Empirical Likelihood Ratio Confidence Interval Estimation of Best Linear Combinations of Biomarkers

Xiwei Chen[a], Albert Vexler[a,*], Marianthi Markatou[a]

[a] *Department of Biostatistics, University at Buffalo, USA*

Abstract

A novel smoothed empirical likelihood (EL) approach that incorporates kernel estimation of the area under the receiver operating characteristic curve (AUC) to construct nonparametric confidence intervals of AUC based on the best linear combination (BLC) of biomarkers is proposed. The method has several advantages including the feasibility to use gradient-based techniques for fast computation of BLC coefficients and to employ powerful likelihood methods without specification of underlying data distributions. Simulation results show that the new method performs well even when the distribution of biomarkers is skewed, a situation commonly met in practice. A data set from a clinical experiment related to atherosclerotic coronary heart disease is used to illustrate the efficiency of the proposed method.

Keywords: Area under the ROC curve; Best linear combination; Empirical likelihood; Kernel; Receiver operating characteristic curve (ROC).

1. Introduction

The receiver operating characteristic (ROC) curve methodology is often applied to evaluate the performance of diagnostic markers that classify subjects into two populations: diseased and nondiseased. Suppose that a biomarker,

*Correspondence to: avexler@buffalo.edu. 715 Kimball Tower, 3435 Main St, Buffalo, NY 14214, USA. Tel: 716-829-2754; fax: 716-829-2200.

Email addresses: xiweiche@buffalo.edu (Xiwei Chen), avexler@buffalo.edu (Albert Vexler), markatou@buffalo.edu (Marianthi Markatou)

measured on a continuous scale, is performed on n diseased subjects, yielding independent, identically distributed (i.i.d.) measurements $\mathbf{X}_i$, $i = 1, ..., n$, with distribution function F, and on m nondiseased subjects, yielding i.i.d. measurements $\mathbf{Y}_j$, $j = 1, ..., m$, with distribution function G. The ROC curve is a plot of $R(t) = 1 - F\{G^{-1}(1-t)\}$ against t, $0 \leq t \leq 1$ (Pepe, 1997; Faraggi and Reiser, 2002). The area under the ROC curve (AUC), as a common measure of the diagnostic performance of a biomarker, is equal to $P(X > Y)$ (Bamber, 1975).

In practice, different markers are usually related to the disease in various magnitudes and different directions. For example, low levels of high density lipoprotein (HDL)-cholesterol and high levels of thiobarbuturic acid reacting substances (TBARS) are biomarkers of oxidative stress and antioxidant status, as indicators of coronary heart disease (Schisterman et al., 2002). When multiple biomarkers are available, we are interested in seeking a simple best linear combination (BLC) of biomarkers such that the combined score achieves the maximum AUC over all possible linear combinations. The maximum AUC measures the ability to discriminate between the control and the disease groups (Pepe and Thompson, 2000; McIntosh and Pepe, 2002).

A variety of publications have focused on the BLC of multiple biomarkers in both parametric and nonparametric cases. Su and Liu (1993) investigated the BLC and the corresponding maximum AUC under multivariate normal assumptions. Based on the point estimator given by Su and Liu (1993), Reiser and Faraggi (1997) derived confidence intervals for the BLC-based AUC under normal assumptions, which is usefull when the sample size in the two groups is small and equal or moderate and unequal, respectively. In practice, however, the distributions of biomarker measurements commonly deviate from the normal assumptions (Limpert et al., 2001). For example, epidemiological studies have demonstrated that TBARS has a distribution with heavy tails (Schisterman et al., 2001). In this case, loss of efficiency of the BLC coefficients constructed under normality assumptions can be expected and alternative approaches are needed to accomodate different data distributions.

Pepe et al. (2006) considered the empirical AUC and noticed that the empirical BLC estimator is consistent under the generalized linear model assumption. They noted that sophisticated computational algorithms are required because the empirical AUC is not a continuous function. Ma and Huang (2005, 2007) also noticed the computational complexity in maximizing the empirical AUC and proposed a corresponding sigmoid approximation. Under the generalized linear model assumption, it is proved that the sigmoid AUC converges to the maximized theoretical AUC and the corresponding BLC coefficients are asymptotically normally distributed (Ma and Huang, 2007). Instead of the specific sigmoid AUC, Vexler et al. (2006) demonstrated consistency, uniqueness, and ease of computation of the kernel-smoothed AUC estimator. Furthermore, they provided an upper confidence bound for the maximum AUC. The distribution-free confidence interval estimation of the maximum AUC remains unsolved.

As a nonparametric alternative to the optimal parametric likelihood method (Markatou et al., 1998), the EL methodology (Owen, 1990, 2001) is an efficient way to construct confidence intervals. In the context of a single biomarker, Qin and Zhou (2006) proposed an EL-based confidence interval for the AUC and demonstrated that it outperforms the existing normal approximation-based intervals and bootstrap intervals, particularly when the AUC is close to one.

In this paper, we extend and modify the method of Qin and Zhou (2006) to adjust for multiple biomarkers. Incorporating a kernel distribution estimation (Azzalini, 1981; Nadaraya, 1964) into the EL-based confidence interval estimation, we propose smoothed EL-based confidence intervals for BLC-based AUCs. The advantages of the proposed method include: (1) the smooth estimate of the AUC is differentiable, making gradient-based methods feasible and the computation of BLC coefficients simple, fast and unique; (2) it is robust to underlying distributions of biomarkers' measurements; (3) it employs the EL methodology, that is known to be an approximate nonparametric most powerful tool based on likelihood ratios (Lazar and Mykland, 1998); (4) the EL-based confidence interval is range preserving (Hall and La Scala, 1990) and therefore always lies between 0.5 and 1. The proposed method is illustrated with a study of the ac-

curacy of biomarkers related to the atherosclerotic coronary heart disease. An EL interval for AUC is derived based on a linear combination of measurements of four important biomarkers related to the atherosclerotic coronary heart disease including lutein, TBARS, HDL cholesterol and uric acid, obtained from a 12-hour fasting blood specimen for biochemical analysis at baseline.

The paper is organized as follows. Section 2 develops the smoothed EL-based confidence interval for the BLC-based AUC. Section 3.2 presents simulation studies to compare the relative performance of the proposed smoothed EL-based interval with the existing intervals for the BLC-based AUC. In Section 4, we apply our method to a real example related to atherosclerotic coronary heart disease. Section 5 presents a broader discussion on deriving linear combinations of biomarkers to improve the diagnostic accuracy. Conditions and proofs are deferred to the Appendix.

2. Methods

Consider a study with d continuous-scale biomarkers yielding measurements $\mathbf{X}_i = (X_{1i}, ..., X_{di})^T$, $i = 1, ..., n$, on n diseased subjects and measurements $\mathbf{Y}_j = (Y_{1j}, ..., Y_{dj})^T$, $j = 1, ..., m$, on m nondiseased patients, respectively. We are interested in constructing effective one-dimensional combined scores of biomarkers measurements, say, $X(\mathbf{a}) = \mathbf{a}^T\mathbf{X}$ and $Y(\mathbf{a}) = \mathbf{a}^T\mathbf{Y}$, such that the AUC based on these scores is maximized over all possible linear combinations of biomarkers. Define $A(\mathbf{a}) = P(X(\mathbf{a}) > Y(\mathbf{a}))$; the statistical problem is to estimate the maximum AUC defined as $A = A(\mathbf{a}_0)$, where $\mathbf{a}_0$ are the BLC coefficients satisfying $\mathbf{a}_0 = \arg\max_{\mathbf{a}} A(\mathbf{a})$. For simplicity, we assume that the first component of the vector $\mathbf{a}$ equals to 1 throughout the paper; see Pepe et al. (2006) and Ma and Huang (2007).

2.1. Confidence interval for a single biomarker-based AUC

Qin and Zhou (2006) developed an EL-approach for constructing confidence intervals for the AUC using a single biomarker (in our notation $d = 1$ and

$\mathbf{a} = \mathbf{a}_0 = 1$). The confidence interval estimation is executed via construction of the empirical likelihood ratio (ELR) test statistic for testing the hypothesis $H_0 : A = A_0$ versus $H_a : A \neq A_0$.

Qin and Zhou (2006) based the ELR test statistic on the concept of placement value of a diseased subject (Pepe and Cai, 2004). The placement value is defined as $U = 1 - G(X)$, where U can be interpreted as the proportion of nondiseased subjects with their biomarker measurements greater than X. Noticing that $E(1 - U) = A$, where A denotes the AUC based on a single marker, EL inference for the AUC for the case of a single marker is developed. Specifically, let $\mathbf{p} = (p_1, p_2, ..., p_n)^T$ be a probability weight vector, $\sum_{i=1}^{n} p_i = 1$ and $p_i \geq 0$ for all $i = 1, ..., n$. The profile EL for the AUC, evaluated at the true value A_0 of AUC, can be defined as

$$\tilde{L}(A_0) = \sup\{\prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i(1 - U_i) = A_0\}, \tag{1}$$

where $U_i = 1 - G(X_i)$, $i = 1, ..., n$. When the distribution function G of the nondiseased population is unknown, the empirical distribution $\hat{G}$ is used. Accordingly, replacing U_i by its estimator $\hat{U}_i = 1 - \hat{G}(X_i)$ in equation (1), and using Lagrange multipliers, one can obtain that

$$p_i = \frac{1}{n} \frac{1}{1 + \lambda(1 - \hat{U}_i - A_0)}, i = 1, ..., n.$$

Furthermore, the Lagrange multiplier λ is the root of

$$\frac{1}{n} \sum_{i=1}^{n} \frac{1 - \hat{U}_i - A_0}{1 + \lambda(1 - \hat{U}_i - A_0)} = 0.$$

The corresponding empirical log-likelihood ratio for the AUC is

$$l(A_0) = 2 \sum_{i=1}^{n} \log(1 + \lambda(1 - \hat{U}_i - A_0)). \tag{2}$$

Thus, an EL-based confidence interval for the AUC of a single biomarker is also constructed based on the empirical log-likelihood ratio, which has an asymptotic scaled Chi-square distribution (Qin and Zhou, 2006).

