UW Biostatistics Working Paper Series

3-30-2005

Semi-Parametric Maximum Likelihood Estimates for ROC Curves of Continuous-Scale Tests

Xiao-Hua Zhou
University of Washington, azhou@u.washington.edu

Hua Zhen Lin

Semi-parametric maximum likelihood estimates

for ROC curves of continuous-scale tests

Xiao Hua Zhou [†] [*] **and Hua Zhen Lin** [*‡]

SUMMARY

In this paper, we propose a semi-parametric maximum likelihood estimate of an ROC curve that satisfies the property of invariance of the ROC curve. In our simulation studies, we demonstrate that the proposed estimator has the best performance among all the existing semi-parametric estimators considered here. Finally, we illustrate the application of the proposed estimator using a real data set.

KEY WORDS: ROC curves; Sensitivity and specificity; Semi-parametric maximum likelihood estimators.

1 Introduction

When the response of a diagnostic test is continuous, its diagnostic accuracy is best represented by the receiver operating characteristic (ROC) curve (Pepe, 2003; Zhou et al., 2002). Let F_1 and F_0 denote distribution functions of the test result Y_1 for a diseased subject and the test result Y_0 for a non-diseased subject, respectively. Then, the ROC curve of the test can be written as

$$ROC(u) = 1 - F_1(F_0^{-1}(1-u)), \tag{1}$$

where F_0^{-1} is the inverse function of F_0, and u is the false positive rate (FPR) corresponding to a cut-off point for positivity. It is well-known that the ROC curve of a test must be invariant to any monotone increasing transformation of test results, a fundamental property of an ROC curve. Hence, any sensible estimation methods should have this property. In the statistical literature,

[†]HSR&D Center of Excellence, VA Puget Sound Health Care System, Seattle, WA 98108
[*]Department of Biostatistics, University of Washington, Seattle, WA 98195
[‡]School of Mathematics, Sichuan University, Chengdu, Sichuan 610064, P. R. China

many parametric, semi-parametric, and non-parametric methods have been proposed for estimating an ROC curve. In general, pure parametric methods do not possess the invariance property; the empirical non-parametric and smoothing non-parametric methods have the property of invariance (Hsieh and Turnbull, 1996; Peng and Zhou, 2004). However, the jagged form of the empirical ROC curve estimator can result in underestimating the true ROC curve as the true ROC curve is a smooth function, and the intensive computation and challenging bandwidth selection of the smoothing non-parametric estimators may effect their application in practice.

An intermediate strategy between pure parametric and non-parametric methods is a semi-parametric approach. The most commonly used semi-parametric method is to assume a parametric form for the ROC curve, but avoid making any additional parametric assumptions about the distributions of test results. This type of semi-parametric methods has the property of invariance. In this paper, we focus on this type of semi-parametric methods.

We assume that the ROC curve has the parametric form,

$$ROC(u) = G(\alpha_0 + \alpha_1 H^{-1}(u)), \tag{2}$$

where G and H are some known cumulative distribution functions. The most common choice for G and H is the binormal form, $G = H = \Phi$, where Φ is the cumulative distribution function of the standard normal random variable.

Under the binormal model, several methods have been proposed by Metz et al. (1998), Alonzo and Pepe (2002), Pepe and Cai (2004), Zou et al. (2000), and Cai and Moskowitz (2004), respectively. The first approach, proposed by Metz et al. (1998) and denoted by MHS, is to first categorize continuous test data into ordinal-scale categorical data and then to apply the maximum likelihood method to estimate the parameters in the binormal model by assuming the ordinal-scale data follow a multinomial distribution. The second approach, proposed by Alonzo and Pepe (2002) and denoted by AP, is to estimate the ROC curve by using procedures for fitting generalized linear models to binary data. The third approach, proposed by Pepe and Cai (2004) and denoted by PC, is to first write the ROC curve as the distribution of placement values and then to estimate the ROC curve by maximizing the pseudo likelihood function of the estimated placement values. The fourth method, proposed by Zou et al. (2000) and denoted by ZH, is to use rank data to estimate the ROC curve

by assuming semi-parametric distributions for test results of diseased and non-diseased subjects. One limitation of these methods is that none are a truly maximum likelihood (ML) estimator, and hence they do not possess the optimal property associated with ML estimators. Recently, Cai and Moskowitz (2004), denoted by CM, have proposed a maximum profile likelihood approach to estimate the ROC curve; however, their computation algorithm requires an input of initial values of a large number of nuisance parameters, which may be difficult in practice when the sample size is large.

In this paper, we propose a new profile likelihood approach to estimate the ROC curve. Our method has a smaller number of nuisance parameters to estimate and hence may be more efficient than the Cai and Moskowitz method. Furthermore, our estimator can be computed by using an algorithm that is based on a recursive relationship among the nuisance parameters, without specifying initial values for a large number of nuisance parameters. Our MLE is asymptotically normal, and our extensive simulation studies show the proposed method is more efficient, more robust, and simpler to compute than the existing estimators.

Since the binormal model is the most commonly used form for an ROC curve, from now on we assume that the true ROC curve is defined by

$$ROC(u) = \Phi(\alpha_0 + \alpha_1 \Phi^{-1}(u)), \tag{3}$$

where $\Phi^{-1}(\cdot)$ is the inverse of the cumulative distribution of the standard normal distribution.

Equivalently, we can derive model (3) by assuming there exists an unknown monotone increasing function $g(.)$ such that $g(Y_0)$ has the standard normal distribution and $g(Y_1)$ has a normal distribution with mean μ and standard deviation σ. The resulting ROC curve satisfies model (3) with $\alpha_0 = \mu/\sigma$ and $\alpha_1 = 1/\sigma$.

This paper is organized as follows. In Section 2, we propose a semi-parametric maximum likelihood estimator of an ROC curve and an algorithm for computing it. In Section 3, we develop asymptotic properties for the resulting estimator. In Section 4, we perform simulation studies to asses efficiency and robustness of our estimator relative to the existing semi-parametric estimators and to verify the validity of the asymptotic inferences in finite samples. In Section 5, we illustrate the application of our method in a real example.

2 Semi-parametric maximum likelihood estimate

Data available for making inferences consist of a random sample of size n_1 from the diseased population with the unknown cumulative distribution function F_1, $\{Y_{1j}, j = 1, \cdots, n_1\}$, and a random sample of size n_0 from the non-diseased population, $\{Y_{0i}, i = 1, \cdots, n_0\}$, with the unknown cumulative distribution function F_0. Denote $n = n_0 + n_1$.

Let f_0 and f_1 be the density functions of F_0 and F_1, respectively. Then, the likelihood function of observations $Y_{0i}, i = 1, \cdots, n_0$, and $Y_{1j}, j = 1, \cdots, n_1$, is given by

$$L = \prod_{i=1}^{n_0} f_0(Y_{0i}) \prod_{j=1}^{n_1} f_1(Y_{1j}). \tag{4}$$

Under model (3), we know that there exists a unknown monotone increasing function $g(.)$ such that $g(Y_0)$ has the standard normal distribution and $g(Y_1)$ has a normal distribution with mean μ and standard deviation σ. Therefore, we have that $f_0(y) = \phi(g(y))g'(y)$ and $f_1(y) = \phi(-\alpha_0 + \alpha_1 g(y))\alpha_1 g'(y)$. Hence we can write the likelihood function (4) as

$$L = \prod_{i=1}^{n_0} \phi(g(Y_{0i}))g'(Y_{0i}) \prod_{j=1}^{n_1} \phi(-\alpha_0 + \alpha_1 g(Y_{1j}))\alpha_1 g'(Y_{1j}), \tag{5}$$

where $\phi(x)$ is the standard normal density function. Consequently, the ML estimation of *ROC* curve parameters α_0 and α_1 requires simultaneous estimation of the unknown function g. In what follows we write $\triangle g(x)$ as the jump of $g(\cdot)$ at x if g is discrete at x and the derivative of $g(\cdot)$ at x if $g(\cdot)$ is continuous at x. We seek to maximize the function L_n given by

$$L_n = \prod_{i=1}^{n_0} \phi(g(Y_{0i}))\triangle g(Y_{0i}) \prod_{j=1}^{n_1} \phi(-\alpha_0 + \alpha_1 g(Y_{1j}))\alpha_1 \triangle g(Y_{1j}). \tag{6}$$

Using a standard argument in the nonparametric maximum likelihood estimation (Murphy and Van der Vaart, 2000), we can restrict the MLE, $\widehat{g}$, of g to be the maximiser of the likelihood function L_n over all discrete functions g and show that the MLE, $\widehat{g}$, has to be a discrete function that only jumps at observations $Y_{0i}, i = 1, \cdots, n_0$, and $Y_{1j}, j = 1, \cdots, n_1$. Denote the distinct ordered test results from the combined sample, Y_{0i}'s and Y_{1j}'s, by $Y^*_{(1)} < \cdots < Y^*_{(I^*_n)}$, where I^*_n is the number of distinct values among Y_{0i}'s and Y_{1j}'s. Then MLE $\widehat{g}$ of g jumps only at $Y^*_{(1)} < \cdots < Y^*_{(I^*_n)}$, and we can write the likelihood function (6) as follows:

$$L_n = \prod_{r=1}^{I^*_n} \left(\phi(g(Y^*_{(r)}))\triangle g(Y^*_{(r)})\right)^{k^*_r} \left(\phi(-\alpha_0 + \alpha_1 g(Y^*_{(r)}))\alpha_1 \triangle g(Y^*_{(r)})\right)^{\ell^*_r}, \tag{7}$$

where frequency counts $k_r^* = \#\{Y_{0i} = Y_{(r)}^*, i = 1, \cdots, n_0\}$ and $\ell_r^* = \#\{Y_{1j} = Y_{(r)}^*, j = 1, \cdots, n_1\}$, corresponding to non-diseased and diseased subjects at distinct ordered test results.

