers are uncorrelated and the second marker is expected to provide further information beyond one marker. Figure 1 shows that $WROC$ curve for bivariate markers behaves quite differently as compared to univariate ROC curve. There is gain in $WROC(q)$ for small values of q (FP) and loss for large values of q . This is due to the right skewness of the weight function $h_0(q)$: the distribution of FP is no longer uniform as in univariate case, but it distributes more probability toward smaller values. By the equivalence between partial concordance probability and partial area under $WROC$ curve, we find that bivariate markers outperform univariate marker for the region with small

FP. The function $\overline{WROC}^*$ for bivariate markers also behaves quite differently as compared to univariate $\overline{ROC}^*$ curve. There is loss in $\overline{WROC}^*(q)$ for small values of q (*FP*) and gain for large values of q , which is due to the left skewness of the weight function $h_1(1 - q)$. When $\rho_0 = \rho_1 = 0.5$, the two markers are moderately correlated, and 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 small and large values of *FP* in Figure 2. When $\rho_0 = \rho_1 = 1$, the two markers are identical and they provide the same information as one marker case. The *ROC* (*WROC*) functions for bivariate case coincides with the *ROC* function for univariate case (Figure 3). Clearly, the univariate case can thus be viewed as a degenerated case of bivariate markers.

More analysis results of estimation of the concordance probability $\widehat{CON}(p, q)$ can be found in Table 1. A clear pattern of comparison across three different scenarios can be found: with p and q fixed, the partial AUC increases as the correlation between the two markers decreases.

6.2 A Data Example

We apply the proposed methods to the Alzheimer's Disease Neuroimaging Initiative (ADNI) data for ROC analysis. The detailed description of the ADNI data is provided in Web Appendix B.

The original disease outcome data include information of possibly censored times to AD. 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 145 subjects, 23 subjects were loss to follow up before 24 months, so we focused on the 122 subjects who have had follow-up time longer than 24 months: there were 44 failures ($D = 1$) and 78 survivors ($D = 0$) at the 24th month. We consider two markers from different biological systems, brain and cognition, for illustration of the proposed methods. Let U be hippocampus volume, a primary brain region of interest in the study of AD. Let V be memory score, an important marker for cognitive function.

Figures 4 and 5 present plots to compare the diagnostic performance of bivariate markers (U, V) and univariate marker U using classifiers $I(U > u \text{ and } V > v)$ and classifier $I(U > u \text{ or } V > v)$. When $I(U > u \text{ and } V > v)$ is used, it is seen that there is gain for small values of *FP* and loss for large values of *FP*. The partial AUC plot, which is estimated by the proposed nonparametric estimates, indicates that bivariate markers produce higher partial concordance summary when $q < 0.8$. In diagnostic testing, it is crucial to maintain the false positive rate to be low to avoid unnecessary monetary costs. Region with *FP* greater than 0.8 may not be of interest. Thus, if the prognostic capacity is evaluated in terms of partial AUC, the bivariate markers hippocampus volume and memory score together would be considered performing much better than hippocampus volume alone. Interestingly, even though the gain in true positive rate is substantial, the loss in true negative rate is negligible in the region that the false negative rate is less than 0.6. When the classifier $I(U > u \text{ or } V > v)$ is employed, it is seen that the gain in true negative rate is more than the loss in true positive rate. Without restriction on the false positive rate, when the classifier is $I(U > u \text{ and } V > v)$ is used, the AUC under the *WROC* curve is 0.479 (SE: 0.044) and the AUC under the $\overline{WROC}^*$ curve is 0.862 (SE: 0.044). When the alternative classifier $I(U > u \text{ or } V > v)$ is used, by symmetry of the concordance probability, the AUC under the *WROC* curve is 0.862 (SE: 0.044), and the AUC under the $\overline{WROC}^*$ curve is 0.479 (SE: 0.044). In contrast with the bivariate marker case, the AUC for hippocampus volume alone is 0.630 (0.061) and the AUC for memory score alone is 0.710 (0.061). The analysis suggests that each of hippocampus volume and

memory score serves as informative marker, and the predictability for AD could be further improved if they are jointly used in prediction.

7. Extension to Multivariate Markers

The inferential results developed in this paper extend to the alternative classifier $I(U > u \text{ or } V > v)$ in a straightforward manner. The classifier $I(U > u \text{ or } V > v)$ describes a different strategy for classification in which either U or V would be sufficient for classifying a test result to be positive. This is proper when, for example, U and V are highly correlated markers (or readings) from same or different biological systems (or examiners).