However, this approach cannot be directly applied when the estimation of best linear combination of multiple biomarkers is involved. In the next section,

we will use kernels to construct the EL-based confidence interval estimation for the BLC-based maximum AUC. Incidentally, Chen and Hall (1993) showed that the kernel approach is very efficient when EL is used to estimate confidence intervals for quantiles.

2.2. Confidence interval for the BLC-based AUC

When multiple biomarkers are available, e.g., $d \geq 2$, we generalize the method of Qin and Zhou (2006) considering a BLC of biomarkers. First the placement values are defined as $U_i(\mathbf{a}_0) = 1 - G_{\mathbf{a}_0}(\mathbf{a_0}^T\mathbf{X}_i)$, $i = 1, ..., n$, where $G_{\mathbf{a}_0}$ is the cumulative distribution function of the combined score $Y(\mathbf{a}_0)$ from the nondiseased group.

In the case that $\mathbf{a}_0$ is known, we can simply replace U_i in equation (1) by $U_i(\mathbf{a}_0)$ or its corresponding empirical distribution estimator. However, the dependence of placement values on not only the unknown cumulative distribution function $G_{\mathbf{a}_0}$ but also the unknown $\mathbf{a}_0$ further complicates construction of the ELR test statistic.

When the best linear combination coefficient $\mathbf{a}_0$ is unknown, the estimation of $\mathbf{a}_0$ will be associated with a maximization of the AUC and we will need a simple and unique solution with respect to the A_0 estimation. The use of the empirical distribution function is problematic because of its discontinuity, so we consider kernel estimation of the distribution function. Let k be a symmetric kernel function and define $K_h(x) = \int_{-\infty}^{x/h} k(u)du$; see details related to the definitions in Appendix A. Then, the kernel estimator of $1 - U_i(\mathbf{a}_0)$ is $v_i(\mathbf{a_0}) = \sum_{j=1} K_h(\mathbf{a_0}^T\mathbf{X}_i - \mathbf{a_0}^T\mathbf{Y}_j)/m$, $i = 1, ..., n$.

The EL for the BLC-based AUC is given as

$$L(A_0) = \sup\{\prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i v_i(\hat{\mathbf{a}}_0) = A_0\}, \tag{3}$$

where $\hat{\mathbf{a}}_0$ satisfies $\sum_{i=1}^{n} p_i \, \partial v_i(\mathbf{a})/\partial \mathbf{a} \mid_{\mathbf{a}=\hat{\mathbf{a}}_0} = \mathbf{0}$. One can show (using Lagrange multipliers) that

$$p_i = \frac{1}{n} \frac{1}{1 + \lambda \, (v_i(\hat{\mathbf{a}}_0) - A_0)}, i = 1, ..., n, \tag{4}$$

where the Lagrange multiplier λ is the root of

$$\frac{1}{n}\sum_{i=1}^{n}\frac{v_i(\hat{\mathbf{a}}_0)}{1+\lambda\left(v_i(\hat{\mathbf{a}}_0)-A_0\right)}=A_0. \tag{5}$$

Under the alternative hypothesis H_a, we have just the constraint $\sum_{i=1}^{n} p_i = 1$, and hence $L(A_0) = (1/n)^n$ at $p_i = 1/n$. Therefore, the empirical log-likelihood ratio test statistic is

$$l(A_0) = -2\log ELR(A_0) = 2\sum_{i=1}^{n}\log(1+\lambda(v_i(\hat{\mathbf{a}}_0)-A_0)). \tag{6}$$

The objective of confidence interval construction based on the empirical likelihood ratio $l(A_0)$ is to include all values of A in the interval $R_\alpha(A) = \{A : l(A) \le c_\alpha\}$ for pre-specified α that satisfy $P(A_0 \in R_\alpha(A)) = 1-\alpha$. The value of the constant $c_\alpha > 0$ depends on the coverage probability of the interval. For future use, we define $A^K_{m,n}(\mathbf{a}) = \sum_{i=1}^{n} v_i(\mathbf{a})/n$ and $\hat{\mathbf{a}}_K = \arg\max_{\mathbf{a}} A^K_{m,n}(\mathbf{a})$. The following proposition establishes the asymptotic distribution of the statistic $l(A_0)$ that is needed to construct the confidence interval for the maximum AUC.

Proposition 1. *Let $\mathbf{X}_i$, $i = 1, ..., n$, and $\mathbf{Y}_j$, $j = 1, ..., m$, be i.i.d. measurements for diseased subjects and nondiseased subjects, respectively. Assume that regularity conditions (C1)-(C7) presented in Appendix A are satisfied. Then, under $H_0 : A = A_0$, the asymptotic distribution of $l(A_0)$ is a scaled Chi-square distribution with one degree of freedom, i.e.,*

$$\gamma(A_0)l(A_0) \xrightarrow{d} \chi^2_1, \text{ as } n, m \to \infty, \tag{7}$$

where

$$\gamma(A_0) = \frac{m\hat{\sigma}^2}{(m+n)s^2}, \hat{\sigma}^2 = \sum_{i=1}^{n}(v_i(\hat{\mathbf{a}}_K) - \sum_{j=1}^{n} v_j(\hat{\mathbf{a}}_K)/n)^2/n, s^2 = \frac{m\hat{\sigma}^2_{10} + n\hat{\sigma}^2_{01}}{m+n},$$

$$\sigma^2_{10} = Cov(K_h(\mathbf{a_0}^T\mathbf{X}_1 - \mathbf{a_0}^T\mathbf{Y}_1), K_h(\mathbf{a_0}^T\mathbf{X}_1 - \mathbf{a_0}^T\mathbf{Y}_2)),$$

$$\sigma^2_{01} = Cov(K_h(\mathbf{a_0}^T\mathbf{X}_1 - \mathbf{a_0}^T\mathbf{Y}_1), K_h(\mathbf{a_0}^T\mathbf{X}_2 - \mathbf{a_0}^T\mathbf{Y}_1)),$$

and $\hat{\sigma}^2_{10}$, $\hat{\sigma}^2_{01}$ are the corresponding estimates.

Proof. It is outlined in the Appendix B. □

Thus, the $100\,(1-\alpha)\,\%$ EL confidence interval for the maximum AUC based on the BLC of multiple markers can be constructed as

$$R_\alpha(A_0) = \left\{A_0 : \gamma(A_0) l(A_0) \leq \chi_1^2\,(1-\alpha)\right\},$$

where $\chi_1^2\,(1-\alpha)$ is the $(1-\alpha)$th quantile of the Chi-square distribution with 1 degree of freedom.

Remark 1. *The conditions that the kernel function needs to satisfy for the results of Proposition 1 to be true are reasonable and generally used in density estimation. An example of a second order kernel that satisfies these conditions is the Gaussian kernel. Here, the random vectors* $\mathbf{X}$ *and* $\mathbf{Y}$ *should satisfy* $E|\mathbf{X}-\mathbf{Y}|^t < \infty$ *and* $\delta \in (1/2t, 1/4)$ *for* $t > 2$.

We further discuss this point in Appendix C.

Remark 2. *Following Vexler et al. (2006), the estimators of the best linear coefficients* $\hat{\mathbf{a}}_K$ *and the corresponding AUC estimator* $A_{m,n}^K(\hat{\mathbf{a}}_K)$ *exist uniquely if (1) the distributions of biomarkers in the case and control group are both absolutely continuous; (2) all biomarkers have discriminatory abilities and there is one biomarker has a better discriminant ability than other markers and (3) the second derivative of* $A_{m,n}^K(\mathbf{a})$ *exists and is nonzero.*

Assumption (2) can be satisfied via variable selection procedures such as stepwise selection, the best subset selection and LASSO (Tibshirani, 1996). Condition (C7) presented in Appendix A guarantees that assumption (3) holds.

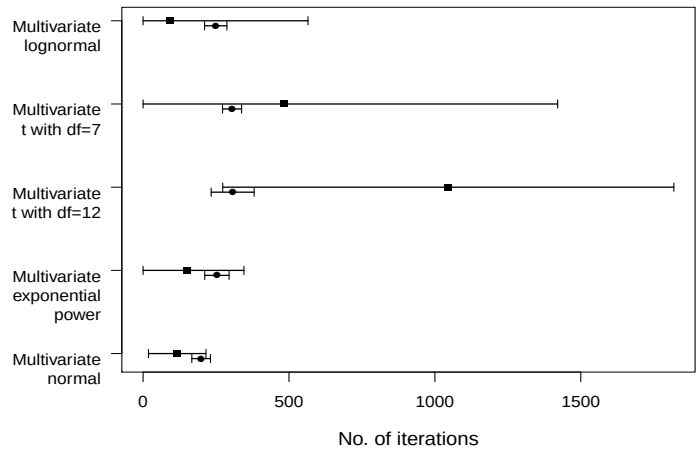
3. Monte Carlo Simulations

3.1. Computational strategy

Computations are conducted using R (R Development Core Team, 2013). The EL log-likelihood ratio test statistic can be obtained using the function *el.test* in the *emplik* package. The maximization of the smooth AUC objective

function is carried out based on gradient-based methods such as the quasi-Newton algorithm (Shanno, 1970) using the basic procedure *optim* in R. The gradient of the kernel-based AUC objective function can be calculated in a simple and fast manner. The related computer codes used to carry out the numerical results presented in this paper are provided in the Supplementary material.

Figure 1: Number of iterations (x-axis) of the maximization procedure with $m = n = 100$ and $d = 10$. The y-axis presents the distributions in ascending order of their empirical skewness. The filled squares represent the number of iterations of the kernel-based AUC method and the filled circles represent the number of iterations used by the empirical AUC.



On the contrary, the optimization of the empirical AUC requires a brutal search or the development of special algorithms. For example, the Nelder-Mead method (Nelder and Mead, 1965) can be used, and it works reasonably well for non-differentiable functions. For comparison purpose, the number of iterations for the maximization of the empirical AUC and the kernel-based AUC are calculated for different data distributions.