When $g = \widehat{g}$, g jumps only at $Y_{(1)}^* < \cdots < Y_{(I_n^*)}^*$, for $r = 1, \ldots, I_n^*$, we have $\triangle\Phi(g(Y_{(r)}^*)) = \Phi(g(Y_{(r)}^*)) - \Phi(g(Y_{(r-1)}^*))$ and $\triangle\Phi(-\alpha_0 + \alpha_1 g(Y_{(r)}^*)) = \Phi(-\alpha_0 + \alpha_1 g(Y_{(r)}^*)) - \Phi(-\alpha_0 + \alpha_1 g(Y_{(r-1)}^*))$, where $g(Y_{(0)}^*) = -\infty$. In addition, $\triangle\Phi(g(Y_{(r)}^*)) = \phi(g(Y_{(r)}^*))\triangle g(Y_{(r)}^*)$ and $\triangle\Phi(-\alpha_0 + \alpha_1 g(Y_{(r)}^*)) = \phi(-\alpha_0 + \alpha_1 g(Y_{(r)}^*))\alpha_1 \triangle g(Y_{(r)}^*)$. Hence, with $C_r^* = g(Y_{(r)}^*)$, $C_0^* = -\infty$, and $C_{I_n^*}^* = +\infty$, we can write (7) as

$$L_n = \prod_{r=1}^{I_n^*} \left(\Phi(C_r^*) - \Phi(C_{r-1}^*)\right)^{k_r^*} \left(\Phi(-\alpha_0 + \alpha_1 C_r^*) - \Phi(-\alpha_0 + \alpha_1 C_{r-1}^*)\right)^{\ell_r^*} \tag{8}$$

when $g = \widehat{g}$. Therefore ML estimation of ROC curve parameters α_0 and α_1, which are of primary interest, requires simultaneous estimation of the $I_n^* - 1$ number of nuisance parameters, $C_1^*, \cdots, C_{I_n^*-1}^*$.

Using the same idea as in Metz (1998), we note that some of the jump points of $\widehat{g}$, $Y_{(r)}^*$'s, can be ignored for estimating α_0 and α_1, which means we can obtain ML estimates of α_0 and α_1 with fewer nuisance parameters. We state the results in Conclusion 1 below.

Denote

$$D(Y_{(r)}^*) = \begin{cases} 2 & \text{if} \quad \#\{Y_{0i} = Y_{(r)}^*, i = 1, \cdots, n_0\} > 0 \quad \text{and} \quad \#\{Y_{1j} = Y_{(r)}^*, j = 1, \cdots, n_1\} > 0 \\ 1 & \text{if} \quad \#\{Y_{0i} = Y_{(r)}^*, i = 1, \cdots, n_0\} > 0 \quad \text{and} \quad \#\{Y_{1j} = Y_{(r)}^*, j = 1, \cdots, n_1\} = 0 \\ 0 & \text{if} \quad \#\{Y_{0i} = Y_{(r)}^*, i = 1, \cdots, n_0\} = 0 \quad \text{and} \quad \#\{Y_{1j} = Y_{(r)}^*, j = 1, \cdots, n_1\} > 0 \end{cases}$$

and

$$\Re = \left\{ Y_{(r)}^* : D(Y_{(r)}^*) = D(Y_{(r+1)}^*) \leq 1, 1 \leq r \leq I_n^* - 2 \right\}.$$

Each jump point in $\Re$ has the same disease status as its next contiguous jump point. Here, $\Re$ includes all jump points of a contiguous sequence with the same disease status except the last point in the sequence.

Conclusion 1. The maximum likelihood estimates of α_0 and α_1 can be determined by some estimating equations that don't depend on those nuisance parameters $C_r^* = g(Y_{(r)}^*)$'s whose $Y_{(r)}^*$ belongs to $\Re$.

See Appendix A.1 for a proof of Conclusion 1. A practical consequence of the conclusion is that we can ignore the jump points in $\Re$ for estimating α_0 and α_1.

After deleting the points in $\Re$, we denote the remaining jump points of $\widehat{g}$ by $Y_{(1)} < \cdots < Y_{(I_n-1)}$ and let $C_r = g(Y_{(r)})$ for $1 \le r \le I_n - 1$, $C_0 = -\infty$ and $C_{I_n} = +\infty$. The MLE of $\theta = (\alpha_0, \alpha_1)^T$ and $\mathbf{C} = (C_1, \cdots, C_{I_n-1})^T$ can be obtained by maximizing

$$\mathcal{L}_n(\theta, \mathbf{C}) = \prod_{r=1}^{I_n} \left(\Phi(C_r) - \Phi(C_{r-1})\right)^{k_r} \left(\Phi(-\alpha_0 + \alpha_1 C_r) - \Phi(-\alpha_0 + \alpha_1 C_{r-1})\right)^{\ell_r}, \tag{9}$$

which is essentially (8) with I_n^* replaced by I_n. Here, for $2 \le r \le I_n - 1$, $k_r = \#\{Y_{(r-1)} < Y_{0i} \le Y_{(r)}, i = 1, \cdots, n_0\}$ and $\ell_r = \#\{Y_{(r-1)} < Y_{1j} \le Y_{(r)}, j = 1, \cdots, n_1\}$; $k_1 = \#\{Y_{0i} \le Y_{(1)}, i = 1, \cdots, n_0\}$, $\ell_1 = \#\{Y_{1j} \le Y_{(1)}, j = 1, \cdots, n_1\}$, $k_{I_n} = \#\{Y_{0i} > Y_{(I_n-1)}, i = 1, \cdots, n_0\}$, and $\ell_{I_n} = \#\{Y_{1j} > Y_{(I_n-1)}, j = 1, \cdots, n_1\}$.

It should be noted that the function (9) is the same as the likelihood proposed by Metz et al (1998). However, Metz et al. (1998) have obtained the likelihood from a parametric viewpoint by assuming that test results could be partitioned into a contingency table with a fixed number of categories and that the resulting contingency table follows a multinomial distribution. In fact, the number of categories, I_n, increases with the sample size, and as a result, the assumed standard multinomial distribution does not hold. Therefore, although Metz et al. (1998) have derived a correct likelihood function form, their justification is not right. Furthermore, to maximize the likelihood function (9) with respect to θ and $\mathbf{C}$, Metz et al. (1998) have used the standard Newton-Raphson iterative method that requires inversion of an $(I_n + 1) \times (I_n + 1)$ matrix; this computation can become a problem if I_n is large. Due to this concern, they have proposed an alternative algorithm by reducing the number of jump points, $I_n - 1$, in an ad hoc way. Hence, their resulting estimate of θ is no longer a ML estimate, and our simulation studies show the ad hoc computation method can lead to some loss of efficiency.

We propose a new algorithm for finding ML estimates based on the function (9). We propose a two-stage iterative procedure for estimating θ and $\mathbf{C}$, alternating the parametric and nonparametric estimation steps. Our idea is from Kvam and Samaniego (1994) on the nonparametric estimation.

Given θ, we can find the MLE of $\mathbf{C}$ by maximizing the likelihood function (9) with respect to

$\mathbf{C}$. The MLE of $\mathbf{C}$ must satisfy the following $(I_n - 1)$ score equations:

$$\begin{aligned}
\frac{\partial \log\{\mathcal{L}_n(\theta, \mathbf{C})\}}{\partial C_1} &= k_1 \frac{\phi(C_1)}{\Phi(C_1)} - k_2 \frac{\phi(C_1)}{\Phi(C_2) - \Phi(C_1)} \\
&\quad + \alpha_1 \ell_1 \frac{\phi(-\alpha_0 + \alpha_1 C_1)}{\Phi(-\alpha_0 + \alpha_1 C_1)} - \alpha_1 \ell_2 \frac{\phi(-\alpha_0 + \alpha_1 C_1)}{\Phi(-\alpha_0 + \alpha_1 C_2) - \Phi(-\alpha_0 + \alpha_1 C_1)} = 0, \\
\frac{\partial \log\{\mathcal{L}_n(\theta, \mathbf{C})\}}{\partial C_r} &= k_r \frac{\phi(C_r)}{\Phi(C_r) - \Phi(C_{r-1})} - k_{r+1} \frac{\phi(C_r)}{\Phi(C_{r+1}) - \Phi(C_r)} \\
&\quad + \alpha_1 \ell_r \frac{\phi(-\alpha_0 + \alpha_1 C_r)}{\Phi(-\alpha_0 + \alpha_1 C_r) - \Phi(-\alpha_0 + \alpha_1 C_{r-1})} \\
&\quad - \alpha_1 \ell_{r+1} \frac{\phi(-\alpha_0 + \alpha_1 C_r)}{\Phi(-\alpha_0 + \alpha_1 C_{r+1}) - \Phi(-\alpha_0 + \alpha_1 C_r)} = 0, \quad 2 \le r \le I_n - 2, \\
\frac{\partial \log\{\mathcal{L}_n(\theta, \mathbf{C})\}}{\partial C_{I_n - 1}} &= k_{I_n - 1} \frac{\phi(C_{I_n - 1})}{\Phi(C_{I_n - 1}) - \Phi(C_{I_n - 2})} - k_{I_n} \frac{\phi(C_{I_n - 1})}{1 - \Phi(C_{I_n - 1})} \\
&\quad + \alpha_1 \ell_{I_n - 1} \frac{\phi(-\alpha_0 + \alpha_1 C_{I_n - 1})}{\Phi(-\alpha_0 + \alpha_1 C_{I_n - 1}) - \Phi(-\alpha_0 + \alpha_1 C_{I_n - 2})} \\
&\quad - \alpha_1 \ell_{I_n} \frac{\phi(-\alpha_0 + \alpha_1 C_{I_n - 1})}{1 - \Phi(-\alpha_0 + \alpha_1 C_{I_n - 1})} = 0. \qquad (10)
\end{aligned}$$

Given θ, the existence and uniqueness of the MLE of $\mathbf{C}$ are established in the following result, and a proof is given in the Appendix A.2.

