

Biomarker Selection in Medical Diagnosis

Man-Jen Hsu, Yuan-Chin Ivan Chang, and Huey-Miin Hsueh

Abstract A biomarker is usually used as a diagnostic or assessment tool in medical research. Finding a single ideal biomarker of a high level of both sensitivity and specificity is not an easy task; especially when a high specificity is required for a population screening tool. Combining multiple biomarkers is a promising alternative and can provide a better overall performance than the use of a single biomarker. It is known that the area under the receiver operating characteristic (ROC) curve is most popular for evaluation of a diagnostic tool. In this study, we consider the criterion of the partial area under the ROC curve (pAUC) for the purpose of population screening. Under the binormality assumption, we obtain the optimal linear combination of biomarkers in the sense of maximizing the pAUC with a pre-specified specificity level. Furthermore, statistical testing procedures based on the optimal linear combination are developed to assess the discriminatory power of a biomarker set and an individual biomarker, respectively. Stepwise biomarker selections, by embedding the proposed tests, are introduced to identify those biomarkers of statistical significance among a biomarker set. Rather than for an exploratory study, our methods, providing computationally intensive statistical evidence, are more appropriate for a confirmatory analysis, where the data has been adequately filtered. The applicability of the proposed methods are shown via several real data sets with a moderate number of biomarkers.

M.-J. Hsu • H.-M. Hsueh (✉)

Department of Statistics, National ChengChi University, Taipei 11605, Taiwan
e-mail: 95354503@nccu.edu.tw; hsueh@nccu.edu.tw

Y.-C.I. Chang

Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan

Department of Statistics, National ChengChi University, Taipei 11605, Taiwan
e-mail: ycchang@sinica.edu.tw

1 Introduction

A biomarker is a biological indicator in showing absence, presence, or the condition of a disease, and it can be used to determine the status of a subject, the effectiveness of a treatment, and so on. Ideally, a biomarker with both high sensitivity and specificity for accurate prediction is expected. However, it is not easy to find such a biomarker in practice. Combining biomarkers provides an alternative to improve the performance of currently available ones. For example, the serum prostate-specific antigen (PSA) is a well-accepted prognostic biomarker to screen for prostate cancer. However, this test has a low specificity and therefore might lead to over-diagnosis. Besides PSA, several potentials are investigated, please see [11]. Nevertheless, no single biomarker among them outperforms PSA, and therefore, more investigators propose the use of a combination of PSA and others. Please see [1, 9].

The receiver operating characteristic (ROC) curve is the most popular graphical tool in evaluation of the diagnostic power of a biomarker. It provides an exhaustive look at the relationship between sensitivity and specificity of a biomarker. The area under the ROC curve (AUC) is proposed for an efficient summarization. In some applications, investigators place their emphasis only on a part of the curve. For example, a high level of specificity is required for a biomarker serving as a population screening tool. As a consequence, a biomarker is assessed on the partial area under the ROC curve (pAUC) in a region of specificity above some level. See [13] and the reviews by [4, 16].

This study focuses on combining multiple continuous-scaled biomarkers into one single diagnostic or predictive rule for disease with emphases on assessment of each biomarker. For better interpretability, we propose the use of a linear combination for information summarization. The discriminatory power of a linear combination of biomarkers is evaluated on the pAUC. The optimal linear combination, which provides the best discriminatory power among all combinations, is the target solution of research interest. In addition to the global predictability, some insights into the importance of an individual biomarker can be obtained from the coefficients. However, it needs to incorporate sampling variation for statistical significance.

In presence of multiple biomarkers, a traditional way is fitting a multiple logistic regression model to the data set for medical diagnosis. For example, see [20]. Alternatively, seeking the maximal discriminatory power, Su and Liu [17] derived the explicit form of the best linear combination in terms of AUC under a binormal model. Following their study, [6] found a solution, that dominates any others in some scenarios. Nevertheless, the dominating scenarios are not universal. Pepe and Thompson [12] and Pepe et al. [14] proposed the use of empirical AUC estimates in finding the optimal linear combination. In our earlier study, we found that not only the analytical derivation but also the computation became much more complicated under the pAUC criterion, please refer [2].