The proposed ROC analysis also extends to multivariate markers $(M_1, M_2, \dots, M_k)$ based on a sequence of arbitrarily combined and-or classifier with positivity specified in either direction of marker values; for example, $I(M_1 > m_1 \text{ or } M_2 < m_2 \text{ and } M_3 > m_3)$. This extension links to potential applications related to classification tree with binary decision outcome. Let $(M_{j1}, M_{j2}, \dots, M_{jk}), j = 0, 1$, be the marker vector for a non-diseased or diseased subject. The test result is classified to be positive if $(M_{j1}, M_{j2}, \dots, M_{jk}) \in D(m_1, m_2, \dots, m_k), j = 0, 1$, where $D(m_1, m_2, \dots, m_k) \in R^k$ is defined as the region for marker-based positivity.

It is seen that the ROC function can be extended and defined as

$$ROC(q) = E[TP(M_{01}, M_{02}, \dots, M_{0k}) | FP(M_{01}, M_{02}, \dots, M_{0k}) = q] \quad (10)$$

Other ROC, WROC and AUC functions together with the related properties can all be formulated or derived with similar framework as those in Sections 3~5, including the nonparametric estimators in Section 4. Note that the ROC function in (10) 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 $TP(M_{01}, M_{02}, \dots, M_{0k})$ and $FP(M_{11}, M_{12}, \dots, M_{1k})$. Thus, the ‘curse of dimensionality’ would not apply to nonparametric estimation of ROC, WROC and other related properties. Thus, the overfitting problem does not become a more serious concern in multivariate marker setting and, as suggested by a referee, a random split of data into training sample (for model fitting) and test sample (for creating ROC curve) would be as proper or efficient as it is for bivariate marker case.

8. Discussion

In this paper we extend tools from univariate marker to bivariate and multivariate markers for evaluating predictive accuracy of markers. 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. The evaluation takes into account the distributions of quantile variables Q_0 and Q_1 in the diseased and non-diseased populations, and 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 future and further research, similar to the considerations for univariate ROC analysis (Tosteson and Begg, 1988; Pepe, 1998), 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, such as those ROC techniques developed for right-censored survival data in univariate marker settings (Etzioni et al., 1999; Heagerty et al., 2000; Slate and Turnbull, 2000; Heagerty and Zheng, 2005).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors wish to thank professor Charles Metz for his help in identifying some of the references cited in this paper. We also thank the associate editor and two reviewers whose insightful comments have led to improvement of this article. The research was partly supported by the National Institute of Health grants R01AI078835 and U01AG033655.

References

- Baker SJ. Identifying combinations of cancer markers for further study as triggers of early intervention. *Biometrics*. 2000; 56:1082–1087. [PubMed: 11129464]
- Baker SJ, Cook NR, Vickers A, Kramer BS. Using relative utility curves to evaluate risk prediction. *Journal of the Royal Statistical Society: Series A*. 2009; 172:729–748.
- Baker SJ. Simple and flexible classification via Swirls and Ripples. *BMC Bioinformatics*. 2010; 11:452. [PubMed: 20825641]
- Campbell G. General methodology I: advances in statistical methodology for the evaluation of diagnostic and laboratory tests. *Statistics in Medicine*. 1994; 13:499–508. [PubMed: 8023031]
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988; 44:837–45. [PubMed: 3203132]
- Dodd L, Pepe M. Partial AUC estimation and regression. *Biometrics*. 2003; 59:614–623. [PubMed: 14601762]
- Etzioni R, Pepe M, Longton G, Hu C, Goodman G. Incorporating the time dimension in receiver operating characteristic curves: A prostate cancer case study. *Medical Decision Making*. 1999; 19:242–251. [PubMed: 10424831]
- Etzioni R, Kooperberg C, Pepe M, Smith R, GANN PH. Combining biomarkers to detect disease with application to prostate cancer. *Biostatistics*. 2003; 4:523–538. [PubMed: 14557109]
- Green, PJ.; Silverman, BW. *Nonparametric Regression and Generalized Linear Models: A Robust Penalty Approach*. London: Chapman and Hall; 1994.
- Gu W, Pepe MS. Measures to summarize and compare the predictive capacity of markers. *International Journal of Biostatistics*. 2009; 5(1)
- Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*. 1983; 148:839–843. [PubMed: 6878708]
- Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*. 2000; 56:337–344. [PubMed: 10877287]
- Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. *Biometrics*. 2005; 61:92–105. [PubMed: 15737082]
- Hoeffding W. A class of statistics with asymptotically normal distributions. *Annals of Statistics*. 1948; 19:293–325.
- Jin H, Lu Y. ROC region of a regression tree. *Statistics and Probability Letters*. 2009; 79:936–942.
- Metz CE. Basic principles of ROC analysis. *Seminars in Nuclear Medicine*. 1978; 8:283–298. [PubMed: 112681]

