UW Biostatistics Working Paper Series

7-30-2004

Non-Parametric Estimation of ROC Curves in the Absence of a Gold Standard

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

Pete Castelluccio
Indiana University, pfcastel@iupui.edu

Chuan Zhou
University of Washington, czhou@u.washington.edu

1. Introduction

To evaluate the accuracy of a diagnostic test, an unbiased estimate for the test accuracy is preferred. But in order to obtain an unbiased estimator for the test accuracy, we need to determine the true disease status for each patient (present or absent) independent of the patient's test result. The procedure that establishes the patient's true disease status is referred to as a gold standard.

For many diseases, it is difficult or impossible to establish a definitive diagnosis. A perfect gold standard may not exist or may be too expensive or impractical to administer. This is especially true for complex clinical conditions in the usual clinical practice setting. For example, a definitive diagnosis of myocardial infarction (MI) is difficult to establish for patients admitted to a hospital for "rule-out MI". Similarly, the definitive diagnosis of Alzheimer's disease cannot be established until a patient has died and a neuropathological examination is performed. Even when the "definitive" diagnosis of a well-defined disease, such as an infection by a known agent, can be performed, it still may require culturing the organism or other detection procedures, any of which may be subject to laboratory and other errors. Consequently, in many diagnostic accuracy studies, an imperfect standard is used to evaluate the test instead. However, when an imperfect standard is used as if it were a gold standard, the accuracy of the new test is often either underestimated or overestimated. This type of bias is called imperfect

reference standard bias.

Hui and Zhou (1998) reviewed available statistical methods for estimating the diagnostic accuracy of one or more new tests, with or without an imperfect standard, when the true disease status is not known for any of the subjects. As noted in Hui and Zhou (1998), almost all available statistical methods focus on binary tests and are based on mixture latent class models; and the majority of those methods require the conditional independence assumption (CIA). Only few published papers dealt with estimation of ROC curves of ordinal or continuous scale tests in the absence of a gold standard. Henkelman, Kay and Bronskill (1990) proposed a maximum likelihood estimation method for the ROC curve of a five-point rating scale using a multivariate normal mixture latent model. One major limitation of this approach is that the latent random variables for multiple ordinal-scale tests are assumed to follow the multivariate normal distribution. In addition, in a published commentary on this paper, Begg and Metz (1990) pointed out three serious potential limitations to this method and called for further research into its properties before they could recommend it for general use. Another paper by Hall and Zhou (2003) proposed a non-parametric estimator for the ROC curves of continuous-scale tests under the conditional independence assumption when the number of tests is more than two.

In this paper, we will apply the ideas in Hall and Zhou (2003) to estimating ROC curve areas of ordinal-scale tests when the number of tests is more than two. As shown in Hall and Zhou (2003), without the conditional

independence assumption the component distributions in a multivariate latent class model are not identifiable non-parametrically. Hence, in this paper we focus on a non-parametric maximum likelihood (ML) method under the conditional independence assumption. In Section 2 we present the method in detail. We show the existence of many local ML estimate solutions under this model, the global ML estimate has a "mirror" solution that yields the same log-likelihood value, and all local ML estimates, including the global ones, give the same sum of squared residuals. In Section 3 we conduct simulation studies to assess the finite-sample properties of the proposed ML estimators, and to compare them with parametric models. We apply the proposed method to a real study in Section 4.

2. A Non-parametric Approach

2.1 *Estimation method*

Here we consider the situation where each of the N patients is scored on an ordinal scale from 1 to J on a battery of K tests. Throughout this paper we will assume that the disease status is unknown for all N patients and will attempt to estimate the ROC curves for each of the K tests without this seemingly necessary piece of information.

Let $T_1, \cdots, T_K$ be the responses from K diagnostic tests for a particular patient whose disease status although unknown is denoted by D, where $D = 1$ if the patient is diseased and $D = 0$ if the patient is non-diseased. Since each test can be scored from 1 to J, we can define its ROC curve in two

ways: (1) the non-parametric ROC curve based on the discrete sensitivity and specificity values, and (2) the continuous ROC curve of a latent variable underlying the observable ordinal data. In this paper, we focus on the non-parametric ROC curve. To compute a discrete ROC curve from the ordinal data, we vary the threshold for a positive test and then calculate $J+1$ pairs of true positive rates (TPR) and false positive rates (FPR). Specifically, for the kth test, if we define a positive test as one with $T_k \geq j$, a corresponding pair of TPR and FPR are

$$TPR_k(j) = P(T_k \geq j \,|\, D = 1),\ FPR_k(j) = P(T_k \geq j \,|\, D = 0),$$

respectively, for $j = 1, \cdots, J+1$. Here, $TPR_k(1) = FPR_k(1) = 1$, and $TPR_k(J+1) = FPR_k(J+1) = 0$. A discrete ROC curve is defined as a discrete function of $(FPR_k(j),\ TPR_k(j))$, $j = 1, \cdots, J+1$. By connecting coordinates with linear lines, we obtain the non-parametric ROC curve. Using the trapezoidal rule for integration (Bamber, 1975), we can obtain the area under the non-parametric ROC curve of the kth test as follows

$$\begin{aligned} A_k = &\sum_{j=1}^{J-1} \left[P(T_k = j \,|\, D = 0) \sum_{l=j+1}^{J} P(T_k = l \,|\, D = 1) \right] \\ &+ \frac{1}{2} \sum_{j=1}^{J} P(T_k = j \,|\, D = 0) P(T_k = j \,|\, D = 1) \end{aligned} \tag{1}$$

If we define $\phi_{0kj} = P(T_k = j \,|\, D = 0)$ and $\phi_{1kj} = P(T_k = j \,|\, D = 1)$, we can express the ROC curve and its area as functions of ϕ_{0kj} and

ϕ_{1kj}. Note that the coordinates of the non-parametric ROC curve of T_k are $(FPR_k(j), TPR_k(j))$, which are related to the parameters ϕ_{1jl} and ϕ_{2jl} in the following form:

$$FPR_k(j) = \sum_{l=j}^{J} \phi_{0kl},\ TPR_k(j) = \sum_{l=j}^{J} \phi_{1kl}. \tag{2}$$

Similarly we can show the area under the ROC curve for the kth test can be written as follows:

$$A_k = \sum_{j=1}^{J-1} \left[\phi_{0kj} \sum_{l=j+1}^{J} \phi_{1kl} \right] + \frac{1}{2} \sum_{j=1}^{J} \phi_{0kj}\phi_{1kj}. \tag{3}$$

We wish to formulate the likelihood for this particular problem in such a way that ϕ_{0kj} and ϕ_{1kj} play central roles. Specifically, we wish to be able to find maximum likelihood estimates (MLEs) for these parameters and to employ them in order to calculate MLEs for the ROC curve and its areas under each of the K tests.

Let y_{ikj} be a binary variable such that $y_{ikj} = 1$ if the response of the kth test is j for the ith patient and $y_{ijk} = 0$ otherwise, where $i = 1, \cdots, N$, $k = 1, \cdots, K$, and $j = 1, \cdots, J$. Then we can construct a $K \times J$ binary vector $\boldsymbol{y}_i$ such that $\boldsymbol{y}_i = (y_{i11}, \cdots, y_{i1J}, \cdots, y_{iK1}, \cdots, y_{iKJ})$. We will call $\boldsymbol{y}_i$ the test score vector for the ith patient. Also let us define the disease status of the ith patient to be $D_i = 1$ if the ith patient is diseased and $D_i = 0$ if the ith patient is not diseased. Now let us define the likelihood function

$g_d(\boldsymbol{y}_i) = P(\boldsymbol{y}_i \,|\, D_i = d)$ to be the conditional probability of ith patient's test score vector $\boldsymbol{y}_i$ given their disease status $D_i = d$. When assuming conditional independence of the K tests, we can write that

$$g_d(\boldsymbol{y}_i) = \prod_{k=1}^{K} \prod_{j=1}^{J} [\phi_{dkj}]^{y_{ikj}}, \tag{4}$$

where $\phi_{dkj} = P(T_k = j|D = d)$. Here we employ the "1/0" property of the vector $\boldsymbol{y}_i$ to turn "on/off" the proper ϕ_{dkj}.

