Statistica Sinica **27** (2017), 357-371
doi:http://dx.doi.org/10.5705/ss.2013.367

A NONPARAMETRIC APPROACH FOR PARTIAL AREAS UNDER ROC CURVES AND ORDINAL DOMINANCE CURVES

Hanfang Yang, Kun Lu and Yichuan Zhao

Renmin University of China, Princeton University and Georgia State University

Abstract: The receiver operating characteristic (ROC) curve is a well-known measure of the performance of a classification method. Interest may only pertain to a specific region of the curve and, in this case, the partial area under the ROC curve (pAUC) provides a useful summary measure. Related measures such as the ordinal dominance curve (ODC) and the partial area under the ODC (pODC) are frequently of interest as well. Based on a novel estimator of pAUC proposed by Wang and Chang (2011), we develop nonparametric approaches to the pAUC and pODC using normal approximation, the jackknife and the jackknife empirical likelihood. A simulation study demonstrates the flaws of the existing method and shows proposed methods perform well. Simulations also substantiate the consistency of our jackknife variance estimator. The Pancreatic Cancer Serum Biomarker data set is used to illustrate the proposed methods.

Key words and phrases: Jackknife empirical likelihood, normal approximation, partial AUC.

1. Introduction

The ROC curve is a well-established graphical tool used to evaluate performance of a classifier in accurately discriminating between subjects from different populations (e.g., diseased and healthy individuals). Let F and G be distribution functions of random variables X and Y corresponding to independent populations. Let $G^{-1}(t) = \inf\{y : G(y) \geq t\}$ be the quantile function of G, $0 < t < 1$. Let $S_F(t)$ and $S_G(t)$ be the corresponding survival functions $S_F(t) = 1 - F(t)$ and $S_G(t) = 1 - G(t)$. For $t \in (0,1)$, the ROC curve is defined as $\mathrm{ROC(t) = 1 - F\{G^{-1}(1-t)\}}$ or $\mathrm{ROC(t) = S_F\{S_G^{-1}(t)\}}$, where t is the value of FPR and $S_G^{-1}(t) = G^{-1}(1-t)$. The ROC curve is not a convenient tool for comparisions, in particular when two ROC curves cross. A summary measure of an ROC curve can be found by integrating the ROC curve over the the range of FPR values to obtain the area under the ROC curve as $AUC = \int_0^1 \mathrm{ROC(t)dt} = \int_{\infty}^{-\infty} \mathrm{S_F(u)dS_G(u)}$. For economical and practical

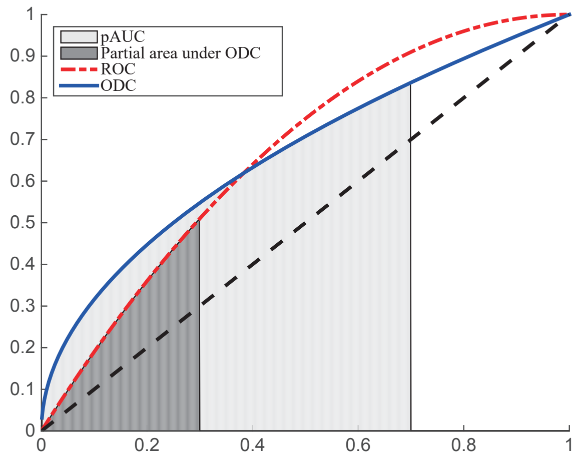


Figure 1. ODC and ROC curve.

purposes, it is common to hold the FPR to a low level. When interest is restricted to a sub-region of the ROC space, the partial area under the ROC curve, $\mathrm{pAUC}(P_0) = \int_0^{P_0} \mathrm{ROC(p)dp}$ for the threshold value of FPR $P_0 \in (0,1)$, can provide a useful summary measure.

The ordinal dominance curve (ODC) introduced by Bamber (1975), see Figure 1, describes the association between true negative rate (TNR) and false negative rate (FNR), $\mathrm{ODC(t)} = \mathrm{G}\{\mathrm{F}^{-1}(\mathrm{t})\}$ where $t \in (0,1)$. The area under the ODC, $\int_0^1 \mathrm{ODC(t)dt} = \int_{-\infty}^{\infty} \mathrm{G(u)dF(u)}$, is a commonly used summary measure. A partial area under the ODC(pODC) from 0 to P_0 is taken as $\mathrm{pODC(P_0)} = \int_0^{\mathrm{P_0}} \mathrm{ODC(t)dt}$.

Nonparametric approaches for statistics based on ROC curves have been extensively investigated. Hsieh and Turnbull (1996) proposed nonparametric estimators for the ODC and AUC, and Wieand, Gail and James (1989) presented nonparametric methods for the difference between ROC curves or AUC's. Based on the jackknife empirical likelihood (Jing, Yuan and Zhou (2009)), Gong, Peng and Qi (2010) proposed a smoothed inference procedure for the ROC curve, and Yang and Zhao (2013, 2015) developed new inference methods for the difference of two ROC curves and ROC curves with missing data. Several researchers have applied the properties of U-statistics (Hoeffding (1948)) to make an inference on AUC and pAUC. For example, DeLong, DeLong and Clarke-Pearson (1988), Sen (1960), and Bamber (1975) employed a multi-dimensional version of Hoeffding (1948) for Mann-Whitney U-statistics to do inference for the AUC and, similarly, Zhang et al. (2002) and Dodd and Pepe (2003) investigated U-statistic theory for the pAUC. More recently, He and Escobar (2008) have pointed out that the Sen-type estimator of the pAUC is not a typical U-statistic, and Hoeffding's theory may not be applicable. Specifically, Hoeffding's theory does not account for the variance of quantile estimates or their correlation with U-statistic kernels derived for these estimators.