Conclusion 2. Given θ, the $\widehat{\mathbf{C}}$ that satisfies the score equations in (10) is unique.

Inspection of (10) shows that finding the MLE of $\mathbf{C}$ in a closed form is a challenge. Hence, an iterative algorithm is required. However, the standard Newton-Raphson iteration requires inversion of an $(I_n - 1) \times (I_n - 1)$ matrix, and this computation can become a problem if I_n is large.

Now we make use of the uniqueness in Conclusion 2 to solve the equations (10). Note that $\ell_r k_r = 0$, for $1 \le r \le I_n$; $\ell_r \ell_{r+1} = 0$ and $k_r k_{r+1} = 0$ for $1 \le r \le I_n - 1$. Suppose that we have selected an initial value of C_1, $\check{C}_1$. Then from the first equation of (10), we obtain an estimate, $\check{C}_2$, of C_2,

$$\check{C}_2 = \begin{cases} \frac{\alpha_0}{\alpha_1} + \frac{1}{\alpha_1} \Phi^{-1} \left(\Phi(-\alpha_0 + \alpha_1 \check{C}_1) + \frac{\alpha_1 \ell_2 \phi(-\alpha_0 + \alpha_1 \check{C}_1) \Phi(\check{C}_1)}{k_1 \phi(\check{C}_1)} \right) & \text{if} \quad k_2 = 0 \\ \Phi^{-1} \left(\Phi(\check{C}_1) + \frac{k_2 \Phi(-\alpha_0 + \alpha_1 \check{C}_1) \phi(\check{C}_1)}{\alpha_1 \ell_1 \phi(-\alpha_0 + \alpha_1 \check{C}_1)} \right) & \text{if} \quad \ell_2 = 0 \end{cases}.$$

For $r = 2, \cdots, I_n - 2$, using the latest estimates, $\check{C}_{r-1}$ and $\check{C}_r$, of C_{r-1} and C_r, we solve the rth equation of (10) to obtain the following estimate of C_{r+1}:

$$\check{C}_{r+1} = \begin{cases} \frac{\alpha_0}{\alpha_1} + \frac{1}{\alpha_1} \Phi^{-1} \left(\Phi(-\alpha_0 + \alpha_1 \check{C}_r) + \frac{\alpha_1 \ell_{r+1} \phi(-\alpha_0 + \alpha_1 \check{C}_r) \left(\Phi(\check{C}_r) - \Phi(\check{C}_{r-1}) \right)}{k_r \phi(\check{C}_r)} \right) & \text{if} \quad k_{r+1} = 0 \\ \Phi^{-1} \left(\Phi(\check{C}_r) + \frac{k_{r+1} \phi(\check{C}_r) \left(\Phi(-\alpha_0 + \alpha_1 \check{C}_r) - \Phi(-\alpha_0 + \alpha_1 \check{C}_{r-1}) \right)}{\alpha_1 \ell_r \phi(-\alpha_0 + \alpha_1 \check{C}_r)} \right) & \text{if} \quad \ell_{r+1} = 0 \end{cases}.$$

Hence, given the initially chosen value of C_1, $\check{C}_1$, we can obtain the estimates, $\check{C}_2, \ldots, \check{C}_{I_n-1}$, of $C_2, \ldots, C_{I_n-1}$ by solving the first $I_n - 2$ equations in (10). Now we are left to check whether these estimates also satisfy the last equation in (10),

$$\Lambda(\check{C}_{I_n-2}, \check{C}_{I_n-1}) = 0, \tag{11}$$

where $\Lambda(C_{I_n-2}, C_{I_n-1}) = \frac{\partial}{\partial C_{I_n-1}} \log\{\mathcal{L}_n(\theta, \mathbf{C})\}$. If $\Lambda(\check{C}_{I_n-2}, \check{C}_{I_n-1}) = 0$, the estimates, $\check{C}_r$, $r = 1, \ldots, I_n - 1$, are the unique solution to equation (10). If $\Lambda(\check{C}_{I_n-2}, \check{C}_{I_n-1}) \neq 0$, we need to update the initially chosen value estimate, $\check{C}_1$, and repeat the whole estimation process until the last equation in (10) is satisfied.

Let θ_0 be the true value of θ and g_0 be the true function of g. Denote $C_{r0} = g_0(\lim_{n\to\infty} Y_{(r)})$ and $\mathbf{C}_0 = (C_{10}, \cdots, C_{(I_n-1)0})^T$. In the following Conclusion 3, we establish the relationship between C_1 and $\Lambda(C_{I_n-2}, C_{I_n-1})$ to help in updating the initially chosen value, $\check{C}_1$. We provide a proof for Conclusion 3 in the Appendix A.3.

Conclusion 3. Let $\theta_n = \theta_0 + o_p(1)$. For any initially chosen value $\check{C}_1$ of C_1, we let $\check{C}_2, \ldots, \check{C}_{I_n-1}$ be the corresponding solution to the first $(I_n - 2)$ equations in (10). Then, when n is large enough,

1. if $\check{C}_1 < C_{10}$, then $\check{C}_r < C_{r0}$ for $r = 2, \cdots, I_n - 1$, and

$$\Lambda(\check{C}_{I_n-2}, \check{C}_{I_n-1}) = \frac{\partial}{\partial C_{I_n-1}} \log\{\mathcal{L}_n(\theta, \mathbf{C})\}|_{\mathbf{C}=\check{\mathbf{C}}, \theta=\theta_n} > 0;$$

2. if $\check{C}_1 > C_{10}$, then $\check{C}_r > C_{r0}$ for $r = 2, \cdots, I_n - 1$, and

$$\Lambda(\check{C}_{I_n-2}, \check{C}_{I_n-1}) = \frac{\partial}{\partial C_{I_n-1}} \log\{\mathcal{L}_n(\theta, \mathbf{C})\}|_{\mathbf{C}=\check{\mathbf{C}}, \theta=\theta_n} < 0,$$

where $\check{\mathbf{C}} = (\check{C}_1, \ldots, \check{C}_{I_n-1})$.

The results of Conclusion 3 provide a mechanism for updating the initially chosen value $\check{C}_1$. If $\Lambda(\check{C}_{I_n-2}, \check{C}_{I_n-1}) > 0$, we should increase our initially chosen value, $\check{C}_1$. On the other hand, if $\Lambda(\check{C}_{I_n-2}, \check{C}_{I_n-1}) < 0$, we should decrease our initially chosen value, $\check{C}_1$.

Given $\mathbf{C}$, we can estimate θ by maximizing (9). We next outline the two-stage iterative procedure for estimating θ and $\mathbf{C}$.

- Step 1. We combine data from the diseased and non-diseased samples and order test results in the combined sample, replace each test result by its true disease status. As a result, we

create a sequence of disease statuses for the combined sample. Denote the number of different sequences with the same consecutive disease status by I_n. Then, count the number of elements in each sequence, denoted by $\mathbf{k} = \{k_1, \cdots, k_{I_n} | \sum_{r=1}^{I_n} k_r = n_0\}$ for non-diseased subjects and $\ell = \{\ell_1, \cdots, \ell_{I_n} | \sum_{s=1}^{I_n} \ell_s = n_1\}$ for diseased subjects. For example, if we have data $\{5.38, 2.1, 4.5\}$ for non-diseased subjects and $\{12.5, 10.4, 16.8, 5.1, 13.5\}$ for diseased subjects, the ordered test results in the combined sample are $\{2.1, 4.5, 5.1, 5.38, 10.4, 12.5, 13.5, 16.8\}$, and their corresponding disease statuses are $\{no, no, di, no, di, di, di, di\}$, where *no* and *di* indicate a non-diseased and diseased subject, respectively. Thus, in the above notation, we have $I_n = 4$ and $k_1 = 2, k_2 = 0, k_3 = 1, k_4 = 0$ and $\ell_1 = 0, \ell_2 = 1, \ell_3 = 0, \ell_4 = 4$.

- Step 2. Given values of α_0, α_1, we estimate $C_1, \cdots, C_{I_n-1}$ by solving (10).
- Step 3. Given estimates of $C_1, \cdots, C_{I_n-1}$, we estimate α_0 and α_1 by maximizing (9) with respect to α_0 and α_1.
- Step 4. Repeat Steps 2 and 3 until two successive values for $(\alpha_0, \alpha_1, C_1, \cdots, C_{I_n-1})$ converge. The convergent values $\widehat{\alpha}_0, \widehat{\alpha}_1, \widehat{C}_1, \cdots, \widehat{C}_{I_n-1}$ are the estimates of $\alpha_0, \alpha_1, C_1, \cdots, C_{I_n-1}$.