Recently, due to newer and better biotechnology, big data are generated easily and related analytical tools are demanding. In developing a binary classification, which is parallel to a diagnostic rule, several algorithm-based approaches have been

proposed by directly using either AUC or pAUC as the objective function, such as [3, 5, 7, 8, 10, 15, 22, 23]. However, these algorithm-based methods are unable to accommodate statistical evidence into variable selection. It motivates our study in developing some stepwise approaches, embedding adequate statistical tests, to identify important biomarkers for data sets of a moderate size. In which, a biomarker is discarded or selected based on the statistical evidence from data, not only on a computational prospect.

The paper is organized as follows. In Sect. 2, the sample version of the optimal linear combination will be defined. In Sect. 3, some testing procedures for the global and individual discriminatory power will be proposed. Furthermore, two biomarker selection approaches adopting the proposed tests will be developed. Real example analyses are given in Sect. 4. We then conclude this paper with a discussion in Sect. 5.

2 Strong Consistency of the Linear Combination Estimator Maximizing the pAUC

Let \mathbf{X} be a random vector of p biomarkers related to the disease of a subject, and D be the binary disease status, where $D = 1$ indicates a subject from the diseased population, $D = 0$ indicates a subject from the non-diseased population. Suppose that

$$\mathbf{X}|D = d \sim MVN(\boldsymbol{\mu}_d, \boldsymbol{\Sigma}_d), \quad d = 0, 1,$$

where $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_1$ are positive definite. For any given real vector $\mathbf{a} \in R^p$,

$$\mathbf{a}^T \mathbf{X}|D = d \sim N(\mathbf{a}^T \boldsymbol{\mu}_d, Q_d),$$

where $Q_d = \mathbf{a}^T \boldsymbol{\Sigma}_d \mathbf{a}$, for $d = 0, 1$. Let $\Phi(\cdot)$ denote the cumulative distribution function of $N(0, 1)$ and $\Phi^{-1}(\cdot)$ be its inverse function. Let $c(u) = \Phi^{-1}(1 - u)$ and $\Delta_\mu = \mu_1 - \mu_0$, then at specificity $(1 - u)$, the sensitivity of $\mathbf{a}^T \mathbf{X}$ is equal to

$$F(\mathbf{a}, u) = \Phi\left(\frac{\mathbf{a}^T \Delta_\mu - c(u)\sqrt{Q_0}}{\sqrt{Q_1}}\right).$$

Therefore, for a false positive rate region $(0, t)$ for some predetermined $t \in (0, 1)$, the pAUC of $\mathbf{a}^T \mathbf{X}$ is equal to

$$pAUC(\mathbf{a}) = \int_0^t F(\mathbf{a}, u) du. \quad (1)$$

Similar to AUC, the pAUC has the scale invariant property. For identification purposes, in this study the search for the optimal linear combination vector is

restricted to the hyper-sphere with an unit radius. Let \mathbf{a}^* be such a pAUC maximizer; that is,

$$\mathbf{a}^* = \arg \max_{\mathbf{a} \in E_p} pAUC(\mathbf{a}),$$

where $E_p = \{\mathbf{a} \mid \|\mathbf{a}\| = 1, \mathbf{a} \in R^p\}$.

When the population parameters are unknown, the maximum likelihood estimates (MLEs) are employed in a sample version of the optimization problem. Assume two independent random samples of n_0, n_1 are drawn from the non-diseased and diseased populations, respectively. The estimated mean vectors and covariance matrices are, respectively, denoted as follows: $\hat{\mu}_0, \hat{\mu}_1$, and $\hat{\Sigma}_0, \hat{\Sigma}_1$. Moreover, let $\hat{\Delta}_\mu = \hat{\mu}_1 - \hat{\mu}_0$ and $\hat{Q}_d = \mathbf{a}^T \hat{\Sigma}_d \mathbf{a}$, for $d = 0, 1$. Replacing the unknown parameters in (1) by their corresponding MLEs defined above, we have a sample version of pAUC below:

$$\widehat{pAUC}_n(\mathbf{a}) = \int_0^1 \hat{F}_n(\mathbf{a}, u) du, \quad (2)$$

where

$$\hat{F}_n(\mathbf{a}, u) = \Phi \left(\frac{\mathbf{a}^T \hat{\Delta}_\mu - c(u) \sqrt{\hat{Q}_0}}{\sqrt{\hat{Q}_1}} \right).$$

Thus, \mathbf{a}^* of the optimal linear combination is estimated by the maximizer of (2):

$$\hat{\mathbf{a}}_n = \arg \max_{\mathbf{a} \in E_p} \widehat{pAUC}_n(\mathbf{a}).$$

The theorem below shows that $\hat{\mathbf{a}}_n$ is a strong consistent estimator of \mathbf{a}^* . A sketch of the proof is provided in Appendix.