Building on the work of He and Escobar (2008), Adimari and Chiogna (2012) introduced the jackknife empirical likelihood (JEL) for the pAUC. However, the effect of an estimated quantile is still unclear since theorems of He and Escobar (2008) do not sufficiently account for the variance of a quantile estimate, and this theoretical result was not established rigorously by Adimari and Chiogna (2012).

In this paper, we present a nonparametric estimator of the pAUC with a variance that correctly accounts for the random error in the estimator. We also derive an interval estimation method based on the pAUC estimator proposed by Wang and Chang (2011). Finally, we develop jackknife and JEL inference procedures for the pAUC and the pODC.

The rest of the paper is organized as follows. In Section 2, we propose the nonparametric approach for the partial area under ODC and pAUC, using normal approximation, the jackknife and the JEL. In Section 3, we report on extensive simulation studies. We show how to apply our methods to a practical problem in Section 4 and add a discussion in Section 5. Proofs are provided in the supplementary material.

2. Main Procedures

Let $\mathbf{X} = \{X_i, i = 1, \ldots, m\}$ and $\mathbf{Y} = \{Y_i, i = 1, \ldots, n\}$ be random samples from the distribution functions $F(x)$ and $G(y)$, respectively. A simple empirical estimator of pODC(P_0) is

$$\widehat{\mathrm{pODC}}(P_0) = \int_{-\infty}^{F_m^{-1}(P_0)} G_n(u)\mathrm{d}F_m(u),$$

where $F_m^{-1}(P_0)$ is an empirical quantile estimate at P_0 and $F_m(\cdot)$ and $G_n(\cdot)$ are the empirical distributions of $F(\cdot)$ and $G(\cdot)$. Alternatively,

$$\widehat{\mathrm{pODC}}(P_0) = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} I(Y_j \le X_i) I\{X_i \le F_m^{-1}(P_0)\}.$$

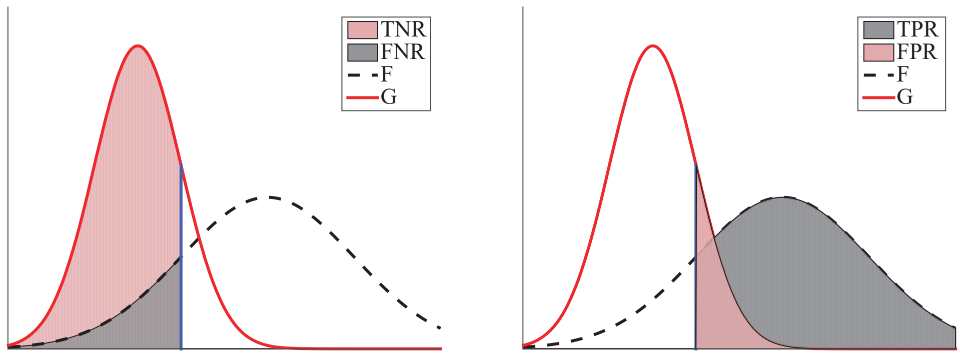
Liu (2006) developed the asymptotic normality for $\widehat{\mathrm{pODC}}(P_0)$. Here are the following conditions that are common in practice.

C.1. $F(t)$ and $G(t)$ are continuous distribution functions.

C.2. $m/(m+n) \to \lambda$, $\lambda \in (0, 1)$.

C.3. $F(t)$ is differentiable, and $F(t)$ is twice differentiable at $F^{-1}(P_0)$, with $F'(F^{-1}(P_0)) > 0$.

C.4. $G(t)$ is differentiable, and $G(t)$ is twice differentiable at $G^{-1}(1 - P_0)$, with $G'(G^{-1}(1 - P_0)) > 0$.



(a) binary classifier for the ODC. (b) binary classifier for the ROC curve

Figure 2. Binary classifier for ODC and ROC curves.

Lemma 1 (Liu (2006))**.** *Under* C.1−C.4,

$$\sqrt{m+n}\{\widehat{\mathrm{pODC}}(P_0) - \mathrm{pODC}(\mathrm{P_0})\} \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, \frac{\sigma_1^2}{1-\lambda} + \frac{\sigma_2^2}{\lambda}\right), \mathrm{m}, \mathrm{n} \to \infty,$$

where

$$\sigma_1^2 = \int_{-\infty}^{F^{-1}(P_0)} \{P_0 - F(t)\}^2 dG(t) - \left\{\int_{-\infty}^{F^{-1}(P_0)} G(t)dF(t)\right\}^2,$$

$$\sigma_2^2 = \int_{-\infty}^{F^{-1}(P_0)} [G(t) - G\{F^{-1}(P_0)\}]^2 dF(t)$$
$$- \left(\int_{-\infty}^{F^{-1}(P_0)} [G(t) - G\{F^{-1}(P_0)\}] dF(t)\right)^2.$$

As shown in Figure 2, points on the ROC curve, ROC $=$ (FPR, TPR) $=$ (1−TNR, 1−FNR) can be obtained from the ODC curve, ODC $=$ (FNR, TNR). We have the empirical estimator

$$\widehat{\mathrm{pAUC}}(P_0) = \int_0^{P_0} S_{F,m}\{S_{G,n}^{-1}(u)\}du = \int_{+\infty}^{S_{G,n}^{-1}(P_0)} S_{F,m}(t)dS_{G,n}(t)$$
$$= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} I(X_i \geq Y_j) I\{Y_j \geq S_{G,n}^{-1}(P_0)\},$$

where $S_{G,n}^{-1}(t) = \inf\{x \in R; t \geq S_{G,n}(x)\}$ and $S_{F,m}(\cdot)$ and $S_{G,n}(\cdot)$ are estimators of S_F and S_G based on empirical distributions. Following Liu (2006), we can extend Lemma 1.

