Biostatistics (2004), **5**, 4, *pp.* 573–586
doi: 10.1093/biostatistics/kxh009

Semi-parametric estimation of the binormal ROC curve for a continuous diagnostic test

TIANXI CAI*
Department of Biostatistics, Harvard University, Boston, MA 02115, USA
tcai@hsph.harvard.edu

CHAYA S. MOSKOWITZ
Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY 10021, USA

SUMMARY

Not until recently has much attention been given to deriving maximum likelihood methods for estimating the intercept and slope parameters from a binormal ROC curve that assesses the accuracy of a continuous diagnostic test. We propose two new methods for estimating these parameters. The first method uses the profile likelihood and a simple algorithm to produce fully efficient estimates. The second method is based on a pseudo-maximum likelihood that can easily accommodate adjusting for covariates that could affect the accuracy of the continuous test.

Keywords: Classification; Diagnostic accuracy; Maximum likelihood; Screening; Semi-parametric transformation model.

1. INTRODUCTION

New medical technologies promise a vast array of tools for diagnosis and screening. In recent years there has been much focus on studying biomarkers that could potentially provide non-invasive and accurate ways of detecting disease, predicting disease progression, and evaluating patients' response to treatment. Tools such as gene expression profiling and protein mass spectrometry provide a plethora of possible biomarkers that must be studied further to assess their accuracy. Many of these biomarkers are measured on a continuous scale. A standard statistical tool for evaluating the accuracy of a continuous diagnostic test is a Receiver Operating Characteristic (ROC) curve (Swets and Pickett, 1982; Hanley, 1989; Begg, 1991).

The ROC curve evaluates the accuracy of diagnostic tests in separating two populations, for example patients with a particular disease ($D = 1$) and patients without the disease ($D = 0$). Let Y be a random variable that denotes the outcome of a diagnostic test or biomarker with the convention that higher values of Y are more indicative of disease. We use Y_{D} and $Y_{\bar{\mathrm{D}}}$ to denote the test result for disease and non-diseased subjects, respectively. The ROC curve is motivated as follows: if a threshold value c is used to classify subjects as diseased or non-diseased on the basis of Y, then the true and false positive rates associated with this classification criterion are respectively $S_D(c) = \Pr(Y \geqslant c \mid D = 1)$ and $S_{\bar{D}}(c) = \Pr(Y \geqslant c \mid D = 0)$. The ROC curve is a plot of the true positive rates versus the false positive rates across all

*To whom correspondence should be addressed.

possible thresholds. Therefore, at an attainable false positive rate u, the corresponding true positive rate is $\text{ROC}(u) = S_D\{S_{\bar{D}}^{-1}(u)\}$. The ROC curve displays the range of possible trade-offs between the true and false positive rates. The choice of the threshold value can be suggested based on the ROC curve, for example, by optimizing a cost or utility function that combines the $\{u, \text{ROC}(u)\}$ values.

There are a number of methods for estimating the ROC curve for a continuous test. A fully parametric approach that results in a smooth curve models the constituent distribution functions parametrically in order to arrive at the induced estimator of the ROC curve. A non-parametric method that results in a step function is to use the empirical estimate whose properties have been derived by Hsieh and Turnbull (1996). An intermediate strategy between these two is a semi-parametric approach.

Several different methods have been used to produce a semi-parametric ROC curve. For example, Li *et al.* (1999) propose using a non-parametric approach to estimate the distribution of test results in non-diseased subjects ($S_{\bar{D}}(c)$), but then assume a parametric model for the distribution of test results in diseased subjects ($S_D(c)$). Without assuming a functional relationship between these two distributions, they use maximum likelihood to estimate the unknown parameters in $S_D(c)$. Qin and Zhang (2003) also assume that the distribution of test results in non-diseased subjects is unknown, but instead assume a functional form for the density (likelihood) ratio function relating the two distributions to each other. Rather than modelling these distribution functions, their semi-parametric approach involves modelling the probability of disease status conditional on the test result.

A more commonly taken strategy to semi-parametric estimation of the ROC curve is to model the ROC curve parametrically, but avoid making additional assumptions about the distribution of test results. These types of approaches have also been calling parametric distribution-free (Pepe, 2000; Alonzo and Pepe, 2002). They produce smooth estimated curves while requiring less stringent assumptions than a fully parametric approach. The binormal ROC curve is perhaps the most popular of these intermediate strategies.

A binormal ROC curve for Y_{D} and $Y_{\bar{\text{D}}}$ assumes that for some (unknown) strictly increasing transformation h, $h(Y_{\text{D}})$ and $h(Y_{\bar{\text{D}}})$ have normal distributions. The binormal ROC model is written as

$$\text{ROC}(u) = \Phi\left\{a + b\Phi^{-1}(u)\right\}, \tag{1.1}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. This ROC curve plays a central role as a classic model in ROC analysis similar to the way that the normal distribution is a classic model for distribution functions (Pepe, 2003). Swets (1986) and Hanley (1996) conclude that it provides a good approximation to a wide range of ROC curves that occur in practice. Further, it has been used extensively in applied research as a simple tool to describe the accuracy of rating data in radiology and psychometric research, to compare tumour markers for various types of cancer, and to compare laboratory blood tests for the screening of prostate cancer. In this paper we consider an example evaluating serum concentration of CA-125, a cancer antigen, as a biomarker for pancreatic cancer in a Mayo Clinic study (Wieand *et al.*, 1989).

This paper focuses on estimation of the binormal ROC curve which involves estimating the parameters a and b in model (1.1). Maximum likelihood (ML) methods for fitting a binormal ROC curve to ordinal test results have long been available (for example, Ogilvie and Creelman, 1968 and Dorfman and Alf, 1969), and an algorithm developed by Dorfman and Alf (1969) for this purpose appears to be a standard way of estimating the binormal ROC curve for ordinal data. Not until somewhat recently, however, have ML methods for continuous test results received much attention. Zhou *et al.* (2002) and more recently Pepe (2003) provide reviews of the existing methods, so we only briefly mention some of these proposed methods below.

Metz *et al.* (1998) suggested that by ranking the continuous data and then arranging them into categories based on truth state runs (that is, the categories are based on the values that result in horizontal

or vertical jumps in an empirical ROC curve), we can obtain the ML estimates of the binormal ROC curve by using the Dorfman and Alf algorithm. There is disagreement, however, on whether this approach provides true ML estimates because the number of parameters that need to be estimated increases as the sample size increases (Metz *et al.*, 1998 and Pepe, 2003). Furthermore, the argument in favour of the validity of these estimates as being true ML estimates is based on using all truth state runs as categories, which can be computationally intensive as Metz *et al.* (1998) point out.

Pepe (2000) followed by Alonzo and Pepe (2002) suggest two similar semi-parametric approaches in the context of developing ROC regression methodology. The binormal ROC curve is included in the class of models they consider. While their approaches do have some nice features, including the fact that they facilitate adjusting the ROC curve for covariates that affect the accuracy of the test and the estimates obtained do appear to be efficient, they are not ML estimates.

Finally, Zou and Hall (2000) develop ML rank-based estimates by ranking the data and numerically solving the score equations derived from the likelihood function of the order statistics using a Monte Carlo procedure. Unfortunately, software is not readily available to implement this method. Furthermore, while this approach does result in fully efficient estimates, it is computationally intensive making it difficult to implement. We expect that even if software were available, the computational time for executing their method would be relatively long. Simply evaluating the score function requires Monte Carlo approximation giving some indication as to the time required for the procedure to yield estimates.

Here, we propose two new approaches to the estimation of the location and scale parameters in the binormal ROC model. The first is a maximum profile likelihood approach which also provides a fully efficient estimator of $\boldsymbol{\theta} = (a, b)^{\mathsf{T}}$. To obtain the estimator numerically, we propose a simple algorithm based on the *method of freezing coefficients* (Golub and Van Loan, 1989). We find that this algorithm is easy to implement and works well in practice. The second approach is a *pseudo* maximum likelihood approach which provides a useful alternative to the first approach with the benefit of easy extension to incorporate covariates.