Figure 1 plots the median of the number of iterations (x-axis) each maximization requires versus ascending ordering of distributions according to their average skewness coefficient (y-axis) over 500 simulations. The simulated data

sets for both the control group and the case group consist of $d = 10$ biomarkers with $m = n = 100$ and with underlying distributions being multivariate normal distribution, multivariate exponential power distribution, multivariate t distribution and multivariate lognormal distribution (Johnson, 2013). As Figure 1 illustrates, the kernel-based AUC requires a smaller number of iterations for most of the selected distributions, indicating the computational advantage of the proposed method. For details about the parameter settings and related concepts, please see section 1 of the supplementary material.

3.2. Simulation studies

In this section, we report the simulation studies for evaluating coverage accuracy and interval length of the newly proposed EL interval for the optimal AUC based on multiple biomarkers. The proposed smoothed EL-based confidence interval is compared with the normality-based confidence intervals derived by Reiser and Faraggi (1997).

To study the performance of our method, we computed the coverage probability and associated lengths of simulated confidence intervals, setting the theoretical coverage probability equal to 95% at a nominal significance level $\alpha = 0.05$. We simulated the cases of two and four biomarkers with different underlying distributions, including multivariate normal distribution, multivariate lognormal distribution and multivariate power exponential distribution (Johnson, 2013). For future use, denote $\mathbf{\Sigma} = \{\sigma_{ij}\}$ where $\sigma_{ii} = 1$ and $\sigma_{ij} = \rho$ for $i \neq j$, $i, j = 1, \ldots, d$.

In each study, we replicated the simulation experiment $MC = 10,000$ times so the simulation error defined as $z_{1-\alpha/2}(\alpha(1-\alpha)/MC)^{1/2}$ is 0.0043, where $z_{1-\alpha/2}$ is the $100(1-\alpha/2)$th quantile of a standard normal distribution. Confidence intervals are calculated for values of AUC equal to 0.75, 0.85, 0.90 and 0.95 in balanced and unbalanced designs with both small samples and large samples. We considered the commonly used Gaussian kernel and the optimal Epanechnikov kernel which yields the minimum asymptotic mean-squared-error for the estimation of the density function.

Table 1: Coverage probabilities and lengths (in parentheses) of the normality-based and EL-based confidence intervals (CI) with a 95% nominal level for various AUC values in the case of four biomarkers with a multivariate normal distribution and $\rho = 0$.

AUC	Normality	EL-based CI with Gaussian kernels			EL-based CI with Epanechnikov kernels		
(m,n)	based CI	$\delta = 1/3$	$\delta = 1/5$	$\delta = 1/7$	$\delta = 1/3$	$\delta = 1/5$	$\delta = 1/7$
0.75							
(25,25)	0.951(0.214)	0.873(0.220)	0.892(0.217)	0.904(0.214)	0.891(0.216)	0.907(0.212)	0.916(0.208)
(75,25)	0.948(0.178)	0.924(0.193)	0.933(0.192)	0.937(0.189)	0.923(0.191)	0.939(0.187)	0.931(0.184)
(25,75)	0.950(0.178)	0.919(0.195)	0.928(0.194)	0.936(0.191)	0.930(0.194)	0.935(0.190)	0.948(0.186)
(75,75)	0.952(0.127)	0.932(0.144)	0.936(0.144)	0.941(0.142)	0.940(0.143)	0.947(0.141)	0.946(0.138)
(200,50)	0.950(0.124)	0.937(0.141)	0.940(0.141)	0.942(0.139)	0.940(0.141)	0.948(0.139)	0.942(0.136)
(200,100)	0.951(0.100)	0.942(0.111)	0.946(0.111)	0.948(0.110)	0.946(0.111)	0.948(0.109)	0.939(0.107)
(200,200)	0.950(0.086)	0.942(0.092)	0.944(0.092)	0.948(0.091)	0.947(0.092)	0.953(0.090)	0.938(0.089)
0.85							
(25,25)	0.951(0.187)	0.853(0.166)	0.874(0.167)	0.887(0.166)	0.874(0.166)	0.891(0.166)	0.905(0.165)
(75,25)	0.953(0.150)	0.911(0.148)	0.922(0.148)	0.929(0.148)	0.916(0.148)	0.913(0.146)	0.918(0.145)
(25,75)	0.950(0.151)	0.921(0.153)	0.927(0.153)	0.932(0.152)	0.933(0.153)	0.933(0.151)	0.931(0.150)
(75,75)	0.950(0.110)	0.931(0.113)	0.933(0.113)	0.935(0.112)	0.935(0.113)	0.940(0.112)	0.925(0.111)
(200,50)	0.948(0.107)	0.941(0.111)	0.943(0.111)	0.940(0.110)	0.937(0.111)	0.941(0.110)	0.940(0.109)
(200,100)	0.951(0.086)	0.942(0.087)	0.944(0.087)	0.941(0.087)	0.951(0.087)	0.950(0.087)	0.923(0.086)
(200,200)	0.950(0.072)	0.942(0.072)	0.945(0.072)	0.944(0.072)	0.948(0.072)	0.946(0.071)	0.917(0.071)
0.90							
(25,25)	0.954(0.157)	0.820(0.128)	0.848(0.130)	0.868(0.131)	0.854(0.131)	0.873(0.132)	0.883(0.132)
(75,25)	0.952(0.127)	0.890(0.116)	0.904(0.117)	0.912(0.117)	0.899(0.116)	0.912(0.117)	0.905(0.117)
(25,75)	0.948(0.127)	0.910(0.121)	0.919(0.122)	0.927(0.122)	0.922(0.122)	0.927(0.123)	0.914(0.122)
(75,75)	0.946(0.091)	0.924(0.090)	0.929(0.090)	0.930(0.009)	0.936(0.091)	0.931(0.091)	0.919(0.090)
(200,50)	0.947(0.088)	0.940(0.088)	0.943(0.088)	0.941(0.088)	0.940(0.088)	0.937(0.088)	0.927(0.088)
(200,100)	0.951(0.070)	0.939(0.070)	0.941(0.070)	0.945(0.070)	0.948(0.070)	0.942(0.070)	0.913(0.070)
(200,200)	0.950(0.058)	0.942(0.058)	0.944(0.058)	0.944(0.058)	0.952(0.058)	0.940(0.058)	0.890(0.058)
0.95							
(25,25)	0.951(0.112)	0.739(0.077)	0.788(0.082)	0.822(0.085)	0.786(0.082)	0.836(0.086)	0.840(0.091)
(75,25)	0.953(0.089)	0.844(0.071)	0.865(0.073)	0.878(0.075)	0.854(0.072)	0.886(0.076)	0.881(0.077)
(25,75)	0.953(0.088)	0.868(0.078)	0.890(0.080)	0.904(0.081)	0.888(0.080)	0.895(0.081)	0.891(0.083)
(75,75)	0.952(0.061)	0.906(0.058)	0.915(0.059)	0.915(0.060)	0.916(0.059)	0.919(0.060)	0.885(0.061)
(200,50)	0.951(0.060)	0.926(0.057)	0.932(0.058)	0.934(0.058)	0.930(0.058)	0.924(0.059)	0.902(0.060)
(200,100)	0.948(0.046)	0.933(0.046)	0.936(0.046)	0.934(0.047)	0.942(0.047)	0.924(0.047)	0.887(0.048)
(200,200)	0.949(0.038)	0.942(0.038)	0.944(0.038)	0.940(0.039)	0.943(0.038)	0.928(0.039)	0.869(0.039)

Choices of bandwidths have already been well studied in the density or distribution function estimation literature (Lloyd, 1998; Lloyd and Yong, 1999; Zou et al., 1997). In the simulations, we considered bandwidths of order $h = (n+m)^{-\delta}$ with $\delta = 1/3$, $1/5$ and $1/7$. For a second order kernel such as the Gaussian kernel, it is easy to check that conditions (C1)-(C7) stated in Appendix A are satisfied when $\delta = 1/5$ and $1/7$. The case that $\delta = 1/3$ shows the results when the conditions are slightly violated.

Table 1 shows the simulation results for four biomarkers ($d = 4$) following multivariate normal distributions with $\boldsymbol{\mu}_{\mathbf{X}} = (\mu, 0.5, 0, -0.2)^T$, $\boldsymbol{\mu}_{\mathbf{Y}} = (0.2, 0, 0.5, 0)^T$ and $\boldsymbol{\Sigma}_{\mathbf{X}} = \boldsymbol{\Sigma}_{\mathbf{Y}} = \boldsymbol{\Sigma}$, where $\rho = 0$. Here the covariance matrices are assumed to be equal for simplicity but without loss of generality. The value of $\mu \in \mathbf{R}$ varies to reach pre-specified values of AUC based on the point estimator of Su and Liu (1993). Table 1 suggest that the Gaussian kernel is not very sensitive to the choice of bandwidths, yielding equivalent performance in terms of achieved coverage probability and associated confidence interval lengths. Furthermore, while one would have expected that, for normally distributed data, the Gaussian kernel would perform considerably better in terms of coverage probability, our simulations indicate that the Epanechnikov kernel shows equivalent performance across all values of bandwidths and sample sizes. Same results can be observed for the case of two biomarkers. For details, please see the supplementary materials. As shown in Tables 2 and 3, these results are also confirmed for the case where $\log \mathbf{X}$ and $\log \mathbf{Y}$ are normally distributed according to the aforementioned normal distributions, making the distribution of the markers a lognormal distribution. However, as expected, normality based confidence intervals have very low coverage probabilities.

In addition, the scenarios where the off-diagonal elements of $\boldsymbol{\Sigma}$ are $\rho = 0.4, 0.7$ for the aforementioned multivariate normal distribution and the multivariate lognormal distribution are also considered. Tables 1 through 8 of the supplementary material present the results. Here we note that higher correlations yield slightly smaller coverage probabilities but the proposed method performs well even in skew data for moderate to large sample sizes.

Table 2: Coverage probabilities and lengths (in parentheses) of the normality-based and EL-based confidence intervals (CI) with a 95% nominal level for various AUC values in the case of two biomarkers with a multivariate lognormal distribution and $\rho = 0$.