3 Asymptotic distribution theory

Our final estimate $\widehat{\theta}$ of θ is a profile likelihood estimate, which maximizes the profile likelihood for θ given by

$$PL(\theta) = \mathcal{L}_n(\theta, \widehat{\mathbf{C}}(\theta)),$$

where $\widehat{\mathbf{C}}(\theta)$ maximizes the likelihood $\mathcal{L}_n(\theta, \mathbf{C})$ for a fixed value of θ. This estimator is a function of the test values only through their ranks. Using the results on the properties of maximum profile likelihood estimates derived by Murphy and Van der Varrt (2000), we can show that $\widehat{\theta}$ is fully efficient and $n^{1/2}(\widehat{\theta} - \theta_0)$ converges in distribution to a zero-mean bivariate normal random vector with covariance matrix Σ, where

$$\begin{aligned}\Sigma \;=\; & \left\{ -\lim_{n\to\infty} \frac{\partial^2 \log \mathcal{L}_n(\theta, \mathbf{C})}{n \partial\theta \partial\theta'} + \left(\lim_{n\to\infty} \frac{\partial^2 \log \mathcal{L}_n(\theta, \mathbf{C})}{n \partial\theta \partial\mathbf{C}'} \right) \right. \\ & \left. \times \left(\lim_{n\to\infty} \frac{\partial^2 \log \mathcal{L}_n(\theta, \mathbf{C})}{n \partial\mathbf{C} \partial\mathbf{C}'} \right)^{-1} \left(\lim_{n\to\infty} \frac{\partial^2 \log \mathcal{L}_n(\theta, \mathbf{C})}{n \partial\mathbf{C} \partial\theta'} \right) \right\}^{-1} |_{\theta=\theta_0, \mathbf{C}=\mathbf{C}_0}. \qquad (12)\end{aligned}$$

Based on the estimates of α_0 and α_1, we can estimate the ROC curve by $\widehat{ROC}(u) = \Phi(\widehat{\alpha}_0 + \widehat{\alpha}_1\Phi^{-1}(u))$. Using the Taylor series expansion and the asymptotically normal result of $\widehat{\theta}$, we can show that $n^{1/2}\left(\widehat{ROC}(u) - ROC(u)\right)$ converges in distribution to a zero-mean normal random variable with variance

$$\phi^2(\alpha_{00} + \alpha_{10}\Phi^{-1}(u)) \begin{pmatrix} 1 \\ \Phi^{-1}(u) \end{pmatrix}^T \Sigma \begin{pmatrix} 1 \\ \Phi^{-1}(u) \end{pmatrix},$$

where Σ is defined by (12), α_{00}, and α_{10} are the true values of α_0 and α_1, respectively.

4 Numerical studies

In this section we conduct several simulation studies to (1) determine if our estimator is more efficient than the five existing estimators, (2) assess the robustness of our estimator against the departure from the binormal model, and (3) evaluate the accuracy of the asymptotical variance estimator of our estimator in finite sample sizes.

4 · 1 *Efficiency*

In this subsection we investigate the statistical efficiency of the six methods: the proposed method (MLE), CM, MHS, AP, PC and ZH methods for estimating α_0 and α_1 in the binormal model and for estimating the corresponding ROC curve by numerical studies. We use the root of mean squared error (RMSE) to measure the performance of the various estimators for α_0 and α_0 and the sum of RMSEs for α_0 and α_1 as an overall performance measure. We evaluate the performance of an estimator $\widehat{ROC}(\cdot)$ for the ROC curve using the average of squared errors (ASE), defined by

$$\text{ASE} = \frac{1}{n_{grid}} \sum_{k=1}^{n_{grid}} \left\{\widehat{ROC}(u_k) - ROC(u_k)\right\}^2, \tag{13}$$

where $\{u_k, k = 1, \cdots, n_{grid}\}$ are the grid points at which the functions $ROC(\cdot)$ are estimated. In the simulation studies, we choose $n_{grid} = 100$ and u_k's to be uniformly distributed over $(0, 1)$. We choose 500 simulations for each scenario. Data for non-diseased subjects are generated from the standard normal distribution, and data for diseased subjects are generated from $N(2, 1.44)$. We choose sizes of the diseased and non-diseased samples to be both equal and unequal, $(n_0, n_1) =$

$\{(100, 100), (200, 100), (200, 200)\}$, to investigate the effect of the sample sizes on the performance of the estimates.

Table 1 gives bias, SD, RMSE and SRMSE of the resulting estimators for α_0 and α_1 by the six methods. From Table 1 we see that in all cases, our new MLE has the smallest SRMSE and is the best choice. Specifically, our MLE consistently has smaller bias, standard error, and RMSE than the CM estimator due to a smaller number of nuisance parameters to estimate. Although the PC can have the smallest variance, its bias is also large and can even be larger than its variance, which means that the bias is significant and could not be ignored; the AP estimator is less biased than the PC estimator but has a larger variance than the PC estimator. The ZH estimator has the largest RMSE and bias.

Figure 1 depicts the distribution of the estimated ASE for the ROC curve over the 500 replications for each method. The MLE and CM estimators have comparable ASE, which is smaller than the other existing methods. The performance of the estimators MHS and AP is close to that of our MLE in this setting. Further simulation study (not reported here) shows that when the accuracy of a diagnostic test is not too high or the sample size is large so that I_n can be large, the computation algorithm in the MHS method, which collapses too many jump points, can lead to some loss of efficiency. The AP estimator has smaller ASE than the PC estimator, and the ZH estimator has the largest ASE.

In summary, the existing CM, MHS, and AP estimators have similar efficiency as our ML estimator with the ML estimators being slightly better. The ZH estimators have the worst performance.

4 · 2 ***Robustness***

In the subsection, we compare the robustness of the six methods against the departure from the binormal assumption.

One anonymous reviewer has suggested that it may be reasonable to expect a transformation to result in approximate normal data for non-diseased subjects, but since the population of diseased subjects is often a mixture of subpopulations of subjects in different stages of the disease/infection, it seems much more reasonable to expect that transformation would result in a

Table 1: Estimates of (α_0, α_1) compared with their actual values over the 500 replications

			$\alpha_0 = 2/1.2$			$\alpha_1 = 1/1.2$			
n_0	n_1	method	Bias	SD	RMSE	Bias	SD	RMSE	SRMSE
100	100	MLE	0.038	0.209	0.212	0.012	0.137	0.137	0.349
		CM	0.074	0.213	0.226	0.043	0.141	0.148	0.373
		MHS	0.002	0.212	0.212	-0.033	0.142	0.146	0.358
		AP	0.025	0.225	0.226	0.041	0.164	0.169	0.395
		PC	-0.105	0.169	0.199	-0.163	0.088	0.185	0.384
		ZH	0.126	0.205	0.241	0.367	0.095	0.379	0.620
200	100	MLE	0.028	0.180	0.182	0.009	0.113	0.113	0.295
		CM	0.039	0.184	0.188	0.022	0.115	0.117	0.305
		MHS	0.022	0.208	0.209	-0.014	0.142	0.143	0.352
		AP	0.027	0.191	0.193	0.034	0.131	0.136	0.328
		PC	-0.105	0.142	0.177	-0.129	0.092	0.158	0.335
		ZH	0.128	0.168	0.211	0.288	0.063	0.295	0.506
200	200	MLE	0.016	0.140	0.141	0.002	0.093	0.093	0.234
		CM	0.024	0.142	0.144	0.012	0.094	0.095	0.239
		MHS	-0.009	0.137	0.137	-0.014	0.102	0.103	0.240
		AP	0.009	0.144	0.145	0.017	0.103	0.104	0.249
		PC	-0.108	0.123	0.164	-0.132	0.087	0.158	0.322
		ZH	0.049	0.152	0.160	0.305	0.078	0.315	0.475

mixture of normals rather than a single normal for diseased subjects. So, to investigate the robustness of the binormal model, we simulate test responses of non-diseased subjects from $N(0,1)$, but test responses of diseased subjects from the mixture of the two normal distributions, $N(1.2, 1.2^2)$ and $N(2.2, 1.5^2)$, with the corresponding mixing proportions of $1/2$ and $1/2$, respectively. We set $(n_0, n_1) = \{(100,100),(200,200)\}$ to investigate the effect of the sample sizes on the performance of the estimates.

Figures 2(A) and 3(A) plot the average of the estimated ROC curves over the 500 replications for each method when the sample sizes are $(100,100)$ and $(200,200)$, respectively. Our MLE has the smallest ASE and hence is the most robust estimate among the six ones considered here. The CM, MHS and AP also have good robustness properties. The PC and ZH estimators have larger bias. Figures 2(B) and 3(B) depict the distribution of the ASE for the estimated ROC curves over the 500 replications for each method when the sample sizes are $(100,100)$ and $(200,200)$, respectively. The AP estimator has better ASE than the PC estimator, which means the AP estimator is more robust than the PC estimator. The ZH estimator has the largest ASE.

We also conduct numerical studies with a larger number of the components in a normal mixture. We generate test results of non-diseased subjects from the standard normal distribution but test results of diseased subjects from a mixture of the three normal distributions, $N(1.2, 1.2^2)$, $N(2.2, 1.5^2)$ and $N(2.2, 1)$, with the corresponding mixing proportions of $1/3$, $1/3$, and $1/3$, respectively. The results (not reported here) are similar to those in Figures 2 and 3 except that the PC estimator seems to have the largest bias and ASE, suggesting that the robustness of the PC estimator may decrease as the number of components in normal mixtures increases.