Theorem 1. *Assume that $pAUC(\mathbf{a})$ in (1) is a continuous function of \mathbf{a} and has a unique maximizer \mathbf{a}^* in E_p . Then $\hat{\mathbf{a}}_n \rightarrow \mathbf{a}^*$ with probability 1 as $n_0, n_1 \rightarrow \infty$.*

In real applications, some of the coefficients in the best linear combination were found to be nearly zero. Numerically, their corresponding biomarkers might have limited contribution to the combination and thus to the disease prediction. In the following section, we will discuss how to assess the significance of biomarkers obtained in our maximizing procedure in terms of their discriminatory power. The proposed testing procedures will be embedded into our biomarker selection approaches in order to find a compact biomarker set, which consists of only significant biomarkers in disease diagnosis.

3 Hypothesis Testing and Biomarker Selection

3.1 Testing the Discriminatory Power

Considering only the class of linear combinations, we evaluate the global discriminatory power of a set of $p \geq 1$ biomarkers, \mathbf{X} , by testing the following hypotheses:

$H_{0,g}$: *The biomarker set has no discriminatory power to the disease*

versus

$H_{1,g}$: *The biomarker set has a discriminatory power to the disease.*

The null hypothesis $H_{0,g}$ is true if and only if the optimal linear combination of the biomarker set has no discriminatory power. Or equivalently, the maximal pAUC that the set can achieve through its linear combinations is not greater than the reference limit $t^2/2$, which is the pAUC value of the non-informative diagnosis with a diagonal ROC curve. That is,

$$H_{0,g} : pAUC(\mathbf{a}^*) \leq \frac{t^2}{2} \text{ versus } H_{1,g} : pAUC(\mathbf{a}^*) > \frac{t^2}{2}.$$

By maximizing the sample pAUC defined in (2), we obtain the maximal sample pAUC and use it as the test statistic. That is,

$$T_g = \max_{\mathbf{a} \in E_p} \widehat{pAUC}_n(\mathbf{a}) = \widehat{pAUC}_n(\hat{\mathbf{a}}_n) = \int_0^t \Phi \left(\frac{\hat{\mathbf{a}}_n^T \hat{\Delta}_\mu - c(u) \sqrt{\hat{Q}_0}}{\sqrt{\hat{Q}_1}} \right) du.$$

The null hypothesis $H_{0,g}$ is rejected if T_g is sufficiently large.

Let $\mathbf{X}^T = (\mathbf{X}_{i^-}^T, X_i)$, we consider to assess the contribution of X_i given the existence of other biomarkers, \mathbf{X}_{i^-} . The following hypothesis is tested:

$H_{0,c}$: *Given \mathbf{X}_{i^-} , X_i has no discriminatory power to the disease.*

The coefficients of the optimal linear combination of \mathbf{X} are written as $\mathbf{a}^{*T} = (\mathbf{a}_{i^-}^{*T}, a_i^*)$, where a_i^* is the corresponding coefficient of X_i . In this problem, we propose evaluating the biomarker X_i from a_i^* . That is, $H_{0,c}$ is equivalent to

$$H_{0,c} : a_i^* = 0.$$

The test statistic is the estimator of a_i^* , denoted by $T_{c,i} = \hat{a}_{n,i}$. The null hypothesis $H_{0,c}$ is then rejected if $T_{c,i}$ is either too small or too large.

Due to the complex formulation of the test statistics, the null distribution and the critical values are estimated by a parametric bootstrapping method. Under the

null hypothesis, the sampling distribution of the test statistic is estimated. Consider drawing two independent random samples of size n_1 and n_0 from the estimated null distribution. Then using the bootstrap samples to find the test statistic. Repeat the sampling B times. The critical value(s) is(are) then equal to the correspondent percentile(s) among these values.

3.2 Biomarker Selection

We now turn to the biomarker selection problem. Assume that all biomarkers are adequately standardized a priori and denoted the full standardized biomarker set by \mathbf{X} . Let $\hat{\mathbf{a}}_n^T = (\hat{a}_{n,1}, \dots, \hat{a}_{n,p})$ be the estimate of the optimal linear combination as before. The magnitude of $|\hat{a}_{n,i}|$ is used as an ordering criterion in the following stepwise biomarker selection approaches. Rearrange the biomarkers according to their corresponding $|\hat{a}_{n,i}|$ values in an ascending order. Denoted the rearranged vector by $\mathbf{X}^T = (X_{(1)}, \dots, X_{(p)})$.