AUC	Normality	EL-based CI with Gaussian kernels			EL-based CI with Epanechnikov kernels		
(m,n)	based CI	$\delta = 1/3$	$\delta = 1/5$	$\delta = 1/7$	$\delta = 1/3$	$\delta = 1/5$	$\delta = 1/7$
0.75							
(25,25)	0.947(0.238)	0.913(0.232)	0.916(0.229)	0.917(0.227)	0.912(0.232)	0.917(0.229)	0.917(0.227)
(75,25)	0.916(0.216)	0.928(0.198)	0.931(0.196)	0.932(0.194)	0.928(0.198)	0.930(0.196)	0.932(0.194)
(25,75)	0.898(0.196)	0.936(0.199)	0.938(0.197)	0.939(0.195)	0.936(0.199)	0.938(0.197)	0.938(0.195)
(75,75)	0.817(0.160)	0.940(0.147)	0.940(0.145)	0.941(0.144)	0.941(0.147)	0.940(0.145)	0.940(0.144)
(200,50)	0.808(0.162)	0.945(0.143)	0.946(0.142)	0.945(0.141)	0.945(0.143)	0.946(0.142)	0.945(0.141)
(200,100)	0.657(0.127)	0.947(0.113)	0.948(0.112)	0.946(0.111)	0.947(0.113)	0.948(0.112)	0.946(0.111)
(200,200)	0.448(0.101)	0.948(0.093)	0.950(0.092)	0.948(0.091)	0.948(0.093)	0.950(0.092)	0.947(0.091)
0.85							
(25,25)	0.807(0.261)	0.893(0.177)	0.900(0.176)	0.904(0.176)	0.894(0.177)	0.901(0.176)	0.905(0.176)
(75,25)	0.758(0.245)	0.913(0.153)	0.916(0.152)	0.916(0.151)	0.913(0.153)	0.916(0.152)	0.916(0.151)
(25,75)	0.431(0.176)	0.930(0.157)	0.932(0.157)	0.934(0.157)	0.930(0.157)	0.933(0.157)	0.935(0.157)
(75,75)	0.274(0.154)	0.942(0.116)	0.945(0.115)	0.944(0.115)	0.942(0.116)	0.945(0.115)	0.944(0.115)
(200,50)	0.380(0.175)	0.947(0.113)	0.946(0.112)	0.942(0.111)	0.947(0.113)	0.946(0.112)	0.943(0.111)
(200,100)	0.085(0.129)	0.950(0.089)	0.947(0.089)	0.944(0.088)	0.950(0.089)	0.947(0.089)	0.944(0.088)
(200,200)	0.005(0.095)	0.946(0.073)	0.947(0.073)	0.948(0.073)	0.946(0.073)	0.947(0.073)	0.948(0.073)
0.90							
(25,25)	0.494(0.258)	0.870(0.137)	0.880(0.137)	0.884(0.138)	0.869(0.137)	0.880(0.138)	0.884(0.138)
(75,25)	0.431(0.249)	0.902(0.119)	0.905(0.119)	0.908(0.119)	0.903(0.119)	0.905(0.119)	0.908(0.119)
(25,75)	0.072(0.161)	0.919(0.125)	0.924(0.126)	0.924(0.126)	0.919(0.125)	0.924(0.126)	0.924(0.126)
(75,75)	0.024(0.150)	0.937(0.092)	0.937(0.092)	0.936(0.092)	0.937(0.092)	0.937(0.092)	0.937(0.092)
(200,50)	0.076(0.177)	0.942(0.089)	0.943(0.089)	0.942(0.089)	0.942(0.089)	0.943(0.089)	0.941(0.089)
(200,100)	0.001(0.128)	0.945(0.071)	0.943(0.071)	0.940(0.071)	0.945(0.071)	0.943(0.071)	0.940(0.071)
(200,200)	0.000(0.093)	0.947(0.059)	0.945(0.059)	0.942(0.059)	0.947(0.059)	0.945(0.059)	0.943(0.059)
0.95							
(25,25)	0.136(0.250)	0.814(0.084)	0.824(0.086)	0.833(0.087)	0.814(0.084)	0.824(0.086)	0.832(0.087)
(75,25)	0.100(0.248)	0.860(0.074)	0.866(0.075)	0.873(0.075)	0.862(0.074)	0.865(0.075)	0.873(0.075)
(25,75)	0.001(0.150)	0.883(0.081)	0.890(0.082)	0.896(0.083)	0.882(0.081)	0.890(0.082)	0.897(0.083)
(75,75)	0.000(0.147)	0.917(0.060)	0.920(0.060)	0.922(0.060)	0.917(0.060)	0.920(0.060)	0.923(0.060)
(200,50)	0.002(0.177)	0.928(0.058)	0.929(0.058)	0.930(0.058)	0.929(0.058)	0.929(0.058)	0.930(0.058)
(200,100)	0.000(0.127)	0.941(0.047)	0.942(0.047)	0.940(0.047)	0.941(0.047)	0.941(0.047)	0.939(0.047)
(200,200)	0.000(0.085)	0.942(0.039)	0.944(0.039)	0.941(0.039)	0.942(0.039)	0.944(0.039)	0.941(0.039)

Table 3: Coverage probabilities and lengths (in parentheses) of the normality-based and EL-based confidence intervals (CI) with a 95% nominal level for various AUC values in the case of four biomarkers with a multivariate lognormal distribution and $\rho = 0$.

AUC	Normality	EL-based CI with Gaussian kernels			EL-based CI with Epanechnikov kernels		
(m,n)	based CI	$\delta = 1/3$	$\delta = 1/5$	$\delta = 1/7$	$\delta = 1/3$	$\delta = 1/5$	$\delta = 1/7$
0.75							
(25,25)	0.957(0.251)	0.87(0.219)	0.875(0.216)	0.881(0.214)	0.870(0.219)	0.877(0.216)	0.882(0.214)
(75,25)	0.920(0.200)	0.916(0.194)	0.92(0.192)	0.924(0.190)	0.916(0.194)	0.920(0.192)	0.925(0.190)
(25,75)	0.937(0.231)	0.912(0.192)	0.917(0.191)	0.923(0.19)	0.912(0.192)	0.919(0.191)	0.922(0.190)
(75,75)	0.855(0.161)	0.93(0.144)	0.934(0.143)	0.935(0.142)	0.931(0.144)	0.934(0.143)	0.934(0.142)
(200,50)	0.799(0.141)	0.939(0.143)	0.943(0.142)	0.943(0.14)	0.940(0.143)	0.942(0.142)	0.942(0.140)
(200,100)	0.686(0.117)	0.941(0.112)	0.945(0.111)	0.947(0.11)	0.941(0.112)	0.945(0.111)	0.948(0.110)
(200,200)	0.567(0.101)	0.939(0.092)	0.941(0.092)	0.946(0.091)	0.939(0.092)	0.941(0.092)	0.945(0.091)
0.85							
(25,25)	0.792(0.282)	0.850(0.162)	0.859(0.162)	0.866(0.162)	0.848(0.163)	0.862(0.163)	0.868(0.163)
(75,25)	0.474(0.183)	0.903(0.147)	0.908(0.146)	0.912(0.146)	0.902(0.147)	0.908(0.146)	0.913(0.146)
(25,75)	0.723(0.262)	0.914(0.150)	0.922(0.150)	0.925(0.150)	0.915(0.150)	0.922(0.150)	0.926(0.150)
(75,75)	0.278(0.156)	0.932(0.112)	0.933(0.112)	0.934(0.112)	0.932(0.112)	0.934(0.112)	0.934(0.112)
(200,50)	0.070(0.117)	0.943(0.111)	0.945(0.110)	0.945(0.110)	0.943(0.111)	0.944(0.110)	0.944(0.110)
(200,100)	0.019(0.103)	0.948(0.088)	0.948(0.087)	0.948(0.087)	0.948(0.088)	0.947(0.087)	0.948(0.087)
(200,200)	0.004(0.095)	0.945(0.072)	0.944(0.072)	0.941(0.072)	0.946(0.072)	0.943(0.072)	0.940(0.072)
0.90							
(25,25)	0.544(0.280)	0.808(0.124)	0.821(0.125)	0.833(0.126)	0.809(0.124)	0.823(0.126)	0.834(0.126)
(75,25)	0.114(0.166)	0.874(0.114)	0.882(0.114)	0.885(0.114)	0.875(0.114)	0.882(0.114)	0.884(0.114)
(25,75)	0.448(0.268)	0.893(0.118)	0.904(0.120)	0.910(0.120)	0.894(0.119)	0.906(0.120)	0.912(0.121)
(75,75)	0.040(0.152)	0.920(0.089)	0.924(0.089)	0.926(0.090)	0.919(0.089)	0.925(0.090)	0.926(0.090)
(200,50)	0.000(0.105)	0.932(0.088)	0.934(0.088)	0.936(0.088)	0.932(0.088)	0.934(0.088)	0.936(0.088)
(200,100)	0.000(0.097)	0.941(0.070)	0.941(0.070)	0.942(0.070)	0.941(0.070)	0.941(0.070)	0.942(0.070)
(200,200)	0.000(0.093)	0.942(0.058)	0.944(0.058)	0.944(0.058)	0.942(0.058)	0.944(0.058)	0.943(0.058)
0.95							
(25,25)	0.161(0.272)	0.744(0.076)	0.764(0.078)	0.779(0.08)	0.748(0.077)	0.765(0.079)	0.781(0.081)
(75,25)	0.002(0.153)	0.846(0.071)	0.851(0.072)	0.859(0.072)	0.846(0.071)	0.852(0.072)	0.858(0.072)
(25,75)	0.112(0.270)	0.861(0.077)	0.874(0.078)	0.883(0.079)	0.861(0.077)	0.875(0.078)	0.884(0.079)
(75,75)	0.000(0.149)	0.905(0.058)	0.910(0.059)	0.912(0.059)	0.905(0.058)	0.910(0.059)	0.912(0.059)
(200,50)	0.000(0.092)	0.925(0.057)	0.926(0.058)	0.926(0.058)	0.926(0.058)	0.926(0.058)	0.927(0.058)
(200,100)	0.000(0.089)	0.935(0.046)	0.935(0.047)	0.933(0.047)	0.935(0.046)	0.935(0.047)	0.934(0.047)
(200,200)	0.000(0.087)	0.942(0.039)	0.941(0.039)	0.939(0.039)	0.943(0.039)	0.940(0.039)	0.939(0.039)

Table 4: Coverage probabilities and lengths (in parentheses) of the normality-based and EL-based confidence intervals (CI) with a 95% nominal level for various AUC values in the case of two biomarkers with marginal Chi-square distributions coupled together via a normal copula with correlation $\rho = 0.7$.

