Statistica Sinica Preprint No: SS-13-367R2	
Title	A nonparametric approach for partial areas under ROC curves and ordinal dominance curves
Manuscript ID	SS-2013-367R2
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.2013.367
Complete List of Authors	Hanfang Yang Kun Lu and Yichuan Zhao
Corresponding Author	Yichuan Zhao
E-mail	dz2007@gmail.com
Notice: Accepted version subject to English editing.	

A nonparametric approach for partial areas under ROC curves and ordinal dominance curves

Hanfang Yang[1], Kun Lu[2] and Yichuan Zhao[3]

[1] School of Statistics, Renmin University of China, Beijing, China

[2] Department of Statistics, University of Chicago, Chicago, U.S.A

[3] Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30303

email: yichuan@gsu.edu

February 10, 2016

Abstract

The receiver operating characteristic (ROC) curve is a well-known technique used to measure the performance of a classification method. For many reasons, 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 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 for the pAUC and pODC using the normal approximation method,

jackknife method and jackknife empirical likelihood method. The simulation study demonstrates the flaws of the existing method and shows three proposed methods perform well. Our simulations also verify the consistency of our jackknife variance estimator as well. The Pancreatic Cancer Serum Biomarker data set is used to illustrate the proposed methods which are useful in medical study.

KEY WORDS: Normal approximation; Jackknife empirical likelihood; 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). For a classification, assume that F and G are two cumulative distribution functions of random variables X and Y corresponding two independent populations. Let $G^{-1}(t) = \inf\{y : G(y) \geq t\}$ be a quantile function of G, for each $0 < t < 1$. Let $S_F(t)$ and $S_G(t)$ be the corresponding survival functions of X and Y, i.e., $S_F(t) = 1 - F(t)$ and $S_G(t) = 1 - G(t)$. For $t \in (0, 1)$, the ROC curve is defined as $ROC(t) = 1 - F\{G^{-1}(1-t)\}$ or $ROC(t) = S_F\{S_G^{-1}(t)\}$, where t is the value of FPR and $S_G^{-1}(t) = G^{-1}(1-t)$. ROC curve is not a convenient tool to make a judgement on which one curve dominates against the other one, in particular when two ROC curves cross. A single 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 (AUC) as $AUC = \int_0^1 ROC(t)\mathrm{d}t = \int_{\infty}^{-\infty} S_F(u)\mathrm{d}S_G(u)$. Due to economical and practical purposes, people usually force the FPR in a low level rather than the entire area under the ROC curve. When our interest is restricted to a sub-region of the ROC space, the partial area under the ROC curve, defined as $pAUC(P_0) = \int_0^{P_0} ROC(p)\mathrm{d}p$ for the threshold value of FPR $P_0 \in (0, 1)$, can provide a more useful summary measure than the AUC.

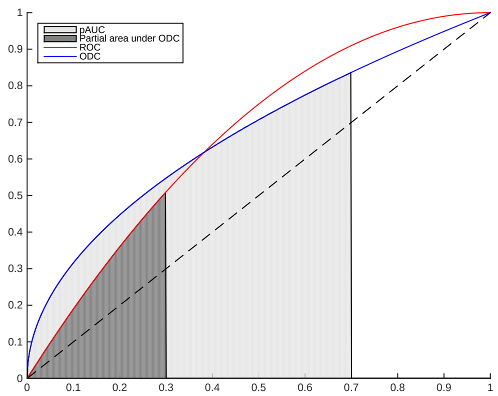


Figure 1: ODC and ROC curve.

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), which is defined by $ODC(t) = G\{F^{-1}(t)\}$ where $t \in (0, 1)$. Similar to the area under the ROC curve, the area under the ODC defined by $\int_0^1 ODC(t)\mathrm{d}t = \int_{-\infty}^{\infty} G(u)\mathrm{d}F(u)$ is a commonly used summary measure. A partial area under the ODC (pODC) from 0 to P_0 is defined as $\int_0^{P_0} ODC(t)\mathrm{d}t$.

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 et al. (1989) presented nonparametric methods for the difference between ROC curves or AUC's. Based on the jackknife empirical likelihood (Jing et al., 2009), Gong et al. (2010) proposed a smoothed inference procedure for the ROC curve and Yang and Zhao (2013) developed new inference method for the difference of two ROC curves. Several researchers have also applied the properties of U-statistics (Hoeffding, 1948) to make an inference for the AUC and pAUC. For example, DeLong et al. (1988), Sen (1960) and Bamber (1975) employed a multi-dimensional version of Hoeffding's theory (1948) for Mann-Whitney U-statistics in an 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 method, jackknife method and JEL method, respectively. In Section 3, we conduct extensive simulation studies in terms of coverage probability and average length of confidence intervals. Furthermore, we show how to apply our proposed methods to a practical problem in Section 4 and make a discussion in Section 5. All the proofs are provided in the supplementary material.