In summary, the existing CM, MHS, and AP estimators have similar robustness as our ML estimators even though the ML estimators are slightly better.

$4 \cdot 3$ *Asymptotic inference in finite sample*

Finally, we assess the accuracy of our variance estimator given in Section 3 in finite sample sizes. We investigate the performance of our variance estimator using the simulated data in Sections 5.1 and 5.2. Based on 500 simulated data sets, we obtain 500 estimates of $\widehat{\alpha}_0$ and $\widehat{\alpha}_1$ and their corresponding

Table 2: Average (SE_{ave}) and standard deviation (SE_{std}) of the standard error estimates over the 500 replications for the binormal simulated data in Section 4.1

		$\alpha_0 = 2/1.2$			$\alpha_1 = 1/1.2$		
n_0	n_1	SD	$SE_{ave}(SE_{std})$	Coverage	SD	$SE_{ave}(SE_{std})$	Coverage
50	100	0.230	0.226(0.049)	0.939	0.160	0.161(0.043)	0.917
100	100	0.209	0.202(0.040)	0.920	0.137	0.133(0.031)	0.922
200	100	0.180	0.186(0.035)	0.931	0.113	0.113(0.024)	0.933
200	200	0.140	0.140(0.020)	0.947	0.093	0.091(0.015)	0.929

Table 3: Average (SE_{ave}) and standard deviation (SE_{std}) of the standard error estimates over the 500 replications for the mixture normal data in Section 4.2

			α_0		α_1	
n_0	n_1	k*	SD	$SE_{ave}(SE_{std})$	SD	$SE_{ave}(SE_{std})$
50	50	3	0.239	0.244(0.034)	0.159	0.167(0.036)
100	100	3	0.164	0.168(0.016)	0.111	0.113(0.018)
100	100	2	0.155	0.161(0.014)	0.097	0.105(0.015)
200	200	2	0.106	0.112(0.006)	0.068	0.071(0.007)

*where k is the number of terms in the mixture of normals for the diseased data.

standard error estimates using our method. From ML estimates of α_0 and α_1, we form the empirical standard deviations, denoted by SD, which can be regarded as an approximation to the true standard deviations. We denote the average and the standard deviation of 500 estimated standard errors for the estimated $\widehat{\alpha}_0$ and $\widehat{\alpha}_1$ by SE_{ave} and SE_{std}, respectively, which summarize the overall performance of the standard error estimator. We report our results in Table 2 for the simulated binormal data and in Table 3 for the simulated mixture normal data, respectively. The standard error estimators are very close to the "true" sample standard errors. The empirical CI coverage probabilities are close to their nominal levels.

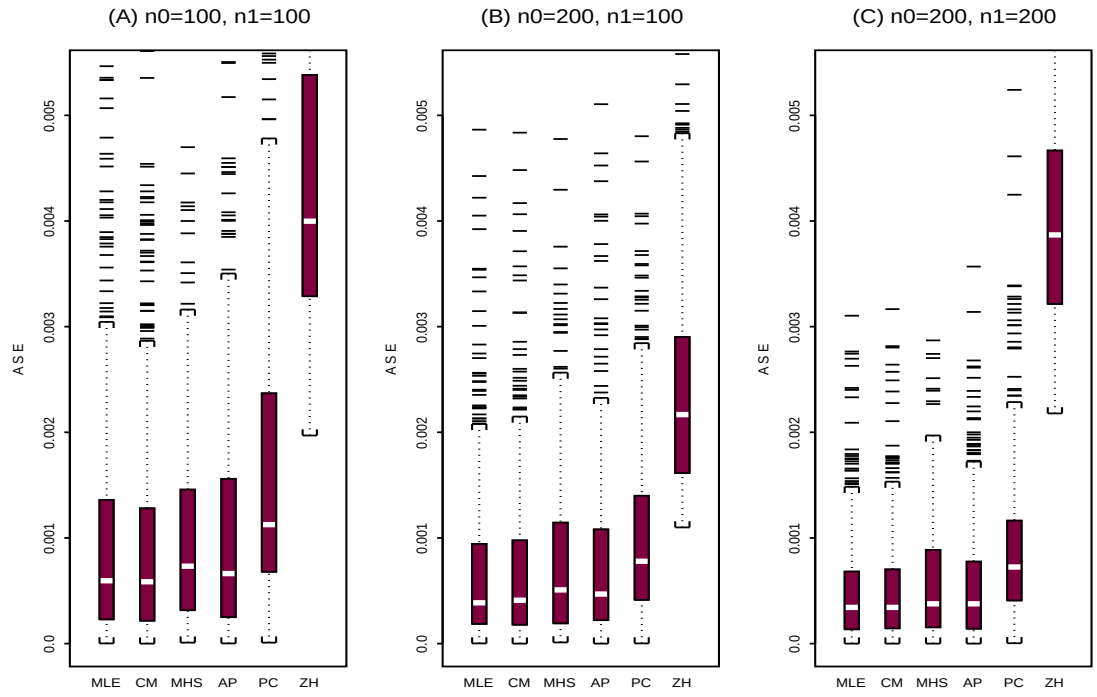

Figure 1: The distribution of ASE for the estimated ROC curves from the binormal model over the 500 replications.

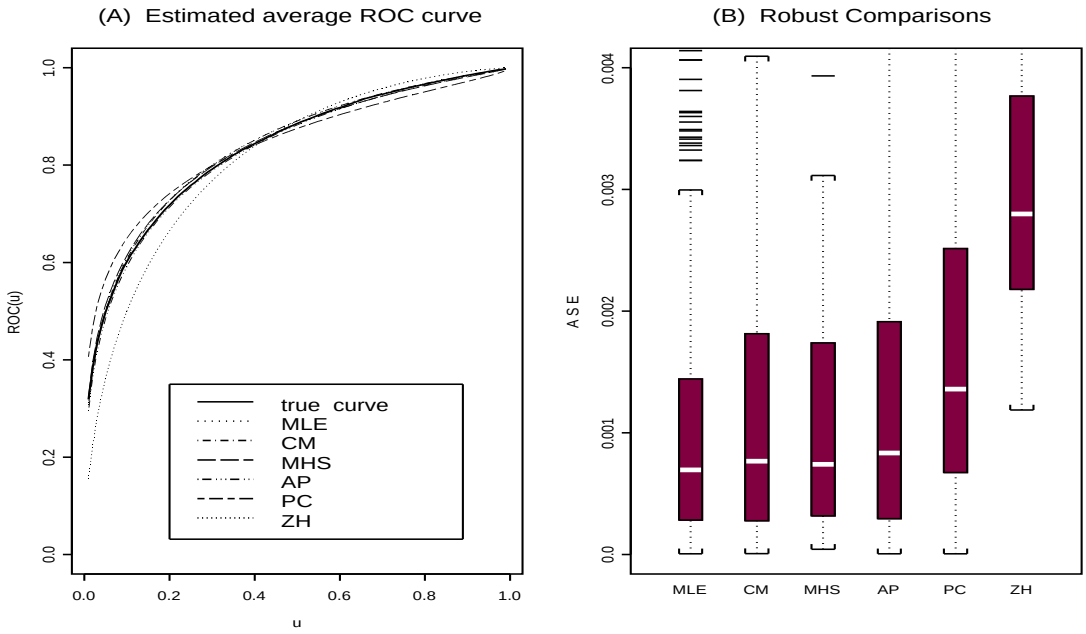

Figure 2: The diseased data are from a mixture of two normal distributions, but modeled with the binormal model when $n_0 = n_1 = 100$. (A) The average of the estimated ROC curves; (B) the distribution of ASE for the estimated ROC curves over the 500 replications.

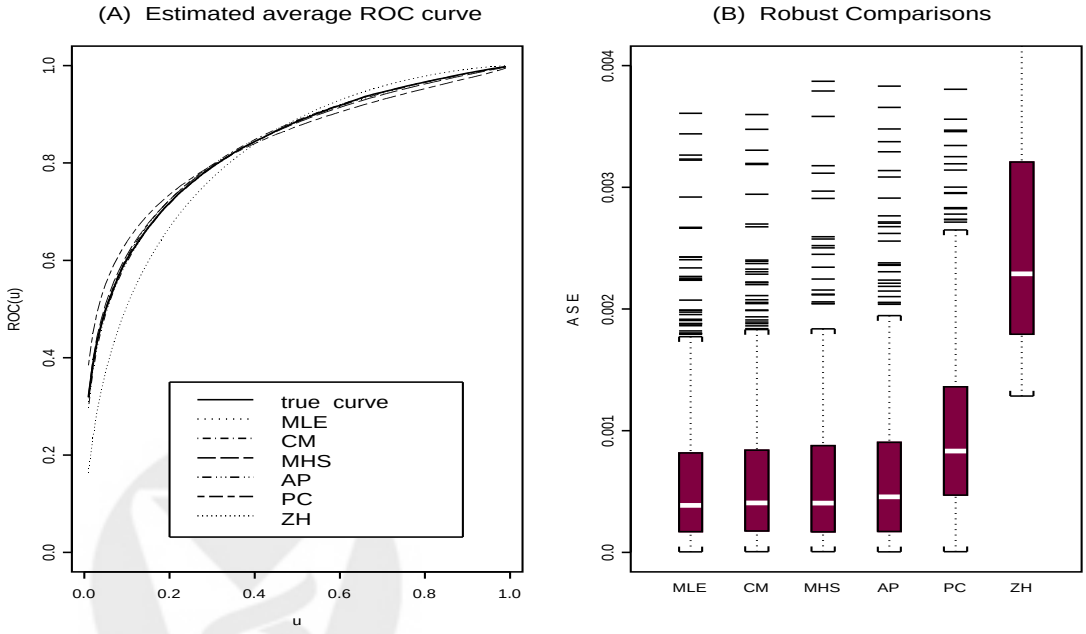

Figure 3: The diseased data are from a mixture of two normal distributions, but modeled with the binormal model when $n_0 = n_1 = 200$. (A) The average of the estimated ROC curves; (B) the distribution of ASE for the estimated ROC curves over the 500 replications.