The paper is organized as follows. In Section 2, we describe the maximum profile likelihood method for estimating $\boldsymbol{\theta}$. The pseudo-maximum likelihood approach and the corresponding procedures for making statistical inference are outlined in Section 3. A graphical method together with a goodness-of-fit statistic for model checking is given in Section 4. Simulation studies were performed to (i) determine if the new estimators have increased efficiency relative to existing methods and (ii) verify the validity of the asymptotic inference in finite samples. Results of these simulation studies are summarized in Section 5. An illustrative example is presented in Section 6 using data from a study to evaluate antigens CA-125 and CA 19-9 as biomarkers for the detection of pancreatic cancer. Some closing remarks and extensions to incorporate covariates are made in Section 7.

2. Maximum profile likelihood estimator

Let $\{Y_i : i = 1, \ldots, n_{\text{D}}\}$ denote n_{D} observations from the diseased population and $\{Y_j : j = n_{\text{D}} + 1, \ldots, n_{\bar{\text{D}}} + n_{\text{D}}\}$ be a random sample from the disease-free population. Let f_{D} and $f_{\bar{\text{D}}}$ denote the respective density functions of Y_{D} and $Y_{\bar{\text{D}}}$ and let $\Lambda_{\bar{\text{D}}} = -\log S_{\bar{\text{D}}}$ be the cumulative hazard function of $Y_{\bar{\text{D}}}$.

Model (1.1) implies that there exists some (unknown) monotonic increasing transformation $h(\cdot)$ such that

$$h(Y_i) \sim N(\alpha, \beta^2) \qquad \text{and} \qquad h(Y_j) \sim N(0, 1), \tag{2.1}$$

for $i = 1, \ldots, n_{\text{D}}$ and $j = n_{\text{D}} + 1, \ldots, n$, where $\alpha = a/b$, $\beta = 1/b$ and $n = n_{\bar{\text{D}}} + n_{\text{D}}$. Note that estimating a and b directly from the mean and variance in (2.1) will produce valid estimates only if the test results are truly normally distributed, rather than the weaker assumption that some transformation of the test results

is normally distributed. Model (1.1) implies that

$$S_{\mathrm{D}}(y) = \Phi\left[a + b\Phi^{-1}\{S_{\bar{\mathrm{D}}}(y)\}\right], \quad \text{and} \quad \frac{f_{\mathrm{D}}(y)}{f_{\bar{\mathrm{D}}}(y)} = \frac{b\phi\{a - bh(y)\}}{\phi\{-h(y)\}},$$

where ϕ is the probability density function for the standard normal distribution. It follows that the likelihood of the data is

$$\mathcal{L}(\boldsymbol{\theta}, \Lambda_{\bar{\mathrm{D}}}) = \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \prod_{i=1}^{n_{\mathrm{D}}} \frac{b\phi(a - bh_i)\lambda_i \mathrm{e}^{-\Lambda_i}}{\phi(-h_i)} \times \prod_{j=n_{\mathrm{D}}+1}^{n} \lambda_j \mathrm{e}^{-\Lambda_j},$$

where $\boldsymbol{\theta} = (a, b)^{\mathsf{T}}$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)^{\mathsf{T}}$, for $l = 1, \ldots, n$, $\lambda_l = \Delta\Lambda_{\bar{\mathrm{D}}}(Y_l)$ is the jump of $\Lambda_{\bar{\mathrm{D}}}$ at Y_l if $\Lambda_{\bar{\mathrm{D}}}$ is discrete and λ_l is the derivative of $\Lambda_{\bar{\mathrm{D}}}$ at Y_l if $\Lambda_{\bar{\mathrm{D}}}$ is continuous, $\Lambda_l = \Lambda_{\bar{\mathrm{D}}}(Y_l)$ and $h_l = h(Y_l) = -\Phi^{-1}(\mathrm{e}^{-\Lambda_l})$. If we restrict the estimate of $\Lambda_{\bar{\mathrm{D}}}$ to the subspace of absolutely continuous estimates of the cumulative hazard function, then there is no maximizer of the likelihood. By restricting the estimate of $\Lambda_{\bar{\mathrm{D}}}$ to the subspace of discrete estimates, however, a maximizer does exist. The ML estimator of $\Lambda_{\bar{\mathrm{D}}}$, $\widehat{\Lambda}_{\bar{\mathrm{D}}}$, will then be a nondecreasing step function with jumps at the observed data points $\{Y_l, l = 1, \ldots, n\}$. See Murphy *et al.* (1997) for a more detailed explanation of this point.

With $\widehat{\boldsymbol{\theta}}_{\mathrm{mle}}$ and $\widehat{\boldsymbol{\lambda}}$ denoting the ML estimators of $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\lambda})$, the likelihood depends smoothly on the unknowns $\boldsymbol{\lambda}$. For any given $\boldsymbol{\lambda}$, maximizing $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\lambda})$ with respect to $\boldsymbol{\theta}$, we obtain a closed form maximizer:

$$\widehat{\boldsymbol{\theta}}(\mathbf{h}) = \begin{pmatrix} \widehat{a}(\mathbf{h}) \\ \widehat{b}(\mathbf{h}) \end{pmatrix} = \begin{pmatrix} \frac{\widehat{\alpha}(\mathbf{h})}{\widehat{\beta}(\mathbf{h})} \\ \frac{1}{\widehat{\beta}(\mathbf{h})} \end{pmatrix}$$

where $\widehat{\alpha}(\mathbf{h}) = \sum_{i=1}^{n_{\mathrm{D}}} h_i / n_{\mathrm{D}}$, $\widehat{\beta}(\mathbf{h}) = \sqrt{\sum_{i=1}^{n_{\mathrm{D}}} \{h_i - \widehat{\alpha}(\mathbf{h})\}^2 / n_{\mathrm{D}}}$ and $\mathbf{h} = (h_1, \ldots, h_{n_{\mathrm{D}}})^{\mathsf{T}}$.

The profile likelihood for $\boldsymbol{\theta}$ is given by $\mathrm{PL}(\boldsymbol{\theta}) = \mathcal{L}\left\{\boldsymbol{\theta}, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\theta})\right\}$, where $\widehat{\boldsymbol{\lambda}}(\boldsymbol{\theta})$ maximizes the likelihood for a fixed $\boldsymbol{\theta}$. The maximum profile likelihood estimator, $\widehat{\boldsymbol{\theta}}_{\mathrm{mle}}$, maximizes $\mathrm{PL}(\boldsymbol{\theta})$. This estimator is a function of the biomarker values only through their ranks. The profile likelihood is the same, whether we use the Y, or replace them by their ranks. Let $\boldsymbol{\theta}_0 = (a_0, b_0)^{\mathsf{T}}$ and $\boldsymbol{\lambda}_0$ denote the true values of $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$, respectively. It follows from the properties of maximum profile likelihood estimators (Murphy and van der Vaart, 2000) that $\widehat{\boldsymbol{\theta}}_{\mathrm{mle}}$ is fully efficient and $n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}_{\mathrm{mle}} - \boldsymbol{\theta}_0)$ converges in distribution to a zero-mean bivariate normal with covariance matrix $\boldsymbol{\Sigma}_{\mathrm{mle}}$, where

$$\boldsymbol{\Sigma}_{\mathrm{mle}} = \left\{ -E\left(\frac{\partial^2 \log \mathcal{L}}{n\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathsf{T}}}\right) + E\left(\frac{\partial^2 \log \mathcal{L}}{n\partial\boldsymbol{\theta}\partial\boldsymbol{\lambda}^{\mathsf{T}}}\right) E\left(\frac{\partial^2 \log \mathcal{L}}{n\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}^{\mathsf{T}}}\right)^{-1} E\left(\frac{\partial^2 \log \mathcal{L}}{n\partial\boldsymbol{\lambda}\partial\boldsymbol{\theta}^{\mathsf{T}}}\right) \right\}^{-1} \Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0, \boldsymbol{\lambda}=\boldsymbol{\lambda}_0}.$$