Corollary 1. *Under* C.1−C.4,

$$\sqrt{m+n}\{\widehat{\mathrm{pAUC}}(P_0) - \mathrm{pAUC}(P_0)\} \xrightarrow{d} N\left\{0, \frac{\sigma_3^2(P_0)}{\lambda} + \frac{\sigma_4^2(P_0)}{1-\lambda}\right\}, m, n \to \infty,$$

where

$$\sigma_3^2(P_0) = \int_{+\infty}^{S_G^{-1}(P_0)} \{P_0 - S_G(t)\}^2 dS_F(t) - \left\{ \int_{+\infty}^{S_G^{-1}(P_0)} S_F(t) dS_G(t) \right\}^2,$$

$$\sigma_4^2(P_0) = \int_{+\infty}^{S_G^{-1}(P_0)} [S_F(t) - S_F\{S_G^{-1}(P_0)\}]^2 dS_G(t) - \left(\int_{+\infty}^{S_G^{-1}(P_0)} [S_F(t) - S_F\{S_G^{-1}(P_0)\}] dS_G(t) \right)^2.$$

Remark 1. We provide the variance of $\widehat{\text{pAUC}}$, accounting for the random error in $S_{G,n}^{-1}(P_0)$ that He and Escobar (2008) have ignored. Our simulation results demonstrate the improved performance of our variance estimators as the sample size becomes large.

Contrary to arguments of Adimari and Chiogna (2012), the quantile estimator is problematic for jackknife variance estimators. Their jackknife variance estimator fails to incorporate the error associated with the quantile estimate, and thus is not a consistent estimator for the variance of $\widehat{\text{pAUC}}(P_0)$. This lack of consistency can be seen in our simulations in Section 3.1.

Jackknife methods will be applied to an alternative estimator of $\text{pAUC}(P_0)$ by Wang and Chang (2011),

$$\widetilde{\text{pAUC}}(P_0) = P_0 - \frac{1}{m}\sum_{i=1}^{m} \min\{S_{G,n}(X_i), P_0\}.$$

Theorem 1. *Under* C.1–C.4,

$$\sqrt{m+n}\{\widetilde{\text{pAUC}}(P_0) - \text{pAUC}(P_0)\} \xrightarrow{d} N\left\{0, \frac{\sigma_3^2(P_0)}{\lambda} + \frac{\sigma_4^2(P_0)}{1-\lambda}\right\}, m, n \to \infty.$$

The estimators $\widehat{\text{pAUC}}(P_0)$ and $\widetilde{\text{pAUC}}(P_0)$ are in close agreement, but $\widetilde{\text{pAUC}}(P_0)$ avoids the use of a quantile estimator. We propose the jackknife method and JEL method based on $\widetilde{\text{pAUC}}(P_0)$. Then

$$\widetilde{\text{pAUC}}_{jack}(P_0) = \frac{1}{n+m}\sum_{h=1}^{n+m} V_h(P_0),$$

where $V_h(P_0) = (n+m)\widetilde{\text{pAUC}}(P_0) - (n+m-1)\widetilde{\text{pAUC}}_h(P_0)$, and

$$\widetilde{\text{pAUC}}_h(P_0) = \begin{cases} P_0 - \frac{1}{m-1}\sum\limits_{i \neq h}^{m} \min\{S_{G,n}(X_i), P_0\} & 1 \le h \le m, \\ P_0 - \frac{1}{m}\sum\limits_{i=1}^{m} \min\{S_{G,n-1,h-m}(X_i), P_0\} & m+1 \le h \le m+n, \end{cases}$$

where

$$S_{G,n-1,h-m}(X_i) = \frac{1}{n-1} \sum_{j=1, j \neq h-m}^{n} I(Y_j > X_i).$$

Lemma 2. *Under* C.1−C.4,

$$\sqrt{m+n}\{\widetilde{\text{pAUC}}_{jack}(P_0) - \text{pAUC}(P_0)\} \xrightarrow{d} N\left\{0, \frac{\sigma_3^2(P_0)}{\lambda} + \frac{\sigma_4^2(P_0)}{1-\lambda}\right\}, m, n \to \infty.$$

Consider the jackknife variance estimator

$$S^2_{\widetilde{\text{pAUC}}} = (m+n)^{-1} \sum_{h=1}^{m+n} \{V_h(P_0) - \widetilde{\text{pAUC}}_{jack}(P_0)\}^2.$$

Lemma 3. *Under* C.1−C.4,

$$S^2_{\widetilde{\text{pAUC}}}(P_0) = \frac{\sigma_3^2(P_0)}{\lambda} + \frac{\sigma_4^2(P_0)}{1-\lambda} + o_p(1).$$

Using Slusky's Theorem, and Lemmas 2 and 3, we have the following.

Theorem 2. *Under* C.1−C.4 *hold,*

$$\frac{\sqrt{m+n}\{\widetilde{\text{pAUC}}_{jack}(P_0) - \text{pAUC}(P_0)\}}{\sqrt{S^2_{\widetilde{\text{pAUC}}}(P_0)}} \xrightarrow{d} N(0,1).$$

In order to derive a Wilks' theorem for the jackknife empirical likelihood ratio, the asymptotic normality and variance consistency of jackknife pseudo-samples are essential. For the JEL, we define the jackknife empirical likelihood ratio for pAUC(P_0) as

$$\begin{aligned} &R_{\text{pAUC}}\{P_0, \text{pAUC}(P_0)\} \\ &= \frac{\sup\left\{\prod_{i=1}^{m+n} p_i : \sum_{i=1}^{m+n} p_i = 1, \sum_{i=1}^{m+n} p_i V_i(P_0) = \text{pAUC}(P_0), p_i > 0, i = 1, \ldots, m+n\right\}}{\sup\left\{\prod_{i=1}^{m+n} p_i, \sum_{i=1}^{m+n} p_i = 1, p_i > 0, i = 1, \ldots, m+n\right\}}. \end{aligned}$$

The empirical log-likelihood ratio for the pAUC(P_0) is

$$\begin{aligned} l_{\text{pAUC}}(P_0, \text{pAUC}(P_0)) &= -2 \log[R_{\text{pAUC}}\{P_0, \text{pAUC}(P_0)\}] \\ &= 2 \sum_{i=1}^{m+n} \log[1 + \lambda_1\{V_i(P_0) - \text{pAUC}(P_0)\}], \end{aligned} \tag{2.1}$$

where the Lagrange multiplier λ_1 satisfies the nonlinear equation

$$\sum_{i=1}^{m+n} \frac{\{V_i(P_0) - \mathrm{pAUC}(P_0)\}}{1 + \lambda_1\{V_i(P_0) - \mathrm{pAUC}(P_0)\}} = 0. \tag{2.2}$$

We derive a Wilks' theorem for $\mathrm{pAUC}(P_0)$ based on the jackknife pseudo-values $V_i(P_0), i = 1, \ldots, m+n$.