If we assume a Bernoulli distribution for disease status with $p_d = P(D = d)$ for $d = 0, 1$, we obtain that the marginal likelihood contributed by the ith patient has the following mixture form: $P(\boldsymbol{y}_i) = p_1 g_1(\boldsymbol{y}_i) + p_0 g_0(\boldsymbol{y}_i)$. The joint log-likelihood of all N patients is given by

$$l(p_1, \boldsymbol{\phi}_0, \boldsymbol{\phi}_1) = \sum_{i=1}^{N} \log[p_0 g_0(\boldsymbol{y}_i) + p_1 g_1(\boldsymbol{y}_i)], \tag{5}$$

where $p_0 = 1 - p_1$, and $\boldsymbol{\phi}_d$ represents the vector of conditional probabilities $(\phi_{d11}, \cdots, \phi_{d1J}, \cdots, \phi_{dK1}, \cdots, \phi_{dKJ})$ for $d = 0, 1$.

Our goal is to find maximum likelihood (ML) estimates for p_1, $\boldsymbol{\phi}_0 = (\phi_{011}, \cdots, \phi_{0KJ})$ and $\boldsymbol{\phi}_1 = (\phi_{111}, \cdots, \phi_{1KJ})$, subjecting to the normalizing conditions $\sum_{j=1}^{J} \phi_{dkj} = 1$ for $d = 0, 1$ and all $k = 1, \cdots, K$. These are precisely the parameters needed in order to estimate the ROC curves and their respective areas for each of the K tests. Here we employ the EM algorithm to find the ML estimates by treating D as missing data. Therefore

our complete data consist of $(\boldsymbol{y}, D)$. The main advantage of the EM algorithm over the directly maximizing the log-likelihood function (5) is that there is an explicit solution in the M step for our non-parametric approach.

Let $\boldsymbol{\theta} = (p_1, \boldsymbol{\phi}_0, \boldsymbol{\phi}_1)$. Note that the complete-data log-likelihood is given by

$$l_c(\boldsymbol{\theta}) = \sum_{i=1}^{N} [D_i \log p_1 g_1(\boldsymbol{y}_i) + (1 - D_i) \log p_0 g_0(\boldsymbol{y}_i)].$$

Let $\boldsymbol{\theta}^{(t)}$ denote the estimate of $\boldsymbol{\theta}$ after the tth iteration of the EM algorithm.

- **E step**: The E step computes the conditional expectation of $l_c(\boldsymbol{\theta})$ given the observed data $\boldsymbol{y}$ and current parameter estimates $\boldsymbol{\theta} = \boldsymbol{\theta}^{(t)}$,

$$\mathrm{E}(l_c(\boldsymbol{\theta}) \,|\, \boldsymbol{y}, \theta = \theta^{(t)}) = \sum_{i=1}^{N} \sum_{d=0}^{1} P(D_i = d \,|\, \boldsymbol{y}_i, \boldsymbol{\theta}^{(t)}) \log p_d g_d(\boldsymbol{y}_i).$$

  If we write

$$q_{id}^{(t)} = P(D_i = d \,|\, \boldsymbol{y}_i, p_1^{(t)}, \boldsymbol{\phi}_0^{(t)}, \boldsymbol{\phi}_1^{(t)}),$$

  and

$$g_d^{(t)}(\boldsymbol{y}_i) = \prod_{k=1}^{K} \prod_{j=1}^{J} [\phi_{dkj}^{(t)}]^{y_{ikj}},$$

  we can show that

$$q_{id}^{(t)} = \frac{p_d^{(t)} g_d^{(t)}(\boldsymbol{y}_i)}{p_0^{(t)} g_0^{(t)}(\boldsymbol{y}_i) + p_1^{(t)} g_1^{(t)}(\boldsymbol{y}_i)}, \tag{6}$$

and

$$\mathrm{E}(l_c(\boldsymbol{\theta}) \,|\, \boldsymbol{y}, \boldsymbol{\theta} = \boldsymbol{\theta}^{(t)}) = \sum_{i=1}^{N} \sum_{d=0}^{1} q_{id}^{(t)} \log g_d(\boldsymbol{y}_i). \tag{7}$$

- **M step**: The M step finds the updated estimate $\boldsymbol{\theta}^{(t+1)}$ for $\boldsymbol{\theta}$ by maximizing $\mathrm{E}(l_c(\boldsymbol{\theta}) \,|\, \boldsymbol{y}, \boldsymbol{\theta} = \boldsymbol{\theta}^{(t)})$ in (7). We can show that $\boldsymbol{\theta}^{(t+1)}$ has the following explicit expression:

$$p_1^{(t+1)} = \frac{1}{N} \sum_{i=1}^{N} q_{i1}^{(t)}, \tag{8}$$

and

$$\phi_{dkj}^{(t+1)} = \frac{\sum_{i=1}^{N} q_{id}^{(t)} y_{ikj}}{\sum_{i=1}^{N} q_{id}^{(t)}}. \tag{9}$$

It is helpful to note that

$$p_0^{(t+1)} \phi_{0kj}^{(t+1)} + p_1^{(t+1)} \phi_{1kj}^{(t+1)} = \frac{1}{N} \sum_{i=1}^{N} [(q_{i0}^{(t+1)} + q_{i1}^{(t+1)}) y_{ikj}] = \frac{1}{N} \sum_{i=1}^{N} y_{ikj} \equiv \bar{y}_{*kj}. \tag{10}$$

Though not required for the initial parameter estimates, the above condition will hold after any iteration and thus will hold for the eventual MLEs, $\hat{\boldsymbol{\theta}}$. Therefore this is a necessary condition for any set of MLEs under our non-parametric model. We call the condition $\hat{p}_0 \hat{\phi}_{0kj} + \hat{p}_1 \hat{\phi}_{1kj} = \bar{y}_{*kj}$ the MLE Mixture Condition. Due to this property we have just reduced the effective parameter space almost in half.

We obtain the estimated covariance matrix for $\boldsymbol{\theta}$ using the Fisher information matrix, which is given in Appendix A.

2.2 *Equal conditional probability solution*

Notice that if one selects the initial parameters such that $\phi_{0kj} = \phi_{1kj}$ for all k and j then $g_0^{(t=0)}(\boldsymbol{y}_i) = g_1^{(t=0)}(\boldsymbol{y}_i)$. By equation (6), $q_{i1}^{(t)}$ does not depend on data $\boldsymbol{y}$ and remains constant for all patients $i = 1, \cdots, N$. So during each iteration $p_1^{(t+1)} = p_1^{(t)} = \cdots = p_1^{(0)}$, and we find that the prevalence rate p_1 stays fixed to whatever value was selected for its initial estimate. It can also be seen that $\phi_{dkj}^{(t+1)} = \frac{1}{N}\sum_{i=1}^{N} y_{ikj} = \bar{y}_{*kj}$ for $d = 0, 1$ and all $1 \leq k \leq K$ and $1 \leq j \leq J$. Thus given any initial parameters such that $\phi_{0kj} = \phi_{1kj}$ for all k and j and any $p_1^{(0)}$ the iterative procedure will stop after just one iteration. It can be shown that for every such case the log-likelihood score function is zero, and thus each such case is a local log-likelihood maximum. Thus we are assured of the existence of an infinite number of local log-likelihood maxima. We hope this set of local maxima do not comprise all local maxima. For then the global maxima would be one such case and not only would p_1 be indeterminate, but we would conclude that $\phi_{0kj} = \phi_{1kj}$ for all k and j. It implies that each test is worthless for determining disease status since for each outcome of any test for a patient is equally likely regardless of disease status. Obviously this problem is sensitive to the selection of the initial parameter estimates and we can see which ones it would probably be worth avoiding. Therefore in seeking the global maximum using the proposed non-parametric approach, we make the following recommendations: 1) avoid equal $\phi_{0kj} = \phi_{1kj}$ for all k and j, this should not be a difficult decision as in practice certain asymmetry in the test scores are often obvious; 2)

try a set of reasonable initial parameter estimates, and compare the local log-likelihood maxima obtained; 3) reasonable initial values can be obtained from similar studies with known disease status; 4) study the likelihood surface using exploratory and simulation techniques such as the Stochastic EM we devise in our simulation study, see Section 4 for more details.

