Statistics and Its Interface Volume 5 (2012) 319–329

Imputation-based empirical likelihood inference for the area under the ROC curve with missing data

Binhuan Wang* and Gengsheng Qin†

In a continuous diagnostic test, the area under the receiver operating characteristic curve (AUC) is commonly used to summarize the diagnostic accuracy of the test. Many current studies on inference of the AUC focus on the complete data case. In this paper, an imputation-based profile empirical likelihood ratio is defined and shown to asymptotically follow a scaled chi-square distribution. Then an empirical likelihood confidence interval for the AUC with missing data is proposed by using the scaled chi-square distribution. The proposed empirical likelihood inference for the AUC is also extended to stratified random samples, and the limiting distribution of the empirical log-likelihood ratio is a weighted summation of independent chi-square distributions with one degree of freedom. Simulation studies are conducted to evaluate the finite sample performance of the proposed method in terms of coverage probability. Additionally, a real example is used to illustrate the proposed method.

Keywords and phrases: AUC, Diagnostic test, Empirical likelihood, Imputation, Missing data, ROC.

1. INTRODUCTION

Diagnostic tests are widely used to detect the occurrence of disease, and monitor disease progression. Sensitivity and specificity are common measures used to evaluate the performance of a diagnostic test. For a continuous-scale test, the disease or non-disease status is dependent upon whether the test result is above or below a specified cut-off point. Let Y and X be the results of a continuous-scale test for a diseased and a non-diseased subject, and assume that F and G are the distribution functions of X and Y, respectively. For a given cut-off point γ, the sensitivity and the specificity of the test are defined by

$$R = P(Y \geq \gamma) = 1 - G(\gamma), \quad Sp = P(X < \gamma) = F(\gamma),$$

respectively. When the cut-off point varies throughout the entire real line, the resulting plot of sensitivity against 1-specificity is called the *Receiver Operating Characteristic* (ROC) curve. Mathematically, the ROC curve can be represented by $R(p) = 1 - G(F^{-1}(p))$, where F^{-1} is the inverse function of F. The area under the curve (AUC), defined as $\delta = \int_0^1 R(p)\mathrm{d}p$, is a commonly used summary measure of the ROC curve. AUC has been frequently used to assess the ability of a diagnostic test to discriminate between individuals with and without a disease.

Bamber [1] showed that the AUC, $\delta = P(Y \geq X)$, which can be interpreted as the probability that in a randomly selected pair of diseased and non-diseased subjects, the test value of the diseased subject is higher than or equal to that of the non-diseased subject. In a more general context, Wolfe and Hogg [22] recommended the use of this index as a general measure for the difference between two distributions. One important problem for the inference on the AUC is how to construct a confidence interval for δ. Let $X_1, \ldots, X_m$ be test results of a random sample of non-diseased subjects and $Y_1, \ldots, Y_n$ be test results of a random sample of diseased subjects. Traditionally, the classical Wilcoxon-Mann-Whitney (WMW) [9] two-sample rank statistic, defined by

$$\delta_{m,n} = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} I(Y_j \geq X_i), \tag{1}$$

is used as a nonparametric estimator of the AUC. Based on the asymptotic normality of the WMW statistic, we can construct a confidence interval (hereafter MW interval) for the AUC. Although the WMW estimator of the AUC is known to be unbiased, the normal approximation-based MW interval suffers from low coverage accuracy for high values of the AUC (e.g., 0.90 to 0.95, which are of most interest in diagnostic tests) when sample sizes of diseased and non-diseased subjects are small and unequal. Therefore, it is desirable to find a reliable alternative approach for constructing a confidence interval of the AUC.

In making statistical inference, samples are usually assumed to be complete. However, due to various reasons, missing data instead of complete data occur commonly in practical situations such as opinion polls, market research surveys, and other scientific and social fields. Missing data

*Corresponding author.
†The project was supported in part by a NSF grant and a NSA grant. We would like to thank the editor and the associate editor for their comments which led to substantial improvements in this article.

are also common in medical diagnostic studies. A naive method is to use complete data and discard missing data. But this method suffers from a loss of efficiency for many reasons, such as small sample size. Other methods include available-case methods and imputation methods. See Little and Rubin [8] for a comprehensive overview. Missing data bring challenges in ROC curve analysis, too. Various methods, including imputation-based methods, have been proposed in order to handle the problems caused by missing data. Geert et al. [4] evaluated five different methods in dealing with missing values in the empirical data from a study among patients suspected of pulmonary embolism, and they found that imputation is relatively better than others. He et al. [6] provided a direct estimate of the AUC in the presence of verification bias.

Empirical likelihood (EL) [10, 11] is a popular nonparametric method traditionally used for providing confidence intervals for the mean. Chen and Van Keilegom [3] provided a general review on empirical likelihood method for regressions. Based on the mean-like form of WMW estimator, Qin and Zhou [13] proposed an EL approach for the inference on the AUC, which was shown to have good small sample performance. Motivated by the asymptotic independence of pseudo-values from jackknife technique, Jing, Yuan and Zhou [7] introduced the jackknife empirical likelihood (JEL) method for U-statistics, and used the AUC as an example to illustrate their method because the WMW estimator is a two-sample U-statistics. Recently, Gong, Peng and Qi [5] proposed a smoothed JEL method for the ROC curve. Empirical likelihood is also a powerful method for handling missing data problems. For example, Wang and Chen [20] applied empirical likelihood to estimating equations with missing data. Qin, Zhang and Leung [14] proposed a unified empirical likelihood approach to missing data problems and explored the use of empirical likelihood to effectively combine unbiased estimating equations when the number of estimating equations is greater than the number of unknown parameters. Wang and Qin [19] constructed imputation-based empirical likelihood confidence intervals for the sensitivity of a continuous-scale diagnostic test with missing data. In this paper, we propose an imputation-based empirical likelihood method to construct confidence interval for the AUC with missing completely at random (MCAR) type of data, which has not been considered in literature. It is necessary to point out that the MCAR assumption is reasonable in practice. For example, in diagnostic studies, a laboratory sample may be dropped, so the resulting observation is missing; some sampled subjects are lost follow-up due to moving and other reasons. In these cases, missing is independent of both observable variables and unobservable variables of interest. The proposed method preserves the advantage of the method in Qin and Zhou [13], which has good small sample performance, and the advantage of the hot deck imputation method, which preserves the distribution of item values whereas the deterministic imputation methods like the ratio imputation and the regression imputation do not have this appealing property [16].

The remainder of the paper is organized as follows. Section 2 presents the imputation-based empirical likelihood method to construct confidence intervals for the AUC with missing data. In section 3, we conduct simulation studies to evaluate the performance of the proposed method. In section 4, we apply the new imputation-based empirical likelihood interval to a real example. A brief discussion is given in section 5. All proofs are deferred until the Appendix.

2. IMPUTATION-BASED EMPIRICAL LIKELIHOOD FOR THE AUC

In this section, we aim at constructing empirical likelihood-based confidence intervals for δ with missing data. We firstly impute the missing data by the hot deck imputation technique, and then apply empirical likelihood method to obtain confidence intervals for the AUC based on the imputed data. Finally we extend the proposed method to stratified random samples with missing data.