AUC	Normality	EL-based CI with Gaussian kernels			EL-based CI with Epanechnikov kernels		
(m,n)	based CI	$\delta = 1/3$	$\delta = 1/5$	$\delta = 1/7$	$\delta = 1/3$	$\delta = 1/5$	$\delta = 1/7$
0.75							
(25,25)	0.840(0.163)	0.899(0.226)	0.900(0.226)	0.899(0.226)	0.899(0.226)	0.900(0.226)	0.900(0.226)
(75,25)	0.832(0.155)	0.941(0.201)	0.941(0.201)	0.942(0.201)	0.940(0.201)	0.941(0.201)	0.942(0.201)
(25,75)	0.642(0.105)	0.918(0.191)	0.918(0.191)	0.918(0.191)	0.917(0.191)	0.918(0.191)	0.918(0.191)
(75,75)	0.532(0.088)	0.947(0.146)	0.947(0.146)	0.947(0.146)	0.947(0.146)	0.947(0.146)	0.947(0.146)
(200,50)	0.636(0.107)	0.941(0.148)	0.941(0.148)	0.941(0.148)	0.941(0.148)	0.941(0.148)	0.941(0.148)
(200,100)	0.361(0.051)	0.942(0.115)	0.942(0.115)	0.942(0.115)	0.942(0.115)	0.942(0.115)	0.942(0.115)
(200,200)	0.052(0.001)	0.945(0.092)	0.945(0.092)	0.946(0.092)	0.945(0.092)	0.945(0.092)	0.946(0.092)
0.85							
(25,25)	0.840(0.163)	0.853(0.166)	0.854(0.166)	0.855(0.166)	0.853(0.166)	0.855(0.166)	0.855(0.166)
(75,25)	0.832(0.155)	0.931(0.158)	0.931(0.158)	0.931(0.158)	0.931(0.158)	0.931(0.158)	0.932(0.158)
(25,75)	0.642(0.105)	0.860(0.140)	0.861(0.140)	0.860(0.140)	0.860(0.140)	0.861(0.140)	0.861(0.140)
(75,75)	0.532(0.088)	0.936(0.115)	0.935(0.115)	0.935(0.115)	0.935(0.115)	0.935(0.115)	0.936(0.115)
(200,50)	0.636(0.107)	0.949(0.123)	0.950(0.123)	0.950(0.123)	0.950(0.123)	0.950(0.123)	0.950(0.123)
(200,100)	0.361(0.051)	0.951(0.096)	0.951(0.096)	0.951(0.096)	0.951(0.096)	0.951(0.096)	0.951(0.096)
(200,200)	0.052(0.001)	0.951(0.074)	0.951(0.074)	0.951(0.074)	0.951(0.074)	0.951(0.074)	0.951(0.074)
0.90							
(25,25)	0.840(0.163)	0.845(0.133)	0.844(0.132)	0.845(0.132)	0.846(0.133)	0.844(0.133)	0.845(0.132)
(75,25)	0.832(0.155)	0.920(0.123)	0.919(0.123)	0.919(0.123)	0.920(0.123)	0.920(0.123)	0.920(0.123)
(25,75)	0.642(0.105)	0.884(0.118)	0.885(0.118)	0.884(0.118)	0.885(0.118)	0.885(0.118)	0.884(0.118)
(75,75)	0.532(0.088)	0.932(0.092)	0.932(0.092)	0.933(0.092)	0.932(0.092)	0.933(0.092)	0.933(0.092)
(200,50)	0.636(0.107)	0.947(0.094)	0.948(0.094)	0.948(0.094)	0.947(0.094)	0.947(0.094)	0.948(0.094)
(200,100)	0.361(0.051)	0.949(0.073)	0.949(0.073)	0.949(0.073)	0.949(0.073)	0.949(0.073)	0.949(0.073)
(200,200)	0.052(0.001)	0.950(0.059)	0.951(0.059)	0.950(0.059)	0.950(0.059)	0.950(0.059)	0.950(0.059)
0.95							
(25,25)	0.840(0.163)	0.782(0.079)	0.782(0.080)	0.781(0.080)	0.784(0.080)	0.783(0.080)	0.783(0.080)
(75,25)	0.832(0.155)	0.905(0.078)	0.905(0.078)	0.905(0.078)	0.905(0.078)	0.905(0.078)	0.905(0.078)
(25,75)	0.642(0.105)	0.818(0.072)	0.819(0.072)	0.819(0.072)	0.818(0.072)	0.819(0.072)	0.818(0.072)
(75,75)	0.532(0.088)	0.909(0.059)	0.909(0.059)	0.909(0.059)	0.909(0.059)	0.910(0.059)	0.909(0.059)
(200,50)	0.636(0.107)	0.940(0.063)	0.940(0.063)	0.940(0.063)	0.940(0.063)	0.940(0.063)	0.940(0.063)
(200,100)	0.361(0.051)	0.940(0.050)	0.940(0.050)	0.940(0.050)	0.940(0.050)	0.940(0.050)	0.940(0.050)
(200,200)	0.052(0.001)	0.943(0.039)	0.943(0.039)	0.943(0.039)	0.943(0.039)	0.943(0.039)	0.943(0.039)

Furthermore, we constructed a bivariate distribution whose marginals are Chi-square distributions, coupled together via a normal copula with correlation ρ. The marginal distributions of the control group are Chi-square distributions with degrees of freedom 2 and 1, respectively, and they are coupled together via a normal copula with $\rho = 0.4$. For the case group, the degrees of freedom of the marginal Chi-square distribution vary to reach pre-specified values of the AUC, and the scenarios $\rho = 0, 0.4, 0.7$ are considered. Table 4 shows the simulation results for the scenario where $\rho = 0.7$. Tables 10 and 11 of the supplementary material present the results for the scenario where $\rho = 0$ and 0.4, respectively. Similar to the multivariate lognormal scenario, the proposed method works well, whereas the method based on the normal assumptions fails.

The family of multivariate power exponential distribution is also investigated. We considered the case $\kappa = 1$ and $\mathbf{\Sigma_X} = \mathbf{\Sigma_Y} = \mathbf{\Sigma}$ where $\rho = 0, 0.4, 0.7$. Note that the multivariate normal distribution is also included in the family of multivariate power exponential distribution, corresponding to $\kappa = 2$. Detailed simulation results are shown in the Supplementary materials (see tables 12 to 17). The extensive Monte Carlo studies confirm the excellent performance and need for the proposed method over normality based confidence intervals when the data distribution is skewed.

4. Data Example

We illustrate the application of the proposed EL interval for AUC based on a linear combination of biomarkers in the study of the accuracy of biomarkers related to atherosclerotic coronary heart disease. In the pathogenesis of various diseases including atherosclerosis, oxygen radicals play an important role. Different biomarkers of individual oxidative stress and antioxidant status quantify different phases of the oxidative stress and antioxidant status process of an individual.

The study sample consists of 105 individuals with myocardial infarction and 437 controls. The sampling frame for adults between the ages of 35 and 65

was the New York State Department of Motor Vehicles drivers license rolls. Also a randomly selected elderly sample (age $65-79$) from the Health Care Financing Administration database was taken (Schisterman et al., 2001). A number of biomarkers were examined from a 12-hour fasting blood specimen for biochemical analysis at baseline.

4.1. Uric acid, TBARS and HDL cholesterol

To improve diagnostic accuracy, we combine three important biomarkers related to the atherosclerotic coronary heart disease in an efficient manner. The measurements are uric acid, TBARS and HDL cholesterol. The p-value of the Shapiro-Wilk test for multivariate normality for the case group is 2.1×10^{-15} and for the control group is 7.6×10^{-27}, indicating the non-normal character of the data. For these three markers, the estimated means and variances for the diseased and the nondiseased are

$$\hat{\boldsymbol{\mu}}_{\mathbf{X}} = (6.106\ 1.385\ 45.655)^T,\ \hat{\boldsymbol{\mu}}_{\mathbf{Y}} = (5.516\ 1.372\ 52.327)^T,$$

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{X}} = \begin{pmatrix} 2.878 & 0.021 & -8.592 \\ 0.021 & 0.251 & -0.316 \\ -8.592 & -0.316 & 162.957 \end{pmatrix},\ \text{and}\ \hat{\boldsymbol{\Sigma}}_{\mathbf{Y}} = \begin{pmatrix} 1.917 & 0.007 & -6.715 \\ 0.007 & 0.156 & -1.322 \\ -6.715 & -1.322 & 218.930 \end{pmatrix},$$

respectively.

The best linear combination estimates derived by the normality-based method of Su and Liu (1993) $\hat{\mathbf{a}}_N$ is uric acid $-0.416 \times$ TBARS $-0.189 \times$ HDL cholesterol, with a maximum AUC estimate of 0.6467. The normality based 95% confidence interval of Reiser and Faraggi (1997) for the maximum AUC is [0.583, 0.695] with a length of 0.112. Recall that the the Shapiro-Wilk test does not justify the normality assumption. Our method yields the best linear combination estimates of uric acid $+ 0.542 \times$ TBARS $- 0.199 \times$ HDL cholesterol. It gives a maximum AUC estimate of 0.6520, and the smooth EL-based 95% confidence interval for the maximum AUC is [0.594, 0.705] with a length of 0.111. Compared to the normality-based method, the proposed method yields similar numerical results for all practical purposes. However, judging from the BLC estimates, the pro-

posed method yields a more realistic result than the normality based method. It is well-known that low levels of HDL-cholesterol and high levels of TBARS as well as uric acid indicate higher risk of coronary heart disease (Schisterman et al., 2002). Different from the normality based result, the signs of the point BLC estimates by the proposed method confirm the negative correlation between HDL-cholesterol and the risk of the disease as well as the positive correlation between TBARS and uric acid and the risk of the disease.