2.3 *Invariance property of log-likelihood function*

Upon looking at the log-likelihood equation (5) one can see that this equation is invariant to the re-labeling of the parameter sets $(p_0, \boldsymbol{\phi}_0, p_1, \boldsymbol{\phi}_1)$ to $(p_1, \boldsymbol{\phi}_1, p_0, \boldsymbol{\phi}_0)$. This implies that there can never exist a unique global maximum likelihood solution, since any such maximum, say $(\hat{p}_1, \hat{\boldsymbol{\phi}}_0, \hat{\boldsymbol{\phi}}_1)$ would imply the existence of a "mirror" maximum of equal likelihood at $((1 - \hat{p}_1), \hat{\boldsymbol{\phi}}_1, \hat{\boldsymbol{\phi}}_0)$. Thus the best case is that we arrive at a presumed global maximum by the use of the EM algorithm, and can distinguish between these two possibilities by a reasonable ordering of the prevalence rate $\hat{p}_1$ or $\hat{p}_0$ or by the plausibility of the resulting areas under the K ROC curves. The latter case may be possible since the area under any test k, say $A_k(\boldsymbol{\phi}_{0k\bullet}, \boldsymbol{\phi}_{1k\bullet})$, is equal to $1 - A_k(\boldsymbol{\phi}_{1k\bullet}, \boldsymbol{\phi}_{0k\bullet})$ where $\boldsymbol{\phi}_{dk\bullet} = (\phi_{dk1}, \ldots, \phi_{dkJ})$. For a proof, see Appendix B. Therefore, for any credible test we could presumably choose the global maximum for which $A_k > 0.5$ for all $k = 1, \ldots, K$ if such a case exists.

2.4 *Sum of squared residuals*

Let us look at the sum of squared residuals as defined by:

$$SS = \sum_{i=1}^{N} \sum_{k=1}^{K} \sum_{j=1}^{J} [y_{ikj} - E(y_{ikj})]^2 \tag{11}$$

Since for any patient i $\mathrm{E}(y_{ikj}) = \hat{p}_0\hat{\phi}_{0kj} + \hat{p}_1\hat{\phi}_{1kj}$, then by the MLE Mixture Condition we find that $E(y_{ikj}) = \bar{y}_{*kj}$. Thus at any local maximum and for every iteration of the EM algorithm the sum of squared residuals remain constant at the value:

$$SS = \sum_{i=1}^{N} \sum_{k=1}^{K} \sum_{j=1}^{J} [y_{ikj} - \bar{y}_{*kj}]^2 \tag{12}$$

Thus at each iteration of the EM algorithm the parameters are being updated in such a manner as to always increase the log-likelihood and meanwhile the sum of squared residuals remains fixed. Just as in the case of the maximum likelihood criterion there are an infinite number of suitable choices among the parameter set but for the case of the sum of squares there is no way of distinguishing between the possible choices for a "best" set.

3. Simulation Study

3.1 *Finite sample performance*

To assess the performance of the proposed likelihood-based approach in finite samples, we carried out simulation studies under different ROC curve

conditions. In the simulations we assessed the bias and mean squared error (MSE) of our estimators. We focused on the prevalence p_1 and the area under curve (AUC). We chose sample size to be $N = 118$ and $N = 500$, and the number of tests to be $K = 7$, as our real data set. In our simulation studies we constructed the experiments by varying the parameters ϕ_{0kj} and ϕ_{1kj} in such a manner that $A_1, \cdots, A_7$ took different values. We also took the prevalence rate p_1 to be 0.5, 0.7, and 0.9.

Tables 1 and 2 summarize the bias and MSE of the estimated ROC areas for equal ROC curve areas ($A_1 = \cdots = A_7 = A$), with $A = 0.7, 0.8$ and 0.9, and the sample size N is 118 and 500, respectively.

[Table 1 about here.]

[Table 2 about here.]

Tables 3 and 4 summarize the bias and MSE of the estimated ROC curve areas for unequal ROC curve areas ($A_1 = 0.9, A_2 = 0.87, A_3 = 0.83, A_4 = 0.80, A_5 = 0.77, A_6 = 0.73, A_7 = 0.7$) when the sample size N is 118 and 500, respectively.

[Table 3 about here.]

[Table 4 about here.]

From the results in Tables 1-4 we see that the proposed method yields ML estimates for ROC curves with small bias and MSE regardless the true

ROC curve areas. In general, the higher the areas under ROC curves are, the smaller their bias and MSEs. In other words, the estimators perform better when the tests distinguish the disease status better. Hence, the proposed method has good finite sample size performance under the conditional independence assumption.

3.2 *Comparison to a parametric approach*

Although the main purpose of this paper is to introduce a non-parametric approach to the ROC curve estimation without gold standard, it is of interest to compare it with existing parametric approaches. Unfortunately most published literature has been on the problem of hypothesis testing, not on estimation based on parametric models (Hui and Zhou, 1998; Qu and Hadgu, 1998). Therefore, for our simulation studies, we had to extend the classic binormal model to the situation where the patient disease status is missing. Following Diebolt and Ip (1996) we adapted the ROCFIT program (Metz et al., 1994) to allow missing indicator of disease status. A stochastic step is added to impute the missing disease indicator D from its conditional density given observation $\boldsymbol{y}$ and current parameter estimate $\boldsymbol{\phi}^{(m)}$ of parameter $\boldsymbol{\phi}$. Since this imputation is based on all our current information, and hence provides us with a plausible pseudo-complete sample. Once we have a pseudo-complete sample, we can directly maximize its log-likelihood to obtain an updated MLE $\boldsymbol{\phi}^{(m+1)}$ using the standard ROCFIT program. This whole process is iterated. This is the stochastic EM (SEM) algorithm introduced

by Celeux and Diebolt (1985). Under mild conditions the SEM algorithm generates a Markov chain $\{\boldsymbol{\phi}^{(m)}\}$ which converges to a stationary distribution $\pi(\cdot)$. The stationary distribution is approximately centered at the MLE of $\boldsymbol{\phi}$ thus provides an alternative for the maximum likelihood estimation. Standard errors of MLE can be easily derived from the simulated samples as well (Diebolt and Ip, 1996). Due to the stochastic nature of SEM algorithm, unlike the EM algorithm the log-likelihood does not decrease monotonically, and the convergence of the Markov chain has to be monitored using convergence diagnostics. As noted by Diebolt and Ip (1996), in most situations convergence is reached reasonably fast. See Biernacki et al. (2003) for more on stopping criteria. For our simulation, we used 400 iterations with 200 as burn-in.