It also follows from Murphy and van der Vaart (2000) that

$$\mathrm{LR}(\boldsymbol{\theta}_0) \equiv 2\log\frac{\mathrm{PL}(\widehat{\boldsymbol{\theta}}_{\mathrm{mle}})}{\mathrm{PL}(\boldsymbol{\theta}_0)} \sim \chi_2^2, \quad \mathrm{LR}_1(a_0) \equiv 2\log\frac{\mathrm{PL}(\widehat{\boldsymbol{\theta}}_{\mathrm{mle}})}{\mathrm{PL}_1(a_0)} \sim \chi_1^2, \quad \mathrm{LR}_2(b_0) \equiv 2\log\frac{\mathrm{PL}(\widehat{\boldsymbol{\theta}}_{\mathrm{mle}})}{\mathrm{PL}_2(b_0)} \sim \chi_1^2,$$

where $\mathrm{PL}_1(a) = \max_b \mathrm{PL}(a, b)$, $\mathrm{PL}_2(b) = \max_a \mathrm{PL}(a, b)$ and χ_k^2 denotes a chi-square distribution with k degrees of freedom. Therefore, $100(1-\gamma)\%$ confidence intervals for the kth component of $\boldsymbol{\theta}_0$ can be obtained based on the Wald-type confidence interval: $\{\theta_k : (\theta_k - \widehat{\theta}_k)^2 \leqslant \sigma_k^2 \chi_{1,1-\gamma}^2\}$ or by inverting the likelihood ratio statistic: $\{\theta_k : \mathrm{LR}_k(\theta_k) \leqslant \chi_{1,1-\gamma}^2\}$, where $\chi_{k,1-\gamma}^2$ is the $100(1-\gamma)$th percentile of χ_k^2, $\widehat{\theta}_k$ is the kth element of $\widehat{\boldsymbol{\theta}}_{\mathrm{mle}}$ and σ_k^2 is the (k, k)th element of $\boldsymbol{\Sigma}_{\mathrm{mle}}$. A joint $100(1-\gamma)$ confidence region for $\boldsymbol{\theta}$ can be obtained as $\left\{\boldsymbol{\theta} : (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})^{\mathsf{T}} \boldsymbol{\Sigma}_{\mathrm{mle}}^{-1} (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}) \leqslant \chi_{2,1-\gamma}^2\right\}$ or $\left\{\boldsymbol{\theta} : \mathrm{LR}(\boldsymbol{\theta}) \leqslant \chi_{2,1-\gamma}^2\right\}$.

In general, with presence of infinite-dimensional parameters, ML estimators are difficult to obtain numerically. Nevertheless, we find that in this particular setting, a simple algorithm works well in practice. We adopt a commonly used method in nonlinear numerical analysis: namely, the *method of freezing coefficients*. Details of the implementation procedure are given in Appendix A, and the code needed to compute these ML estimates is available upon request. This algorithm is a computationally non-intensive way to obtain the described ML estimates.

3. Pseudo maximum likelihood estimator (PMLE)

An alternative approach to the estimation of $\boldsymbol{\theta}$ is through a *pseudo* ML estimation procedure. If h is given, we can obtain the ML estimator for $\boldsymbol{\theta}$ by solving the score equations for $\boldsymbol{\theta}$ in (A.2) of Appendix A. To estimate the unknown transformation h, let $\delta_i(y) = I(Y_i > y) + \frac{1}{2}I(Y_i = y)$ and $\widehat{S}(y) = n^{-1}\sum_{i=1}^{n}\delta_i(y)$. It follows from (2.1) that the expected value of $\widehat{S}(y)$ is $S\{h_0(y), \boldsymbol{\theta}_0\}$, where $S(h, \boldsymbol{\theta}) = p_{\text{D}}\Phi(a - bh) + p_{\bar{\text{D}}}\Phi(-h)$, h_0 is the true underlying transformation function, $p_{\bar{\text{D}}} = n_{\bar{\text{D}}}/n$ and $p_{\text{D}} = n_{\text{D}}/n$. This motivates the following estimating equation for $h(y)$ at any given $\boldsymbol{\theta}$:

$$p_{\text{D}}\Phi(a - bh_y) + p_{\bar{\text{D}}}\Phi(-h_y) = \widehat{S}(y). \tag{3.1}$$

Note that we estimate h with data from both diseased and non-diseased subjects whereas Alonzo and Pepe (2002) estimate h with data from non-diseased subjects only.

Let $\widehat{h}(y, \boldsymbol{\theta})$ denote the solution to (3.1) for a given $\boldsymbol{\theta}$. Plugging $\widehat{h}$ to the score equation for $\boldsymbol{\theta}$, $\frac{\partial\mathcal{L}(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\theta}} = 0$, we obtain an estimating equation for $\boldsymbol{\theta}_0$:

$$\sum_{i=1}^{n_{\text{D}}}\begin{pmatrix} -a + b\widehat{h}(Y_i, \boldsymbol{\theta}) \\ b^{-1} + a\widehat{h}(Y_i, \boldsymbol{\theta}) - b\widehat{h}(Y_i, \boldsymbol{\theta})^2 \end{pmatrix} = \mathbf{0}. \tag{3.2}$$

Let $\widehat{\boldsymbol{\theta}}_{\text{pmle}}$ denote the resulting estimator for $\boldsymbol{\theta}_0$. Note that for any given $\boldsymbol{\theta}$, the Jacobian matrix for (3.1) is diagonal. $\widehat{\boldsymbol{\theta}}_{\text{pmle}}$ can easily be obtained either by solving (3.1) for $\mathbf{h} = (h_1, \ldots, h_{n_{\text{D}}})^{\mathsf{T}}$ and solving (3.2) for $\boldsymbol{\theta}$ iteratively or by solving for them simultaneously using the Newton–Raphson algorithm.

To make inference about $\boldsymbol{\theta}_0$ based on the PMLE, we show in Appendix B that $\widehat{\boldsymbol{\theta}}_{\text{pmle}}$ is consistent and the distribution of $n_{\text{D}}^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}_{\text{pmle}} - \boldsymbol{\theta}_0)$ can be approximated by a zero-mean normal random vector with covariance matrix $\boldsymbol{\Sigma}_{\text{pmle}}$, where

$$\boldsymbol{\Sigma}_{\text{pmle}} = \mathbb{A}^{-1}\left\{n_{\text{D}}^{-1}\sum_{i=1}^{n_{\text{D}}}(\mathbf{v}_{1i} + \mathbf{v}_{2i})(\mathbf{v}_{1i} + \mathbf{v}_{2i})^{\mathsf{T}} + n_{\text{D}}^{-1}\sum_{j=n_{\text{D}}+1}^{n_{\text{D}}+n_{\bar{\text{D}}}}\mathbf{v}_{2j}\mathbf{v}_{2j}^{\mathsf{T}}\right\}\mathbb{A}^{-1}, \tag{3.3}$$

$$\mathbf{v}_{1i} = \begin{pmatrix} -a_0 + b_0h_0(Y_{\text{D}i}) \\ b_0^{-1} + a_0h_0(Y_{\text{D}i}) - b_0h_0^2(Y_{\text{D}i}) \end{pmatrix} \qquad \mathbf{v}_{2i} = -\int\begin{pmatrix} b_0 \\ a_0 + 2b_0h_0(y) \end{pmatrix}\frac{p_{\text{D}}e_i(y)\,\text{d}S_{\text{D}}(y)}{\dot{s}_h\{h_0(y), \boldsymbol{\theta}_0\}},$$

$e_i(y) = \delta_i(y) - S\{h_0(y), \boldsymbol{\theta}_0\}$, $\dot{s}_h$ is the limit of $\partial S(h, \boldsymbol{\theta})/\partial h$ and $\mathbb{A}$ is defined in Appendix B. A consistent estimate of $\boldsymbol{\Sigma}_{\text{pmle}}$ can be obtained by replacing all the theoretical quantities in $\boldsymbol{\Sigma}_{\text{pmle}}$ by their empirical counterparts.