Theorem 3. *Under* C.1−C.4*,*

$$l_{\mathrm{pAUC}}\{P_0, \mathrm{pAUC}(P_0)\} \xrightarrow{d} \chi_1^2. \tag{2.3}$$

From Theorem 3, an asymptotic $100(1-\alpha)\%$ JEL confidence interval for $\mathrm{pAUC}(P_0)$ is $I_{\mathrm{pAUC}}(P_0) = \left\{\tilde{V} : l_{\mathrm{pAUC}}(P_0, \tilde{V}) \le \chi_1^2(\alpha)\right\}$, where $\chi_1^2(\alpha)$ is the upper α-quantile of χ_1^2.

Because the ODC curve is reversed from the ROC curve, we may apply results for the pAUC to the pODC. Following Wang and Chang (2011), let

$$\widetilde{\mathrm{pODC}}(P_0) = P_0 - \frac{1}{n}\sum_{j=1}^{n} \min\{F_m(Y_j), P_0\}.$$

Corollary 2. *Under* C.1−C.4*, as* $m, n \to \infty$

$$\sqrt{m+n}\{\widetilde{\mathrm{pODC}}(P_0) - \mathrm{pODC}(\mathrm{P}_0)\} \xrightarrow{\mathrm{d}} \mathrm{N}\left\{0, \frac{\sigma_1^2(\mathrm{P}_0)}{1-\lambda} + \frac{\sigma_2^2(\mathrm{P}_0)}{\lambda}\right\}.$$

For the jackknife procedure of $\widetilde{\mathrm{pODC}}$, we take

$$\widetilde{\mathrm{pODC}}_{jack}(P_0) = \frac{1}{n+m}\sum_{h=1}^{n+m} \check{U}_h(P_0),$$

where $\check{U}_h(P_0) = (n+m)\widetilde{\mathrm{pODC}}(P_0) - (n+m-1)\widetilde{\mathrm{pODC}}_h(P_0)$ and

$$\widetilde{\mathrm{pODC}}_h(P_0) = \begin{cases} P_0 - \frac{1}{n-1}\sum\limits_{j\neq h}^{n} \min\{F_m(Y_j), P_0\} & 1 \le h \le n, \\ P_0 - \frac{1}{n}\sum\limits_{j=1}^{n} \min\{F_{m-1,h-n}(Y_j), P_0\} & n+1 \le h \le m+n, \end{cases}$$

where

$$F_{m-1,h-n}(Y_j) = \frac{1}{m-1}\sum_{i=1, i\neq h-n}^{m} I(X_i \le Y_j).$$

Let

$$S^2_{\widetilde{\mathrm{pODC}}} = (m+n)^{-1}\sum_{h=1}^{m+n}\{\check{U}_h(P_0) - \widetilde{\mathrm{pODC}}_{jack}(P_0)\}^2.$$

Corollary 3. *Under* C.1–C.4, *as* $m, n \to \infty$,

$$\sqrt{m+n}\{\widetilde{\mathrm{pODC}}_{jack}(P_0) - \mathrm{pODC}(\mathrm{P}_0)\} \xrightarrow{d} N\left(0, \frac{\sigma_1^2}{1-\lambda} + \frac{\sigma_2^2}{\lambda}\right),$$

$$S^2_{\widetilde{\mathrm{pODC}}}(P_0) = \frac{\sigma_1^2(P_0)}{1-\lambda} + \frac{\sigma_2^2(P_0)}{\lambda} + o_p(1),$$

$$\frac{\sqrt{m+n}\{\widetilde{\mathrm{pODC}}_{jack}(P_0) - \mathrm{pODC}(\mathrm{P}_0)\}}{\sqrt{S^2_{\widetilde{\mathrm{pODC}}}(P_0)}} \xrightarrow{d} N(0,1).$$

We define the empirical likelihood ratio $R_{\mathrm{pODC}}\{P_0, \mathrm{pODC}(\mathrm{P}_0)\}$,

$$R_{\mathrm{pODC}}\{P_0, \mathrm{pODC}(\mathrm{P}_0)\} = \frac{\sup\left\{\prod_{i=1}^{m+n} p_i : \sum_{i=1}^{m+n} p_i = 1, \sum_{i=1}^{m+n} p_i \check{U}_i(P_0) = \mathrm{pODC}(\mathrm{P}_0), \mathrm{p_i} > 0, \mathrm{i} = 1, \ldots, \mathrm{m}+\mathrm{n}\right\}}{\sup\left\{\prod_{i=1}^{m+n} p_i, \sum_{i=1}^{m+n} p_i = 1, p_i > 0, i = 1, \ldots, m+n\right\}}.$$

The empirical log-likelihood ratio is

$$l_{\mathrm{pODC}}\{P_0, \mathrm{pODC}(\mathrm{P}_0)\} = -2\log[\mathrm{R}_{\mathrm{pODC}}\{\mathrm{P}_0, \mathrm{pODC}(\mathrm{P}_0)\}].$$

Corollary 4. *Under* C.1–C.4, $l_{\mathrm{pODC}}\{P_0, \mathrm{pODC}(\mathrm{P}_0)\} \xrightarrow{\mathrm{d}} \chi_1^2$.