Table 5 lists the MLEs from both parametric and non-parametric approaches when the data are actual binormal. The data were generated from two overlapping Gaussian distributions with different degree of separation ($A_1 = 0.95, A_2 = 0.90, A_3 = 0.90, A_4 = 0.85, A_5 = 0.85, A_6 = 0.80, A_7 = 0.75$), with various ratios of SDs of distributions for the non-diseased to diseased, and various decision thresholds. It appears the results from our non-parametric approach are quite comparable to those from parametric models. The validity of the parametric and non-parametric methods relies on one common assumption and some unique assumptions. Both the methods make the conditional independence assumption. While the non-parametric method makes the irreducibility assumption, the parametric method makes paramet-

ric assumptions on the distributions of latent variables.

But the non-parametric approach has additional advantages: it is potentially more robust since no distributional assumptions are required, and it is easier to implement when there is no gold standard.

[Table 5 about here.]

Table 6 lists the results from a simulation study where the true parametric distributions are not binormal. The distribution for non-diseased subjects was chosen to be Gaussian, but the distribution for diseased subjects was formed from a mixture of two Gaussian distributions to create right-skewed bimodal distributions. Again the AUCs were fixed a priori, with various ratios of SDs of component distributions for the non-diseased to diseased subjects. It can be seen under such the setting, both approaches perform worse than those in Table 5. The results from both non-parametric approach and parametric approach are quite comparable. And for $p_1 = 0.9$, the non-parametric approach again appears to perform better, but for smaller p_1 parametric approach appears to perform better. It has been shown that binormal model is quite robust to model mis-specification (Hajian-Tilaki, Hanley, Joseph and Collet, 1997; Walsh, 1997), the close estimates of diagnostic accuracy (AUC) and the corresponding precision in our simulations are consistent with the findings by those authors.

[Table 6 about here.]

4. A Real Example

Holmquist, McMahan and Williams (1967) studied variability in detection of carcinoma in situ of the Uterine Cervix among seven pathologists under the study. These seven pathologists were all senior staff pathologists who were involved in diagnoses of surgical pathologic specimens during 1963 at Louisiana State University Medical Center. During the period July 1, 1964 through June 30, 1965, these seven pathologists independently evaluated and classified lesions on each of the 118 randomly ordered slides into five category ordinal-scale, ranging from 1 (negative) to 5 (invasive carcinoma) categories. In this study, there was a clinical definition on carcinoma in situ of the uterine cervix. However, due to technological limitations, diagnosis based on the clinical definition was not available.

Landis and Koch (1977) assessed variability in detection of carcinoma among the seven readers using agreement measures. However, the agreement information cannot translated into the accuracy information. For example, the seven readers might agree on the disease status of a patient, they all could be wrong. In this section we apply the proposed method in Sections 2 to assess the variability in the diagnostic accuracy of each reader in detecting the carcinoma in situ of the uterine cervix, in terms of the empirical ROC curve and its area under the curve. We summarize the estimated non-parametric ROC curves for the seven readers in Figure 1, and the corresponding areas under the ROC curves are 0.94, 0.92, 0.90, 0.93, 0.95, 0.87, and 0.98, respectively. The estimated prevalence is 0.61. The estimated areas under the

curves under binormal model assumption are 0.84, 0.83, 0.83, 0.86, 0.84, 0.84 and 0.85, respectively, the prevalence is estimated to be 0.65. The computation was carried out using the SEM algorithm discussed above. It appears the parametric approach gave more conservative estimates of the test accuracy, but the validity of the estimates relies on the binormal assumption and the convergence of the Markov chain. The log-likelihood under non-parametric MLEs is -779.23, whereas the log-likelihood under MLEs from the binormal model is -841.15. The likelihood ratio test statistic is 123.84 with degrees of freedom 14, larger than $\chi^2_{0.95,14} = 23.68$, so the non-parametric model gives significant better fit to the data than the binormal model.

[Figure 1 about here.]

5. Discussion

In this paper we have proposed a ML method for estimating the accuracy of ordinal-scale diagnostic tests with the EM algorithm, based on a latent class model. To avoid controversies on the use of a latent class model in the problem of imperfect gold standard (Alonzo and Pepe, 1999;Hadgu and Miller, 2001; Pepe and Alonzo, 2001), in this paper we assume that we are dealing with clinical studies in which a gold standard exists but is not available.

Our simulation result has shown the proposed estimators have good small bias and mean squared error (MSE). However, the global ML estimate is not unique; there is a "mirror" solution. If $(\hat{p}_1, \hat{\boldsymbol{\phi}}_0, \hat{\boldsymbol{\phi}}_1)$ is ML estimates for

$(p_1, \boldsymbol{\phi}_0, \boldsymbol{\phi}_1)$, $(1 - \hat{p}_1, \hat{\boldsymbol{\phi}}_1, \hat{\boldsymbol{\phi}}_0)$ is also ML estimates. This result is consistent with the imperfect gold standard bias problem in binary-scale tests as observed by Hui and Walter (1980). One additional complication in finding a global ML estimate is that there are many local ML estimates. To overcome this problem we recommend randomly perturbing the starting point, or recomputing the ML estimates based on a set of plausible initial values.

An advantage with our proposed non-parametric ML approach is that it does not require specific modeling assumptions, therefore it is likely to be more robust. Another advantage is that there is explicit solution at the M-step, so it is much easier to implement the corresponding EM algorithm. Our simulation studies show that non-parametric estimates are comparable to those from parametric models. But because the missing disease status, the parametric likelihood involves complicated mixture form and differentiation, which makes it harder to carry out the computation. Our experience also suggests the computation for parametric models is more sensitive to initial values and less stable when the tables are close to degeneracy.

The proposed non-parametric method requires the conditional independence and irreducibility assumptions. One future research is to develop a parametric ML method without assuming conditional independence. For example, we may use a log-linear model without higher order interactions or a random-effect latent class model as done in Hadgu and Qu (1998) and Qu and Hadgu (1998).

Acknowledgements

We would like to thank the editor and an associate editor for many helpful comments that resulted in an improved version of this manuscript.

References

Alonzo, T. A. and Pepe, M. S. (1999). Using a combination of reference tests to assess the accuracy of a new diagnostic test. *Statistics in Medicine* **18**, 2987–3003.

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

Begg, C. B. and Metz, C. E. (1990). Consensus diagnosis and "gold standards". *Medical Decision Making* **10**, 29–30.

Biernacki, C., Celeux, G. and Govaert, G. (2003). Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. *Computational statistics and data analysis* **41**, 561–575.

Celeux, G. and Diebolt, J. (1985). The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Comp. Statist. Quart.* **2**, 73–82.

Diebolt, J. and Ip, E. H. (1996). *Markov Chain Monte Carlo in practice*, chapter 15. Stochastic EM: method and application, pages 259–273. Chapman&Hall/CRC.

Hadgu, A. and Miller, W. (2001). Comment on: "Using a combination of reference tests to assess the accuracy of a diagnostic test" (1999 V18

p2987-3003). *Statistics in Medicine* **20**, 656–658.

Hadgu, A. and Qu, Y. (1998). A biomedical application of latent class models with random effects. *Applied Statistics* **47**, 603–616.

Hajian-Tilaki, K. O., Hanley, J. A., Joseph, L. and Collet, J.-P. (1997). A comparison of parametric and nonparametric approaches to ROC analysis of quantitiative diagnostic tests. *Medical Decision Making* **17**, 94–102.

Hall, P. and Zhou, X.-H. (2003). Nonparametric estimation of component distributions in a multivariate mixture. *Annals of Statistics* **31**, 201–224.

Henkelman, R. M., Kay, I. and Bronskill, M. J. (1990). Receiver operator characteristic (roc) analysis without truth. *Medical Decision Making* **10**, 24–29.

Holmquist, N. D., McMahan, C. A. and Williams, O. D. (1967). Variablity in classification of carcinoma in situ of the uterien cervix. *Archives of Pathology* **84**, 334–345.