To explain the numerical results obtained above and to evaluate the obtained confidence interval estimators, we fixed the intervals [0.583, 0.695] and [0.594, 0.705], and bootstrap 1,000 times from the original case and control data sets with the same sample sizes as the original case and control data, calculating the empirical estimator

$$\tilde{\theta}_{n,m} = \underset{a_1,a_2}{argmax} \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} I \left\{ \text{uric acid}_i + a_1 \text{TBARS}_i + a_2 \text{HDL cholesterol}_i \right. \\ \left. > \text{uric acid}_j + a_1 \text{TBARS}_j + a_2 \text{HDL cholesterol}_j \right\},$$

where $n = 105$ and $m = 437$. The proportion of the event that the proposed CI contains $\tilde{\theta}_{105,437}$ was observed to be 0.943, whereas the normality-based Reiser and Faraggi (1997)'s method provided a frequency of 0.890. This proportion is expected to be 95%.

Additionally, in order to evaluate the proposed confidence interval estimation versus Reiser and Faraggi (1997)'s approach, we conduct the following cross-validation type study. In each stage of this study, we randomly splitted the original data sets into two parts. One with fixed samples sizes, consists of 40 observations for the case group and 177 observations for the control group, and the proposed confidence interval estimator and Reiser and Faraggi (1997)'s confidence interval estimator are obtained. The second part of data sets consists of the remaining 65 observations for the case group and 260 observations for the control group. The empirical estimator $\tilde{\theta}_{65,260}$ is calculated based on the second part of data sets. Then we checked the event that $\tilde{\theta}_{65,260}$ belongs to the corresponding confidence intervals. This stage was repeated in a cross-validation

manner 1,000 times. The frequency of $\tilde{\theta}_{65,260}$ belonging to the proposed CI was observed to be 0.945, whereas for Reiser and Faraggi (1997)'s method, we observed the frequency as 0.890. This proportion is expected to be 95%. Thus, despite almost numerical equivalence in the estimation of the AUC and confidence intervals, the problem of the Reiser and Faraggi (1997)'s application may be described by the invalidation of the normal assumptions.

4.2. Blood glucose, vitamin E and body mass index

For one additional simple example, we randomly selected the data with 25 observations in the case group and 100 observations in the control group, and evaluate markers including blood glucose, vitamin E (a fat-soluble antioxidant vitamin) and body mass index (BMI). For these three markers, the estimated means and variances for the diseased and the nondiseased are

$$\hat{\boldsymbol{\mu}}_{\mathbf{X}} = (111.400\ 13.806\ 27.954)^T,\ \hat{\boldsymbol{\mu}}_{\mathbf{Y}} = (109.260\ 13.695\ 27.947)^T,$$

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{X}} = \begin{pmatrix} 680.583 & -7.781 & 4.815 \\ -7.781 & 28.975 & 3.652 \\ 4.815 & 3.652 & 14.771 \end{pmatrix}, \text{ and } \hat{\boldsymbol{\Sigma}}_{\mathbf{Y}} = \begin{pmatrix} 2184.881 & 4.284 & 50.747 \\ 4.284 & 30.417 & 3.469 \\ 50.747 & 3.469 & 24.169 \end{pmatrix},$$

respectively.

The Shapiro-Wilk test for multivariate normality applied to data from the case group and control group rejects normality with p-values of 1.1×10^{-5} and 6.7×10^{-28}, respectively. Based on the normality assumption that is not justified by the Shapiro-Wilk multivariate normality test, the maximum AUC estimate is 0.517, whereas the maximum AUC estimate by the proposed method is 0.659. Though the proposed method is not focused on the point estimator of the maximum AUC, it yields a much higher maximum AUC than Su and Liu (1993)'s estimate. The 95% confidence interval based on the normality assumption (Reiser and Faraggi, 1997) is [0.500, 0.539], and the proposed smooth EL-based 95% confidence interval for the maximum AUC is [0.533, 0.762]. In addition, the best linear combination estimates derived by Su and Liu (1993) is glucose $+ 2.675 \times$ vitamin E $- 1.681 \times$ BMI, whereas the BLC derived by the

proposed method is glucose $+ 0.098 \times$ vitamin E $+ 0.351 \times$ BMI. The sign of BMI in the BLC estimates between the proposed method and Su and Liu (1993)'s method is reversed. It is shown that higher BMI measures are associated with a higher risk of a heart disease (Lamon-Fava et al., 1996), indicating that a negative coefficient associated with BMI goes against scientific findings. Judging from the BLC estimates, the proposed method yields a realistic result that is not provided by the use of the normality-based methods.

Remark 3. *A natural question pertains to the use of all markers to compute the AUC and associated confidence interval. This aspect of the problem is related to biomarker selection and, given the mathematical challenges associated with it constitutes a topic of our future work.*

5. Discussion

In this paper, we have proposed an EL-based confidence interval for the maximum area under the ROC curve based on a linear combination of biomarkers measured on a continuous scale. We show that the proposed method has good theoretical properties regardless of the data distribution. Simulation results indicate the proposed EL-based interval has a similar performance with normality based confidence intervals when normality is present, and outperforms normality based confidence intervals in the case of skewed data.

Future work may focus on approaches to extend the proposed method to categorical biomarkers. Since empirical likelihood regions are Bartlett correctable, additional future work may consider the Bartlett correction to reduce the coverage error of EL confidence intervals.

Supplementary Materials

The supplementary material includes some details considered in the computational strategy, simulation results reporting the coverage probability and the length of the confidence intervals based on the proposed method for a variety of

distributions. Furthermore, the R code for obtaining the EL-based confidence interval for the maximum AUC based on multiple biomarkers is provided.

Acknowledgments

We gratefully acknowledge the co-editor and the two anonymous referees for their very helpful and constructive comments and suggestions, which led to significant improvements to the paper.

Appendix

Appendix A. Conditions

Throughout the paper, we assume the following:

(C1) Let $h = h_{n,m} > 0$ be a bandwidth sequence converging to zero as n and m converge to infinity and $k(\cdot)$ is a compactly supported rth order kernel ($r \geq 2$). That is, for some integer $r \geq 2$ and a finite constant $\xi \neq 0$,

$$\int u^j k(u)du = \begin{cases} 1, & j = 0, \\ 0, & j \in [1, r-1], \\ \xi, & j = r. \end{cases}$$

(C2) The kernel function $k(u)$ is symmetric about $u = 0$ with the first derivative being bounded and continuous.

(C3) For the density function $g_{\mathbf{a}}(x) := \partial G_{\mathbf{a}}/\partial x$, $g_{\mathbf{a}}^{(r-1)}(x)$ exists and is continuous.

(C4) As m and n converge to infinity, $m/(m+n) \to q$, where $0 < q < 1$.

(C5) (i) The true parameter value $\mathbf{a}_0$ is an interior point of Λ, which is a compactor subset of $\mathbf{R}^d$; (ii) The partial derivative of the marginal density of ${\mathbf{a}_0}^T(\mathbf{X}^T, \mathbf{Y}^T)^T$ exists and is bounded.

(C6) $E|\mathbf{X} - \mathbf{Y}|^2 < \infty$ and $(m+n)^{1/2}h^2 \to \infty$ as $m, n \to \infty$.

(C7) There exist $t > 2$ such that $(m+n)^{1/2}h^t \to 0$, $E|\mathbf{X} - \mathbf{Y}|^t < \infty$, the tth derivative $K^{(t)}$ exists and is continuous with the lth derivative $\left|K^{(l)}(u)\right| \leq c$ for all $l \leq t$.

The kernel function given in (C1) is commonly used in nonparametric density estimation. Condition (C2) allows the corresponding scaled symmetric distribution function as an approximation to the indicator function. Condition (C3) requires that the density of the combined scores in the nondiseased group be sufficiently smooth. Condition (C4) states the growth rate of two sample sizes should be of equal order. Condition (C5) is needed for establishing the consistency of $\hat{\mathbf{a}}_K$ and the uniform convergence of $\mathrm{A}^K_{\mathrm{m,n}}(\mathbf{a})$ to its population version. Condition (C6) ensures that $\mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K) - \mathrm{A}^K_{\mathrm{m,n}}(\mathbf{a}_0)$ converges to 0 at the rate of at least $(\mathrm{m}+\mathrm{n})^{-1/2}$. Condition (C7) guarantees that $\mathrm{EA}^K_{\mathrm{m,n}}(\mathbf{a}_0) - A_0$ convergse to 0 at the rate of at least $(\mathrm{m}+\mathrm{n})^{-1/2}$.

Appendix B. Proofs

The proof of Proposition 1 will be based on the following steps: (1) Lemma 1 gives the derivation of $l(A_0)$ with respect to A_0 up to order 3. The first and second derivative show the equivalence between $\mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K)$ and A_0, as well as $\hat{\mathbf{a}}_{\mathbf{0}} = \hat{\mathbf{a}}_K$; (2) we will show $\hat{\mathbf{a}}_K$ converges to $\mathbf{a}_0$ at the rate of at least $(\mathrm{m}+\mathrm{n})^{-1/2}$ in Lemma 2 and $\mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K) - \mathrm{A}^K_{\mathrm{m,n}}(\mathbf{a}_0)$ converges to 0 at the rate of at least $(\mathrm{m}+\mathrm{n})^{-1/2}$ with a proper choice of bandwidths in Lemma 3; (3) the asymptotic distribution of $\mathrm{A}^K_{\mathrm{m,n}}(\mathbf{a}_0) - \mathrm{EA}^K_{\mathrm{m,n}}(\mathbf{a}_0)$ is shown in Lemma 4 based on the theory of U-statistics.