We consider two stepwise selection methods: the Forward and the Backward approaches. Define A as the set of biomarkers under consideration in each step for convenience. The Forward procedure starts from a null A and tests the contribution of the potentially most discriminatory biomarker $X_{(p)}$. The biomarker is added to A if it is significant. Then it consecutively assesses $X_{(p-1)}$, $X_{(p-2)}$ and so on. On the other hand, the Backward procedure starts from testing the overall discriminatory power of $A = \{\mathbf{X}\}$. If an insignificance is obtained, we stop the selection and conclude that the full biomarker set is independent of the disease. With a significant global effect, one further determines whether the potentially least discriminatory biomarker $X_{(1)}$ is significant. Remove the biomarker from A if an insignificant result is present. Given the result, this procedure consecutively assesses the conditional contribution of $X_{(2)}$, of $X_{(3)}$, and so on. After evaluating the contribution of every individual biomarker, we conclude that the biomarkers remaining in A have a significant contribution to the linear combination in terms of pAUC. The details are presented below:

Forward Method

Step 1. Set $A = \emptyset$. Test the marginal effect of $X_{(p)}$ with respect to

$$H_{0,(p)} : X_{(p)} \text{ has no discriminatory power.}$$

If $H_{0,(p)}$ is rejected, add $X_{(p)}$ to A . Go to the next step.

Step 2. Test the significance of $X_{(p-1)}$ with respect to

$$H_{0,(p-1)} : \text{Given } A, X_{(p-1)} \text{ has no discriminatory power.}$$

If $H_{0,(p-1)}$ is rejected, add $X_{(p-1)}$ to A . Go to the next step.

⋮

Step p . Test the significance of $X_{(1)}$ with respect to

$$H_{0,(1)} : \text{Given } A, X_{(1)} \text{ has no discriminatory power.}$$

If $H_{0,(1)}$ is rejected, add $X_{(1)}$ to A . Stop.

Backward Method

Step 0. Set $A = \{\mathbf{X}\}$. Test the global effect of A with respect to

$$H_{0,(0)} : A \text{ has no discriminatory power.}$$

If $H_{0,(0)}$ is rejected, go to the next step; otherwise, stop and conclude $A = \emptyset$.

Step 1. Assess $X_{(1)}$ by removing $X_{(1)}$ from A and test the hypothesis,

$$H_{0,(1)} : \text{Given } A, X_{(1)} \text{ has no discriminatory power.}$$

If $H_{0,(1)}$ is rejected, add $X_{(1)}$ to A . Go to the next step.

Step 2. Assess $X_{(2)}$ by removing $X_{(2)}$ from A and test the hypothesis,

$$H_{0,(2)} : \text{Given } A, X_{(2)} \text{ has no discriminatory power.}$$

If $H_{0,(2)}$ is rejected, add $X_{(2)}$ to A . Go to the next step.

⋮

Step p . Assess the effect of $X_{(p)}$. If $A = \{X_{(p)}\}$, stop; otherwise, remove $X_{(p)}$ from A and test the following null hypothesis,

$$H_{0,(p)} : \text{Given } A, X_{(p)} \text{ has no discriminatory power.}$$

If $H_{0,(p)}$ is rejected, add $X_{(p)}$ to A . Stop.

Note that except in the initial step of the Backward method, there is no early stopping criterion in both approaches in order to minimize the risk of not taking the variation of $|\hat{a}_{n,i}|$ into the ordering criterion at the beginning of the procedure. Note that at every step, the biomarker set involved is likely to differ, thus the optimal linear combination should be recalculated for the hypothesis testing. Further, the two biomarker selection approaches assess the conditional discriminatory power of one target biomarker at every step and hence the related null hypothesis is $H_{0,c}$. For simplicity, one can consider a fixed significance level stepwisely in the procedure. To control the global type I error rate, a suitable multiplicity adjustment can be employed. For example, the Bonferroni adjustment suggests a α/p stepwise significance level in the Forward method, and $\alpha/(p+1)$ level in the Backward method for the global type I error rate to be controlled at α level.