Hui, S. L. and Walter, S. D. (1980). Estimating the error rates of diagnostic tests. *Biometrics* **36**, 167–171.

Hui, S. L. and Zhou, X. H. (1998). Evaluation of diagnostic tests without gold standards. *Statistical Methods in Medical Research* **7**, 354–370.

Landis, J. R. and Koch, G. G. (1977). An application of hierarchical Kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics* **33**, 363–374.

Metz, C. E. et al. (1994). LABROC and ROCFIT software. Technical report, University of Chicago.

Pepe, M. S. and Alonzo, T. A. (2001). Reply to comment on: "Using a combination of reference tests to assess the accuracy of a diagnostic test" (1999 V18 p2987-3003). *Statistics in Medicine* **20**, 658–660.

Qu, Y. and Hadgu, A. (1998). A model for evaluating sensitivity and specificity for correlated diagnostic tests in efficacy studies with an imperfect reference test. *Journal of the American Statistical Association* **93**, 920–928.

Walsh, S. J. (1997). Limitations to the robustness of binormal ROC curves: effects of model misspecification and location of decision thresholds on bias, precision, size and power. *Statistics in Medicine* **16**, 669–679.

Appendix A

Expected Fisher's Information Matrix

We summarize Expected Fisher's information matrix of the log-likelihood function in the following.

The Expected Fisher's information matrix equals to

$$\mathrm{E}\left[-\frac{\partial^2 l(p_1,\boldsymbol{\phi}_0,\boldsymbol{\phi}_1)}{\partial(p_1,\boldsymbol{\phi}_0,\boldsymbol{\phi}_1)^2}\right].$$

Here

$$\mathrm{E}\left[-\frac{\partial^2 l(p_1,\boldsymbol{\phi}_0,\boldsymbol{\phi}_1)}{\partial p_1^2}\right]=\sum_{j_1=1}^{J}\cdots\sum_{j_K=1}^{J}\left[\mathrm{E}[n(j_1,\cdots,j_K)]\left(\frac{\pi_1(j_1,\cdots,j_K)}{p_1}-\frac{\pi_0(j_1,\cdots,j_K)}{p_0}\right)^2\right],$$

$$\mathrm{E}\left[-\frac{\partial^2 l(p_1,\boldsymbol{\phi}_0,\boldsymbol{\phi}_1)}{\partial p_1\partial\phi_{0kj}}\right]=\sum_{j_1=1}^{J}\cdots\sum_{j_{k-1}=1}^{J}\sum_{j_{k+1}=1}^{J}\cdots\sum_{j_K=1}^{J}\left[\mathrm{E}[n(j_k=j)]\left(\frac{\pi_0(j_k=j)\pi_1(j_k=j)}{p_0p_1\phi_{0kj}}\right)-\mathrm{E}[n(j_k=J)]\left(\frac{\pi_0(j_k=J)\pi_1(j_k=J)}{p_0p_1\phi_{0kJ}}\right)\right],$$

$$\mathrm{E}\left[-\frac{\partial^2 l(p_1,\boldsymbol{\phi}_0,\boldsymbol{\phi}_1)}{\partial p_1\partial\phi_{1kj}}\right]=\sum_{j_1=1}^{J}\cdots\sum_{j_{k-1}=1}^{J}\sum_{j_{k+1}=1}^{J}\cdots\sum_{j_K=1}^{J}\left[-E[n(j_k=j)]\left(\frac{\pi_0(j_k=j)\pi_1(j_k=j)}{p_0p_1\phi_{1kj}}\right)+\mathrm{E}[n(j_k=J)]\left(\frac{\pi_0(j_k=J)\pi_1(j_k=J)}{p_0p_1\phi_{1kJ}}\right)\right],$$

$$\mathrm{E}\left[-\frac{\partial^2 l(p_1,\boldsymbol{\phi}_0,\boldsymbol{\phi}_1)}{\partial\phi_{0kj}\partial\phi_{0kj}}\right]=\sum_{j_1=1}^{J}\cdots\sum_{j_{k-1}=1}^{J}\sum_{j_{k+1}=1}^{J}\cdots\sum_{j_K=1}^{J}\left[\mathrm{E}[n(j_k=j)]\left(\frac{\pi_0(j_k=j)}{\phi_{0kj}}\right)^2 +\mathrm{E}[n(j_k=J)]\left(\frac{\pi_0(j_k=J)}{\phi_{0kJ}}\right)^2\right],$$

$$\mathrm{E}\left[-\frac{\partial^2 l(p_1,\boldsymbol{\phi}_0,\boldsymbol{\phi}_1)}{\partial\phi_{0kj_1}\partial\phi_{0kj_2}}\right]=\sum_{j_1=1}^{J}\cdots\sum_{j_{k-1}=1}^{J}\sum_{j_{k+1}=1}^{J}\cdots\sum_{j_K=1}^{J}\left[\mathrm{E}[n(j_k=J)]\left(\frac{\pi_0(j_k=J)}{\phi_{0kJ}}\right)^2\right],$$

$$\begin{aligned}E\left[-\frac{\partial^2 l(p_1,\boldsymbol{\phi}_0,\boldsymbol{\phi}_1)}{\partial\phi_{0k_1j_1}\partial\phi_{0k_2j_2}}\right]=&\sum_{j_1=1}^{J}\cdots\sum_{j_{k_1-1}=1}^{J}\sum_{j_{k_1+1}=1}^{J}\cdots\sum_{j_{k_2-1}=1}^{J}\sum_{j_{k_2+1}=1}^{J}\cdots\sum_{j_K=1}^{J}\Bigg[\\ &\mathrm{E}[n(j_{k_1}=j_1,j_{k_2}=J)]\left(\frac{\pi_0(j_{k_1}=j_1,j_{k_2}=J)\pi_1(j_{k_1}=j_1,j_{k_2}=J)}{\phi_{0k_1j_1}\phi_{0k_2J}}\right)\\ &+\mathrm{E}[n(j_{k_1}=J,j_{k_2}=j_2)]\left(\frac{\pi_0(j_{k_1}=J,j_{k_2}=j_2)\pi_1(j_{k_1}=J,j_{k_2}=j_2)}{\phi_{0k_1J}\phi_{0k_2j_2}}\right)\\ &-\mathrm{E}[n(j_{k_1}=j_1,j_{k_2}=j_2)]\left(\frac{\pi_0(j_{k_1}=j_1,j_{k_2}=j_2)\pi_1(j_{k_1}=j_1,j_{k_2}=j_2)}{\phi_{0k_1j_1}\phi_{0k_2j_2}}\right)\\ &-\mathrm{E}[n(j_{k_1}=J,j_{k_2}=J)]\left(\frac{\pi_0(j_{k_1}=J,j_{k_2}=J)\pi_1(j_{k_1}=J,j_{k_2}=J)}{\phi_{0k_1J}\phi_{0k_2J}}\right)\Bigg],\end{aligned}$$

$$\mathrm{E}\left[-\frac{\partial^2 l(p_1,\boldsymbol{\phi}_0,\boldsymbol{\phi}_1)}{\partial\phi_{0kj}\partial\phi_{1kj}}\right]=\sum_{j_1=1}^{J}\cdots\sum_{j_{k-1}=1}^{J}\sum_{j_{k+1}=1}^{J}\cdots\sum_{j_K=1}^{J}\left[\mathrm{E}[n(j_k=j)]\left(\frac{\pi_0(j_k=j)\pi_1(j_k=j)}{\phi_{0kj}\phi_{1kj}}\right) +E[n(j_k=J)]\left(\frac{\pi_0(j_k=J)\pi_1(j_k=J)}{\phi_{0kJ}\phi_{1kJ}}\right)\right],$$