Thus, the asymptotic $100(1-\alpha)\%$ JEL confidence interval for $\mathrm{pODC}(\mathrm{P}_0)$ is

$$I_{\mathrm{pODC}}(P_0) = \left\{\tilde{V} : l_{\mathrm{pODC}}(P_0, \tilde{V}) \le \chi_1^2(\alpha)\right\}.$$

3. Numerical Studies

In this section, we report on simulations to evaluate the estimators of Section 2. In the first simulations, we compared our normal approximation method with that of He and Escobar (2008) based on the empirical variance estimator. In the second simulations, we compare the performance of the normal approximation (NA), jackknife, and JEL methods for both pAUC and pODC.

3.1. Comparison of Corollary 1's method with the existing method

Using the same settings as He and Escobar (2008), we generated data sets consisting of samples, $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ where $X_i \sim N(0,1)$ and $Y_j \sim N(1,1)$ for $i = 1, \ldots, m$ and $j = 1, \ldots, n$. We used samples sizes (m, n) of (50, 50), (100, 100), (150, 150), (200, 200), (200, 3,000), (3,000, 200), (200, 6,000) and (6,000, 200). For each data set we computed $\widehat{\mathrm{pAUC}}(P_0)$ for P_0 of 0.6 and

Table 1. Coverage probability (cp) of 95% NA confidence intervals for pAUC(P_0) and standard deviation (s).

P_0	m	n	Corollary 1's Method		He and Escobar's Method	
			cp	s	cp	s
0.6	50	50	0.897	1.24	0.877	1.44
0.6	100	100	0.925	1.09	0.897	1.27
0.6	150	150	0.934	1.06	0.911	1.21
0.6	200	200	0.943	1.04	0.897	1.20
0.6	200	3,000	0.939	0.98	0.946	1.04
0.6	3,000	200	0.951	1.00	0.808	1.56
0.6	200	6,000	0.947	1.01	0.942	1.02
0.6	6,000	200	0.953	1.00	0.758	1.70
0.8	50	50	0.895	1.28	0.907	1.23
0.8	100	100	0.930	1.08	0.915	1.14
0.8	150	150	0.933	1.07	0.909	1.17
0.8	200	200	0.946	1.02	0.918	1.14
0.8	200	3,000	0.951	1.01	0.938	1.01
0.8	3,000	200	0.949	1.00	0.808	1.56
0.8	200	6,000	0.950	1.01	0.942	0.99
0.8	6,000	200	0.951	1.00	0.821	1.48

0.8, and 95% confidence interval (CI) for pAUC(P_0). For each setting we then computed the coverage probability (cp) of CI and sample standard deviation (s) for 1,000 data sets.

As shown in Table 1, our proposed estimator performs better than that of He and Escobar. For our estimator, coverage probabilities and estimates of standard deviation are close to expected values of 0.95 and 1 in small and moderate samples, while those of He and Escobar do not. For example, for sample sizes (3,000, 200) and (6,000, 200), the He and Escobar method had low coverage and inaccurate standard deviation estimates. Both methods are acceptable for imbalanced samples when n is much larger than m such as $(m, n) = (200,\ 3{,}000)$ or (200, 6,000).

With similar arguments as in equations (3.15), (3.16), (3.18), and (3.19) in Liu (2006), $\sqrt{m+n}\{\widehat{\mathrm{pAUC}}(P_0) - \mathrm{pAUC}(P_0)\}$ can be represented as a sum of two terms, where the second term $\sigma_4^2/(1-\lambda)$ is estimated by including the variance from sample quantile. By ignoring the trimmed effect on the variance estimator, the method of He and Escobar fails to correctly estimate the contribution of $\sigma_4^2/(1-\lambda)$ to the variance of $\widehat{\mathrm{pAUC}}(P_0)$: settings where n is much larger than m, $\lambda \to 0$, the estimate of $\sigma_4^2/(1-\lambda)$ from the sample quantile $G_n(y)$ becomes negligible and this explains the good performance of their estimator in these settings. However, if m is much larger than n and in their balanced samples, their estimator performs poorly.

3.2. Comparison of NA, jackknife, and JEL methods for pAUC and pODC

Table 2. Coverage probability of 95% confidence interval for the pAUC(P_0).

P_0	m	n	JEL (A)	JKN (A)	NA (A)	JEL (B)	JKN (B)	NA (B)	JEL (C)	JKN (C)	NA (C)
0.5	20	20	0.924	0.943	0.937	0.856	0.898	0.885	0.920	0.933	0.924
0.5	30	30	0.924	0.939	0.934	0.876	0.914	0.906	0.931	0.945	0.935
0.5	40	40	0.925	0.927	0.922	0.894	0.916	0.912	0.932	0.941	0.941
0.5	50	50	0.944	0.945	0.942	0.917	0.927	0.922	0.920	0.928	0.924
0.5	80	80	0.945	0.946	0.943	0.914	0.926	0.923	0.942	0.950	0.948
0.5	100	100	0.939	0.944	0.941	0.932	0.936	0.934	0.953	0.952	0.952
0.6	20	20	0.929	0.938	0.929	0.907	0.923	0.915	0.934	0.937	0.927
0.6	30	30	0.948	0.954	0.950	0.927	0.936	0.926	0.932	0.936	0.935
0.6	40	40	0.941	0.951	0.948	0.937	0.935	0.930	0.940	0.946	0.943
0.6	50	50	0.941	0.943	0.943	0.923	0.931	0.930	0.933	0.939	0.936
0.6	80	80	0.944	0.947	0.947	0.939	0.944	0.941	0.945	0.948	0.943
0.6	100	100	0.959	0.957	0.957	0.940	0.946	0.944	0.940	0.947	0.945

Table 3. Average length of 95% confidence interval for the pAUC(P_0).