4 Applications to Real Data Sets

We apply our procedures to two real examples in [6, 19]. By using the raw data and the standardized data, the optimal linear combinations of the full biomarker set are both reported in Table 1. We consider the following standardization: every biomarker in the raw data subtracts the non-diseased group mean and then divides by its pooled sample standard deviation from the two groups for a more uniform unit across biomarkers. The two proposed biomarker selection methods with 5% stepwise significance level are applied on the standardized data. The optimal linear combinations of the reduced biomarker set are given in Table 1.

The first example is a study of Duchenne Muscular Dystrophy (DMD).The DMD carriers generally are elevated by certain serum enzymes, not by physical symptoms. The measurements of three biomarkers of DMD of 87 normal and 38 carrier females were collected in this data set. The sample means of the three biomarkers in the normal and carrier groups are, respectively, $\hat{\mu}_0^T = (3.3932, 4.5213, 2.4863)$, $\hat{\mu}_1^T = (4.7615, 4.5228, 3.0105)$; and the sample covariance matrices are

$$\hat{\Sigma}_0 = \begin{pmatrix} .0316 & -.0039 & .0024 \\ -.0039 & .0065 & .0006 \\ .0024 & .0006 & .0113 \end{pmatrix}, \hat{\Sigma}_1 = \begin{pmatrix} .7683 & -.0050 & .3054 \\ -.0050 & .0094 & -.0064 \\ .3054 & -.0064 & .2268 \end{pmatrix}.$$

We observe from Table 1 that the contribution of the second biomarker is greatly downsized by the standardization. In fact, we find that the marginal distributions of the second biomarker of the two groups do not vary much. Consequently, it should have a limited discriminatory power. The reason that it has an inflated coefficient in the optimal linear combination based on the raw data is due to the fact that it has relatively small variances, which means that it is measured by a greater unit than other biomarkers. The standardization makes the units of the biomarkers more uniform. It leads to a more fair comparison across the biomarkers. After data

Table 1 The estimated best linear combination and the corresponding pAUC for the specificity range (0.9,1) in DMD and atherosclerotic coronary heart disease examples.

Case	Method	\hat{a}_1	\hat{a}_2	\hat{a}_3		\widehat{pAUC}_n
DMD	Full set (raw)	0.8350	0.5116	0.2026		0.0888
	Full set (Standardized)	0.9895	0.0653	0.1292		0.0888
	Forward (Standardized)	0.9657	0.0000	0.2597		0.0885
	Backward (Standardized)	0.9657	0.0000	0.2597		0.0885
Case	Method	lutein	TBARS	HDL_C	U_A	\widehat{pAUC}_n
Heart disease	Full set (raw)	0.9447	0.3258	0.0265	0.0274	0.0165
	Full set (Standardized)	0.7079	0.6754	0.0834	0.1890	0.0165
	Forward (Standardized)	1.0000	0.0000	0.0000	0.0000	0.0099
	Backward (Standardized)	1.0000	0.0000	0.0000	0.0000	0.0099

standardization, Table 1 shows that both Forward and Backward approaches select the first and the third biomarkers. We find that the decrement in the pAUC by removing the second biomarker is slim.

In another real example, we consider four biomarkers lutein, TBARS, HDL cholesterol (HDL_C), and uric acid (U_A) for construction of a classification tool for atherosclerotic coronary heart disease. A cohort of 434 subjects were selected for the analysis yielding 72 cases and 362 controls. One obtains an insignificant conclusion in testing the null hypothesis of normality. For the non-diseased and diseased groups, the estimated means of the four markers are $\hat{\mu}_0^T = (0.1275, 0.8845, 4.0766, 6.7724)$, $\hat{\mu}_1^T = (0.1402, 0.9337, 4.1225, 6.9112)$; and the two sample covariance matrices are

$$\hat{\Sigma}_0 = \begin{pmatrix} .0034 & -.0004 & -.0002 & -.0051 \\ -.0004 & .0285 & .0039 & .0417 \\ -.0002 & .0039 & .0488 & .0268 \\ -.0051 & .0417 & .0268 & .2846 \end{pmatrix}, \quad \hat{\Sigma}_1 = \begin{pmatrix} .0043 & .0033 & .0006 & .0067 \\ .0033 & .0415 & .0019 & .0426 \\ .0006 & .0019 & .0389 & .0010 \\ .0067 & .0426 & .0010 & .1504 \end{pmatrix}.$$

From Table 1, the impact of the first biomarker lutein, which has relatively small variances in the raw data, is downsized by the standardization. Before the biomarker selection, the first two biomarkers, lutein and TBARS, seem important to the disease from the magnitudes of their coefficients. However, the two stepwise selections produce the same conclusion that only the biomarker lutein achieves statistical significance, although there is a moderate reduction in the pAUC by discarding other three biomarkers.