$$\mathrm{E}\left[-\frac{\partial^2 l(p_1,\boldsymbol{\phi}_0,\boldsymbol{\phi}_1)}{\partial\phi_{0kj_1}\partial\phi_{1kj_2}}\right]= \sum_{j_1=1}^{J}\cdots\sum_{j_{k-1}=1}^{J}\sum_{j_{k+1}=1}^{J}\cdots\sum_{j_K=1}^{J}\left[\mathrm{E}[n(j_k=J)]\left(\frac{\pi_0(j_k=J)\pi_1(j_k=J)}{\phi_{0kJ}\phi_{1kJ}}\right)\right],$$

$$\begin{aligned}\mathrm{E}\left[-\frac{\partial^2 l(p_1,\boldsymbol{\phi}_0,\boldsymbol{\phi}_1)}{\partial\phi_{0k_1j_1}\partial\phi_{1k_2j_2}}\right]=&\sum_{j_1=1}^{J}\cdots\sum_{j_{k_1-1}=1}^{J}\sum_{j_{k_1+1}=1}^{J}\cdots\sum_{j_{k_2-1}=1}^{J}\sum_{j_{k_2+1}=1}^{J}\cdots\sum_{j_K=1}^{J}\Bigg[\\ &-\mathrm{E}[n(j_{k_1}=j_1,j_{k_2}=J)]\left(\frac{\pi_0(j_{k_1}=j_1,j_{k_2}=J)\pi_1(j_{k_1}=j_1,j_{k_2}=J)}{\phi_{0k_1j_1}\phi_{1k_2J}}\right)\\ &-\mathrm{E}[n(j_{k_1}=J,j_{k_2}=j_2)]\left(\frac{\pi_0(j_{k_1}=J,j_{k_2}=j_2)\pi_1(j_{k_1}=J,j_{k_2}=j_2)}{\phi_{0k_1J}\phi_{1k_2j_2}}\right)\\ &+\mathrm{E}[n(j_{k_1}=j_1,j_{k_2}=j_2)]\left(\frac{\pi_0(j_{k_1}=j_1,j_{k_2}=j_2)\pi_1(j_{k_1}=j_1,j_{k_2}=j_2)}{\phi_{0k_1j_1}\phi_{1k_2j_2}}\right)\\ &+\mathrm{E}[n(j_{k_1}=J,j_{k_2}=J)]\left(\frac{\pi_0(j_{k_1}=J,j_{k_2}=J)\pi_1(j_{k_1}=J,j_{k_2}=J)}{\phi_{0k_1J}\phi_{1k_2J}}\right)\Bigg],\end{aligned}$$

$$\begin{aligned}\mathrm{E}\left[-\frac{\partial^2 l(p_1,\boldsymbol{\phi}_0,\boldsymbol{\phi}_1)}{\partial\phi_{1kj}\partial\phi_{1kj}}\right]=&\sum_{j_1=1}^{J}\cdots\sum_{j_{k-1}=1}^{J}\sum_{j_{k+1}=1}^{J}\cdots\sum_{j_K=1}^{J}\Bigg[\mathrm{E}[n(j_k=j)]\left(\frac{\pi_0(j_k=j)}{\phi_{0kj}}\right)^2\\ &+\mathrm{E}[n(j_k=J)]\left(\frac{\pi_0(j_k=J)}{\phi_{0kJ}}\right)^2\Bigg],\end{aligned}$$

$$\mathrm{E}\left[-\frac{\partial^2 l(p_1,\boldsymbol{\phi}_0,\boldsymbol{\phi}_1)}{\partial\phi_{1kj_1}\partial\phi_{1kj_2}}\right]=\sum_{j_1=1}^{J}\cdots\sum_{j_{k-1}=1}^{J}\sum_{j_{k+1}=1}^{J}\cdots\sum_{j_K=1}^{J}\left[\mathrm{E}[n(j_k=J)]\left(\frac{\pi_0(j_k=J)}{\phi_{0kJ}}\right)^2\right],$$

and

$$\begin{aligned}
\mathrm{E}\left[-\frac{\partial^2 l(p_1,\boldsymbol{\phi}_0,\boldsymbol{\phi}_1)}{\partial\phi_{1k_1j_1}\partial\phi_{1k_2j_2}}\right] = \sum_{j_1=1}^{J}\cdots\sum_{j_{k_1-1}=1}^{J}\sum_{j_{k_1+1}=1}^{J}\cdots\sum_{j_{k_2-1}=1}^{J}\sum_{j_{k_2+1}=1}^{J}\cdots\sum_{j_K=1}^{J}\Bigg[\\
\mathrm{E}[n(j_{k_1}=j_1,j_{k_2}=J)]\left(\frac{\pi_0(j_{k_1}=j_1,j_{k_2}=J)\pi_1(j_{k_1}=j_1,j_{k_2}=J)}{\phi_{0k_1j_1}\phi_{0k_2J}}\right) \\
+\mathrm{E}[n(j_{k_1}=J,j_{k_2}=j_2)]\left(\frac{\pi_0(j_{k_1}=J,j_{k_2}=j_2)\pi_1(j_{k_1}=J,j_{k_2}=j_2)}{\phi_{0k_1J}\phi_{0k_2j_2}}\right) \\
-\mathrm{E}[n(j_{k_1}=j_1,j_{k_2}=j_2)]\left(\frac{\pi_0(j_{k_1}=j_1,j_{k_2}=j_2)\pi_1(j_{k_1}=j_1,j_{k_2}=j_2)}{\phi_{0k_1j_1}\phi_{0k_2j_2}}\right) \\
-\mathrm{E}[n(j_{k_1}=J,j_{k_2}=J)]\left(\frac{\pi_0(j_{k_1}=J,j_{k_2}=J)\pi_1(j_{k_1}=J,j_{k_2}=J)}{\phi_{0k_1J}\phi_{0k_2J}}\right)\Bigg].
\end{aligned}$$

We can prove this using calculus and algebra. Since the proof is straightforward but requires tedious calculus and algebra operations, we omit the proof.

Appendix B

Proof of Invariance Property of ROC Curve Areas

From the expression (3) on an ROC curve area, we obtain that

$$\begin{aligned}
&A_k(\boldsymbol{\phi}_{0k\bullet},\boldsymbol{\phi}_{1k\bullet}) + A_k(\boldsymbol{\phi}_{1k\bullet},\boldsymbol{\phi}_{0k\bullet}) \\
&= \sum_{j=1}^{J-1}\left[\phi_{0kj}\sum_{l=j+1}^{J}\phi_{1kl}\right] + \frac{1}{2}\sum_{j=1}^{J}\phi_{0kj}\phi_{1kj} + \sum_{l=1}^{J-1}\left[\phi_{1kl}\sum_{j=l+1}^{J}\phi_{0kj}\right] + \frac{1}{2}\sum_{j=1}^{J}\phi_{1kj}\phi_{0kj} \\
&= \sum_{j=1}^{J-1}\left[\phi_{0kj}\sum_{l=j+1}^{J}\phi_{1kl}\right] + \sum_{l=1}^{J-1}\left[\phi_{1kl}\sum_{j=l+1}^{J}\phi_{0kj}\right] + \sum_{j=1}^{J}\phi_{0kj}\phi_{1kj}.
\end{aligned}$$

Note that

$$\sum_{l=1}^{J-1}\phi_{1kl}\sum_{j=l+1}^{J}\phi_{0kj}=\sum_{j=2}^{J}\phi_{0kj}\sum_{l=1}^{j-1}\phi_{1kl}$$