5 A real data example

In this section we illustrate the application of our newly proposed method in a real example on the accuracy of biomarkers for detecting pancreatic cancer (Wieand et al., 1989). This study examined two biomarkers, the antigenic determinant, designated as CA125, and carbohydrate antigen designated as CA19-9. The data consist of 51 measurements on subjects free of disease and 90 measurements on diseased subjects using the two biomarkers. Here, we used the results with CA125 to illustrate the application of our methodology.

Although the binormal ROC model (3) possesses a certain degree of robustness against normal mixtures, as shown in the simulation study, it is also important to assess whether the binormal model (3) is appropriate for the data before we make inferences on the ROC curve of the CA125 using the binormal model. Here, we present a graphical method to test model (3).

To detect any large discrepancies in fit, we compare the empirical ROC curve with the MLE of the ROC curve obtained by substituting ML estimates of α_0, α_1 into model (3).

Figure 6(a) plots the empirical ROC curve, the maximum likelihood estimate of the ROC curve and its 95% pointwise confidence intervals (denoted CI in Figure 6(a)), showing no obvious difference between the empirical ROC curve and the estimated ROC curve based on the binormal model. So, the binormal model is reasonable.

Table 4 lists the estimates for the coefficients α_0 and α_1 using the six methods. Both the PC and ZH estimates are different from the others for the estimation of α_0; ZH is very different with the other for estimation of α_1. These results are consistent with the simulation results, which have shown that the PC and ZH estimates have large bias. Figure 6(b) plots the estimated ROC curves using the six methods. The MLE, CM, and MHS estimates are quite similar and are distinct from the others. We also note that the ZH estimate is substantially different from the rest.

6 Discussion

In this paper we have proposed a semi-parametric MLE for the ROC curve under the binormal ROC curve model (3). The ML estimator is asymptotically normal. The asymptotic results also hold for

Table 4: Estimates of (α_0, α_1) for CA125 as a diagnostic market of pancreatic cancer

method	$\widehat{\alpha}_0(SD)$	$\widehat{\alpha}_1(SD)$
MLE	1.192(0.158)	0.431(0.081)
MHS	1.177(0.160)	0.399(0.082)
CM	1.235(0.129)	0.480(0.074)
AP	1.142(0.153)	0.468(0.110)
PC	1.343(0.192)	0.490(0.040)
ZH	1.240(0.161)	0.911(0.048)

a more general specification of the parametric ROC curve model given by (2), for example, when G and H are symmetric distributions and when H belongs to a location-scale family. Our simulation results have indicated that the proposed ML estimators also have good finite-sample properties and have similar efficiency and robustness as the existing CM, MHS, and AP estimators with the ML estimators being slightly better than all the existing estimators considered here.

Hanley (1988) has shown that the binormal ROC curve model for ordinal-scale tests enjoys a certain degree of robustness against departure from the bi-normality assumption. Our own simulation studies have also demonstrated this result. However, given limitations of any simulation study, we want to emphasize that it is important to check the assumption of the bi-normality in any application; for example one may use the graphical method suggested in Section 5.

ACKNOWLEDGMENTS

We like to thanks the editor, an associate editor and anonymous referees for many helpful suggestions and comments that result in an improved version of this manuscript. This paper presents the findings and conclusions of the authors. It does not necessarily represent those of VA HSR&D Service.

Appendix

We first define some additional notations that are needed to prove Conclusions 1-3. Define $D_r = (-1, C_r)^T$, $D_r^* = (-1, C_r^*)^T$, $D_{r0} = (-1, C_{r0})^T$, $\check{D}_r = (-1, \check{C}_r)^T$ and $b_0 = \lim n_1/(n_0 + n_1)$.

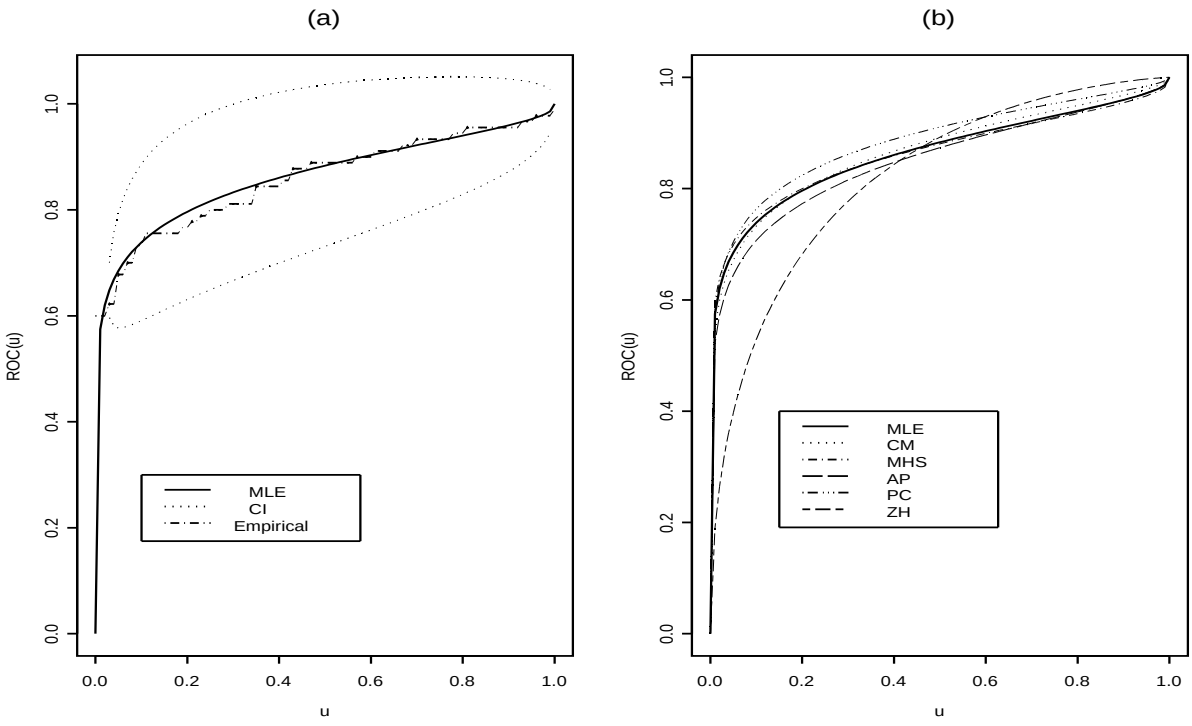

Figure 4: Estimated ROC curves using MLE, CM, MHS, AP, PC, and ZH for CA125 as a diagnostic marker of pancreatic cancer.

A.1 *Proof of Conclusion 1*

Let $\lambda_n(\theta, \mathbf{C}^*) = \frac{1}{n}\log L_n$, where L_n is defined by (8), $\mathbf{C}^* = (C_1^*, \cdots, C_{I_n^*-1}^*)^T$, $C_r^* = g(Y_{(r)}^*)$, and $Y_{(1)}^* < \cdots < Y_{(I_n^*)}^*$ are distinct ordered test results of $Y_{0i}, i = 1, \cdots, n_0$ and $Y_{1j}, j = 1, \cdots, n_1$. It can be shown that the MLE of θ and $\mathbf{C}^*$ must satisfy the following equations:

$$\frac{\partial \lambda_n(\theta, \mathbf{C}^*)}{\partial C_r^*} = \frac{1}{n}\left(\frac{k_r^*}{\Phi(C_r^*) - \Phi(C_{r-1}^*)} - \frac{k_{r+1}^*}{\Phi(C_{r+1}^*) - \Phi(C_r^*)}\right)\phi(C_r^*)$$
$$+\frac{1}{n}\left(\frac{\ell_r^*}{\Phi(\theta^T D_r^*) - \Phi(\theta^T D_{r-1}^*)} - \frac{\ell_{r+1}^*}{\Phi(\theta^T D_{r+1}^*) - \Phi(\theta^T D_r^*)}\right)\alpha_1\phi(\theta^T D_r^*) = 0, \tag{14}$$

for $1 \le r \le I_n^* - 1$, and

$$\frac{\partial \lambda_n(\theta, \mathbf{C}^*)}{\partial \theta} = \frac{1}{n}\sum_{r=1}^{I_n^*} \ell_r^* \frac{\phi(\theta^T D_r^*)D_r^* - \phi(\theta^T D_{r-1}^*)D_{r-1}^*}{\Phi(\theta^T D_r^*) - \Phi(\theta^T D_{r-1}^*)}$$
$$+\frac{1}{n}\ell_1^* \frac{\phi(\theta^T D_1^*)D_1^*}{\Phi(\theta^T D_1^*)} - \frac{1}{n}\ell_{I_n^*}^* \frac{\phi(\theta^T D_{I_n^*-1}^*)D_{I_n^*-1}^*}{1 - \Phi(\theta^T D_{I_n^*-1}^*)} = 0, \tag{15}$$