2 Inference procedure

Recall that X and Y are two independent random variables with distribution functions $F(x)$ and $G(y)$. Let $F^{-1}(t)$ be a quantile function of F. Recall $0 < P_0 < 1$. The partial area under ODC, i.e., $ODC(t) = G\{F^{-1}(t)\}$ is $pODC(P_0) = \int_0^{P_0} ODC(t)\mathrm{d}t = \int_{-\infty}^{F^{-1}(P_0)} G(u)\mathrm{d}F(u)$. Let $\mathbf{X} = \{X_i, i = 1, ..., m\}$ and $\mathbf{Y} = \{Y_i, i = 1, ..., n\}$ be random samples from the distribution functions $F(x)$ and $G(y)$, respectively. A simple empirical estimator of $\widehat{pODC}(P_0)$ is given by

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

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

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

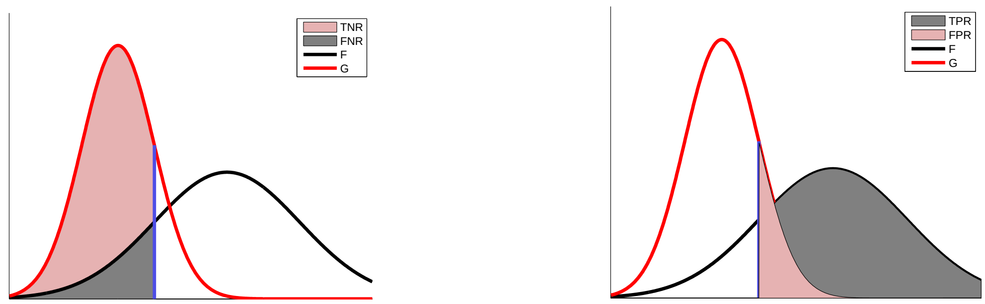
Liu (2006) developed the asymptotic normality for the empirical estimator $\widehat{pODC}(P_0)$. Before giving the asymptotic normality of $\widehat{pODC}(P_0)$, we display the following conditions which are common in practice:

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 $F^{-1}(P_0)$, and $F'(F^{-1}(P_0)) > 0$;

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



(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 the above conditions C.1-C.4, we have that*

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

where

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

and

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

As shown in Figure 2, points on the ROC curve, $ROC = (FPR, TPR) = (1 - TNR, 1 - FNR)$ can be obtained from the ODC curve, $ODC = (FNR, TNR)$. Recall that $pAUC(P_0) = \int_0^{P_0} S_F\{S_G^{-1}(u)\}du$ and obtain its empirical estimator as

$$\widehat{pAUC}(P_0) = \int_0^{P_0} S_{F,m}\{S_{G,n}^{-1}(u)\}du = \int_{+\infty}^{S_{G,n}^{-1}(P_0)} S_{F,m}(t)dS_{G,n}(t),$$

or

$$\widehat{pAUC}(P_0) = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} I(X_i \geq Y_j) I\{Y_j \geq S_{G,n}^{-1}(P_0)\},$$

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. Using the similar methods in Liu (2006), we extend Lemma 1 for the pODC to Corollary 1 for the pAUC as follows.

Corollary 1. *Assume the conditions C.1-C.4 hold. Then*

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

where

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

and

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

Remark 1: In this corollary, we provide the variance of $\widehat{pAUC}$ and explicitly account for the random error due to $S_{G,n}^{-1}(P_0)$, which He and Escobar (2008) have ignored. Our simulation study in Section 3.1 follows from Corollary 1, where the variance of $\widehat{pAUC}$ consists of two parts σ_3^2/λ and $\sigma_4^2/(1-\lambda)$, representing two independent processes. The simulation results demonstrates the improved performance of variance estimators based on our method as the sample size becomes large.

Also, contrary to arguments presented by Adimari and Chiogna (2011), the quantile estimator is problematic for jackknife variance estimators (see Miller (1974) and Shao

and Wu (1989)). Thus jackknifing procedures cannot be applied to $\widehat{pAUC}(P_0)$ directly. Adimari and Chiogna (2011) attempt to overcome this limitation of the jackknife by plugging in quantile estimate $S_{G,n}^{-1}(P_0)$ determined only from the full sample. However, their resulting jackknife variance estimator fails to incorporate the error associated with the quantile estimate, and thus it is not a consistent estimator for the variance of $\widehat{pAUC}(P_0)$. The lack of consistency in the approach of Adimari and Chiogna (2011) is demonstrated from our simulation study in Section 3.1.

Jackknife methods can be applied to an alternative estimator of $pAUC(P_0)$ by Wang and Chang (2011), which is defined as follows.