Then we find

$$\begin{aligned}
&A_k(\boldsymbol{\phi}_{0k\bullet},\boldsymbol{\phi}_{1k\bullet})+A_k(\boldsymbol{\phi}_{1k\bullet},\boldsymbol{\phi}_{0k\bullet})\\
=&\sum_{j=1}^{J-1}\left[\phi_{0kj}\sum_{l=j+1}^{J}\phi_{1kl}\right]+\sum_{j=2}^{J}\left[\phi_{0kj}\sum_{l=1}^{j-1}\phi_{1kl}\right]+\sum_{j=1}^{J}\phi_{0kj}\phi_{1kj}\\
=&\phi_{0k1}\left(\sum_{l=2}^{J}\phi_{1kl}\right)+\sum_{j=2}^{J-1}\left[\phi_{0kj}\sum_{l=j+1}^{J}\phi_{1kl}\right]+\phi_{0kJ}\left(\sum_{l=1}^{J-1}\phi_{1kl}\right)+\sum_{j=2}^{J-1}\left[\phi_{0kj}\sum_{l=1}^{j-1}\phi_{1kl}\right]+\sum_{j=1}^{J}\phi_{0kj}\phi_{1kj}\\
=&\phi_{0k1}\left(\sum_{l=2}^{J}\phi_{1kl}\right)+\phi_{0kJ}\left(\sum_{l=1}^{J-1}\phi_{1kl}\right)+\sum_{j=2}^{J-1}\left[\phi_{0kj}\left(\sum_{l=1}^{j-1}\phi_{1kl}+\sum_{l=j+1}^{J}\phi_{1kl}\right)\right]+\sum_{j=1}^{J}\phi_{0kj}\phi_{1kj}\\
=&\phi_{0k1}\left(1-\phi_{1k1}\right)+\phi_{0kJ}\left(1-\phi_{1kJ}\right)+\sum_{j=2}^{J-1}\phi_{0kj}\left(1-\phi_{1kj}\right)+\sum_{j=1}^{J}\phi_{0kj}\phi_{1kj}\\
=&\sum_{j=1}^{J}\phi_{0kj}\left(1-\phi_{1kj}\right)+\sum_{j=1}^{J}\phi_{0kj}\phi_{1kj}\\
=&\sum_{j=1}^{J}\phi_{0kj}-\sum_{j=1}^{J}\phi_{0kj}\phi_{1kj}+\sum_{j=1}^{J}\phi_{0kj}\phi_{1kj}=\sum_{j=1}^{J}\phi_{0kj}\\
=&1,
\end{aligned}$$

where we have repeatedly used the fact $\sum_{j=1}^{J}\phi_{0kj}=\sum_{j=1}^{J}\phi_{1kj}=1$.

Thus we conclude that

$$A_k(\boldsymbol{\phi}_{0k\bullet},\boldsymbol{\phi}_{1k\bullet})+A_k(\boldsymbol{\phi}_{1k\bullet},\boldsymbol{\phi}_{0k\bullet})=1$$

List of Figures

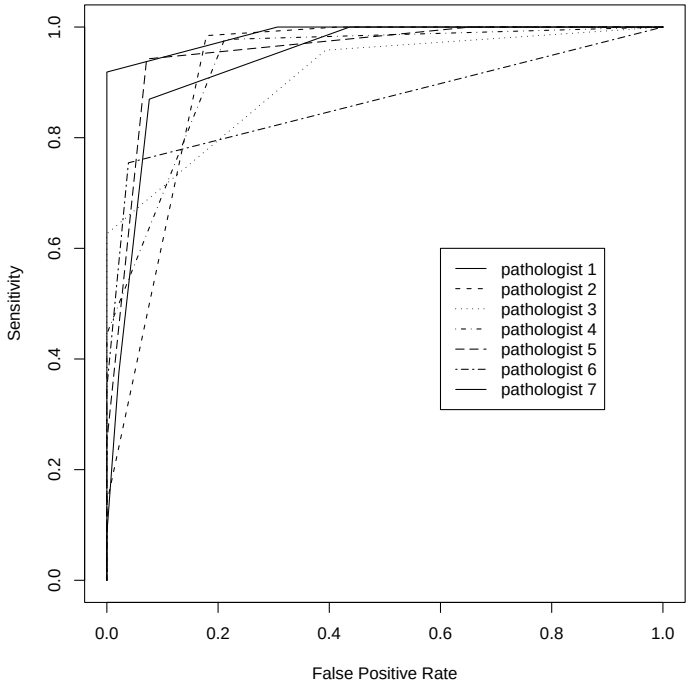

Figure 1. Estimated ROC curves for each of the seven pathologist based on non-parametric model

List of Tables

Table 1

Results From 500 Simulations With N=118, J=5, K=7 Under Various Parameter Settings

True	True Areas		0.7	0.7	0.7	0.7	0.7	0.7	0.7
Prevalence	Statistic	p_1	A_1	A_2	A_3	A_4	A_5	A_6	A_7
0.5	Bias	-0.0186	-0.0127	-0.0109	-0.0140	-0.0135	-0.0094	-0.0142	-0.0164
	MSE	0.0305	0.0089	0.0083	0.0095	0.0090	0.0082	0.0087	0.0091
0.7	Bias	-0.1098	-0.0266	-0.0214	-0.0262	-0.0321	-0.0254	-0.0352	-0.0272
	MSE	0.0497	0.0127	0.0109	0.0127	0.0121	0.0106	0.0117	0.0128
0.9	Bias	-0.3398	-0.1065	-0.1021	-0.0960	-0.1082	-0.0980	-0.1059	-0.0983
	MSE	0.1604	0.0277	0.0259	0.0241	0.0262	0.0235	0.0254	0.0240
True	True Areas		0.8	0.8	0.8	0.8	0.8	0.8	0.8
Prevalence	Statistic	p_1	A_1	A_2	A_3	A_4	A_5	A_6	A_7
0.5	Bias	-0.0067	0.0012	0.0008	-0.0042	0.0022	-0.0008	-0.0059	0.0007
	MSE	0.0039	0.0024	0.0025	0.0025	0.0026	0.0024	0.0027	0.0025
0.7	Bias	-0.0109	0.0000	-0.0057	-0.0035	-0.0041	-0.0037	-0.0079	0.0014
	MSE	0.0042	0.0031	0.0032	0.0031	0.0031	0.0033	0.0033	0.0033
0.9	Bias	-0.0947	-0.0593	-0.0595	-0.0596	-0.0615	-0.0602	-0.0605	-0.0543
	MSE	0.0263	0.0170	0.0169	0.0169	0.0162	0.0159	0.0167	0.0157
True	True Areas		0.9	0.9	0.9	0.9	0.9	0.9	0.9
Prevalence	Statistic	p_1	A_1	A_2	A_3	A_4	A_5	A_6	A_7
0.5	Bias	0.0022	0.0007	0.0008	0.0007	0.0005	0.0005	-0.0001	0.0006
	MSE	0.0023	0.0007	0.0007	0.0007	0.0006	0.0007	0.0007	0.0007
0.7	Bias	0.0021	0.0009	0.0001	0.0000	0.0013	0.0000	-0.0003	0.0001
	MSE	0.0021	0.0009	0.0008	0.0008	0.0009	0.0008	0.0009	0.0008
0.9	Bias	0.0019	-0.0004	-0.0019	-0.0002	-0.0023	-0.0011	0.0008	-0.0042
	MSE	0.0007	0.0024	0.0023	0.0021	0.0026	0.0024	0.0023	0.0023