where $k_r^* = \#\{Y_{0i} = Y_{(r)}^*, i = 1, \cdots, n_0\}$ and $\ell_r^* = \#\{Y_{1j} = Y_{(r)}^*, j = 1, \cdots, n_1\}$. If both $Y_{(r)}^*$ and $Y_{(r+1)}^*$ correspond to non-diseased subjects, then $\ell_r^* = \ell_{r+1}^* = 0$, $k_r^* > 0$ and $k_{r+1}^* > 0$. Hence, from (14), we have $\frac{k_r^*}{\Phi(C_r^*) - \Phi(C_{r-1}^*)} = \frac{k_{r+1}^*}{\Phi(C_{r+1}^*) - \Phi(C_r^*)}$. Extending this argument to a sequence of M contiguous jump points which only involve non-diseased subjects, we have

$$\frac{k_r^*}{\Phi(C_r^*) - \Phi(C_{r-1}^*)} = \cdots = \frac{k_{r+M-1}^*}{\Phi(C_{r+M-1}^*) - \Phi(C_{r+M-2}^*)},$$

which is equal to

$$\frac{\sum_{j=r}^{r+M-1} k_j^*}{\Phi(C_{r+M-1}^*) - \Phi(C_{r-1}^*)}.$$

Similar arguments indicate that for a sequence of M contiguous jump points which only involves diseased subjects, we have

$$\frac{\ell_r^*}{\Phi(\theta^T D_r^*) - \Phi(\theta^T D_{r-1}^*)} = \cdots = \frac{\ell_{r+M-1}^*}{\Phi(\theta^T D_{r+M-1}^*) - \Phi(\theta^T D_{r+M-2}^*)},$$

which is equal to

$$\frac{\sum_{j=r}^{r+M-1} \ell_j^*}{\Phi(\theta^T D_{r+M-1}^*) - \Phi(\theta^T D_{r-1}^*)}.$$

Therefore, if we denote $Y_{(1)} < \cdots < Y_{(I_n-1)}$ to be the last points of contiguous sequences with same disease statuses, and $C_r = g(Y_{(r)}), r = 1, \cdots, I_n - 1$, for $1 \le r \le I_n - 1$, we can write (14) and (15)

as

$$\frac{\partial \lambda_n(\theta, \mathbf{C}^*)}{\partial C_r} = \frac{1}{n}\left(\frac{k_r}{\Phi(C_r) - \Phi(C_{r-1})} - \frac{k_{r+1}}{\Phi(C_{r+1}) - \Phi(C_r)}\right)\phi(C_r)$$
$$+\frac{1}{n}\left(\frac{\ell_r}{\Phi(\theta^T C_r) - \Phi(\theta^T C_{r-1})} - \frac{\ell_{r+1}}{\Phi(\theta^T C_{r+1}) - \Phi(\theta^T C_r)}\right)\alpha_1\phi(\theta^T C_r) = 0 \tag{16}$$

and

$$\frac{\partial \lambda_n(\theta, \mathbf{C}^*)}{\partial \theta} = \frac{1}{n}\sum_{r=2}^{I_n-1}\ell_r\frac{\phi(\theta^T D_r)D_r - \phi(\theta^T D_{r-1})D_{r-1}}{\Phi(\theta^T D_r) - \Phi(\theta^T D_{r-1})}$$
$$+\frac{1}{n}\ell_1\frac{\phi(\theta^T D_1)D_1}{\Phi(\theta^T D_1)} - \frac{1}{n}\ell_{I_n}\frac{\phi(\theta^T D_{I_n-1})D_{I_n-1}}{1 - \Phi(\theta^T D_{I_n-1})} = 0, \tag{17}$$

respectively, where k_r and ℓ_r are defined in Section 2. Note that (16) and (17) do not depend on the nuisance parameters $C_r^* = g(Y_{(r)}^*)$'s with $Y_{(r)}^* \in \Re$. Hence Conclusion 1 follows.

A.2 *Proof of Conclusion 2*

The Hessian matrix corresponding to the log-likelihood function is tridiagonal; that is, the Hessian matrix has nonzero entries only along its diagonal and in elements adjacent to its diagonal. If we denote Q to be the $(I_n - 1) \times (I_n - 1)$ Hessian matrix, and z to be any $(I_n - 1)$ dimensional vector, we have

$$\begin{aligned}
z^T Q z &= \frac{1}{n}\sum_{r=1}^{I_n-1} z_r^2\left\{k_r\frac{\phi'(C_r)\left(\Phi(C_r) - \Phi(C_{r-1})\right) - \phi^2(C_r)}{\left(\Phi(C_r) - \Phi(C_{r-1})\right)^2} - k_{r+1}\frac{\phi'(C_r)\left(\Phi(C_{r+1}) - \Phi(C_r)\right) + \phi^2(C_r)}{\left(\Phi(C_{r+1}) - \Phi(C_r)\right)^2}\right. \\
&\quad +\alpha_1^2\ell_r\frac{\phi'(\theta^T D_r)\left(\Phi(\theta^T D_r) - \Phi(\theta^T D_{r-1})\right) - \phi^2(\theta^T D_r)}{\left(\Phi(\theta^T D_r) - \Phi(\theta^T D_{r-1})\right)^2} \\
&\quad \left.-\alpha_1^2\ell_{r+1}\frac{\phi'(\theta^T D_r)\left(\Phi(\theta^T D_{r+1}) - \Phi(\theta^T D_r)\right) + \phi^2(\theta^T D_r)}{\left(\Phi(\theta^T D_{r+1}) - \Phi(\theta^T D_r)\right)^2}\right\} \\
&\quad +\frac{1}{n}\sum_{r=1}^{I_n-2} z_r z_{r+1}\left\{k_{r+1}\frac{\phi(C_r)\phi(C_{r+1})}{\left(\Phi(C_{r+1}) - \Phi(C_r)\right)^2} + \alpha_1^2\ell_{r+1}\frac{\phi(\theta^T D_r)\phi(\theta^T D_{r+1})}{\left(\Phi(\theta^T D_{r+1}) - \Phi(\theta^T D_r)\right)^2}\right\} \\
&\quad +\frac{1}{n}\sum_{r=2}^{I_n-1} z_r z_{r-1}\left\{k_r\frac{\phi(C_r)\phi(C_{r-1})}{\left(\Phi(C_r) - \Phi(C_{r-1})\right)^2} + \alpha_1^2\ell_r\frac{\phi(\theta^T D_r)\phi(\theta^T D_{r-1})}{\left(\Phi(\theta^T D_r) - \Phi(\theta^T D_{r-1})\right)^2}\right\} \\
&= \varpi_1 + \varpi_2 - \varpi_3.
\end{aligned}$$

Some computations show that

$$\begin{aligned}
\varpi_1 &= \frac{1}{n}\sum_{r=1}^{I_n-1} z_r^2\phi'(C_r)\left\{\frac{k_r}{\Phi(C_r)-\Phi(C_{r-1})}-\frac{k_{r+1}}{\Phi(C_{r+1})-\Phi(C_r)}\right\},\\
\varpi_2 &= \frac{1}{n}\sum_{r=1}^{I_n-1} z_r^2\phi'(\theta^T D_r)\left\{\frac{\alpha_1^2\ell_r}{\Phi(\theta^T D_r)-\Phi(\theta^T D_{r-1})}-\frac{\alpha_1^2\ell_{r+1}}{\Phi(\theta^T D_{r+1})-\Phi(\theta^T D_r)}\right\},\\
\varpi_3 &= \frac{1}{n}\left\{\sum_{r=1}^{I_n-1}\frac{\left(z_r\phi(C_r)-z_{r-1}\phi(C_{r-1})\right)^2 k_r}{\left(\Phi(C_r)-\Phi(C_{r-1})\right)^2}+\sum_{r=1}^{I_n-1}\frac{\left(z_r\phi(\theta^T D_r)-z_{r-1}\phi(\theta^T D_{r-1})\right)^2\alpha_1^2\ell_r}{\left(\Phi(\theta^T D_r)-\Phi(\theta^T D_{r-1})\right)^2}\right.\\
&\quad\left.+\frac{z_{I_n-1}^2 k_{I_n}\phi^2(C_{I_n-1})}{\left(1-\Phi(C_{I_n-1})\right)^2}+z_{I_n-1}^2\frac{\alpha_1^2\ell_{I_n}\phi^2(\theta^T D_{I_n-1})}{\left(1-\Phi(\theta^T D_{I_n-1})\right)^2}\right\}.
\end{aligned}$$

Note that $k_r=\sum_{i=1}^{n_0} I(C_{r-1}<g(Y_{0i})\le C_r)$ and $\ell_r=\sum_{j=1}^{n_1} I(C_{r-1}<g(Y_{1j})\le C_r)$, by the central limit theorem, we can show that $\varpi_1=O(n^{-1/2})$ and $\varpi_2=O(n^{-1/2})$. Since $\varpi_3\ge 0$ and equals zero if and only if $z=0$, $z^TQz\le 0$ and equals zero if and only if $z=0$ when n is large enough. Therefore, given θ, the log-likelihood function is a concave function of $\mathbf{C}$, which implies that a unique maximum exists.