4. Model checking procedures

The inference procedures proposed in the previous sections are valid only if the binormal ROC model (1.1) is correctly specified. Here, we present a graphical method as well as a statistical test to assess

whether model (1.1) is appropriate for the data in hand. Specifically, we consider the residual process

$$\widehat{Q}(y) \equiv \Phi^{-1}\left\{\widehat{S}_{\text{D}}(y)\right\} - \widehat{a} - \widehat{b}\Phi^{-1}\left\{\widehat{S}_{\bar{\text{D}}}(y)\right\},$$

where $\widehat{S}_{\bar{\text{D}}}(y) = n_{\bar{\text{D}}}^{-1}\sum_{j=n_{\text{D}}+1}^{n_{\text{D}}+n_{\bar{\text{D}}}}\delta_j(y)$ and $\widehat{S}_{\text{D}}(y) = n_{\text{D}}^{-1}\sum_{i=1}^{n_{\text{D}}}\delta_i(y)$. This process is motivated by an equivalent form of the ROC model: $\Phi^{-1}\{S_{\text{D}}(y)\} = a + b\Phi^{-1}(S_{\bar{\text{D}}}(y))$.

We show in Appendix C that under H_0, $\widehat{Q}(y) \to 0$ almost surely, uniformly in y, and $n_{\text{D}}^{\frac{1}{2}}\widehat{Q}(y)$ is asymptotically equivalent to $\widetilde{Q}(y) \equiv n_{\text{D}}^{-\frac{1}{2}}\sum_{i=1}^{n}O(y, Y_i)$, where $O(y, Y_i)$ is defined in Appendix C. The asymptotic equivalence between $\widehat{Q}(\cdot)$ and $\widetilde{Q}(\cdot)$ allows us to approximate the limiting distribution of $\widehat{Q}(\cdot)$ using the re-sampling techniques (Parzen *et al.*, 1994) in practice. In essence, one can simulate independent random samples $\mathcal{Z} = \{\mathcal{Z}_l, l = 1, \ldots, n\}$ from the standard normal distribution, and for each set of $\mathcal{Z}$, obtain

$$\widehat{Q}_{\mathcal{Z}}(y) = n_{\text{D}}^{-\frac{1}{2}}\sum_{l=1}^{n}\widehat{O}(y, Y_l)\mathcal{Z}_l$$

where $\widehat{O}(y, Y_l)$ is obtained by replacing all the theoretical quantities in $O(y, Y_l)$ by their empirical counterparts. It is also shown in Appendix C that $\widehat{Q}_{\mathcal{Z}}(\cdot)$ converges weakly to the same limiting process as that of $\widehat{Q}(\cdot)$.

To assess how unusual the observed process $\widehat{Q}(\cdot)$ is under H_0, one may plot the observed process along with a few realizations from $\widehat{Q}_{\mathcal{Z}}(\cdot)$ and supplement the graphical display with an estimated p-value from a supremum-type test statistic $q \equiv \sup_y|\widehat{Q}(y)|$. We note that the test based on q is consistent against the alternative that (1.1) does not hold. An unusually large observed value q would suggest improper specification of (1.1). In practice, the p-value can be approximated by $P(\widehat{q} \geqslant q)$, where $\widehat{q} = \sup_y|\widehat{Q}_{\mathcal{Z}}(y)|$. We estimate $P(\widehat{q} \geqslant q)$ by generating a large number, say 5000, of realizations from $\widehat{Q}_{\mathcal{Z}}(\cdot)$.

5. SIMULATION STUDIES

5.1 *Relative efficiency*

To investigate the efficiency of the new estimators relative to the existing estimators, we simulated data from the same setting as given in Zou and Hall (2000). Specifically, Y is generated from a standard normal for the disease-free population and for the diseased population, Y follows a normal distribution with mean $\alpha = 1.868$ and standard error $\beta = 1.5$. The induced ROC curve is

$$\text{ROC}(u) = \Phi\{1.245 + 0.667\Phi^{-1}(u)\}.$$

The area under this ROC curve is AUC $= 0.85$.

We obtain both our proposed ML estimator and PMLE of $\boldsymbol{\theta} = (a, b)^{\text{T}}$. For comparison, the LABROC procedure described in Metz *et al.* (1998) and the Alonzo and Pepe (2002) binary regression approach are also implemented. The Zou and Hall approach is not included here for the reasons explained in the introduction. We expect their approach to yield similar results to those seen for our proposed approach.

In Table 1, we present the empirical bias and mean square error of $\hat{\boldsymbol{\theta}}$ based on 1000 simulated datasets. All methods yield estimators with little bias, at least for sample size $n_{\text{D}} = n_{\bar{\text{D}}} > 50$. The GLM and LABROC estimators do appear to have less bias than the ML estimate and PMLE. Interestingly, the ML estimates of α and β (the mean and standard deviation in (2)) are less biased than the GLM and LABROC estimators (results not shown). It is possible that the re-parametrization from α and β to a and b leads to additional bias. The increase in bias is most noticeable for smaller sample sizes, but is minimal in all the settings we considered.

Table 1. *Estimates of a and b compared with their actual values, $a = 1.245$ and $b = 0.667$, based on four different approaches: maximum likelihood (MLE), pseudo-maximum likelihood (PMLE), Alonzo and Pepe's binary regression approach (GLM) and Metz's LABROC (LAB). The results are based on 1000 simulated data sets*

$n_{\bar{D}}$	n_D		Bias				Mean square error $\times 10$			
			MLE	PMLE	GLM	LAB	MLE	PMLE	GLM	LAB
50	50	a	0.079	0.064	0.034	0.032	0.633	0.641	0.586	0.589
		b	0.043	0.062	−0.009	0.002	0.215	0.221	0.223	0.215
200	200	a	0.026	0.018	0.013	0.002	0.125	0.138	0.126	0.128
		b	0.014	0.017	−0.004	0.001	0.048	0.051	0.052	0.049
1000	500	a	0.005	0.004	0.001	−0.003	0.047	0.048	0.049	0.047
		b	0.005	0.007	0.002	−0.004	0.012	0.013	0.015	0.014
1000	1000	a	0.004	0.003	0.002	−0.001	0.024	0.025	0.025	0.028
		b	0.004	0.004	0.001	−0.002	0.008	0.009	0.010	0.011

The ML estimate and the PMLE are more efficient than the other estimators, however, and this increased efficiency yields MSEs that are comparable to or smaller than the GLM and LABROC estimators. The ML estimate and the LABROC achieve similar efficiencies in small samples. But as the sample size increases, the LABROC estimator has a reduced efficiency relative to the ML estimate in estimating $\boldsymbol{\theta}$. For example, when $n_D = n_{\bar{D}} = 1000$, the efficiency of the LABROC estimator relative to the ML estimate is 85% for a and 76% for b. The PMLE has efficiencies comparable to the ML estimate. The binary regression approach is slightly less efficient. Relative to the ML estimate, the efficiency of the PMLE is about 98% for both a and b and the efficiency of the binary regression approach is about 96% for a and 80% for b.