Table 2

Results From 500 Simulations With N=500, J=5, K=7 Under Various Parameter Settings

true	True Areas		0.7	0.7	0.7	0.7	0.7	0.7	0.7
Prevalence	Statistic	p_1	A_1	A_2	A_3	A_4	A_5	A_6	A_7
0.5	Bias	-0.0003	0.0003	0.0010	0.0021	0.0015	0.0004	0.0007	0.0008
	MSE	0.0065	0.0013	0.0015	0.0014	0.0015	0.0014	0.0014	0.0013
0.7	Bias	0.0006	0.0027	0.0051	0.0005	0.0040	-0.0015	0.0013	0.0047
	MSE	0.0066	0.0018	0.0019	0.0021	0.0022	0.0020	0.0021	0.0018
0.9	Bias	-0.1308	-0.0371	-0.0452	-0.0360	-0.0364	-0.0336	-0.0360	-0.0325
	MSE	0.0818	0.0125	0.0138	0.0116	0.0120	0.0124	0.0125	0.0123
True	True Areas		0.8	0.8	0.8	0.8	0.8	0.8	0.8
Prevalence	Statistic	p_1	A_1	A_2	A_3	A_4	A_5	A_6	A_7
0.5	Bias	0.0001	-0.0017	0.0010	0.0006	-0.0005	-0.0002	-0.0003	0.0007
	MSE	0.0008	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005
0.7	Bias	0.0005	-0.0017	0.0013	0.0006	-0.0005	0.0012	-0.0005	0.0016
	MSE	0.0007	0.0006	0.0007	0.0006	0.0006	0.0007	0.0007	0.0006
0.9	Bias	-0.0022	-0.0019	0.0026	0.0031	-0.0006	-0.0001	-0.0046	0.0009
	MSE	0.0004	0.0020	0.0019	0.0018	0.0020	0.0021	0.0021	0.0019
True	True Areas		0.9	0.9	0.9	0.9	0.9	0.9	0.9
Prevalence	Statistic	p_1	A_1	A_2	A_3	A_4	A_5	A_6	A_7
0.5	Bias	0.0014	-0.0009	0.0007	0.0000	-0.0008	-0.0005	-0.0004	0.0001
	MSE	0.0005	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
0.7	Bias	0.0018	-0.0006	0.0006	0.0001	-0.0007	0.0004	0.0002	0.0011
	MSE	0.0004	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
0.9	Bias	-0.0011	0.0011	-0.0002	0.0007	-0.0008	-0.0001	-0.0006	0.0007
	MSE	0.0002	0.0004	0.0004	0.0004	0.0005	0.0004	0.0005	0.0004

Table 3

Results From 500 Simulations With N=118, J=5, K=7 Under Various Parameter Settings

True	True Areas		0.90	0.87	0.83	0.80	0.77	0.73	0.70
Prevalence	Statistic	p_1	A_1	A_2	A_3	A_4	A_5	A_6	A_7
0.5	Bias	-0.0009	-0.0045	-0.0059	0.0052	0.0039	0.0049	0.0007	0.0042
	MSE	0.0027	0.0009	0.0012	0.0017	0.0019	0.0022	0.0025	0.0027
0.7	Bias	-0.0053	-0.0057	-0.0082	0.0044	0.0074	0.0072	0.0037	0.0032
	MSE	0.0029	0.0012	0.0015	0.0023	0.0025	0.0025	0.0032	0.0039
0.9	Bias	-0.0389	-0.0500	-0.0515	-0.0181	-0.0181	-0.0087	-0.0078	-0.0055
	MSE	0.0082	0.0124	0.0114	0.0085	0.0102	0.0088	0.0088	0.0096

Table 4

Results From 500 Simulations With N=500, J=5, K=7 Under Various Parameter Settings

True	True Areas		0.90	0.87	0.83	0.80	0.77	0.73	0.70
True Prevalence	Statistic	p_1	A_1	A_2	A_3	A_4	A_5	A_6	A_7
0.5	Bias	0.0018	-0.0014	0.0000	0.0016	-0.0001	0.0004	0.0001	0.0006
	MSE	0.0006	0.0002	0.0002	0.0004	0.0004	0.0005	0.0005	0.0005
0.7	Bias	0.0015	-0.0013	0.0001	0.0019	0.0002	0.0016	0.0004	0.0017
	MSE	0.0005	0.0002	0.0003	0.0005	0.0006	0.0006	0.0007	0.0006
0.9	Bias	0.0000	-0.0022	0.0005	0.0058	0.0025	0.0037	-0.0008	0.0040
	MSE	0.0003	0.0007	0.0009	0.0011	0.0015	0.0016	0.0019	0.0017

Table 5

Results From 500 Simulations With N=118, J=5, K=7 Under BiNormal Models

True	True Areas			0.95	0.90	0.90	0.85	0.85	0.80	0.75
Prevalence	Statistic	Model*	p_1	A_1	A_2	A_3	A_4	A_5	A_6	A_7
0.5	Bias	P	0.0034	-0.0046	-0.0024	-0.0070	0.0002	0.0009	0.0002	-0.0014
		NP	-0.0008	-0.0124	-0.0118	-0.0092	-0.0046	-0.0059	0.0038	0.0112
	MSE	P	0.0023	0.0005	0.0012	0.0011	0.0012	0.0014	0.0022	0.0029
		NP	0.0023	0.0007	0.0011	0.0011	0.0014	0.0015	0.0019	0.0020
0.7	Bias	P	-0.0041	-0.0119	-0.0388	-0.0507	0.0103	0.0017	-0.0040	-0.0271
		NP	-0.0064	-0.0106	-0.0125	-0.0104	-0.0074	-0.0077	0.0038	0.0141
	MSE	P	0.0017	0.0004	0.0026	0.0037	0.0010	0.0014	0.0017	0.0037
		NP	0.0020	0.0009	0.0013	0.0012	0.0014	0.0013	0.0014	0.0021
0.9	Bias	P	-0.0205	-0.0663	-0.0736	-0.0649	-0.0494	-0.0346	-0.0569	-0.0497
		NP	-0.0143	-0.0401	-0.0316	-0.0295	-0.0205	-0.0196	-0.0057	0.0022
	MSE	P	0.0012	0.0064	0.0091	0.0065	0.0074	0.0039	0.0067	0.0071
		NP	0.0016	0.0069	0.0054	0.0054	0.0044	0.0038	0.0036	0.0042

$*$ P = parametric model; NP = non-parametric model.

Table 6

Results From 500 Simulations With N=118, J=5, K=7 Under Non-BiNormal Models

True	True Areas			0.95	0.90	0.90	0.85	0.85	0.80	0.75
Prevalence	Statistic	Model*	p_1	A_1	A_2	A_3	A_4	A_5	A_6	A_7
0.5	Bias	P	0.0020	-0.0113	-0.0053	-0.0054	-0.0041	0.0024	0.0019	0.0012
		NP	-0.0039	-0.0118	-0.0117	-0.0127	-0.0070	-0.0043	0.0011	0.0149
	MSE	P	0.0026	0.0005	0.0011	0.0008	0.0015	0.0018	0.0025	0.0021
		NP	0.0021	0.0007	0.0011	0.0012	0.0015	0.0015	0.0018	0.0026
0.7	Bias	P	-0.0033	-0.0107	-0.0083	-0.0074	-0.0031	-0.0023	-0.0001	0.0005
		NP	-0.0006	-0.0122	-0.0136	-0.0099	-0.0066	-0.0047	-0.0018	0.0123
	MSE	P	0.0025	0.0005	0.0010	0.0009	0.0013	0.0016	0.0021	0.0017
		NP	0.0016	0.0008	0.0013	0.0013	0.0015	0.0014	0.0018	0.0022
0.9	Bias	P	-0.0147	-0.0566	-0.0416	-0.0318	-0.0363	-0.0361	-0.0186	-0.0096
		NP	-0.0149	-0.0297	-0.0298	-0.0260	-0.0234	-0.0138	-0.0052	0.0001
	MSE	P	0.0008	0.0043	0.0036	0.0026	0.0035	0.0033	0.0028	0.0041
		NP	0.0012	0.0043	0.0046	0.0040	0.0042	0.0029	0.0040	0.0041

$*$ P = parametric model; NP = non-parametric model.