5 Discussion

In this study, we focus on disease diagnosis with the presence of multiple biomarkers. We consider the class of linear combinations for an effective and easy-to-interpret summarization of the multiple biomarkers. The diagnostic power of a linear combination is evaluated upon its pAUC over a clinically relevant threshold region. In specific, we consider the requirement of a high specificity for the purpose of population screening.

Under the binormality assumption, the pAUC of a linear combination is estimated by employment of MLEs of the population parameters. In addition, the strong consistency of the estimated optimal linear combination is proved. We introduce a testing procedure to assess the overall diagnostic power of a set of biomarkers from the greatest pAUC it can achieve in the class of linear combinations. Furthermore, a testing procedure for determining the conditional contribution of a single biomarker given the existence of other biomarkers is developed. The parametric bootstrap method is applied to find the critical value(s) of the tests. These proposed tests are then embedded in two biomarker selection approaches. The applicability of the proposed methods is illustrated by two real data sets.

Differing with other algorithm-based marker selection approaches, the proposed methods select or discard a biomarker based upon the evidence of statistical significance. As a trade-off, to acquire statistical evidence, our methods necessarily involve many computations. As such, it decreases the feasibility of these methods for big data sets. Consequently, our methods are less appropriate in an exploratory study. We suggest the application of adequate data filtering for dimension reduction prior to advanced statistical confirmatory analysis, such as the construction of a diagnostic rule.

Appendix

Proof of Theorem 1. Since $E(\mathbf{X}|D)^2 < \infty$, by SLLN, as $n \rightarrow \infty$,

$$\hat{\mu}_0 \xrightarrow{a.s.} \mu_0, \hat{\mu}_1 \xrightarrow{a.s.} \mu_1, \hat{\Sigma}_0 \xrightarrow{a.s.} \Sigma_0, \hat{\Sigma}_1 \xrightarrow{a.s.} \Sigma_1.$$

Consequently, for any fixed $\mathbf{a} \in E_p$,

$$\mathbf{a}^T \hat{\Delta}_\mu \xrightarrow{a.s.} \mathbf{a}^T \Delta_\mu, \hat{Q}_0 \xrightarrow{a.s.} Q_0, \hat{Q}_1 \xrightarrow{a.s.} Q_1, \text{ and } \hat{F}_n(\mathbf{a}, u) \xrightarrow{a.s.} F(\mathbf{a}, u),$$

since $\Phi(\cdot)$ is a continuous function ([18], Theorem 1.10 (i)).

Further, since $\Phi(\cdot)$ is bounded, by the dominated convergence theorem,

$$\lim_{n \rightarrow \infty} \widehat{pAUC}_n(\mathbf{a}) = \lim_{n \rightarrow \infty} \int_0^1 F_n(\mathbf{a}, u) du = \int_0^1 F(\mathbf{a}, u) du = pAUC(\mathbf{a}).$$

Hence, for any fixed $\mathbf{a} \in E_p$, $\widehat{pAUC}_n(\mathbf{a}) \rightarrow pAUC(\mathbf{a})$ with probability 1. Since $\mathbf{a}^* = \arg \max_{\mathbf{a} \in E_p} pAUC(\mathbf{a})$ is assumed unique, it implies that for every $\epsilon > 0$, $\sup_{\mathbf{a}: d(\mathbf{a}, \mathbf{a}^*) \geq \epsilon} pAUC(\mathbf{a}) < pAUC(\mathbf{a}^*)$. As a consequence, according to Theorem 5.7 ([21]),

$$\hat{\mathbf{a}}_n \rightarrow \mathbf{a}^*, \text{ with probability 1.}$$

References

- [1] ETZIONI, R., KOOPERBERG, C., PEPE, M., SMITH, R. AND GANN, P. H. (2003). Combining biomarkers to detect disease with application to prostate cancer. *Biostatistics* **4**, 523-538.
- [2] HSU, M.-J. AND HSUEH, H.-M. (2012). The linear combinations of biomarkers which maximize the partial area under the ROC curves. *Computational Statistics*, to appear. DOI: 10.1007/s00180-012-0321-5.