5.2 *Asymptotic inference in finite samples*

We also conducted simulation studies to examine the validity of the large sample approximations for making inference in finite samples. We simulated 1000 sets of data with $(n_{\bar{D}}, n_D) = (50, 50)$, $(100, 50)$, and $(200, 200)$ from the same model as described in Section 5.1. In Table 2, we present the bias, the sampling standard error, the average of the standard error estimators and the coverage probability of the 95% confidence intervals for a and b (a log transformation was used when obtaining the confidence intervals for b). The standard error estimators are reasonably close to the true sampling standard errors, at least for sample size $n > 100$. For small samples, $n_D = n_{\bar{D}} = 50$, it appears that for the PMLE approach, the estimated standard errors based on large sample approximations tend to be slightly larger than the sampling standard errors when estimating b. Nevertheless, the empirical coverage probabilities for the confidence intervals are close to their nominal counterparts, even in small samples. When the sample size is 50 for each group, the empirical coverage probabilities for a and b are 95.3% and 96.4% based on the ML estimate and are 97.4% and 96.8% based on the PMLE.

6. EXAMPLE: PANCREATIC CANCER BIOMARKER

As reported by Wieand *et al.* (1989), two antigens, CA-125 and CA 19-9, were studied at the Mayo Clinic as possible biomarkers of pancreatic cancer. These two biomarkers were measured in the sera of

Table 2. *Bias, sampling standard error (Sampling SE), average of the estimated standard error estimator (Ave($\widehat{SE}$)), and the coverage probability (Coverage) of the 95% confidence interval. Each entry is based on 1000 simulation samples*

$n_{\bar{D}}$	n_D		Bias		Sampling SE		Ave($\widehat{SE}$)		Coverage	
			MLE	PMLE	MLE	PMLE	MLE	PMLE	MLE	PMLE
50	50	a	0.079	0.064	0.239	0.245	0.234	0.248	0.953	0.974
		b	0.043	0.062	0.140	0.135	0.149	0.193	0.964	0.968
100	50	a	0.062	0.063	0.227	0.246	0.218	0.248	0.957	0.963
		b	0.033	0.054	0.129	0.126	0.126	0.150	0.945	0.941
200	200	a	0.026	0.018	0.109	0.116	0.111	0.113	0.950	0.945
		b	0.014	0.017	0.068	0.070	0.067	0.073	0.943	0.949

Table 3. *Estimates (standard errors) of parameters a and b in the binormal ROC model for CA-125*

	MLE	PMLE	Zou and Hall	GLM	LABROC
a	0.761(0.191)	0.719(0.198)	0.727(0.190)	0.778(0.197)	0.720(0.185)
b	1.065(0.140)	1.020(0.148)	1.007(0.142)	1.017(0.167)	1.002(0.137)

$n_D = 90$ patients with pancreatic cancer and $n_{\bar{D}} = 51$ healthy patients with pancreatitis. We consider fitting a binormal ROC curve to the data for the CA-125 antigen.

Table 3 lists the resulting estimates of (a, b) and their standard errors based on our proposed ML estimator, the PMLE, the Zou and Hall (2000) estimator, Alonzo and Pepe (2002)'s binary regression approach and the LABROC procedure. All five methods give similar estimates of a and b. The ML estimator provides a slightly larger estimate of $\widehat{a} = 0.761$ (s.e. $= 0.191$) and $\widehat{b} = 1.065$ (s.e. $= 0.140$) compared to $\widehat{a} = 0.727$ (s.e. $= 0.190$) and $\widehat{b} = 1.007$ (s.e. $= 0.142$) based on the Zou and Hall (2000) estimator. The observed difference in these parameter estimates is not unexpected considering that the sample sizes are relatively small and these two methods deal with the nuisance parameter h differently even though they are both likelihood based.

The resulting estimates of the ROC curve and their 95% confidence bands based on the ML estimate and the PMLE are shown in Figure 1. While Figure 1 shows that the fitted binormal ROC curves overlap the empirically estimated one, there is a slight difference between the two. To investigate whether this discrepancy is simply due to randomness or to a true lack of fit of the binormal model (1.1), we apply the model checking procedure discussed in Section 4. As shown in Figure 2, the observed residual process $\widehat{Q}(\cdot)$ does not appear to be unusual when compared with the realizations of $\widehat{Q}_Z(\cdot)$ (using the PMLE of $\boldsymbol{\theta}$). Testing the hypothesis H_0 using the supreme-type statistic, we obtain a p-value of 0.78 if using the PMLE for $\boldsymbol{\theta}$ and 0.36 if using the MLE for $\boldsymbol{\theta}$ (based on 1000 realizations) indicating that the binormal model is a reasonable choice.

7. DISCUSSION AND EXTENSION

In this paper, we proposed two semi-parametric approaches to the estimation of the location and scale parameter in the binormal ROC model. The estimators of $\boldsymbol{\theta}$ produced by both approaches are invariant to monotonic increasing transformations. Simulation results suggest that the proposed pseudo ML method has efficiency comparable to the ML estimate in estimating $\boldsymbol{\theta}$. However, since the ML estimate is fully

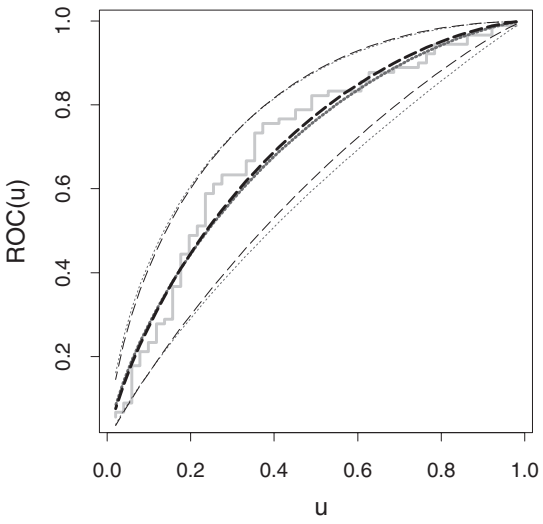

Fig. 1. Estimated ROC curves (thicker curves) and their 95% confidence intervals (thinner curves) based on the MLE (dashed curve) and PMLE (dotted curve) for CA-125. Shown also is the empirical estimate of the ROC curve (grey solid curve). (The upper 95% confidence interval curves for the MLE and PLME ROC curves lie directly on top of each other.)

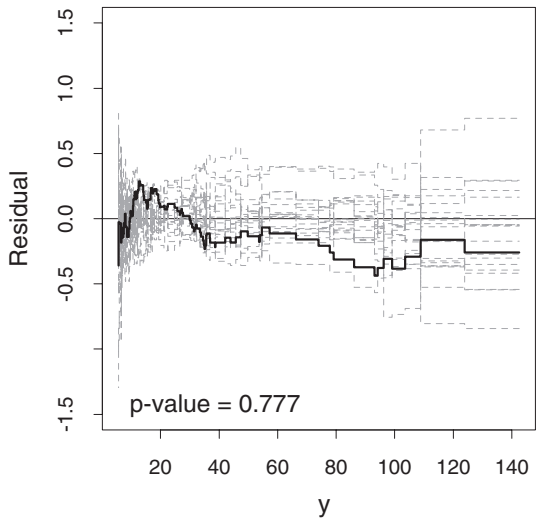

Fig. 2. Residual plot for testing the goodness of fit of the binormal ROC model. The observed pattern is shown by the thick solid curve, and 20 simulated realizations under the null are shown by the dotted curves.

efficient and is not hard to obtain numerically, we recommend using the ML estimate when fitting the binormal model without covariates.

To obtain the ML estimate numerically, we find that the algorithm based on the method of freezing coefficients performs well in practice. The program for the estimation procedure is currently written in *S-Plus* and the computation time with the current version of the software is about 4 seconds for a sample

size of $n_{\bar{\mathrm{D}}} = n_{\mathrm{D}} = 200$. Based on our experience, this algorithm rarely fails to converge. The PMLE can be obtained based on the Newton–Raphson algorithm. Note that, for this case, the inverse of the Jacobian matrix can easily be obtained by inverting a $n_{\mathrm{D}} \times n_{\mathrm{D}}$ diagonal matrix and a 2×2 matrix.