A.3 *Proof of Conclusion 3*

Let $\check{C}_1=C_{10}+\varepsilon$ for any $\varepsilon>0$ and $\check{C}_2,\cdots,\check{C}_{I_n-1}$ be the solution to the first I_n-2 score equations in (10) given $C_1=\check{C}_1$. Define $\check{\Phi}_r=\Phi(\check{C}_r)$ for $r=1,\cdots,I_n-1$. Let $\check{\Phi}_{I_n}$ be the solution to the following equation:

$$\begin{aligned}
G_n(x) &\equiv k_{I_n-1}\frac{\phi(\check{C}_{I_n-1})}{\Phi(\check{C}_{I_n-1})-\Phi(\check{C}_{I_n-2})}-k_{I_n}\frac{\phi(\check{C}_{I_n-1})}{x-\Phi(\check{C}_{I_n-1})}\\
&\quad+\alpha_1\ell_{I_n-1}\frac{\phi(\theta^T\check{D}_{I_n-1})}{\Phi(\theta^T\check{D}_{I_n-1})-\Phi(\theta^T\check{D}_{I_n-2})}-\alpha_1\ell_{I_n}\frac{\phi(\theta^T\check{D}_{I_n-1})}{x-\Phi(\theta^T\check{D}_{I_n-1})}=0.
\end{aligned}$$

Since

$$\frac{k_r}{n}-(1-b_0)\left[\Phi(C_{r0})-\Phi(C_{(r-1),0})\right]=O_p(n^{-1/2}) \tag{18}$$

and

$$\frac{\ell_r}{n}-b_0\left[\Phi(\theta_0^T D_{r0})-\Phi(\theta_0^T D_{r-1,0})\right]=O_p(n^{-1/2}), \tag{19}$$

for $1\le r\le I_n$, we have

$$G_n(x)=g_n(x)\left(1+o_p(1)\right), \tag{20}$$

where

$$\begin{aligned} g_n(x) &= (1-b_0)\phi(\check{C}_{I_n-1})\left[\frac{\Phi(C_{I_n-1,0})-\Phi(C_{I_n-2,0})}{\Phi(\check{C}_{I_n-1})-\Phi(\check{C}_{I_n-2})}-\frac{1-\Phi(C_{I_n-1,0})}{x-\Phi(\check{C}_{I_n-1})}\right] \\ &\quad +b_0\alpha_1\phi(\theta^T\check{D}_{I_n-1})\left[\frac{\Phi(\theta^T D_{I_n-1,0})-\Phi(\theta^T D_{I_n-2,0})}{\Phi(\theta^T\check{D}_{I_n-1})-\Phi(\theta^T\check{D}_{I_n-2})}-\frac{1-\Phi(\theta^T D_{I_n-1,0})}{x-\Phi(\theta^T\check{D}_{I_n-1})}\right]. \end{aligned} \quad (21)$$

Since $G_n(\check{\Phi}_{I_n})=0$, we have

$$g_n(\check{\Phi}_{I_n}) = o_p(1). \quad (22)$$

Furthermore, we have,

$$\begin{aligned} &\frac{1}{n}\frac{\partial}{\partial C_{I_n-1}}\log\{L_n\}|_{\mathbf{C}=\check{\mathbf{C}}} \\ =\ &\left\{(1-b_0)\phi(\check{C}_{I_n-1})\left[\frac{\Phi(C_{I_n-1,0})-\Phi(C_{I_n-2,0})}{\Phi(\check{C}_{I_n-1})-\Phi(\check{C}_{I_n-2})}-\frac{1-\Phi(C_{I_n-1,0})}{1-\Phi(\check{C}_{I_n-1})}\right]\right. \\ &\left.+b_0\alpha_1\phi(\theta^T\check{D}_{I_n-1})\left[\frac{\Phi(\theta^T D_{I_n-1,0})-\Phi(\theta^T D_{I_n-2,0})}{\Phi(\theta^T\check{D}_{I_n-1})-\Phi(\theta^T\check{D}_{I_n-2})}-\frac{1-\Phi(\theta^T D_{I_n-1,0})}{1-\Phi(\theta^T\check{D}_{I_n-1})}\right]\right\}(1+o_p(1)) \\ =\ &g_n(1)\,(1+o_p(1)) = g_n(\Phi(C_{I_n,0}))\,(1+o_p(1)), \end{aligned}$$

where $\check{\mathbf{C}}=(\check{C}_1,\ldots,\check{C}_{I_n-1})$. Hence, if the assumption that

$$\check{\Phi}_r \ > \ \Phi(C_{r0}) \quad (23)$$

holds for $r=I_n$, by (22) and the fact that $g_n(x)$ is a strict increasing function of x, we obtain

$$\frac{1}{n}\frac{\partial}{\partial C_{I_n-1}}\log\{L_n\}|_{\mathbf{C}=\check{\mathbf{C}}} < 0$$

for any given $\varepsilon>0$. Hence the second part of Conclusion 3 follows.

Now we prove that the assumption (23) holds for $r=2,\cdots,I_n$. We use the inductive method to prove (23). The inductive method relies on I_n-1 steps. The first step consists of an conclusion for $r=2$. From (10), we see that $\check{C}_2$ satisfies

$$\begin{aligned} G_1(x) &= \frac{1}{n}\frac{\partial}{\partial C_1}\log\{\mathcal{L}_n(\theta,\mathbf{C})\}|_{C_1=\check{C}_1,C_2=x} \\ &= \frac{k_1}{n}\frac{\phi(\check{C}_1)}{\Phi(\check{C}_1)}-\frac{k_2}{n}\frac{\phi(\check{C}_1)}{\Phi(x)-\Phi(\check{C}_1)} \\ &\quad +\alpha_1\frac{\ell_1}{n}\frac{\phi(\theta^T\check{D}_1)}{\Phi(\theta^T\check{D}_1)}-\alpha_1\frac{\ell_2}{n}\frac{\phi(\theta^T\check{D}_1)}{\Phi(\theta^T\tilde{x})-\Phi(\theta^T\check{D}_1)}=0. \end{aligned} \quad (24)$$

By (18) and (19), we have

$$G_1(x) = g_1(x)\,(1 + o_p(1)) \tag{25}$$

where

$$\begin{aligned} g_1(x) &= (1-b_0)\phi(\check{C}_1)\left\{\frac{\Phi(C_{10})}{\Phi(\check{C}_1)} - \frac{\Phi(C_{20})-\Phi(C_{10})}{\Phi(x)-\Phi(\check{C}_1)}\right\} \\ &\quad + b_0\alpha_1\phi(\theta^T\check{D}_1)\left\{\frac{\Phi(\theta^T D_{10})}{\Phi(\theta^T\check{D}_1)} - \frac{\Phi(\theta^T D_{20})-\Phi(\theta^T D_{10})}{\Phi(\theta^T\tilde{x})-\Phi(\theta^T\check{D}_1)}\right\}, \end{aligned}$$

and $D_r = (-1, C_r)^T$. Thus by (24) and (25), we have $g_1(\check{C}_2) = o_p(1)$. In addition, $g_1(C_2)$ is an increasing function of C_2 and $g_1(C_{20}) < 0$ since $\check{C}_1 > C_{10}$. Hence

$$\check{C}_2 > C_{20},$$

and (23) holds for $r = 2$.

The step j consists of showing that the conclusion (23) for $r = j+1$ is true given that the conclusions from the step $1, \cdots, j$. Using the same argument as before with $r = 2$, we can prove (23) hold for $r = j+1$ given $\check{\Phi}_r > \Phi(C_{r0}), r = 2, \cdots, j$. Hence (23) hold for $r \le I_n$.

Using the same argument as before with $\check{C}_1 = C_{10} + \varepsilon$, we can obtain the first part of Conclusion 3.

References

Alonzo, T.A. and Pepe, M.S.(2002). Distribution-free ROC analysis using binary regression techniques. *Biostatistics*, **3**, 421–432.

Cai, T. and Moskowitz, C.S.(2004). Semi-parametric estimation of the binormal ROC curve for a continuous diagnostic test. *Biostatistics*, **5**, 573–586.

J.A. Hanley. (1988). The robustness of the "binormal" assumptions used in fitting ROC curves. *Medical Decision Making*, **8**, 197-203.

Hsieh, F. and Turnbull, B.W.(1996). Nonparametric estimation of the receiver operating characteristic curve. *Ann. Statis.*, **24**, 25-40.

Kvam, P. H. and Samaniego, F. J.(1994), Nonparametric maximum likelihood estimation based on ranked set samples. *Journal of the American Statistical Association*,**89**, 526-537.

Metz, C.E., Herman, B.A. and Shen, J.H.(1998). Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data. *Statistics in Medicine*, **17**, 1033-1053.

Murphy, S.A. and Van der Vaart, A.W. (2000). On profile likelihood. *Journal of the American Statistical Association*, **95**, 449-465.

Pepe, M.S. (2003). The Statistical Evaluation of Medical tests for Classification and Prediction. *Oxford University Press*, Oxford, UK.

Pepe, M.S and Cai,T.(2004) The analysis of placement values for evaluating discriminatory measures.

Zou, K.H. and Hall, W.J.(2000). Two transformation models for estimating an ROC curve derived from continuous data. *J.Appl.Statist.*, **27**, 621-31.

Walsh S.J.(1999). Goodness-of-fit issues in ROC curve estimation. *Med Decis Making*, **19(2)**, 193-201.

Wieand, S., Gail, M.H., James, B.R., James,K.L. (1989). A family of nonparametric statistics for comparing diagnostic markers with paired or unpaired data. *Biometrika*, **76**,585-592.

Zhou, X. H., Obuchowski, N. A., and McClish, D. K. (2002). Statistical Methods in Diagnostic Medicine. *Wiley & Sons*, New York, USA.