- Metz CE, Shen JH. Gains in Accuracy from Replicated Readings of Diagnostic Images: Prediction and Assessment in Terms of ROC Analysis. *Medical Decision Making*. 1992;60–75. [PubMed: 1538634]
- Mets, CE.; Wang, PL.; Kronman, HB. A new approach for testing the significance of differences between ROC curves measured from correlated data. In: Deconinck, F., editor. *Information Processing in Medical Imaging*. The Hague, The Netherlands: Nijhoff; 1984. p. 432–45.
- McIntosh MW, Pepe MS. Combining several screening tests: Optimality of the risk score. *Biometrics*. 2002; 58(3):657–664. [PubMed: 12230001]
- Pepe MS. Three Approaches to Regression Analysis of Receiver Operating Characteristic Curves for Continuous Test Results. *Biometrics*. 1998; 54:124–135. [PubMed: 9544511]
- Pepe, MS. *The statistical evaluation of medical tests for classification and prediction*. New York: Oxford; 2003.
- Risacher SL, Saykin AJ, West JD, Shen L, Firpi HA, McDonald BC. the Alzheimer's Disease Neuroimaging Initiative (ADNI). Baseline MRI Predictors of Conversion from MCI to Probable AD in the ADNI Cohort. *Current Alzheimer Research*. 2009; 6:347–361. [PubMed: 19689234]
- Slate EH, Turnbull BW. Models for longitudinal biomarkers of disease onset. *Statistics in Medicine*. 2000; 19(4):617–637. [PubMed: 10694740]
- Tosteson AN, Begg CB. A General Regression Methodology for ROC Curve Estimation. *Medical Decision Making*. 1988; 8:205–215.
- Wieand S, Gail MH, James BR, James KL. A family of nonparametric statistics for comparing diagnostic markers with paired or unpaired data. *Biometrika*. 1989; 76:585–592.
- Zhou XH, Gatsonis CA. A simple method for comparing correlated ROC curves using incomplete data. *Statistics in Medicine*. 1996; 15:1687–1693. [PubMed: 8858790]

Appendix

Proof of Theorem 3.1

Define the kernel function of the U-statistic (Hoeffding, 1948) as

$$h((U_{0i}, V_{0i}), (U_{1j}, V_{1j}); FP, TP) = I(U_{1j} > U_{0i}, V_{1j} > V_{0i}) \cdot I(FP(U_{0i}, V_{0i}) \leq p, TP(U_{1j}, V_{1j}) > q).$$

Note that

$$\begin{aligned} \widehat{AUC}(p, q) &= \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n h((U_{0i}, V_{0i}), (U_{1j}, V_{1j}); \widehat{FP}, \widehat{TP}) \\ &= \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n h((U_{0i}, V_{0i}), (U_{1j}, V_{1j}); FP, TP) + \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n [h((U_{0i}, V_{0i}), (U_{1j}, V_{1j}); \widehat{FP}, \widehat{TP}) - h((U_{0i}, V_{0i}), (U_{1j}, V_{1j}); FP, TP)] \\ &= I + II \end{aligned}$$

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

$$II = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n I(U_{1j} > U_{0i}, V_{1j} > V_{0i}) \{I(\widehat{FP}(U_{0i}, V_{0i}) \leq p, \widehat{TP}(U_{1j}, V_{1j}) > q) - I(FP(U_{0i}, V_{0i}) \leq p, TP(U_{1j}, V_{1j}) > q)\}$$