2.1 Point estimation of the AUC with missing data

Let $(X_1, \delta_{X_1}), \ldots, (X_m, \delta_{X_m})$ and $(Y_1, \delta_{Y_1}), \ldots, (Y_n, \delta_{Y_n})$ be the simple random sample sequences of incomplete data associated with the populations (X, δ_X) and (Y, δ_Y) respectively, where

$$\delta_{X_i} = \begin{cases} 0, & \text{if } X_i \text{ is missing} \\ 1, & \text{if } X_i \text{ is observed} \end{cases}, \quad i = 1, \ldots, m,$$

$$\delta_{Y_j} = \begin{cases} 0, & \text{if } Y_j \text{ is missing} \\ 1, & \text{if } Y_j \text{ is observed} \end{cases}, \quad j = 1, \ldots, n.$$

Missing data are common in many situations. For example, patients involved in a regular blood or urine test in a medical diagnosis might quit the research because they moved to other districts, or missed a visit to the hospital due to bad weather or schedule conflicts. Also, in some kinds of clinical trials, the measurement of genes related to a specific cancer could be missing due to the limit of equipment or cost requirements. These situations have an insight that such kinds of missingness are unrelated to any patient's characteristics. This class of missingness is classified as missing completely at random (MCAR, see [15, 8]). Throughout this paper, motivated by these observations, we assume X and Y are MCAR, i.e.,

$$P(\delta_X|X) = \pi_1, \quad \text{and} \quad P(\delta_Y|Y) = \pi_2,$$

where both π_1 and π_2 are constants belonging to $(0, 1)$.

For convenience, some standard notations are needed. Let $r_X = \sum_{i=1}^{m} \delta_{X_i}$, $r_Y = \sum_{j=1}^{n} \delta_{Y_j}$, $m_X = m - r_X$ and $m_Y = n - r_Y$. Denote the sets of observed data with respect to X and Y as S_{r_X} and S_{r_Y} respectively, and the sets of missing data with respect to X and Y as S_{m_X} and S_{m_Y} respectively. Then the means of the observed data with respect to X and Y are denoted as $\bar{X}_r = \frac{1}{r_X}\sum_{i\in S_{r_X}} X_i$ and $\bar{Y}_r = \frac{1}{r_Y}\sum_{j\in S_{r_Y}} Y_j$, respectively. Furthermore, let X_i^* and Y_j^* be the imputed values for the missing data with respect to X and Y, respectively.

An imputation method is useful in dealing with missing data. With MCAR data, we prefer a random hot deck imputation method to impute the missing values rather than the deterministic imputation, because the latter one is not appropriate in making an inference of distribution functions [2]. The idea of random hot deck imputation [16] is natural. For set $\{(X_i, \delta_{X_i}), i = 1, \ldots, m\}$, the random hot deck imputation draws a simple random sample of size m_X with replacement from S_{r_X}, and then let $X_i^* = X_k$ for some $k \in S_{r_X}$. After imputation, a sample of so-called "complete data" is obtained as follows:

$$\tilde{X}_i = \delta_{X_i} X_i + (1-\delta_{X_i})X_i^*, \quad i = 1, \ldots, m.$$

Similarly, the imputed "complete data" of Y could be obtained as follows:

$$\tilde{Y}_j = \delta_{Y_j} Y_j + (1-\delta_{Y_j})Y_j^*, \quad j = 1, \ldots, n.$$

Wang and Qin [19] have proved that based on the imputed data $\tilde{X}_i$'s and $\tilde{Y}_j$'s, the empirical distributions $\widetilde{F}(x) = \frac{1}{m}\sum_{i=1}^m I(\tilde{X}_i \le x)$, and $\widetilde{G}(y) = \frac{1}{n}\sum_{j=1}^n I(\tilde{Y}_j \le y)$ are still consistent and asymptotically normal.

We define the imputed version of WMW estimator for the AUC as follows:

$$\widetilde{\delta} = \frac{1}{mn}\sum_{i=1}^m \sum_{j=1}^n I(\tilde{Y}_j \ge \tilde{X}_i). \tag{2}$$

2.2 Empirical likelihood for the AUC

In order to obtain better confidence intervals for the AUC, Qin and Zhou [13] proposed an empirical likelihood-based interval for the AUC. This interval has good coverage accuracy for high values of the AUC when sample sizes for diseased and non-diseased subjects are small and unequal.

For a given test value Y from a diseased subject, let $U = 1 - F(Y)$. The value U can be interpreted as the proportion of the non-diseased population with a test value greater than Y [12]. It is easy to obtain the following equality:

$$E(1-U) = E(F(Y)) = P(Y \ge X) = \delta.$$

Based on the relationship between δ and U, an empirical likelihood procedure for the inference of the AUC was derived by Qin and Zhou [13]. Let $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ be a probability vector, i.e., $\sum_{j=1}^n p_j = 1$ and $p_j \ge 0$ for all j. The empirical likelihood for the AUC, evaluated at the true value δ_0 of δ, is defined as follows:

$$L(\delta_0) = \sup\left\{\prod_{j=1}^n p_j : \sum_{j=1}^n p_j = 1, \sum_{j=1}^n p_j W_j(\delta_0) = 0\right\},$$

where $W_j(\delta_0) = 1 - U_j - \delta_0$ with $U_j = 1 - F(Y_j)$, $j = 1, 2, \ldots, n$. Since the unknown distribution function F of the non-diseased population can be replaced by its empirical distribution $F_m(x) = \frac{1}{m}\sum_{i=1}^m I(X_i \le x)$, then a profile empirical likelihood (PEL) for δ_0 can be given by

$$\hat{L}(\delta_0) = \sup\left\{\prod_{j=1}^n p_j : \sum_{j=1}^n p_j = 1, \sum_{j=1}^n p_j \hat{W}_j(\delta_0) = 0\right\},$$

where $\hat{W}_j(\delta_0) = 1 - \hat{U}_j - \delta_0$ with $\hat{U}_j = 1 - F_m(Y_j)$, $j = 1, 2, \ldots, n$. By the standard procedure of empirical likelihood method, the empirical likelihood ratio for δ_0 can be defined as follows:

$$R(\delta_0) = \prod_{j=1}^n (np_j) = \prod_{j=1}^n \left\{1 + \hat{\lambda}\hat{W}_j(\delta_0)\right\},$$

where $\hat{\lambda}$ is the solution of

$$\frac{1}{n}\sum_{j=1}^n \frac{\hat{W}_j(\delta_0)}{1 + \lambda \hat{W}_j(\delta_0)} = 0. \tag{3}$$

Then the corresponding log-EL ratio is

$$l(\delta_0) \equiv -2\log R(\delta_0) = 2\sum_{j=1}^n \log\left\{1 + \hat{\lambda}\hat{W}_j(\delta_0)\right\}. \tag{4}$$

Qin and Zhou [13] proved that the limiting distribution of $l(\delta_0)$ is a scaled chi-square distribution.

2.3 Imputation-based empirical likelihood interval for the AUC

Based on the imputed data $\tilde{X}_i$'s and $\tilde{Y}_j$'s, we could substitute all complete data X_i's and Y_j's in the previous part and obtain the similar log-EL ratio for δ_0 as follows:

$$\widetilde{l}(\delta_0) = 2\sum_{j=1}^n \log\left\{1 + \widetilde{\lambda}\widetilde{W}_j(\delta_0)\right\}. \tag{5}$$

where $\widetilde{W}_j(\delta_0) = 1 - \widetilde{U}_j - \delta_0$ with $\widetilde{U}_j = 1 - \widetilde{F}(\tilde{Y}_j)$, $j = 1, 2, \ldots, n$, and $\widetilde{\lambda}$ is the solution of

$$\frac{1}{n}\sum_{j=1}^n \frac{\widetilde{W}_j(\delta_0)}{1 + \widetilde{\lambda}\widetilde{W}_j(\delta_0)} = 0. \tag{6}$$

The following theorem establishes the asymptotic distribution of the imputation-based empirical log-likelihood ratio for the AUC.