It is difficult to extend the maximum profile likelihood approach to allow for covariates, while the modifications needed to extend the pseudo ML approach to incorporate covariates are straightforward. In particular, the transformation models given in (2.1) can be generalized as follows:

$$h(Y_i) = \boldsymbol{\gamma}_{\mathrm{D}}^{\mathsf{T}} \mathbf{Z}_i + \alpha + \beta \epsilon_{\mathrm{D}i} \quad \text{and} \quad h(Y_j) = \boldsymbol{\gamma}_{\bar{\mathrm{D}}}^{\mathsf{T}} \mathbf{Z}_j + \epsilon_{\bar{\mathrm{D}}j} \tag{7.1}$$

where $\epsilon_{\mathrm{D}i}$ and $\epsilon_{\bar{\mathrm{D}}j}$ are standard normal random variables, $\mathbf{Z}_i$ and $\mathbf{Z}_j$ are covariates associated with Y_i and Y_j, respectively, for $i = 1, \ldots, n_{\mathrm{D}}$ and $j = n_{\mathrm{D}} + 1, \ldots, n$. The induced ROC curve associated with covariates $\mathbf{Z}$ is

$$\mathrm{ROC}_{\mathbf{Z}}(u) = \Phi\left\{a + b\Phi^{-1}(u) + \boldsymbol{\gamma}^{\mathsf{T}}\mathbf{Z}\right\}$$

where $a = \alpha/\beta$, $b = 1/\beta$ and $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_{\mathrm{D}} - \boldsymbol{\gamma}_{\bar{\mathrm{D}}})/\beta$. As usual the parameters α and β can be interpreted as the location and shape of the ROC curve. The parameter vector $\boldsymbol{\gamma}$ quantifies the covariate effect on the ROC curve by the difference between the covariate effect in the diseased and non-diseased populations (up to a scalar β).

Pepe (2000) incorporates covariates by modelling how the covariates in the diseased population and the covariates in the non-diseased population affect the ROC curve. In contrast, Alonzo and Pepe (2002) consider the comparison between the diseased and non-diseased populations with the same value for the covariate. Our approach is similar to the Pepe (2000) approach in that we also allow the comparison between the two populations when they are associated with different covariate levels. We estimate the covariate effects on the distribution of Y in the two populations to induce an estimate for the covariate effect on the ROC curve.

For a given $\mathbf{h} = (h_1, \ldots, h_n)^{\mathsf{T}}$, we estimate $\boldsymbol{\theta} = (\alpha, \boldsymbol{\gamma}_{\mathrm{D}}^{\mathsf{T}}, \boldsymbol{\gamma}_{\bar{\mathrm{D}}}, \beta)^{\mathsf{T}}$ by maximizing the likelihood of the data as a function of $\boldsymbol{\theta}$ and obtain

$$\widehat{\boldsymbol{\theta}}(\mathbf{h}) = \begin{pmatrix} \widehat{\alpha}(\mathbf{h}) \\ \widehat{\boldsymbol{\gamma}}_{\mathrm{D}}(\mathbf{h}) \\ \widehat{\boldsymbol{\gamma}}_{\bar{\mathrm{D}}}(\mathbf{h}) \\ \widehat{\beta}(\mathbf{h}) \end{pmatrix} = \begin{pmatrix} \left\{\sum_{i=1}^{n_{\mathrm{D}}} \vec{\mathbf{Z}}_i \vec{\mathbf{Z}}_i^{\mathsf{T}}\right\}^{-1} \sum_{i=1}^{n_{\mathrm{D}}} \vec{\mathbf{Z}}_i h_i \\ \left\{\sum_{j=n_{\mathrm{D}}+1}^{n_{\mathrm{D}}+n_{\bar{\mathrm{D}}}} \mathbf{Z}_j \mathbf{Z}_j^{\mathsf{T}}\right\}^{-1} \sum_{j=n_{\mathrm{D}}+1}^{n_{\mathrm{D}}+n_{\bar{\mathrm{D}}}} \mathbf{Z}_j h_j \\ n_{\mathrm{D}}^{-1} \sum_{i=1}^{n_{\mathrm{D}}} \left\{h_i - \widehat{\alpha}(h) - \widehat{\boldsymbol{\gamma}}_{\mathrm{D}}(h)^{\mathsf{T}} \mathbf{Z}_i\right\}^2 \end{pmatrix}, \tag{7.2}$$

where $\vec{\mathbf{Z}}_i = (1, \mathbf{Z}_i)^{\mathsf{T}}$, $h_i = h(Y_{\mathrm{D}i})$ for $i = 1, \ldots, n_{\mathrm{D}}$ and $h_j = h(Y_{\bar{\mathrm{D}}j})$ for $j = n_{\mathrm{D}} + 1, \ldots, n$. Analogous to (3.1), the following estimating equation can be used to estimate $h(y)$ for any given y:

$$n^{-1} \sum_{j=n_{\mathrm{D}}+1}^{n_{\mathrm{D}}+n_{\bar{\mathrm{D}}}} \Phi(-h_y + \boldsymbol{\gamma}_{\bar{\mathrm{D}}}^{\mathsf{T}} \mathbf{Z}_j) + n^{-1} \sum_{i=1}^{n_{\mathrm{D}}} \Phi\left(\frac{-h_y + \alpha + \boldsymbol{\gamma}_{\mathrm{D}}^{\mathsf{T}} \mathbf{Z}_i}{\beta}\right) = \widehat{S}(y). \tag{7.3}$$

The solution to (7.2) and (7.3) provides an estimator for $\boldsymbol{\theta}$ and hence the ROC curve. This approach utilizes information from both the diseased and non-diseased subjects for the estimation of h and therefore the FPR function. In contrast, the binary regression approach proposed by Pepe (2000) and Alonzo and Pepe (2002) estimates the FPR function based on observations from the non-diseased population only. Simulation results suggest that this approach has higher efficiency in estimating $\boldsymbol{\theta}$ compared to the binary regression approach. For example, in one setting, we generated data from models (7.1) with one continuous covariate Z from uniform(0, 1), $h(y) = y$ and we set $\alpha = 1.868$, $\beta = 1.5$, $\gamma_{\mathrm{D}} = 1$ and $\gamma_{\bar{\mathrm{D}}} = 0.5$. This configuration for the data induces ROC curve

$$\mathrm{ROC}_Z(u) = \Phi\left\{1.245 + 0.667\Phi^{-1}(u) + 0.333Z\right\}.$$

Table 4. *Estimates of a, b and γ compared with their actual values, $a = 1.245$, $b = 0.667$ and $\gamma = 0.333$, based on the pseudo-maximum likelihood approach (PMLE) and Alonzo and Pepe's binary regression approach (GLM). The results are based on 1000 simulated data sets*

$n_{\bar{\mathrm{D}}}$	n_{D}		Bias		Mean Square Error	
			PMLE	GLM	PMLE	GLM
100	100	a	0.046	0.028	0.082	0.095
		b	0.042	0.018	0.012	0.013
		γ	0.011	0.022	0.191	0.260
200	200	a	0.025	0.012	0.038	0.049
		b	0.016	0.004	0.005	0.006
		γ	−0.005	0.008	0.095	0.138

For each simulated data set, we obtained point estimates of $\boldsymbol{\theta}$ with our pseudo-maximum profile likelihood approach and the binary regression approach of Alonzo and Pepe (2002). In Table 4, we present the bias and mean square error of $\widehat{\boldsymbol{\theta}}$ based on 1000 simulations. The results in Table 4 show that the new estimator outperforms the Alonzo and Pepe (2002) estimator. At sample size of $n_{\mathrm{D}} = n_{\bar{\mathrm{D}}} = 200$, the empirical efficiencies of the Alonzo and Pepe (2002) estimator relative to the new estimator are 78% for a, 83% for b and 69% for γ.