- [3] KOMORI, O. AND EGUCHI, S. (2010). A boosting method for maximizing the partial area under the ROC curve. *BFC Bioinformatics* **11**, 314-330.
- [4] LASKO, T. A., BHAGWAT, J. G., ZOU, K. H. AND OHNO-MACHADO L. (2005). The use of receiver operating characteristic curves in biomedical informatics. *Journal of Biomedical Informatics* **38**, 404-415.
- [5] LIN, H., ZHOU, L., PENG, H. AND ZHOU, X.-H. (2011). Selection and combination of biomarkers using ROC method for disease classification and prediction. *The Canadian Journal of Statistics* **39**, 324-343.
- [6] LIU, A., SCHISTERMAN, E. F. AND ZHU, Y. (2005). On linear combinations of biomarkers to improve diagnostic accuracy. *Statistics in Medicine* **24**, 37-47.
- [7] MA, S. AND HUANG, J. (2005). Regularized ROC method for disease classification and biomarker selection with microarray data. *Bioinformatics* **21**, 4356-4362.
- [8] MA, S. AND HUANG, J. (2007). Combining multiple markers for classification Using ROC. *Biometrics* **63**, 751-757.
- [9] MADU, C. O. AND LU, Y. (2010). Novel diagnostic biomarkers for prostate cancer. *Journal of Cancer* **1**, 150-177.
- [10] MARROCCO, C., DUIN, R. P. W. AND TORTORELLA, F. (2008). Maximizing the area under the ROC curve by pairwise feature combination. *Pattern Recognition* **41**, 1961-1974.
- [11] National Cancer Institute: PDQ@Prostate Cancer Screening. Bethesda, MD: National Cancer Institute. Date last modified /06/08/2012). Available at: <http://www.cancer.gov/cancertopics/pdq/screening/prostate/HealthProfessional/Page3#Section.67>
- [12] PEPE, M. S. AND THOMPSON, M. L. (2000). Combining diagnostic test results to increase accuracy. *Biostatistics* **1**, 123-140.
- [13] PEPE, M. S., LONGTON, G., ANDERSON, G. L. AND SCHUMMER, M. (2003). Selecting differentially expressed genes from microarray experiments. *Biometrics* **59**, 133-142.
- [14] PEPE, M. S., CAI, T. AND LONGTON, G. (2006). Combining predictors for classification using the area under the receiver operating characteristic curve. *Biometrics* **62**, 221-229.
- [15] RICAMATO, M. T. AND TORTORELLA, F. (2011). Partial AUC maximization in a linear combination of dichotomizers. *Pattern Recognition* **44**, 2669-2677.
- [16] ROBIN, X., TURCK, N., HAINARD, A., TIBERTI, N., LISACEK, F., SANCHEZ, J.-C. AND MULLER, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BFC Bioinformatics* **12**, 77-84.
- [17] SU, J. Q. AND LIU, J. S. (1993). Linear combinations of multiple diagnostic markers. *Journal of the American Statistical Association* **88**, 1350-1355.
- [18] SHAO, J. (1999). *Mathematical Statistics*. Springer-Verlag Inc.
- [19] TIAN, L. (2010). Confidence interval estimation of partial area under curve based on combined biomarkers. *Computational Statistics & Data Analysis* **54**, 466-472.
- [20] TURCK, N., VUTSKITS, L., SANCHEZ-PENA, P., ROBIN, X., HAINARD, A., GEX-FABRY, M., FOUADA, C., BASSEM, H., MUELLER, M., LISACEK, F., PUYBASSET, L. AND SANCHEZ, J.-C. (2010). A multiparameter panel method for outcome prediction following aneurysmal subarachnoid hemorrhage. *Intensive Care Medicine* **36**, 107-115.
- [21] VAART, A. W. VAN DER (1998). *Asymptotic Statistics*. Cambridge University Press.
- [22] WANG, Z. AND CHANG, Y.-C. I. (2011). Marker selection via maximizing the partial area under the ROC curve of linear risk scores. *Biostatistics* **12**, 369-385.
- [23] ZHOU, X. H., CHEN, B., XIE, Y. M., TIAN, F., LIU, H. AND LIANG, X. (2012). Variable selection using the optimal ROC curve: An application to a traditional Chinese medicine study on osteoporosis disease. *Statistics in Medicine* **31**, 628-635.