Theorem 1. *Let δ_0 be the true value of the AUC. If $\lim_{m,n\to\infty} \frac{n}{m} = \tau < \infty$, a fixed quantity, then the asymptotic distribution of $\widetilde{l}(\delta_0)$, defined by (5), is a scaled χ^2 distribution with degree of freedom one, i.e.,*

$$r(\delta_0)\widetilde{l}(\delta_0) \xrightarrow{d} \chi_1^2, \tag{7}$$

where the scale constant $r(\delta_0)$ is

$$r(\delta_0) = \frac{m}{m+n} \frac{\sum_{j=1}^n \widetilde{W}_j^2(\delta_0)}{nS^2}$$

with

$$S^2 = \frac{m(1-\pi_2+\pi_2^{-1})S_{01}^2 + n(1-\pi_1+\pi_1^{-1})S_{10}^2}{m+n},$$

$$S_{10}^2 = \frac{1}{(m-1)n^2}\left[\sum_{i=1}^m (R_i - i)^2 - m\left(\bar{R} - \frac{m+1}{2}\right)^2\right],$$

$$S_{01}^2 = \frac{1}{(n-1)m^2}\left[\sum_{j=1}^n (S_j - j)^2 - n\left(\bar{S} - \frac{n+1}{2}\right)^2\right],$$

$$\bar{R} = \frac{1}{m}\sum_{i=1}^m R_i, \quad \text{and} \quad \bar{S} = \frac{1}{n}\sum_{j=1}^n S_j.$$

Here R_i is the rank of $\tilde{X}_{(i)}$ (the i-th ordered value among $\tilde{X}_i$'s) in the combined sample of $\tilde{X}_i$'s and $\tilde{Y}_j$'s, and S_j is the rank of $\tilde{Y}_{(j)}$ (the j-th ordered value among $\tilde{Y}_j$'s) in the combined sample of $\tilde{X}_i$'s and $\tilde{Y}_j$'s.

If only complete observations are used without applying random hot deck imputation, asymptotic distributions of empirical distributions with observed data only were obtained in [19]. Define

$$\widetilde{F}^*(x) = \frac{1}{r_X}\sum_{i\in S_{r_X}} I(X_i \le x),$$

$$\widetilde{G}^*(y) = \frac{1}{r_Y}\sum_{j\in S_{r_Y}} I(Y_j \le y).$$

Then we have that

$$\sqrt{m}(\widetilde{F}^*(x) - F(x)) \xrightarrow{d} \mathcal{N}(0, \sigma_X^{*2})$$

where $\sigma_X^{*2} = \pi_1^{-1}F(x)(1-F(x))$, and

$$\sqrt{n}(\widetilde{G}^*(y) - G(y)) \xrightarrow{d} \mathcal{N}(0, \sigma_Y^{*2})$$

where $\sigma_Y^{*2} = \pi_2^{-1}G(y)(1-G(y))$.

The above results for $\widetilde{F}^*(x)$ and $\widetilde{G}^*(y)$ are slightly different from Lemma 1 in Appendix. Without the random hot deck imputation, some terms in Lemma 1 are absent. Actually, it is equivalent to disregard missing data and apply the method based on complete data to the observed data only. When sample sizes are small and missing proportion is high, the performances of the method with observed data only may be unstable because missingness results in an even smaller sample size. However, the proposed method will benefit from the imputation. Similar results were observed in simulation studies in [19].

The confidence interval for the AUC could be constructed based on Theorem 1. Intuitively, by plugging in the consistent estimates of all unknown quantities, we could get the plug-in form confidence interval. Let $\tilde{\pi}_1 = \frac{r_X}{m}$, $\tilde{\pi}_2 = \frac{r_Y}{n}$, and

$$\widetilde{S}^2 = \frac{m(1-\tilde{\pi}_2+\tilde{\pi}_2^{-1})S_{01}^2 + n(1-\tilde{\pi}_1+\tilde{\pi}_1^{-1})S_{10}^2}{m+n},$$

$$r(\widetilde{\delta}) = \frac{m}{m+n}\frac{\sum_{j=1}^n \widetilde{W}_j(\widetilde{\delta})}{n\widetilde{S}^2},$$

where $\widetilde{\delta}$ is defined by (2). Then a $(1-\alpha)100\%$ imputation-based profile empirical likelihood confidence interval for δ_0, denoted by IPEL interval, can be defined as follows:

$$R_\alpha(\delta) = \{\delta : r(\widetilde{\delta})\widetilde{l}(\delta) \le \chi_1^2(1-\alpha)\}, \tag{8}$$

where $\chi_1^2(1-\alpha)$ is the $(1-\alpha)100\%$ quantile of the chi-square distribution with degree of freedom one.

2.4 Imputation-based EL intervals for the AUC with stratified samples

In this section, we extend the IPEL method in the previous section to stratified samples. Suppose L institutions participate in a ROC study of continuous-scale diagnostic test, which are indexed by l. Let X_l and Y_l be the results of a continuous-scale test for a non-diseased and a diseased subject in the lth institution, and F_l and G_l be the corresponding distribution functions, respectively. Let $X_{l1},\ldots,X_{lm_l}$ be the test results of a random sample of non-diseased patients, $Y_{l1},\ldots,Y_{ln_l}$ be results of a random sample of diseased subjects in the l-th institution, and the observation rate pairs of each institution be (π_{l1},π_{l2}), $1\le l\le L$. Based on the MCAR assumption and the hot deck imputation technique, the imputed data $\widetilde{X}_{l1},\ldots,\widetilde{X}_{lm_l}$ and $\widetilde{Y}_{l1},\ldots,\widetilde{Y}_{ln_l}$ could be obtained for each institution.

Similar to Qin and Zhou [13], we do not assume that F_l's and G_l's are homogeneous institutions. Instead, we only assume $\delta_1 = \cdots = \delta_l = \delta$, where δ_l denotes the AUC for the l-th institution.

Let $\mathbf{p}_l = (p_{l1},\ldots,p_{ln_l})$ be a probability vector for $l = 1,\ldots,L$. Similarly, the profile empirical likelihood for the common AUC, evaluated at the true value δ, is defined as follows:

$$\widetilde{L}(\delta) = \sup\left\{\prod_{l=1}^L\prod_{j=1}^{n_l} p_{lj} : \sum_{j=1}^{n_l} p_{lj} = 1, \sum_{j=1}^{n_l} p_{lj}\widetilde{W}_{lj}(\delta) = 0, \right.$$
$$\left. l = 1,\ldots,L\right\},$$

where $\widetilde{W}_{lj}(\delta) = 1 - \widetilde{U}_{lj} - \delta$ with $\widetilde{U}_{lj} = 1 - \widetilde{F}_l(\widetilde{Y}_{lj})$, $l = 1, \ldots, L$, $j = 1, 2, \ldots, n_l$, and the $\widetilde{F}_l$ is the imputation-based empirical distribution of F_l. Then, the corresponding empirical log-likelihood ratio is

$$\widetilde{l}(\delta) = 2\sum_{l=1}^{L}\sum_{j=1}^{n_l} \log\left\{1 + \widetilde{\lambda}_l \widetilde{W}_{lj}(\delta)\right\}, \tag{9}$$

where $\widetilde{\lambda}_l$ is the solution of

$$\frac{1}{n_l}\sum_{j=1}^{n_l} \frac{\widetilde{W}_{lj}(\delta)}{1 + \widetilde{\lambda}_l \widetilde{W}_{lj}(\delta)} = 0, \quad l = 1, \ldots, L. \tag{10}$$

The following theorem establishes the asymptotic distribution of the imputation-based empirical log-likelihood ratio for the AUC with stratified samples.