Note that

$$\begin{aligned}
|II| &\leq \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n \left| I(\widehat{FP}(U_{0i}, V_{0i}) \leq p, \widehat{TP}(U_{1j}, V_{1j}) > q) - I(FP(U_{0i}, V_{0i}) \leq p, TP(U_{1j}, V_{1j}) > q) \right| \\
&\leq \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n \left| I(\widehat{FP}(U_{0i}, V_{0i}) \leq p) - I(FP(U_{0i}, V_{0i}) \leq p) \right| + \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n \left| I(\widehat{TP}(U_{1j}, V_{1j}) > q) - I(TP(U_{1j}, V_{1j}) > q) \right| \\
&= \frac{1}{mn} \sum_{i=1}^m \left| I(\widehat{FP}(U_{0i}, V_{0i}) \leq p) - I(FP(U_{0i}, V_{0i}) \leq p) \right| + \frac{1}{n} \sum_{j=1}^n \left| I(\widehat{TP}(U_{1j}, V_{1j}) > q) - I(TP(U_{1j}, V_{1j}) > q) \right| \\
&= o_p(m^{-1/2}) + o_p(n^{-1/2}) = o_p(N^{-1/2})
\end{aligned}$$

The consistency result, (i), in Theorem 3.1 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

$$\tau_{1,0} = \text{COV}[h\{(U_{01}, V_{01}), (U_{11}, V_{11})\}, h\{(U_{01}, V_{01}), (U_{12}, V_{12})\}]$$

and

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

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 4

- 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

$$\begin{aligned}
ROC(q) + \overline{ROC}(1-q) &= E[TP(U_0, V_0) | FP(U_0, V_0) = q] + E[FN(U_0, V_0) | FP(U_0, V_0) = q] \\
&= E[TP(U_0, V_0) + FN(U_0, V_0) | FP(U_0, V_0) = q] \\
&= E[1 | FP(U_0, V_0) = q] = 1,
\end{aligned}$$

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).

Figure 1. Simulation results for markers with correlations $\rho_0 = \rho_1 = 0$ based on classifier $I(U > u \text{ and } V > v)$.

Figure 2. Simulation results for markers with correlations $\rho_0 = \rho_1 = 0.5$ based on classifier $I(U > u \text{ and } V > v)$.

Figure 3. Simulation results for markers with correlations $\rho_0 = \rho_1 = 1$ based on classifier $I(U > u \text{ and } V > v)$.

Figure 4. ROC analysis with bivariate markers (U,V)=(hippocampus, memory) based on classifier $I(U > u \text{ and } V > v)$.

Figure 5. ROC analysis with bivariate markers (U,V)=(hippocampus, memory) based on classifier $I(U > u \text{ or } V > v)$.

Table 1

Average of bootstrap mean (SD) of $AUC(p, q)$ in simulation study

q	p				
	0.2	0.3	0.4	0.5	0.6
$\rho_0 = \rho_1 = 0$					
0	0.193 (0.025)	0.287 (0.031)	0.369 (0.033)	0.434 (0.034)	0.487 (0.036)
0.1	0.185 (0.026)	0.276 (0.015)	0.353 (0.035)	0.416 (0.037)	0.466 (0.039)
0.2	0.174 (0.026)	0.258 (0.033)	0.331 (0.037)	0.390 (0.040)	0.437 (0.042)
0.3	0.158 (0.027)	0.236 (0.034)	0.303 (0.039)	0.356 (0.042)	0.399 (0.046)
0.4	0.139 (0.026)	0.208 (0.034)	0.266 (0.040)	0.314 (0.044)	0.351 (0.048)
$\rho_0 = \rho_1 = 0.5$					
0	0.145 (0.025)	0.230 (0.030)	0.313 (0.033)	0.389 (0.036)	0.459 (0.036)
0.1	0.141 (0.026)	0.223 (0.031)	0.304 (0.036)	0.377 (0.037)	0.445 (0.038)
0.2	0.134 (0.027)	0.212 (0.032)	0.288 (0.037)	0.357 (0.039)	0.422 (0.041)
0.3	0.124 (0.026)	0.197 (0.032)	0.268 (0.038)	0.332 (0.041)	0.393 (0.043)
0.4	0.114 (0.025)	0.180 (0.032)	0.245 (0.038)	0.304 (0.040)	0.358 (0.045)
$\rho_0 = \rho_1 = 1$					
0	0.078 (0.011)	0.141 (0.015)	0.214 (0.019)	0.295 (0.022)	0.381 (0.024)
0.1	0.077 (0.011)	0.139 (0.031)	0.211 (0.019)	0.292 (0.023)	0.377 (0.025)
0.2	0.075 (0.012)	0.137 (0.016)	0.208 (0.020)	0.286 (0.024)	0.370 (0.027)
0.3	0.073 (0.012)	0.132 (0.016)	0.201 (0.021)	0.277 (0.025)	0.358 (0.029)
0.4	0.069 (0.012)	0.126 (0.017)	0.192 (0.023)	0.264 (0.028)	0.341 (0.032)