In conclusion, we have presented a theoretically rigorous method to obtain maximum likelihood estimates of the intercept and slope parameters of the binormal ROC curve. This is a semi-parametric approach to estimating the ROC curve in the sense that it assumes a parametric binormal form for the ROC curve, but does not assume that the test results follow any particular distribution. As discussed in the introduction, the binormal ROC curve is not the only semi-parametric approach for obtaining an estimate of the ROC curve. It is, however, the most popular of these semi-parametric approaches and means of estimating it easily and efficiently deserve attention. While others have proposed methods to estimate the necessary parameters, we feel that the methodology presented here distinguishes itself as a computationally feasible way to obtain the fully efficient maximum likelihood estimates.

APPENDIX A

Algorithm for computing $\widehat{\boldsymbol{\theta}}_{mle}$

Let $\mathbf{I}$ be the n by n matrix with the $[i, j]$th element $\mathbf{I}[i, j] = I_{ij} = I(Y_j \leqslant Y_i)$, $\mathbf{I}_{\mathrm{D}}$ be the first n_{D} rows of $\mathbf{I}$. Then $\Lambda_i = \sum_{l=1}^{n} I_{il}\lambda_l$ and $h_i = -\Phi^{-1}(\mathrm{e}^{-\sum_{l=1}^{n} I_{il}\lambda_l})$. For any vector $\mathbf{x}$, we use the notation x_l to denote the lth element of $\mathbf{x}$. Maximizing $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\lambda})$ with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ is equivalent to solving the following equations for $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$:

$$\frac{\partial \log \mathcal{L}}{\partial \lambda_l} = \frac{1}{\lambda_l} - \sum_{j=1}^{n} I_{jl} - \sum_{i=1}^{n_{\mathrm{D}}} D_1(\boldsymbol{\theta}, h_i) I_{il} = 0, \qquad l = 1, \ldots, n, \tag{A.1}$$

$$\frac{\partial \log \mathcal{L}}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{n_{\mathrm{D}}} \mathbf{D}_2(\boldsymbol{\theta}, h_i) = \mathbf{0}, \tag{A.2}$$

where

$$D_1(\boldsymbol{\theta}, h) = -\frac{(h - b^2 h + ab)\Phi(-h)}{\phi(-h)}, \quad \text{and} \quad \mathbf{D}_2(\boldsymbol{\theta}, h) = \begin{pmatrix} -a + bh \\ \frac{1}{b} + ah - bh^2 \end{pmatrix}.$$

To obtain the solutions to (A.1) and (A.2) numerically, we use the following algorithm:

Step 1: Choose some convergence threshold value δ_0 and $\widetilde{\delta}_0$. Let $\delta = 1$, $m = 1$, and let $\widehat{\boldsymbol{\lambda}}^{[0]}$ and $\widehat{\boldsymbol{\theta}}^{[0]}$ be the respective initial values for $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$ and $\boldsymbol{\theta}$. We recommend the following initial values:

$$\widehat{\lambda}_l^{[0]} = \frac{1}{\sum_{j=1}^n I_{jl}}, \quad \text{for } l = 1, \ldots, n, \qquad \widehat{\mathbf{h}}^{[0]} = -\Phi^{-1}(\mathrm{e}^{-\mathbf{I}_{\mathrm{D}}\widehat{\boldsymbol{\lambda}}^{[0]}}), \quad \text{and} \quad \widehat{\boldsymbol{\theta}}^{[0]} = \widehat{\boldsymbol{\theta}}(\widehat{\mathbf{h}}^{[0]}).$$

Step 2: Compute

$$\widehat{\lambda}_l^{[m]} = \frac{1}{\sum_{j=1}^n I_{jl} + \sum_{i=1}^{n_{\mathrm{D}}} D_1(\widehat{\boldsymbol{\theta}}^{[m-1]}, \widehat{h}_i^{[m-1]}) I_{il}}, \qquad \text{for } l = 1, \ldots, n,$$

$$\widehat{\mathbf{h}}^{[m]} = -\Phi^{-1}(\mathrm{e}^{-\mathbf{I}_{\mathrm{D}}\widehat{\boldsymbol{\lambda}}^{[m]}}), \quad \text{and} \quad \widehat{\boldsymbol{\theta}}^{[m]} = \widehat{\boldsymbol{\theta}}(\widehat{\mathbf{h}}^{[m]}).$$

Step 3: Calculate

$$\delta = \left\| \widehat{\boldsymbol{\lambda}}^{[m]} - \widehat{\boldsymbol{\lambda}}^{[m-1]} \right\| + \left\| \widehat{\boldsymbol{\theta}}^{[m]} - \widehat{\boldsymbol{\theta}}^{[m-1]} \right\|$$

If $\delta > \delta_0$, set $m = m + 1$ and go to step 2; otherwise calculate

$$\widetilde{\delta} = \sum_{l=1}^n \left\| \frac{1}{\widehat{\lambda}_l^{[m]}} - \sum_{j=1}^n I_{jl} - \sum_{i=1}^{n_{\mathrm{D}}} D_1(\widehat{h}_i^{[m]}, \boldsymbol{\theta}^{[m]}) I_{il} \right\| + \left\| \sum_{i=1}^{n_{\mathrm{D}}} \mathbf{D}_2(\widehat{\boldsymbol{\theta}}^{[m]}, \widehat{h}_i^{[m]}) \right\|$$

If $\widetilde{\delta} > \widetilde{\delta}_0$, go to step 1 and choose a different initial value for $\boldsymbol{\lambda}$ and $\boldsymbol{\theta}$; otherwise $\widehat{\mathbf{h}} = \widehat{\mathbf{h}}^{[m]}$ and $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}^{[m]}$.

APPENDIX B

Consistency and asymptotic distribution of $\widehat{\boldsymbol{\theta}}_{pmle}$

By the uniform law of large numbers (Pollard, 1990), $\widehat{S}(y) - S\{h_0(y), \boldsymbol{\theta}_0\} \to 0$ almost surely, uniformly in y. It follows from the monotonicity of $\Phi(\cdot)$ that there exists a unique $\widehat{h}(y, \boldsymbol{\theta})$ to equation (3.1) for any given $\boldsymbol{\theta}$. Let $\mathbf{V}(\boldsymbol{\theta})$ denote the left-hand side of (3.2). Since $\mathbf{V}(\boldsymbol{\theta}_0) \to 0$ and there exists a unique solution to $\mathbf{V}(\boldsymbol{\theta}) = 0$, $\widehat{\boldsymbol{\theta}}_{\mathrm{pmle}}$ is consistent.

To show the large sample distribution of $\widehat{\boldsymbol{\theta}}_{\mathrm{pmle}}$, we let $\dot{S}_h(h, \boldsymbol{\theta}) = \partial S(h, \boldsymbol{\theta})/\partial h$ and $\dot{S}_{\boldsymbol{\theta}}(h, \boldsymbol{\theta}) = \partial S(h, \boldsymbol{\theta})/\partial \boldsymbol{\theta}$. By a Taylor series expansion of $\mathbf{V}(\widehat{\boldsymbol{\theta}}_{\mathrm{pmle}})$ around $\boldsymbol{\theta}_0$, we obtain

$$n_{\mathrm{D}}^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}_{\mathrm{pmle}} - \boldsymbol{\theta}_0) \approx \mathbb{A}^{-1} n_{\mathrm{D}}^{-\frac{1}{2}} \mathbf{V}(\boldsymbol{\theta}_0),$$

where $\mathbb{A}$ is the limit of $-\partial \mathbf{V}(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}$. Again, by Taylor's series expansion of $\widehat{h}(y, \boldsymbol{\theta}_0)$ around $h_0(y)$,

$$n_{\mathrm{D}}^{\frac{1}{2}} \left\{ \widehat{h}(y, \boldsymbol{\theta}_0) - h_0(y) \right\} \approx \dot{S}_h \left\{ h_0(y), \boldsymbol{\theta}_0 \right\}^{-1} n_{\mathrm{D}}^{\frac{1}{2}} \sum_{i=1}^n e_i(y)$$

and

$$n_{\mathrm{D}}^{-\frac{1}{2}}\mathbf{V}(\boldsymbol{\theta}_0) \approx n_{\mathrm{D}}^{-\frac{1}{2}}\sum_{i=1}^{n_{\mathrm{D}}}(\mathbf{v}_{1i}+\mathbf{v}_{2i}) + n_{\mathrm{D}}^{-\frac{1}{2}}\sum_{j=n_{\mathrm{D}}+1}^{n_{\mathrm{D}}+n_{\bar{\mathrm{D}}}}\mathbf{v}_{2j}.$$

Therefore the distribution of $n_{\mathrm{D}}^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}_{\mathrm{pmle}} - \boldsymbol{\theta}_0)$ can be approximated by a zero-mean normal random vector with covariance matrix $\boldsymbol{\Sigma}_{\mathrm{pmle}}$.