Theorem 2. *Let δ_0 be the true value of the common AUC. If $\lim_{m_l,n_l\to\infty} \frac{n_l}{m_l} = \tau_l < \infty$, a fixed quantity, for $l = 1, \ldots, L$, then the asymptotic distribution of $\widetilde{l}(\delta_0)$, defined by (9), is a weighted summation of independent χ^2 distribution with degree of freedom one, i.e.,*

$$\widetilde{l}(\delta_0) \xrightarrow{d} w_1\chi^2_{1,1} + \cdots + w_L\chi^2_{L,1}, \tag{11}$$

where the weights $w_l = \lim_{m_l,n_l\to\infty} \widetilde{w}_l(\delta_0)$, $1 \le l \le L$, with

$$\widetilde{w}_l(\delta_0) = \frac{m_l + n_l}{m_l} \frac{n_l S_l^2}{\sum_{j=1}^{n_l} \widetilde{W}^2_{lj}(\delta_0)},$$

$$S_l^2 = \frac{m_l(1 - \pi_{l2} + \pi_{l2}^{-1})S_{01}(l)^2 + n_l(1 - \pi_{l1} + \pi_{l1}^{-1})S^2_{10}(l)}{m_l + n_l},$$

$$S^2_{10}(l) = \frac{1}{(m_l - 1)n_l^2}\left[\sum_{i=1}^{m_l}(R_i(l) - i)^2 - m_l\left(\bar{R}_l - \frac{m_l + 1}{2}\right)^2\right],$$

$$S_{01}(l)^2 = \frac{1}{(n_l - 1)m_l^2}\left[\sum_{j=1}^{n_l}(S_j(l) - j)^2 - n_l\left(\bar{S}_l - \frac{n_l + 1}{2}\right)^2\right],$$

$$\bar{R}_l = \frac{1}{m_l}\sum_{i=1}^{m_l} R_i(l), \quad \text{and} \quad \bar{S}_l = \frac{1}{n_l}\sum_{j=1}^{n_l} S_j(l).$$

Here $R_i(l)$ is the rank of $\tilde{X}_{l(i)}$ (the i-th ordered value among $\tilde{X}_{li}$'s) in the combined sample of $\tilde{X}_{li}$'s and $\tilde{Y}_{lj}$'s, and $S_j(l)$ is the rank of $\tilde{Y}_{l(j)}$ (the j-th ordered value among $\tilde{Y}_{lj}$'s) in the combined sample of $\tilde{X}_{li}$'s and $\tilde{Y}_{lj}$'s.

Then the EL-based confidence interval for the common AUC can be constructed as follows:

$$R_\alpha(\delta) = \left\{\delta : \widetilde{l}(\delta) \le c_{1-\alpha}\right\}, \tag{12}$$

where $c_{1-\alpha}$ is the $(1-\alpha)100\%$th quantile of the weighted chi-square distribution $w_1\chi^2_{1,1} + \cdots + w_L\chi^2_{L,1}$. The quantile $c_{1-\alpha}$ could be calculated using a simple Monte Carlo simulation by plugging in consistent estimates of all unknown quantities. Therefore, $R_\alpha(\delta)$ defined by (12) offers an approximate confidence interval for the common AUC with asymptotically correct coverage probability $1 - \alpha$.

3. SIMULATION STUDIES

In this section simulation studies are conducted to evaluate the finite-sample performance of the proposed IPEL interval for the AUC in terms of coverage probability when the AUC is taken to be 0.8 (moderate accuracy), 0.9, and 0.95 (high accuracy). For simplicity, we take $L = 1$ in simulation studies. Here two typical settings of distribution are considered, one for symmetric distribution and the other for asymmetric distribution:

(1) $X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}(\sqrt{5}\Phi^{-1}(\delta), 2^2)$;
(2) $X \sim \exp(1)$ and $Y \sim \exp(\frac{\delta}{1-\delta})$.

Note that in the first simulation setting, δ is related to the mean and the standard deviation by the following relationship:

$$\delta = \Phi\left(\frac{\mu - \mu_0}{\sqrt{\sigma^2 + \sigma_0^2}}\right),$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution, if $X \sim \mathcal{N}(\mu_0, \sigma_0^2)$ and $Y \sim \mathcal{N}(\mu, \sigma^2)$. Meanwhile, if $X \sim \exp(\theta_1)$ and $Y \sim \exp(\theta_2)$, then

$$\delta = \frac{\theta_2}{\theta_1 + \theta_2}.$$

For each setting, 2,000 random samples of incomplete data $(X_i, \delta_{X_i}), i = 1, \ldots, m$ and $(Y_j, \delta_{Y_j}), j = 1, \ldots, n$ are generated from the underlying non-diseased distribution F and diseased distribution G, respectively. The sample size ranges from 50 to 200 with both $m = n$ and $m \ne n$ two cases for the two settings. We also consider different observation rates: $(\pi_1, \pi_2) = 90\%$ (high), 80% or 70% (moderate), and 60% (low) with $\pi_1 = \pi_2$ and $\pi_1 \ne \pi_2$. For comparison, the full observation case is also included in the study. Note that when $\pi_1 = \pi_2 = 1$, the proposed method will be reduced to the method developed by Qin and Zhou (2006), which has been shown to have good finite sample performance.

In Tables 1–4, we present the coverage probabilities of 90% and 95% IPEL intervals for various values of the AUC based on the proposed imputation-based empirical likelihood method under two model settings. The simulation results in these tables indicate that the proposed method works well in moderate accuracy cases even with small sample sizes (i.e., $m = n = 50$). In high accuracy cases, the proposed method seems to be conservative in a small sample size case, and the performance improves as the sample

Table 1. Model setting (1): Coverage probabilities of the IPEL interval for the AUC with nominal confidence level 90% and various observation rates (π_1, π_2)

AUC	(m,n)	Observation rates (π_1,π_2)					
		$(1,1)$	$(0.9,0.9)$	$(0.9,0.8)$	$(0.8,0.8)$	$(0.8,0.7)$	$(0.6,0.6)$
0.80	$(50,50)$	0.9159	0.9101	0.9050	0.9085	0.9078	0.9189
	$(50,80)$	0.9060	0.9084	0.9049	0.9003	0.9018	0.9120
	$(80,80)$	0.9040	0.9115	0.9090	0.9035	0.8958	0.9111
	$(80,100)$	0.8955	0.8960	0.8945	0.8985	0.8989	0.9068
	$(100,100)$	0.8955	0.8920	0.8955	0.8984	0.8965	0.8968
	$(100,150)$	0.9065	0.8925	0.9010	0.9000	0.9045	0.9010
	$(200,200)$	0.8890	0.8940	0.8980	0.9030	0.8939	0.9025
0.90	$(50,50)$	0.9315	0.9410	0.9367	0.9344	0.9319	0.9467
	$(50,80)$	0.9153	0.9252	0.9209	0.9229	0.9253	0.9432
	$(80,80)$	0.9061	0.9160	0.9148	0.9196	0.9201	0.9276
	$(80,100)$	0.8974	0.9022	0.8970	0.9048	0.9179	0.9152
	$(100,100)$	0.9018	0.9030	0.8995	0.9000	0.9056	0.9052
	$(100,150)$	0.9020	0.8933	0.9009	0.9009	0.9058	0.9091
	$(200,200)$	0.8904	0.8939	0.8979	0.8955	0.8958	0.9013
0.95	$(50,50)$	0.9498	0.9589	0.9521	0.9478	0.9463	0.9543
	$(50,80)$	0.9346	0.9455	0.9456	0.9465	0.9462	0.9617
	$(80,80)$	0.9256	0.9419	0.9479	0.9475	0.9479	0.9548
	$(80,100)$	0.9240	0.9333	0.9327	0.9297	0.9507	0.9459
	$(100,100)$	0.9072	0.9163	0.9299	0.9297	0.9354	0.9369
	$(100,150)$	0.9019	0.8951	0.9113	0.9140	0.9267	0.9333
	$(200,200)$	0.8882	0.9012	0.9003	0.9012	0.9065	0.9075