Lemma 1. *The maximum smoothed empirical likelihood estimator (MELE) of the maximum AUC is* $\mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K)$. *At the smoothed MELE* $\mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K)$, *we have* $l(A_0)|_{A_0=\mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K)} = l'(A_0)|_{A_0=\mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K)} = 0$, $l''(A_0)|_{A_0=\mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K)} = 2n\hat{\sigma}^{-2}$, $l'''(A_0)|_{A_0=\tilde{A}} = O(n)$, *where* $\hat{\sigma}^2$ *is stated in the Proposition 1, and* $\tilde{A}$ *is some real number between* $\mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K)$ *and* A_0.

Proof. By the constraints defined in equation (3), the definition of $\hat{\mathbf{a}}_{\mathbf{0}}$, and equation (6), we have

$$l'(A_0) = 2n\frac{\partial\lambda}{\partial A_0}\sum\nolimits_{i=1}^{n} p_i\left(v_i(\hat{\mathbf{a}}_0) - A_0\right) + 2n\lambda\sum\nolimits_{i=1}^{n} p_i\frac{\partial v_i(\hat{\mathbf{a}}_0)}{\partial\hat{\mathbf{a}}_0^T}\frac{\partial\hat{\mathbf{a}}_0}{\partial A_0} - 2n\lambda = -2n\lambda.$$

Since the MELE minimizes $l(A_0)$, we have $\lambda = 0$ and $p_i = n^{-1}$ by equation (4). Therefore, by equation (5), the MELE of the maximum AUC is $\mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K)$ and the corresponding $\hat{\mathbf{a}}_\mathbf{0}$ is equivalent to $\hat{\mathbf{a}}_K$. Therefore, we have

$$l(A_0)|_{A_0=\mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K)} = l'(A_0)|_{A_0=\mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K)} = 0.$$

To derive the second derivative of $l(A_0)$ with respect to A_0 at the MELE $\mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K)$, we take the first derivative of both sides of the equation (5) with respect to A_0. It leads to

$$-n\partial\lambda/\partial A_0 \sum\nolimits_{i=1}^{n} p_i^2(v_i(\hat{\mathbf{a}}_0) - A_0)^2 - n\lambda \sum\nolimits_{i=1}^{n} {p_i}^2 \partial v_i(\hat{\mathbf{a}}_0)/\partial A_0 (v_i(\hat{\mathbf{a}}_\mathbf{0}) - A_0) = 1$$

At the MELE $\mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K)$, we have

$$\partial\lambda/\partial A_0|_{A_0=\mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K)} = -n \Big/ \sum\nolimits_{i=1}^{n} [v_i(\hat{\mathbf{a}}_K) - A_0]^2 = -\hat{\sigma}^{-2}.$$

The second derivative $l''(A_0)|_{A_0=\mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K)} = 2n\hat{\sigma}^{-2} > 0$ confirms that the MELE minimizes $l(A_0)$. Based on the fact $\max\limits_{1\le i\le n} |v_i(\hat{\mathbf{a}}_0)| = O(1)$, we have $l'''(A_0)|_{A_0=\tilde{A}} = -2n\,\partial^2\lambda/\partial A^2\big|_{A_0=\tilde{A}} = O(n)$ for some real number $\tilde{A}$ between $\mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K)$ and A_0. □

Lemma 2. *Suppose (C1)-(C3) and (C5) hold. Then* $\hat{\mathbf{a}}_K - \mathbf{a}_0 = O_p((m+n)^{1/2}\mathbf{1})$, *as* $m, n \to \infty$.

Proof. This lemma follows from Theorem 2 of Ma and Huang (Biometrics, 2007). □

Lemma 3. *Suppose (C1)-(C3) and (C6) hold. Then*

$$\sqrt{\mathrm{m}+\mathrm{n}}\left(\mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K) - \mathrm{A}^K_{\mathrm{m,n}}(\mathbf{a}_0)\right) \to 0, \textit{ as } m, n \to \infty.$$

Proof. The first derivative of $\mathrm{A}^K_{\mathrm{m,n}}(\mathbf{a})$ with respect to $\mathbf{a}$ at $\mathbf{a} = \hat{\mathbf{a}}_K$ is 0 and $K(\cdot)$ has continuous second partial derivatives. Apply the Taylor expansion of $\mathrm{A}^K_{\mathrm{m,n}}(\mathbf{a}_0)$ at $\hat{\mathbf{a}}_K$ to obtain

$$\mathrm{A}^K_{\mathrm{m,n}}(\mathbf{a}_0) = \mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K) + (\mathbf{a}_0 - \hat{\mathbf{a}}_K)^T\, \partial \mathrm{A}^K_{\mathrm{m,n}}(\mathbf{a})/\partial\mathbf{a}\big|_{\mathbf{a}=\hat{\mathbf{a}}_K} + R_{m,n} = \mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K) + R_{m,n},$$

where $R_{m,n} = \frac{1}{2}(\hat{\mathbf{a}}_K - \mathbf{a}_0)^T \left.\frac{\partial^2 \mathrm{A}^K_{\mathrm{m,n}}(\mathbf{a})}{\partial \mathbf{a} \partial \mathbf{a}^T}\right|_{\mathbf{a}=\tilde{\mathbf{a}}} (\hat{\mathbf{a}}_K - \mathbf{a}_0)$ with $\tilde{\mathbf{a}} \in (\mathbf{a}_0, \hat{\mathbf{a}}_K)$. By (C1)-(C3) and (C6), we have

$$\begin{aligned}\left|\left.\frac{\partial^2 \mathrm{A}^K_{\mathrm{m,n}}(\mathbf{a})}{\partial \mathbf{a} \partial \mathbf{a}^T}\right|_{\mathbf{a}=\tilde{\mathbf{a}}}\right| &\leq \frac{1}{nm} \sum\nolimits_{i=1}^{n} \sum\nolimits_{j=1}^{m} \left| k'\left(\frac{\tilde{\mathbf{a}}^T(\mathbf{X}_i - \mathbf{Y}_j)}{h}\right)\right| \left|\frac{\mathbf{X}_i - \mathbf{Y}_j}{h}\right|^2 \\ &\leq \frac{c}{nm} \sum\nolimits_{i=1}^{n} \sum\nolimits_{j=1}^{m} \left|\frac{\mathbf{X}_i - \mathbf{Y}_j}{h}\right|^2.\end{aligned}$$

In addition, Lemma 1 states that $\hat{\mathbf{a}}_K - \mathbf{a}_0 = O_p((\mathrm{m}+\mathrm{n})^{-1/2}\mathbf{1})$ and thus $\sqrt{\mathrm{m}+\mathrm{n}} R_{m,n} = O_p((m+n)^{-1/2}h^{-2})$. Lemma 3 follows when $(m+n)^{1/2}h^2 \to \infty$ as $m, n \to \infty$. □

Lemma 4. *Suppose that (C4) holds, then*

$$\sqrt{\frac{mn}{m+n}} \frac{\mathrm{A}^K_{\mathrm{m,n}}(\mathbf{a}_0) - \mathrm{EA}^K_{\mathrm{m,n}}(\mathbf{a}_0)}{\mathrm{s}} \xrightarrow{d} N(0,1),$$

where s is defined in Proposition 1.

Proof. Let $\psi(x,y) = K_h(x-y)$. Based on the theory of U statistics (Serfling, 2009), we have

$$\frac{\sqrt{m+n}(\mathrm{A}^K_{\mathrm{m,n}}(\mathbf{a}_0) - \mathrm{EA}^K_{\mathrm{m,n}}(\mathbf{a}_0))}{(\sigma_{10}^2/(1-q) + \sigma_{01}^2/q)^{1/2}} \xrightarrow{d} N(0,1),$$

and therefore Lemma 4 holds. □

Proof of Proposition 1. Apply Taylor expansion of $l(A_0)$ at $A_0 = \mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K)$. Based on the first and the second derivative of $l(A_0)$ as shown in Lemma 1, we have

$$\begin{aligned}l(A_0) &= l(\mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K)) + (A_0 - \mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K)) l'(\mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K)) \\ &+ (A_0 - \mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K))^2 l''(\mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K))/2 + r_{m,n} \\ &= (\mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K) - A_0)^2 n\hat{\sigma}^{-2} + r_{m,n},\end{aligned}$$

where $\hat{\sigma}^2$ is defined in Proposition 1, and $r_{m,n} = O_p((A_0 - \mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K))^3 l'''(A_0)|_{A_0=\tilde{A}})$ for some real number $\tilde{A}$ between $\mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K)$ and A_0.

To derive the asymptotic distribution of $\mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K) - A_0$, we use the following decomposition,

$$\mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K) - A_0 = (\mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K) - \mathrm{A}^K_{\mathrm{m,n}}(\mathbf{a}_0)) + (\mathrm{A}^K_{\mathrm{m,n}}(\mathbf{a}_0) - \mathrm{EA}^K_{\mathrm{m,n}}(\mathbf{a}_0)) + (\mathrm{EA}^K_{\mathrm{m,n}}(\mathbf{a}_0) - A_0).$$