P_0	m	n	JEL (A)	JKN (A)	NA (A)	JEL (B)	JKN (B)	NA (B)	JEL (C)	JKN (C)	NA (C)
0.5	20	20	0.227	0.238	0.229	0.257	0.268	0.258	0.199	0.210	0.201
0.5	30	30	0.181	0.191	0.186	0.211	0.222	0.216	0.158	0.168	0.164
0.5	40	40	0.154	0.165	0.162	0.181	0.192	0.188	0.136	0.146	0.143
0.5	50	50	0.138	0.148	0.146	0.163	0.174	0.171	0.119	0.129	0.127
0.5	80	80	0.106	0.116	0.115	0.126	0.137	0.135	0.092	0.101	0.100
0.5	100	100	0.093	0.103	0.103	0.113	0.123	0.122	0.081	0.091	0.090
0.6	20	20	0.264	0.277	0.268	0.301	0.314	0.304	0.245	0.257	0.248
0.6	30	30	0.212	0.224	0.219	0.246	0.258	0.252	0.195	0.206	0.201
0.6	40	40	0.182	0.194	0.190	0.213	0.224	0.220	0.167	0.179	0.175
0.6	50	50	0.161	0.173	0.170	0.188	0.199	0.196	0.148	0.158	0.156
0.6	80	80	0.125	0.136	0.135	0.147	0.158	0.157	0.114	0.124	0.123
0.6	100	100	0.111	0.122	0.121	0.130	0.141	0.140	0.101	0.111	0.110

In these simulations, we evaluated the performance of our NA method, the jackknife method (JKN), and JEL for pAUC and pODC. We generated samples, $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$. As (A), $X_i \sim N(0.2, 0.5^2)$ and $Y_j \sim N(0, 0.5^2)$ for $i = 1, \ldots, m$ and $j = 1, \ldots, n$. As (B), $X_i \sim Exp(1)$ and $Y_j \sim N(1, 0.5^2)$. As (C), $X_i \sim Exp(1)$ and $Y_j \sim Exp(1)$. We used samples sizes (m, n) of (20, 20), (30, 30), (40, 40), (50, 50), (80, 80) and $(100, 100)$. For each data set we computed 95% confidence interval (CI) for either pAUC(P_0) or pODC(P_0) at $P_0 = 0.5$ or 0.6. For each setting we computed coverage probability and average length of confidence intervals for 1,000 data sets.

Simulation results for the pAUC in Table 2 show that coverage probabilities are good for all three methods with P_0 of 0.5 and 0.6. Coverage probability increases with increasing sample sizes and it approaches the nominal 0.95 level

Table 4. Coverage probability of 95% confidence interval for the pODC(P_0).

P_0	m	n	JEL (A)	JKN (A)	NA (A)	JEL (B)	JKN (B)	NA (B)	JEL (C)	JKN (C)	NA (C)
0.5	20	20	0.913	0.930	0.916	0.856	0.886	0.879	0.916	0.927	0.916
0.5	30	30	0.918	0.919	0.914	0.896	0.929	0.925	0.934	0.949	0.946
0.5	40	40	0.929	0.935	0.930	0.918	0.929	0.925	0.912	0.925	0.922
0.5	50	50	0.939	0.941	0.937	0.914	0.927	0.927	0.945	0.952	0.950
0.5	80	80	0.917	0.925	0.924	0.942	0.942	0.937	0.945	0.947	0.944
0.5	100	100	0.956	0.961	0.960	0.929	0.930	0.929	0.931	0.940	0.936
0.6	20	20	0.920	0.915	0.911	0.882	0.919	0.909	0.932	0.936	0.931
0.6	30	30	0.943	0.943	0.937	0.907	0.939	0.931	0.934	0.939	0.935
0.6	40	40	0.941	0.939	0.935	0.918	0.936	0.933	0.929	0.932	0.930
0.6	50	50	0.933	0.944	0.941	0.926	0.936	0.934	0.932	0.937	0.931
0.6	80	80	0.930	0.933	0.933	0.939	0.953	0.950	0.932	0.938	0.936
0.6	100	100	0.945	0.943	0.942	0.929	0.938	0.937	0.938	0.942	0.940

Table 5. Average length of 95% confidence interval for the pODC(P_0).

P_0	m	n	JEL (A)	JKN (A)	NA (A)	JEL (B)	JKN (B)	NA (B)	JEL (C)	JKN (C)	NA (C)
0.5	20	20	0.162	0.166	0.159	0.146	0.148	0.141	0.205	0.210	0.202
0.5	30	30	0.131	0.134	0.131	0.116	0.119	0.115	0.164	0.169	0.165
0.5	40	40	0.111	0.115	0.113	0.098	0.101	0.099	0.139	0.144	0.141
0.5	50	50	0.098	0.102	0.100	0.084	0.087	0.086	0.124	0.130	0.128
0.5	80	80	0.075	0.080	0.079	0.065	0.069	0.069	0.097	0.102	0.101
0.5	100	100	0.067	0.072	0.071	0.056	0.060	0.060	0.086	0.091	0.090
0.6	20	20	0.205	0.211	0.203	0.208	0.212	0.205	0.249	0.256	0.248
0.6	30	30	0.165	0.170	0.166	0.164	0.169	0.165	0.199	0.205	0.200
0.6	40	40	0.143	0.148	0.145	0.141	0.145	0.143	0.172	0.178	0.175
0.6	50	50	0.127	0.132	0.131	0.124	0.128	0.127	0.152	0.158	0.156
0.6	80	80	0.098	0.103	0.102	0.096	0.101	0.100	0.119	0.125	0.124
0.6	100	100	0.087	0.092	0.091	0.084	0.089	0.089	0.106	0.111	0.110

for the largest sample size across all three estimators under A, B, and C. Results in Table 3 demonstrate that CI length decreases with increasing sample sizes for all three methods, and the JEL method produces slightly narrower CIs compared with the jackknife and NA methods in most cases. Simulation results in Tables 4 and 5 show that the proposed JEL method also has a similar advantage over the NA and jackknife methods for the pODC.