Table 2. Model setting (1): Coverage probabilities of the IPEL interval for the AUC with nominal confidence level 95% and various observation rates (π_1, π_2)

AUC	(m,n)	Observation rates (π_1,π_2)					
		$(1,1)$	$(0.9,0.9)$	$(0.9,0.8)$	$(0.8,0.8)$	$(0.8,0.7)$	$(0.6,0.6)$
0.80	$(50,50)$	0.9600	0.9603	0.9593	0.9582	0.9620	0.9742
	$(50,80)$	0.9510	0.9545	0.9494	0.9559	0.9544	0.9608
	$(80,80)$	0.9500	0.9550	0.9575	0.9640	0.9544	0.9623
	$(80,100)$	0.9490	0.9485	0.9560	0.9550	0.9540	0.9494
	$(100,100)$	0.9510	0.9515	0.9510	0.9530	0.9560	0.9464
	$(100,150)$	0.9535	0.9485	0.9530	0.9530	0.9585	0.9525
	$(200,200)$	0.9480	0.9445	0.9515	0.9525	0.9490	0.9490
0.90	$(50,50)$	0.9760	0.9754	0.9743	0.9747	0.9751	0.9739
	$(50,80)$	0.9617	0.9641	0.9660	0.9686	0.9675	0.9734
	$(80,80)$	0.9553	0.9572	0.9662	0.9611	0.9608	0.9704
	$(80,100)$	0.9540	0.9518	0.9543	0.9597	0.9650	0.9619
	$(100,100)$	0.9509	0.9515	0.9523	0.9533	0.9609	0.9577
	$(100,150)$	0.9505	0.9469	0.9530	0.9575	0.9579	0.9568
	$(200,200)$	0.9485	0.9460	0.9515	0.9545	0.9479	0.9519
0.95	$(50,50)$	0.9833	0.9757	0.9760	0.9712	0.9697	0.9729
	$(50,80)$	0.9761	0.9823	0.9733	0.9717	0.9763	0.9787
	$(80,80)$	0.9762	0.9751	0.9748	0.9748	0.9721	0.9761
	$(80,100)$	0.9640	0.9721	0.9723	0.9733	0.9748	0.9711
	$(100,100)$	0.9592	0.9635	0.9668	0.9672	0.9711	0.9642
	$(100,150)$	0.9527	0.9544	0.9661	0.9654	0.9680	0.9698
	$(200,200)$	0.9429	0.9531	0.9607	0.9567	0.9573	0.9673

size increases. Reasonably, the proposed method works better in a symmetric distribution case. Also, the performance of the proposed method under missing data cases is comparable with that under the complete data cases in terms of coverage probability.

4. A REAL EXAMPLE

In this section, we evaluate the diagnostic accuracy of the proposed method by applying it to the data set of carbohydrate antigenic determinant CA19-9 in the detection

Table 3. Model setting (2): Coverage probabilities of the IPEL interval for the AUC with nominal confidence level 90% and various observation rates (π_1, π_2)

AUC	(m, n)	Observation rates (π_1, π_2)					
		$(1, 1)$	$(0.9, 0.9)$	$(0.9, 0.8)$	$(0.8, 0.8)$	$(0.8, 0.7)$	$(0.6, 0.6)$
0.80	$(50, 50)$	0.9137	0.9189	0.9296	0.9260	0.9319	0.9368
	$(50, 80)$	0.9100	0.9058	0.9087	0.9147	0.9177	0.9078
	$(80, 80)$	0.9065	0.9043	0.9094	0.9125	0.9129	0.9177
	$(80, 100)$	0.9080	0.8974	0.9034	0.9014	0.9040	0.9134
	$(100, 100)$	0.9085	0.9070	0.9130	0.9119	0.9209	0.9013
	$(100, 150)$	0.9000	0.9080	0.9105	0.9140	0.9174	0.9199
	$(200, 200)$	0.8950	0.8940	0.8990	0.9045	0.8995	0.8949
0.90	$(50, 50)$	0.9260	0.9340	0.9366	0.9387	0.9455	0.9499
	$(50, 80)$	0.9091	0.9181	0.9110	0.9214	0.9342	0.9318
	$(80, 80)$	0.9008	0.9125	0.9102	0.9126	0.9234	0.9394
	$(80, 100)$	0.9042	0.9031	0.9100	0.9093	0.9170	0.9256
	$(100, 100)$	0.8998	0.9045	0.9134	0.9144	0.9177	0.9240
	$(100, 150)$	0.9035	0.9084	0.9124	0.9194	0.9197	0.9323
	$(200, 200)$	0.8940	0.8890	0.8984	0.8993	0.9018	0.8959
0.95	$(50, 50)$	0.9434	0.9460	0.9494	0.9494	0.9489	0.9519
	$(50, 80)$	0.9326	0.9361	0.9398	0.9436	0.9448	0.9559
	$(80, 80)$	0.9204	0.9400	0.9435	0.9396	0.9530	0.9554
	$(80, 100)$	0.9206	0.9267	0.9241	0.9295	0.9403	0.9494
	$(100, 100)$	0.9109	0.9285	0.9350	0.9320	0.9450	0.9497
	$(100, 150)$	0.9076	0.9109	0.9129	0.9184	0.9274	0.9414
	$(200, 200)$	0.8954	0.8951	0.9060	0.9051	0.9059	0.9145

Table 4. Model setting (2): Coverage probabilities of the IPEL interval for the AUC with nominal confidence level 95% and various observation rates (π_1, π_2)

AUC	(m, n)	Observation rates (π_1, π_2)					
		$(1, 1)$	$(0.9, 0.9)$	$(0.9, 0.8)$	$(0.8, 0.8)$	$(0.8, 0.7)$	$(0.6, 0.6)$
0.80	$(50, 50)$	0.9579	0.9657	0.9666	0.9645	0.9716	0.9771
	$(50, 80)$	0.9565	0.9564	0.9549	0.9584	0.9634	0.9627
	$(80, 80)$	0.9510	0.9564	0.9617	0.9568	0.9622	0.9687
	$(80, 100)$	0.9560	0.9485	0.9509	0.9464	0.9530	0.9597
	$(100, 100)$	0.9505	0.9565	0.9600	0.9585	0.9599	0.9592
	$(100, 150)$	0.9565	0.9610	0.9530	0.9615	0.9635	0.9625
	$(200, 200)$	0.9475	0.9520	0.9470	0.9455	0.9480	0.9460
0.90	$(50, 50)$	0.9655	0.9714	0.9804	0.9748	0.9787	0.9791
	$(50, 80)$	0.9593	0.9631	0.9670	0.9742	0.9750	0.9754
	$(80, 80)$	0.9567	0.9631	0.9665	0.9660	0.9739	0.9759
	$(80, 100)$	0.9534	0.9573	0.9527	0.9582	0.9651	0.9740
	$(100, 100)$	0.9494	0.9568	0.9627	0.9678	0.9652	0.9709
	$(100, 150)$	0.9550	0.9630	0.9604	0.9609	0.9679	0.9649
	$(200, 200)$	0.9520	0.9485	0.9465	0.9454	0.9464	0.9527
0.95	$(50, 50)$	0.9754	0.9727	0.9760	0.9710	0.9702	0.9727
	$(50, 80)$	0.9717	0.9742	0.9720	0.9754	0.9737	0.9779
	$(80, 80)$	0.9702	0.9752	0.9767	0.9761	0.9791	0.9800
	$(80, 100)$	0.9674	0.9698	0.9711	0.9715	0.9725	0.9747
	$(100, 100)$	0.9618	0.9722	0.9794	0.9758	0.9827	0.9809
	$(100, 150)$	0.9593	0.9587	0.9676	0.9696	0.9739	0.9815
	$(200, 200)$	0.9445	0.9478	0.9527	0.9548	0.9562	0.9624

of pancreatic cancer. Pancreatic cancer is a disease of the tissues of pancreas where cancer cells are found. It is hard to diagnose the pancreatic cancer because this organ is hidden behind other organs. Furthermore, its early detection is poor or almost impossible. Therefore, the death rate of pancreatic cancer patients is extremely high. By the end of 2010 in the United States, it is estimated about 43,140 individuals will be diagnosed with this condition, and 36,800 will die from the disease, reported by the National Cancer Institute.