APPENDIX C

Large sample properties of $\widehat{Q}(y)$ under model (1.1)

We now use $\widehat{\boldsymbol{\theta}}$ to represent any consistent estimator of $\boldsymbol{\theta}_0$ that has a square root n_{D} convergence rate. Without loss of generality, we further assume that there exist some deterministic functions $\boldsymbol{\eta}_{\mathrm{D}}(\cdot)$ and $\boldsymbol{\eta}_{\bar{\mathrm{D}}}(\cdot)$ such that $n_{\mathrm{D}}^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$ is asymptotically equivalent to $n_{\mathrm{D}}^{-\frac{1}{2}}\sum_{i=1}^{n_{\mathrm{D}}}\boldsymbol{\eta}_{\mathrm{D}}(Y_i)+n_{\mathrm{D}}^{-\frac{1}{2}}\sum_{j=n_{\mathrm{D}}+1}^{n_{\mathrm{D}}+n_{\bar{\mathrm{D}}}}\boldsymbol{\eta}_{\bar{\mathrm{D}}}(Y_j)$. It follows from the strong consistency of $\widehat{\boldsymbol{\theta}}$, $\widehat{S}_{\mathrm{D}}(\cdot)$ and $\widehat{S}_{\bar{\mathrm{D}}}(\cdot)$ that $\widehat{Q}(y) \to 0$ uniformly in y.

The weak convergence of $n_{\mathrm{D}}^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$, coupled with the functional central limit theorem (Pollard, 1990), implies that

$$n_{\mathrm{D}}^{\frac{1}{2}}\widehat{Q}(y) = n_{\mathrm{D}}^{-\frac{1}{2}}\sum_{l=1}^{n} O(y, Y_l) + o_p(1)$$

and it converges weakly to a zero-mean Gaussian process $\mathcal{Q}(y)$, where

$$O(y, Y_i) = \frac{\delta_i(y) - S_{\mathrm{D}}(y)}{\phi\left\{\Phi^{-1}(S_{\mathrm{D}}(y))\right\}} - \left(1\ \Phi^{-1}\{S_{\bar{\mathrm{D}}}(y)\}\right)\boldsymbol{\eta}_{\mathrm{D}}(Y_i) \quad \text{for } i = 1, \ldots, n_{\mathrm{D}},$$

$$Q(y, Y_j) = -\frac{\{\delta_j(y) - S_{\bar{\mathrm{D}}}(y)\}bp_{\mathrm{D}}}{\phi\left\{\Phi^{-1}(S_{\bar{\mathrm{D}}}(y))\right\}p_{\bar{\mathrm{D}}}} - \left(1\ \Phi^{-1}\{S_{\bar{\mathrm{D}}}(y)\}\right)\boldsymbol{\eta}_{\bar{\mathrm{D}}}(Y_j) \quad \text{for } j = n_{\mathrm{D}}+1, \ldots, n.$$

It is straightforward to show that the process $n_{\mathrm{D}}^{\frac{1}{2}}\widehat{Q}_{\mathcal{Z}}(y)$ has the same limiting covariance function as that of $n_{\mathrm{D}}^{\frac{1}{2}}\widehat{Q}(y)$. Furthermore, conditional on the data, the process $n_{\mathrm{D}}^{\frac{1}{2}}\widehat{Q}_{\mathcal{Z}}(y)$ is tight (Shorack and Wellner, 1986). It follows that the distribution of $n_{\mathrm{D}}^{\frac{1}{2}}\widehat{Q}(y)$ can be approximated by the conditional distribution of $n_{\mathrm{D}}^{\frac{1}{2}}\widehat{Q}_{\mathcal{Z}}(y)$.

REFERENCES

ALONZO, T. A. AND PEPE, M. S. (2002). Distribution-free ROC analysis using binary regression techniques. *Biostatistics* **3**, 421–433.

BEGG, C. B. (1991). Advances in statistical methodology for diagnostic medicine in the 1980s. *Statistics in Medicine* **10**, 1887–1895.

DORFMAN, D. AND ALF, E. (1969). Maximum likelihood estimation of parameters of signal detection theory and determination of confidence intervals-rating method data. *Journal of Mathematical Psychology* **6**, 487–496.

GOLUB, G. AND VAN LOAN, C. (1989). *Matrix Computations*, 2nd edn. Baltimore, MD: The John Hopkins University Press.

HANLEY, J. A. (1989). Receiver Operating Characteristic (ROC) methodology : the state of the art. *Clinical Reviews in Diagnostic Imaging* **29**, 307–335.

HANLEY, J. A. (1996). The use of the 'binormal' model for parametric ROC analysis of quantitative diagnostic tests. *Statistics in Medicine* **15**, 1575–1585.

HSIEH, F. AND TURNBULL, B. W. (1996). Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *The Annals of Statistics* **24**, 25–40.

LI, G., TIWARI, R. C. AND WELLS, M. T. (1999). Semiparametric inference for a quantile comparison function with applications to receiver operating characteristic curves. *Biometrika* **86**, 487–502.

METZ, C. E. (1986). ROC methodology in radiologic imaging. *Investigative Radiology* **21**, 720–733.

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., ROSSINI, A. J. AND VAN DER VAART, A. W. (1997). Maximum likelihood estimation in the proportional odds model. *Journal of the American Statistical Association* **92**, 968–976.

MURPHY, S. A. AND VAN DER VAART, A. W. (2000). On profile likelihood (C/R: p. 466–485). *Journal of the American Statistical Association* **95**, 449–465.

OGILVIE, J. C. AND CREELMAN, C. D. (1968). Maximum likelihood estimation of receiver operating characteristic curve parameters. *Journal of Mathematical Psychology* **5**, 377–391.

PARZEN, M. I., WEI, L. J. AND YING, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81**, 341–350.

PEPE, M. S. (2000). An interpretation for the ROC curve and inference using GLM procedures. *Biometrics* **56**, 352–359.

PEPE, M. S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. New York: Oxford Univeristy Press.

POLLARD, D. (1990). *Empirical Processes: Theory and Applications*. Hayward, CA: Institute of Mathematical Statistics.

QIN, J. AND ZHANG, B. (2003). Using logistic regression procedures for estimating receiver operating characteristic curves. *Biometrika* **90**, 585–596.

SHORACK, G. R. AND WELLNER, J. A. (1986). *Empirical Processes With Applications to Statistics*. New York: Wiley.

SWETS, J. A. (1986). Indices of discrimination or diagnostic accuracy: their ROCs and implied models. *Psychological Bulletin* **99**, 100–117.

SWETS, J. A. AND PICKETT, R. M. (1982). *Evaluation of Diagnostic Systems: Methods from Signal Detection theory*. New York: Academic.

WIEAND, S., GAIL, M. H., JAMES, B. R. AND 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*. New York: Wiley.

ZOU, K. AND HALL, W. J. (2000). Two transformation models for estimating an ROC curve derived from continuous data. *Journal of Applied Statistics* **27**, 621–633.

[*Received December 2, 2003; revised March 10, 2004; accepted for publication March 26, 2004*]