We also used simulations to substantiate the consistency of jackknife variance estimators (Lemma 3 and Corollary 3). We generated data from the normal and exponential distributions of A, B, and C. For each setting, we generated 50 repetitions and computed the mean squared error (MSE) of jackknife variance estimators of $\widetilde{\mathrm{pAUC}}(P_0)$ and $\widetilde{\mathrm{pODC}}(P_0)$ at $P_0 = 0.4$ or 0.6. The plots in Figures 3 and 4 show a decrease in MSE as the sample size increases. The Matlab code for these simulations is available from the authors upon request.

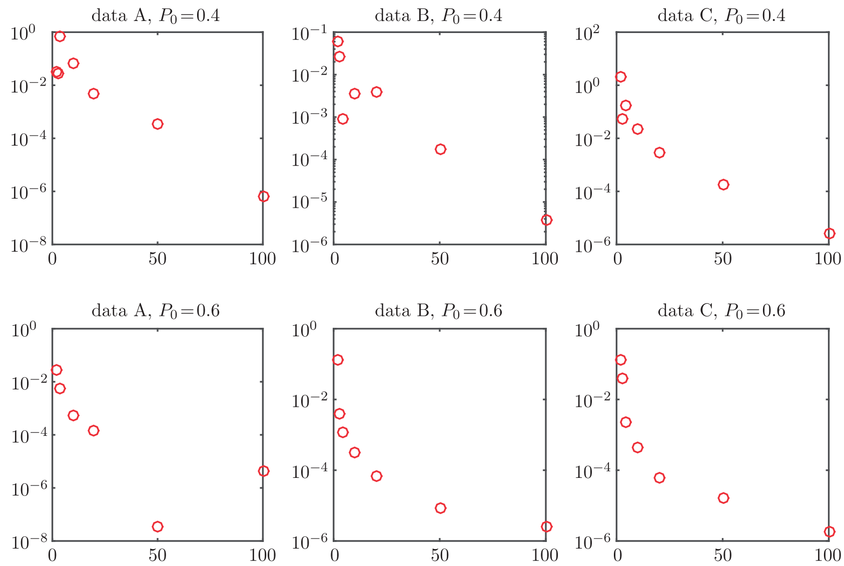


Figure 3. MSE for the jackknife variance estimator for the partial AUC.

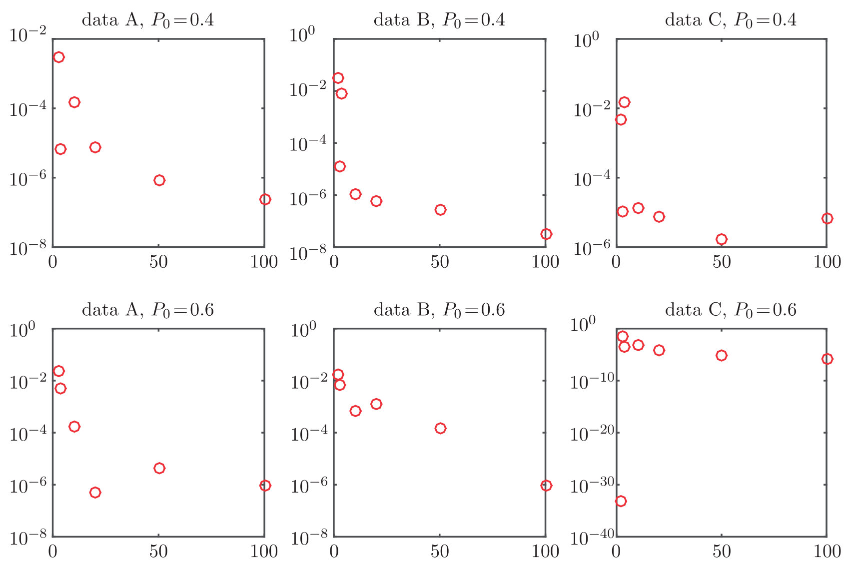


Figure 4. MSE for the jackknife variance estimator for the partial area under ODC.

4. An Application

In this section, we illustrate the proposed approaches for the partial AUC using data from the Pancreatic Cancer Serum Biomarkers study. We calculated 95% NA and JEL confidence intervals for the pAUC at varying levels of P_0

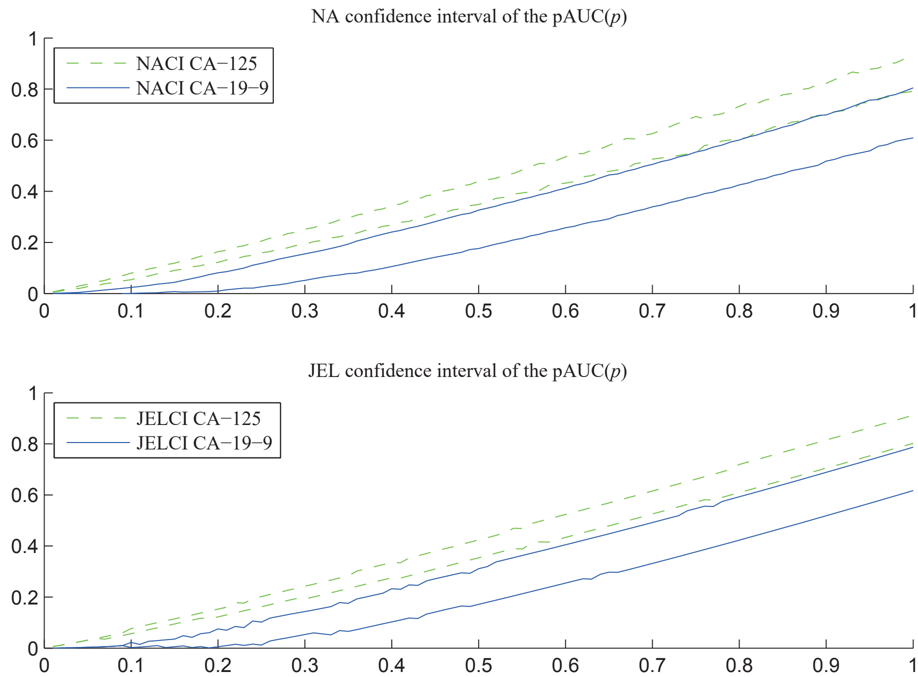


Figure 5. 95% point-wise JEL and NA confidence intervals for partial AUC's with Pancreatic Cancer Serum Biomarkers data.