Table 5. A real example: 95% IPEL confidence intervals for the AUC of CA19-9 with various observation rates

(π_1,π_2)	$\widetilde{\delta}$	Confidence interval	r_X	r_Y
(1.0, 1.0)	0.862	(0.793, 0.913)	51	90
(0.9, 0.9)	0.874	(0.803, 0.924)	46	86
(0.9, 0.8)	0.873	(0.787, 0.931)	47	68
(0.8, 0.8)	0.876	(0.793, 0.931)	42	72
(0.8, 0.7)	0.811	(0.704, 0.891)	39	56
(0.6, 0.6)	0.835	(0.717, 0.916)	27	48

We apply the newly proposed IPEL method to the data set studied by Wieand et al. [21] on the diagnostic accuracy of CA19-9 in detecting pancreatic cancer. The data set consists of 51 patients in the control group and 90 patients with pancreatic cancer. We simulated the missing mechanism MCAR to obtain missing data with different observation rates of (π_1, π_2), because the original data set is complete. The WMW estimates and IPEL intervals for the AUC are calculated. The results are presented in Table 5. These intervals indicate that CA19-9 has moderate to high levels of diagnostic accuracy in detecting patients with pancreatic cancer. Under different observation rates, $\widetilde{\delta}$ is close to the estimate with complete data, and all confidence intervals contain $\widetilde{\delta}$ based on complete data.

5. DISCUSSION

This paper focuses on missing data under the MCAR assumption. An imputation-based empirical likelihood method is proposed to construct confidence interval for the AUC with MCAR data. The random hot deck imputation is needed for the proposed method. Instead of conditional missing probability, MCAR assumption assures constant missing probability which could be estimated efficiently. Multiple imputation could also be applied in this paper as in Wang and Chen [20]. Since our simulation results have shown that the newly proposed methods work well, for simplicity, multiple imputation is not applied in this paper. Sometimes, the missing mechanism depends on the observed data, called missing at random (MAR, see [15, 8]). The MAR case is a more general case and of more interest. Direct application of complete data inference procedures to MAR problems may result in biased estimation and loss of efficiency. He et al. [6] provided a direct estimate of the AUC in the presence of verification bias under the MAR assumption. But their inference method required the estimation of the variance, which is not an easy job. Future work could be done on the confidence interval for the AUC under the MAR assumption, which could be free of the estimation of variance by applying empirical likelihood technique.

APPENDIX: PROOFS

In order to prove Theorem 1, a few lemmas are necessary.

Lemma 1 (Wang and Qin, [19])**.** *$\widetilde{F}(x)$ and $\widetilde{G}(y)$ defined by (2.1) and (2.1) are uniformly consistent for $F(x)$ and $G(y)$, respectively. Furthermore, they are asymptotically normal, i.e.,*

$$\frac{1}{\sqrt{m}}\sum_{i=1}^{m} I(\tilde{X}_i \le x) \xrightarrow{d} \mathcal{N}(F(x), \sigma_X^2)$$

where $\sigma_X^2 = (1-\pi_1+\pi_1^{-1})F(x)(1-F(x))$, and

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n} I(\tilde{Y}_j \le y) \xrightarrow{d} \mathcal{N}(G(y), \sigma_Y^2)$$

where $\sigma_Y^2 = (1-\pi_2+\pi_2^{-1})G(y)(1-G(y))$.

Lemma 2. *Under the same conditions as in Theorem 1, we have that:*

(i) $\frac{1}{n}\sum_{j=1}^{n}\widetilde{W}_j^2(p) \xrightarrow{P} \sigma_0^2$, *where* $\sigma_0^2 = E[F^2(Y)] - \delta_0^2$;

(ii) $(\frac{mn}{m+n})^{1/2}\frac{\widetilde{\delta}-\delta_0}{S} \xrightarrow{d} \mathcal{N}(0,1)$, *where $\widetilde{\delta}$ is defined by (2).*

Proof. (i) From the uniform consistency of $\widetilde{F}$ in Lemma 1, it follows that

$$\frac{1}{n}\sum_{j=1}^{n}\widetilde{F}^2(\widetilde{Y}_j) - \frac{1}{n}\sum_{j=1}^{n}F^2(\widetilde{Y}_j) \xrightarrow{P} 0.$$

By using the similar technique employed in the proof of Lemma 1, we get that

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^{n}F^2(\widetilde{Y}_j) &\\ &= \frac{1}{r_Y}\sum_{j\in S_{r_Y}}F^2(Y_j) + \frac{m_Y}{n}\frac{1}{m_Y}\sum_{j\in S_{m_Y}}(F^2(Y_j^*) - \bar{F}_{1r})\\ &\xrightarrow{P} E[F^2(Y)],\end{aligned}$$

where $\bar{F}_{1r} = \frac{1}{r_Y}\sum_{j\in S_{r_Y}}F^2(Y_j)$. Therefore,

$$\frac{1}{n}\sum_{j=1}^{n}\widetilde{F}^2(\widetilde{Y}_j) \xrightarrow{P} E[F^2(Y)].$$

Similarly, we can prove that $\frac{1}{n}\sum_{j=1}^{n}\widetilde{F}(\widetilde{Y}_j) \xrightarrow{P} E[F(Y)] = \delta_0$. Combining the above results, from Lemma 2(i), it follows that:

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^{n}\widetilde{W}_j^2(\delta_0) &= \frac{1}{n}\sum_{j=1}^{n}\left(\widetilde{F}(\widetilde{Y}_j)-\delta_0\right)^2\\ &= \frac{1}{n}\sum_{j=1}^{n}\widetilde{F}^2(\widetilde{Y}_j) - \frac{2\delta_0}{n}\sum_{j=1}^{n}\widetilde{F}(\widetilde{Y}_j) + \delta_0^2\\ &\xrightarrow{P} E[F^2(Y)] - \delta_0^2 = \sigma_0^2.\end{aligned}$$

(ii) If the data set is complete, Sen [18] has proved a similar result. Based on imputed data, some necessary modifications are needed. Let

$$\alpha_\delta = \int_0^1 F^2(y)\mathrm{d}G(y), \qquad \beta_\delta = \int_0^1 [1-G(x)]^2\mathrm{d}F(x),$$
$$n_0 = \frac{mn}{m+n},$$
$$\mathscr{B}_n = \sigma(\tilde{Y}_j, j=1,\ldots,n), \qquad \mathscr{A}_m = \sigma(\tilde{X}_i, i=1,\ldots,m).$$

Then, the variance of $\sqrt{n_0}\widetilde{\delta}$ can be calculated as follows:

$$\begin{aligned}(13)\quad &\mathrm{VAR}(\sqrt{n_0}\widetilde{\delta})\\ &= \mathrm{VAR}\left(E(\sqrt{n_0}\widetilde{\delta}|\mathscr{B}_n)\right) + E\left(\mathrm{VAR}(\sqrt{n_0}\widetilde{\delta}|\mathscr{B}_n)\right).\end{aligned}$$

For the first term of the right-hand side in (13), from

$$\begin{aligned}E(\sqrt{n_0}\widetilde{\delta}|\mathscr{B}_n) &= \frac{\sqrt{n_0}}{mn}\sum_{j=1}^n\sum_{i=1}^m E[I(\tilde{X}_i \le \tilde{Y}_j)|\mathscr{B}_n]\\ &= \frac{\sqrt{n_0}}{n}\sum_{j=1}^n [\widetilde{F}(\tilde{Y}_j)],\end{aligned}$$

it follows that

$$\begin{aligned}\mathrm{VAR}\left(E(\sqrt{n_0}\widetilde{\delta}|\mathscr{B}_n)\right) &= \mathrm{VAR}\left(\frac{\sqrt{n_0}}{n}\sum_{j=1}^n[\widetilde{F}(\tilde{Y}_j)]\right)\\ &\to \frac{1}{1+\tau}(1-\pi_2+\pi_2^{-1})(\alpha_\delta - \delta_0^2),\end{aligned}$$

where the last step follows from Lemma 1.