Apply Taylor theorem of $K_h(u - Y(\mathbf{a}_0))$ around $h = 0$ to the tth order. Under assumption (C7), we have

$$|K_h(u - Y(\mathbf{a}_0)) - 1| \, I\{u - Y(\mathbf{a}_0) > 0\} = h^t \left| K^{(t)}_{\theta_1 h}(u - Y(\mathbf{a}_0)) \right| I\{u - Y(\mathbf{a}_0) > 0\} \leq h^t c,$$

$$|K_h(u - Y(\mathbf{a}_0))| \, I\{u - Y(\mathbf{a}_0) \leq 0\} = h^t \left| K^{(t)}_{\theta_2 h}(u - Y(\mathbf{a}_0)) \right| I\{u - Y(\mathbf{a}_0) \leq 0\} \leq h^t c,$$

and

$$\begin{aligned} &|K_h(u - Y(\mathbf{a}_0)) - I\{Y(\mathbf{a}_0) < u\}| \\ &= |K_h(u - Y(\mathbf{a}_0)) - 1| \, I\{Y(\mathbf{a}_0) < u\} + |K_h(u - Y(\mathbf{a}_0))| I\{Y(\mathbf{a}_0) \geq u\} \leq 2h^t c, \end{aligned}$$

where $0 < \theta_1, \ \theta_2 < 1$, and thus

$$\begin{aligned} &\left| A_0 - \mathrm{A}^K_{\mathrm{m,n}}(\mathbf{a}_0) \right| = \left| \int_{-\infty}^{\infty} G_{\mathbf{a}_0} dF_{\mathbf{a}_0} - \int_{-\infty}^{\infty} G^K_{\mathbf{a}_0} dF^n_{\mathbf{a}_0} \right| \\ &= \left| \int_{-\infty}^{\infty} G_{\mathbf{a}_0} d\left(F_{\mathbf{a}_0} - F^n_{\mathbf{a}_0}\right) + \int_{-\infty}^{\infty} \left(G_{\mathbf{a}_0} - G^m_{\mathbf{a}_0}\right) dF^n_{\mathbf{a}_0} + \int_{-\infty}^{\infty} \left(G^m_{\mathbf{a}_0} - G^K_{\mathbf{a}_0}\right) dF^n_{\mathbf{a}_0} \right| \\ &\leq \sup_{-\infty < u < \infty} \left| F_{\mathbf{a}_0}(u) - F^n_{\mathbf{a}_0}(u) \right| + \sup_{-\infty < u < \infty} \left| G_{\mathbf{a}_0}(u) - G^m_{\mathbf{a}_0}(u) \right| + 2h^t c, \end{aligned}$$

where $F_{\mathbf{a}_0}$ and $G_{\mathbf{a}_0}$ is the distribution function of $X(\mathbf{a}_0)$ and $Y(\mathbf{a}_0)$, $F^n_{\mathbf{a}_0}$ and $G^m_{\mathbf{a}_0}$ is the empirical distribution function of $X(\mathbf{a}_0)$ and $Y(\mathbf{a}_0)$, and $G^K_{\mathbf{a}_0}(u) = \int_{-\infty}^{\infty} \int_{-\infty}^{(u-y)/h} k(w) dw dG^m_{\mathbf{a}_0}(y)$. Since

$$\sqrt{m+n} \left| A_0 - E\mathrm{A}^K_{\mathrm{m,n}}(\mathbf{a}_0) \right| \leq \sqrt{m+n} E \left| A_0 - \mathrm{A}^K_{\mathrm{m,n}}(\mathbf{a}_0) \right|,$$

it follows that when $\sqrt{m+n} h^t \to 0$, we have $\sqrt{m+n} \left| E\mathrm{A}^K_{\mathrm{m,n}}(\mathbf{a}_0) - A_0 \right| \to 0$.

In addition, by Lemma 1-Lemma 3, it follows that $\sqrt{\mathrm{m+n}} \left(\mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K) - A_0 \right)$ and $\sqrt{\mathrm{m+n}} \left(\mathrm{A}^K_{\mathrm{m,n}}(\mathbf{a}_0) - \mathrm{EA}^K_{\mathrm{m,n}}(\mathbf{a}_0) \right)$ shares the same distribution. Thus by Lemma 4, we have

$$\sqrt{\frac{mn}{m+n}} \frac{\mathrm{A}^K_{\mathrm{m,n}}(\mathbf{a}_0) - \mathrm{EA}^K_{\mathrm{m,n}}(\mathbf{a}_0)}{\mathrm{s}} \xrightarrow{d} N(0,1).$$

Since $|A_0 - \mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K)| = O_p(n^{-1/2})$ and $l'''(A_0)|_{A_0 = \tilde{A}} = O(n)$ shown in Lemma 1, we have $r_{m,n} = O_p(n^{-1/2})$ and hence,

$$\gamma(A_0) l(A_0) \doteq \left(\sqrt{\frac{mn}{m+n}} \frac{\mathrm{A}^K_{\mathrm{m,n}}(\hat{\mathbf{a}}_K) - A_0}{s} \right)^2.$$

Therefore, $\gamma(A_0) l(A_0)$ has the χ^2_1 distribution asymptotically. The proof of Proposition 1 is thus completed. □

Appendix C. Discussion on the Remark

For the Gaussian kernel, the first derivative is $k'(u) = -\frac{u}{\sqrt{2\pi}} \exp\left(-u^2/2\right)$. To show the boundedness for $k'(u)$, we calculate the first and the second derivative of $k'(u)$, obtaining

$$k''(u) = (u^2 - 1) \exp\left(-u^2/2\right) / \sqrt{2\pi},$$

and

$$k'''(u) = (3u - u^3) \exp\left(-u^2/2\right) / \sqrt{2\pi}.$$

Set $k''(u)$ to be 0. At $u = -1$, $k'''(u) < 0$, therefore k' reaches to the maximum of $\exp(-1/2)/\sqrt{2\pi}$. At $u = 1$, $k'''(u) > 0$, k' reaches the minimum of $-\exp(-1/2)/\sqrt{2\pi}$. Similarly, we can show the lth derivative $K^{(l)}(u)$ is bounded for all $l \geq 2$. If there exist $t > 2$ such that $t = sup_l\{l : E|\mathbf{X} - \mathbf{Y}|^l < \infty\}$, then Proposition 1 is satisfied with proper selections of bandwidths where $\delta \in (1/2t, 1/4)$.

References

Azzalini, A., 1981. A note on the estimation of a distribution function and quantiles by a kernel method. Biometrika 68, 326–328.

Bamber, D., 1975. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. Journal of Mathematical Psychology 12, 387–415.

Chen, S.X., Hall, P., 1993. Smoothed empirical likelihood confidence intervals for quantiles. The Annals of Statistics 21, 1166–1181.

Faraggi, D., Reiser, B., 2002. Estimation of the area under the ROC curve. Statistics in Medicine 21, 3093–3106.

Hall, P., La Scala, B., 1990. Methodology and algorithms of empirical likelihood. International Statistical Review/Revue Internationale de Statistique 58, 109–127.

Johnson, M.E., 2013. Multivariate Statistical Simulation: A Guide to Selecting and Generating Continuous Multivariate Distributions. John Wiley & Sons.

Lamon-Fava, S., Wilson, P.W., Schaefer, E.J., 1996. Impact of body mass index on coronary heart disease risk factors in men and women the Framingham offspring study. Arteriosclerosis, Thrombosis, and Vascular Biology 16, 1509–1515.

Lazar, N., Mykland, P.A., 1998. An evaluation of the power and conditionality properties of empirical likelihood. Biometrika 85, 523–534.

Limpert, E., Stahel, W.A., Abbt, M., 2001. Log-normal distributions across the sciences: Keys and clues on the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability–normal or log-normal: That is the question. BioScience 51, 341–352.

Lloyd, C.J., 1998. Using smoothed receiver operating characteristic curves to summarize and compare diagnostic systems. Journal of the American Statistical Association 93, 1356–1364.

Lloyd, C.J., Yong, Z., 1999. Kernel estimators of the ROC curve are better than empirical. Statistics and Probability Letters 44, 221–228.

Ma, S., Huang, J., 2005. Regularized ROC method for disease classification and biomarker selection with microarray data. Bioinformatics 21, 4356–4362.

Ma, S., Huang, J., 2007. Combining multiple markers for classification using ROC. Biometrics 63, 751–757.

Markatou, M., Basu, A., Lindsay, B.G., 1998. Weighted likelihood equations with bootstrap root search. Journal of the American Statistical Association 93, 740–750.

McIntosh, M.W., Pepe, M.S., 2002. Combining several screening tests: optimality of the risk score. Biometrics 58, 657–664.

Nadaraya, E.A., 1964. Some new estimates for distribution functions. Theory of Probability & Its Applications 9, 497–500.

Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. The Computer Journal 7, 308–313.

Owen, A.B., 1990. Empirical likelihood ratio confidence regions. The Annals of Statistics 18, 90–120.

Owen, A.B., 2001. Empirical Likelihood. CRC press.

Pepe, M.S., 1997. A regression modelling framework for receiver operating characteristic curves in medical diagnostic testing. Biometrika 84, 595–608.

Pepe, M.S., Cai, T., 2004. The analysis of placement values for evaluating discriminatory measures. Biometrics 60, 528–535.

Pepe, M.S., Cai, T., Longton, G., 2006. Combining predictors for classification using the area under the receiver operating characteristic curve. Biometrics 62, 221–229.

Pepe, M.S., Thompson, M.L., 2000. Combining diagnostic test results to increase accuracy. Biostatistics 1, 123–140.

Qin, G., Zhou, X.H., 2006. Empirical likelihood inference for the area under the ROC curve. Biometrics 62, 613–622.

R Development Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: `http://www.R-project.org/`. ISBN 3-900051-07-0.

Reiser, B., Faraggi, D., 1997. Confidence intervals for the generalized ROC criterion. Biometrics 53, 644–652.

Schisterman, E.F., Faraggi, D., Browne, R., Freudenheim, J., Dorn, J., Muti, P., Armstrong, D., Reiser, B., Trevisan, M., 2001. Tbars and cardiovascular disease in a population-based sample. Journal of Cardiovascular Risk 8, 219–225.

Schisterman, E.F., Faraggi, D., Browne, R., Freudenheim, J., Dorn, J., Muti, P., Armstrong, D., Reiser, B., Trevisan, M., 2002. Minimal and best linear combination of oxidative stress and antioxidant biomarkers to discriminate cardiovascular disease. Nutrition, Metabolism, and Cardiovascular Diseases: NMCD 12, 259–266.

Serfling, R.J., 2009. Approximation Theorems of Mathematical Statistics. Volume 162. John Wiley & Sons.

Shanno, D.F., 1970. Conditioning of quasi-newton methods for function minimization. Mathematics of Computation 24, 647–656.

Su, J.Q., Liu, J.S., 1993. Linear combinations of multiple diagnostic markers. Journal of the American Statistical Association 88, 1350–1355.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B 58, 267–288.

Vexler, A., Liu, A., Schisterman, E.F., Wu, C., 2006. Note on distribution-free estimation of maximum linear separation of two multivariate distributions. Nonparametric Statistics 18, 145–158.

Zou, K.H., Hall, W., Shapiro, D.E., 1997. Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests. Statistics in Medicine 16, 2143–2156.