from 0 to 1 for the biomarkers CA-125 (V1) and CA-19-9 (V2). From 95% JEL confidence interval in Figure 5, we can distinguish the two biomarkers. Due to the overlapping NA confidence intervals for two biomarkers in the right tail, the normal approximation method cannot do it. The proposed jackknife empirical likelihood with a slightly narrower confidence interval for the pAUC provides a more accurate interval estimate than the normal approximation method does in practice.

5. Discussion

Properties of U-statistics have been widely employed in inference procedures for ROC-related estimators including the pAUC. Since the pAUC involves sample-dependent quantile estimator, an application of U-statistic theory and jackknife procedures is not straightforward. Our proposed jackknife and JEL methods based on the estimator from Wang and Chang (2011) avoids these difficulties. We have results about the normal approximation and jackknife empirical likelihood methods. The proposed jackknife variance estimator is straightforward to implement.

Our simulations find that our proposed interval estimation methods are robust and relatively simple to carry out. The jackknife empirical likelihood method provides data-driven and asymmetric confidence intervals, but the jackknifing process may carry a large computational burden that we will address in future studies.

The variance estimation of He and Escobar (2008) originated from Sen (1960) and Bamber (1975) according to U-statistics properties. Their estimation equation of pAUC is a trimmed U-statistics instead of a typical two-sample U-statistics; it is not consistent, failing to include the trimmed effect for estimating the sample quantiles. Their method is innovative in the application of two-sample trimmed U-statistic (Janssen, Serfling and Veraverbeke (1987)) to pAUC analysis. Arvesen (1969) also derived several theorems for jackknifing trimmed U-statistics and they can provide a foundation for developing jackknife empirical likelihood methods for trimmed U-statistics.

Motivated by DeLong, DeLong and Clarke-Pearson (1988), it will be useful to apply jackknifing and JEL methods to a linear combination of partial AUC's, and theorems for multi-variable trimmed U-statistics would be helpful. The study of the jackknife empirical likelihood approach for the difference in two correlated pAUC's is a natural extension of the JEL approach in this paper.

Supplementary Materials

Proofs of the main results in this paper are provided in the online supplementary material.

Acknowledgements

Hanfang Yang is supported by the National Natural Science Foundation of China (No. 11501567). Yichuan Zhao is grateful to the support from the NSA Grant (H98230-12-1-0209) and NSF Grant (DMS-1406163). The authors would like to express the sincere thanks to the two referees, an associate editor, and the Co-Editor for their suggestions leading to a significant improvement of the paper. The authors thank Anna Moss for the useful discussions.

References

Adimari, G. and Chiogna, M. (2012). Jackknife empirical likelihood based confidence intervals for partial areas under ROC curves. *Statist. Sinica* **22**, 1457-1477.

Arvesen, J. (1969). Jackknifing U-Statistics. *Ann. Math. Statist.* **40**, 2076-2100.

Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *J. Math. Psychol.* **12**, 387-415.

DeLong, E. R., DeLong, D. and Clarke-Pearson, D. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* **44**, 837-845.

Dodd, L. E. and Pepe, M. S. (2003). Partial AUC estimation and regression. *Biometrics* **59**, 614-623.

Gong, Y., Peng, L. and Qi, Y. (2010). Smoothed jackknife empirical likelihood method for ROC curve. *J. Multivariate Anal.* **101**, 1520-1531.

He, Y. and Escobar, M. (2008). Nonparametric statistical inference method for partial areas under receiver operating characteristic curves, with application to genomic studies. *Stat. Med.* **27**, 5291-5308.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19**, 293-325.

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

Janssen, P., Serfling, R. and Veraverbeke, N. (1987). Asymptotic normality of U-statistics based on trimmed samples. *J. Statist. Plann. Inference* **16**, 63-74.

Jing, B., Yuan, J. and Zhou, W. (2009). Jackknife empirical likelihood. *J. Amer. Statist. Assoc.* **104**, 1224-1232.

Liu, H. (2006). Asymptotic properties of partial areas under the receiver operating characteristic curve with applications in microarray experiments. University of Kentucky, Doctoral Dissertation.

Sen, P. K. (1960). On some convergence properties of U-statistics. *Calcutta Statist. Assoc. Bull.* **10**, 1-18.

Wang, Z. and Chang, Y. (2011). Marker selection via maximizing the partial area under the ROC curve of linear risk scores. *Biostatistics* **12**, 369-385.

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

Yang, H. and Zhao, Y. (2013). Smoothed jackknife empirical likelihood for the difference of ROC curve. *J. Multivariate Anal.* **115**, 270-284.

Yang, H. and Zhao, Y. (2015). Smoothed jackknife empirical likelihood inference for ROC curves with missing data, *J. Multivariate Anal.* **115**, 123-138.

Zhang, D. Z., Zhou, X. H., Freeman, D. H. and Freeman, J. L. (2002). A non-parametric method for the comparison of partial areas under ROC curves and its application to large health care data sets. *Statist. Medicine* **21**, 701-715.

School of Statistics, Renmin University of China, Beijing 100872, China.

E-mail: hyang@ruc.edu.cn

Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544, USA.

E-mail: kun_lu1992@hotmail.com

Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30303, USA.

E-mail: yichuan@gsu.edu

(Received November 2013; accepted January 2016)