As for the second term of the right-hand side in (13), from

$$\begin{aligned}&\mathrm{VAR}\left(\sqrt{n_0}\widetilde{\delta}|\mathscr{B}_n\right)\\ &= \frac{n_0}{m^2n^2}\mathrm{VAR}\left(\sum_{j=1}^n\sum_{i=1}^m I(\tilde{X}_i \le \tilde{Y}_j)|\mathscr{B}_n\right)\\ &= \frac{n_0 m}{m^2n^2}\Bigg[(1-\pi_1+\pi_1^{-1})\mathrm{VAR}\left(\sum_{j=1}^n I(X \le \tilde{Y}_j|\mathscr{B}_n)\right)\\ &\quad + o_P(1)\Bigg]\\ &= \frac{n_0}{mn^2}\Bigg[(1-\pi_1+\pi_1^{-1})\Bigg(\sum_{j=1}^n F(\tilde{Y}_j)\\ &\quad + 2\sum_{j\le k} E(I(X\le\tilde{Y}_j)I(X\le\tilde{Y}_k)|\mathscr{B}_n)\\ &\quad - \left(\sum_{j=1}^n F(\tilde{Y}_j)\right)^2\Bigg) + o_P(1)\Bigg],\end{aligned}$$

it follows that

$$\begin{aligned}&E\left(\mathrm{VAR}(\sqrt{n_0}\widetilde{\delta}|\mathscr{B}_n)\right)\\ &= \frac{n_0}{mn^2}\Bigg[(1-\pi_1+\pi_1^{-1})\Bigg(\sum_{j=1}^n EF(\tilde{Y}_j)\\ &\quad + 2\sum_{j\le k} E\big(E(I(X\le\tilde{Y}_j)I(X\le\tilde{Y}_k)|\mathscr{B}_n)\big)\\ &\quad - E\left(\sum_{j=1}^n F(\tilde{Y}_j)\right)^2\Bigg) + o(1)\Bigg]\\ &= \frac{n_0}{mn^2}\Bigg[(1-\pi_1+\pi_1^{-1})\Bigg(n\delta_0\\ &\quad + 2\sum_{j\le k} E\big(E(I(X\le\tilde{Y}_j)I(X\le\tilde{Y}_k)|\mathscr{B}_n)\big)\\ &\quad - \left(\mathrm{VAR}\left(\sum_{j=1}^n F(\tilde{Y}_j)\right) + E^2\left(\sum_{j=1}^n F(\tilde{Y}_j)\right)\right)\Bigg) + o(1)\Bigg]\\ &= \frac{n_0}{mn^2}\Bigg[(1-\pi_1+\pi_1^{-1})\Bigg(n\delta_0\\ &\quad + 2\sum_{j\le k} E\big(E(I(X\le\tilde{Y}_j)I(X\le\tilde{Y}_k)|\mathscr{B}_n)\big)\\ &\quad - \left((1-\pi_2+\pi_2^{-1})n(\alpha_\delta-\delta_0^2) + (n\delta_0)^2 + o(n)\right)\Bigg)\\ &\quad + o(1)\Bigg].\end{aligned}$$

From

$$\begin{aligned}&\sum_{j\le k} E\left(E(I(X\le\tilde{Y}_j)I(X\le\tilde{Y}_k)|\mathscr{B}_n)\right)\\ &= E\left[E\left(\sum_{j\le k} I(X\le\tilde{Y}_j)I(X\le\tilde{Y}_k)|\mathscr{B}_n\right)\right]\\ &= EE\Bigg[\Bigg(\sum_{\substack{j\le k\\ j,k\in S_{r_Y}}} + \sum_{\substack{j\le k\\ j\in S_{r_Y},k\in S_{m_Y}}} + \sum_{\substack{j\le k\\ j\in S_{m_Y},k\in S_{r_Y}}} + \sum_{\substack{j\le k\\ j,k\in S_{m_Y}}}\Bigg)\\ &\quad \times I(X\le\tilde{Y}_j)I(X\le\tilde{Y}_k)\Bigg|\mathscr{B}_n\Bigg]\\ &= EE\Bigg[\Bigg(\sum_{\substack{j\le k\\ j,k\in S_{r_Y}}} + \sum_{\substack{j\le k\\ j\in S_{r_Y},k\in S_{m_Y}}} + \sum_{\substack{j\le k\\ j\in S_{m_Y},k\in S_{r_Y}}} + \sum_{\substack{j\le k\\ j,k\in S_{m_Y}}}\Bigg)\\ &\quad \times I(X\le\tilde{Y}_j)I(X\le\tilde{Y}_k)\Bigg|X)\Bigg]\\ &= EE\Bigg[\sum_{\substack{j\le k\\ j,k\in S_{r_Y}}} \big(1-G(X)\big)^2\\ &\quad + \sum_{\substack{j\le k\\ j\in S_{r_Y},k\in S_{m_Y}}} \left(\frac{1}{r_Y}(1-G(X)) + \frac{r_Y-1}{r_Y}(1-G(X))^2\right)\end{aligned}$$

$$+ \sum_{\substack{j \le k \\ j \in S_{m_Y}, k \in S_{r_Y}}} \left(\frac{1}{r_Y}(1 - G(X)) + \frac{r_Y - 1}{r_Y}(1 - G(X))^2 \right)$$

$$+ \sum_{\substack{j \le k \\ j,k \in S_{m_Y}}} \frac{1}{r_Y^2} \left(r_Y(1 - G(X)) + r_Y(r_Y - 1)(1 - G(X))^2 \right)$$

$$\left| X, \sigma(\delta_{Y_j}, j = 1, \dots, n) \right]$$

$$= EE\left[\frac{n(n-1)}{2}(1 - G(X))^2 \right.$$

$$+ \sum_{\substack{j \le k \\ j \text{ or } k \in S_{m_Y}}} \frac{1}{r_Y} \left((1 - G(X)) - (1 - G(X))^2 \right)$$

$$\left| X, \sigma(\delta_{Y_j}, j = 1, \dots, n) \right]$$

$$= EE\left[\frac{n(n-1)}{2}(1 - G(X))^2 \right.$$

$$+ \frac{n(n-1) - r_Y(r_Y - 1)}{2r_Y} \left((1 - G(X)) - (1 - G(X))^2 \right)$$

$$\left| X, \sigma(\delta_{Y_j}, j = 1, \dots, n) \right],$$

it follows that

$$2 \sum_{j \le k} E\left(E(I(X \le \tilde{Y}_j) I(X \le \tilde{Y}_k) | \mathscr{B}_n) \right)$$

$$= EE\left[2(1 - G(X))^2 + \left(\frac{n(1 - \pi_2^2)}{\pi_2} + O_P(n) \right) \right.$$

$$\left. \times \left((1 - G(X)) - (1 - G(X))^2 \right) \middle| X \right]$$

$$= n(n-1)\beta_\delta$$

$$+ EE\left[\left(\frac{n(1 - \pi_2^2)}{\pi_2} + O_P(n) \right) \left((1 - G(X)) \right. \right.$$

$$\left. \left. - (1 - G(X))^2 \right) \middle| X \right].$$

Therefore,

$$E\left(\mathrm{VAR}(\sqrt{n_0}\widetilde{\delta} | \mathscr{B}_n) \right) \to \frac{\tau}{1+\tau}(1 - \pi_1 + \pi_1^{-1})(\beta_\delta - \delta_0^2),$$

$$\mathrm{VAR}(\sqrt{n_0}\widetilde{\delta}) \to \frac{1}{1+\tau}(1 - \pi_2 + \pi_2^{-1})(\alpha_\delta - \delta_0^2) + \frac{\tau}{1+\tau}(1 - \pi_1 + \pi_1^{-1})(\beta_\delta - \delta_0^2).$$

In order to prove Lemma 2(ii), we need to show that S and $\mathrm{VAR}(\sqrt{n_0}\widetilde{\delta})$ converge to the same limit. Let

$$V_{10}(\tilde{X}_i) = \frac{1}{n} \sum_{j=1}^{n} I(\tilde{X}_i \le \tilde{Y}_j), \quad i = 1, \dots, m;$$

$$V_{01}(\tilde{Y}_j) = \frac{1}{m} \sum_{i=1}^{m} I(\tilde{X}_i \le \tilde{Y}_j), \quad j = 1, \dots, n.$$

It follows that

$$S_{10}^2 = \frac{1}{m-1} \sum_{i=1}^{m} \left[V_{10}(\tilde{X}_i) - \widetilde{\delta} \right]^2 = \frac{1}{m-1} \sum_{i=1}^{m} \left[V_{10}^2(\tilde{X}_i) - 2V_{10}(\tilde{X}_i)\widetilde{\delta} + \widetilde{\delta}^2 \right].$$

By Lemma 1, we have

$$\widetilde{\delta} \xrightarrow{P} \delta_0, \quad \text{and} \quad V_{10}(\tilde{X}_i) \xrightarrow{P^*} 1 - G(\tilde{X}_i)$$

where P^* is the probability measure on $\mathscr{A}_m$. Thus,

$$S_{10}^2 \xrightarrow{P} \beta_\delta - \delta_0^2.$$

Similarly, we have

$$S_{01}^2 \xrightarrow{P} \alpha_\delta - \delta_0^2.$$

Therefore,

$$S^2 = \frac{m(1 - \pi_2 + \pi_2^{-1})S_{01}^2 + n(1 - \pi_1 + \pi_1^{-1})S_{10}^2}{m+n}$$

$$\xrightarrow{P} \frac{1}{1+\tau}(1 - \pi_2 + \pi_2^{-1})(\alpha_\delta - \delta_0^2) + \frac{\tau}{1+\tau}(1 - \pi_1 + \pi_1^{-1})(\beta_\delta - \delta_0^2).$$

Based on Sen (1967), S_{01}^2 and S_{10}^2 have the alternative algebraic expressions in Theorem 1. Finally, from Lemma 1, the Slutsky's theorem and the similar procedures of structural convergence of U-statistics in Sen [17, 18], it follows that

$$\sqrt{n_0}\frac{\widetilde{\delta} - \delta_0}{S} = \left(\frac{mn}{m+n} \right)^{1/2} \frac{\widetilde{\delta} - \delta_0}{S} \xrightarrow{d} \mathcal{N}(0, 1). \qquad \square$$

Proof of Theorem 1. Based on Lemma 2 and the same procedure of the proof of Theorem 1 in Qin and Zhou [13], it is straightforward to obtain the result. □

Proof of Theorem 2. Based on Theorem 1 and the same procedure of the proof of Theorem 2 in Qin and Zhou [13], it is straightforward to obtain the result. □

Received 9 August 2011

REFERENCES

[1] Bamber, D. C. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic curve graph. *Journal of Mathematical Psychology* **12** 387–415. MR0384214

[2] Chen, J., Rao, J. N. K. and Sitter, R. R. (2000). Efficient random imputation for missing data in complex surveys. *Statistica Sinica* **10** 1153–1169. MR1804553

[3] Chen, S. X. and Van Keilegom, I. (2009). A review on empirical likelihood methods for regressions (with discussions). *Test* **3** 415–447. MR2566404

[4] Geert, J. M. G., Van der, H., Rogier, A., Donders, T., Theo, S. and Karel, G. M. M. (2006). *Journal of Clinical Epidemiology* **59** 1102–1109.

[5] Gong, Y., Peng, L. and Qi, Y. C. (2010). Smoothed jackknife empirical likelihood method for ROC curve. *Journal of Multivariate Analysis* **101** 1520–1531. MR2609511

[6] He, H., Lyness, J. M. and McDermott, M. P. (2009). Direct estimation of the area under the receiver operating characteristic curve in the presence of verification bias. *Statistics in Medicine* **28** 361–376. MR2655685

[7] Jing, B. Y., Yuan, J. and Zhou, W. (2009). Jackknife empirical likelihood. *Journal of the American Statistical Association* **104** 1224–1232. MR2562010

[8] Little, R. J. A. and Rubin, D. B. (2002). *Statistical analysis with missing data*, 2nd ed. John Wiley & Sons, New York. MR1925014

[9] Mann, H. B. and Whitney, D. R. (1947). On a test whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* **18** 50–60. MR0022058

[10] Owen, A. (1990). Empirical likelihood ratio confidence regions. *Annals of Statistics* **18** 90–120. MR1041387

[11] Owen, A. (2001). *Empirical Likelihood*. Chapman & Hall/CRC, New York.

[12] Pepe, M. S. and Cai, T. (2004). The analysis of placement values for evaluating discriminatory measures. *Biometrics* **60** 528–535. MR2067011

[13] Qin, G. S. and Zhou, X. H. (2006). Empirical likelihood inference for the area under the ROC curve. *Biometrics* **62** 612–622. MR2236844

[14] Qin, J., Zhang, B. and Leung, D. H. Y. (2009). Empirical likelihood in missing data problems. *Journal of the American Statistical Association* **104** 1492–1503. MR2750574

[15] Rubin, D. B. (1976). Inferences and missing data. *Biometrika* **63** 581–590. MR0455196

[16] Rao, J. N. K. and Shao, J. (1992). Jackknife variance estimation with survey data under hot deck imputation. *Biometrika* **79** 811–822. MR1209480

[17] Sen, P. K. (1960). On some convergence properties of U-statistics. *Calcutta Stat. Assoc. Bull.* **10** 1–18. MR0119286

[18] Sen, P. K. (1967). A note on asymptotically distribution-free confidence bounds for $P(X < Y)$ based on two independent samples. *Sankhyã, Series A* **29** 95–102. MR0226772

[19] Wang, B. H. and Qin, G. S. (2011). Empirical likelihood-based confidence intervals for the sensitivity of a continuous-scale diagnostic test with missing data. Technical Report, Department of Mathematics and Statistics, Georgia State University, USA.

[20] Wang, D. and Chen, S. (2009). Empirical likelihood for estimating equations with missing values. *The Annals of Statistics* **37** 490–517. MR2488360

[21] Wieand, S., Gail, M. H., James, B. R. and James, K. R. (1989). A family of non-parametric statistics for comparing diagnostic markers with paired and unpaired data. *Biometrika* **76** 585–592. MR1040651

[22] Wolfe, D. A. and Hogg, R. V. (1971). On constructing statistics and reporting data. *American Statistician* **25** 27–30.

Binhuan Wang
Department of Mathematics and Statistics
Georgia State University
Atlanta, GA 30303
USA
E-mail address: matbxwx2@langate.gsu.edu

Gengsheng Qin
Department of Mathematics and Statistics
Georgia State University
Atlanta, GA 30303
USA
E-mail address: gqin@gsu.edu