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

We established the following theorem:

Theorem 1. *Suppose that the conditions C.1-C.4 hold. One has that*

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

The two estimators $\widehat{pAUC}(P_0)$ and $\widetilde{pAUC}(P_0)$ closely agree with each other, but $\widetilde{pAUC}(P_0)$ avoids the use of a quantile estimator. Thus we can derive jackknife and JEL methods based on the estimator $\widetilde{pAUC}(P_0)$. We propose the jackknife method and JEL method based on the estimator $\widetilde{pAUC}(P_0)$. For $\widetilde{pAUC}(P_0)$, the jackknife estimator $\widetilde{pAUC}_{jack}(P_0)$ is given by

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

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

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

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 the conditions C.1-C.4, we have that*

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

The following lemma establishes the consistency of the jackknife variance estimator $S^2_{\widetilde{pAUC}}$ given as

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

Lemma 3. *Suppose the conditions C.1-C.4 hold. Then,*

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

From Slusky's theorem, and Lemmas 2 and 3, we have the following theorem, which establishes the jackknife based method for $\widetilde{pAUC}(P_0)$.

Theorem 2. *Assume the conditions C.1-C.4 hold. We have that*

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

In order to derive the 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

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

The empirical log-likelihood ratio for the $pAUC(P_0)$ is obtained as follows.

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

where the Lagrange multiplier λ_1 satisfies the nonlinear equation

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

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

Theorem 3. *Under the conditions C.1-C.4, one has that*

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

From Theorem 3, the asymptotic $100(1-\alpha)\%$ JEL confidence interval for $pAUC(P_0)$ is given by

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

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. Using arguments similar to Corollary 1, we derive the following corollaries for the pODC. Following Wang and Chang (2011), we define an estimator $\widetilde{pODC}(P_0)$ as

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

Corollary 2. *Assume that the conditions C.1-C.4 hold. We have that as* $m, n \to \infty$

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

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

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

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

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

where

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

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

Corollary 3. *Under the conditions C.1-C.4, one has that as $m, n \to \infty$*

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

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

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

We define the empirical likelihood ratio $R_{pODC}\{P_0, pODC(P_0)\}$ as follows,

$$\begin{aligned}&R_{pODC}\{P_0, pODC(P_0)\}\\ =&\frac{\sup\{\prod_{i=1}^{m+n} p_i : \sum_{i=1}^{m+n} p_i = 1, \sum_{i=1}^{m+n} p_i\check{U}_i(P_0) = pODC(P_0), p_i > 0, i = 1, \ldots, m+n\}}{\sup\{\prod_{i=1}^{m+n} p_i, \sum_{i=1}^{m+n} p_i = 1, p_i > 0, i = 1, \ldots, m+n\}}.\end{aligned}$$

The empirical log-likelihood ratio is represented as follows,

$$l_{pODC}\{P_0, pODC(P_0)\} = -2\log[R_{pODC}\{P_0, pODC(P_0)\}].$$

Then we establish the Wilks' theorem for $pODC(P_0)$.

Corollary 4. *Assume the conditions C.1-C.4 hold. We have*

$$l_{pODC}\{P_0, pODC(P_0)\} \xrightarrow{d} \chi_1^2.$$

Thus, the asymptotic $100(1-\alpha)\%$ JEL confidence interval for $pODC(P_0)$ is

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

3 Numerical Studies

In this section, we perform simulation studies to evaluate the estimators derived in Section 2. In the first simulation study, we conduct comprehensive simulation studies to compare our normal approximation method with He and Escobar (2008)'s method based on the empirical variance estimator. In the second simulation study, 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

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	3000	0.939	0.98	0.946	1.04
0.6	3000	200	0.951	1.00	0.808	1.56
0.6	200	6000	0.947	1.01	0.942	1.02
0.6	6000	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	3000	0.951	1.01	0.938	1.01
0.8	3000	200	0.949	1.00	0.808	1.56
0.8	200	6000	0.950	1.01	0.942	0.99
0.8	6000	200	0.951	1.00	0.821	1.48

In this simulation study, we perform the NA method comparing our variance estimator based on Corollary 1 with the variance estimator from variance formula (3) in He and Escobar (2008). Using the same settings as He and Escobar (2008), we generat-

ed data sets consisting of two 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 $(50, 50), (100, 100), (150, 150), (200, 200), (200, 3000), (3000, 200), (200, 6000)$ and $(6000, 200)$. For each data set we computed $\widehat{pAUC}(P_0)$ for P_0 of 0.6 and 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 1000 data sets.

As shown in Table 1, our proposed estimator performs better than the existing one from He and Escobar. For our estimator, coverage probabilities and estimates of standard deviation are close to expected values of 0.95 and 1. Our estimator performs well in small and moderate samples, but the He and Escobar method has poor performance. For example, for sample sizes $(3000, 200)$ and $(6000, 200)$, 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 $(m, n) = (200, 3000)$ or $(200, 6000)$.

With similar arguments as equations (3.15), (3.16), (3.18) and (3.19) in Liu (2006), $\sqrt{m+n}\{\widehat{pAUC}(P_0) - pAUC(P_0)\}$ 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 $\widehat{pAUC}(P_0)$.

In settings where n is much larger than m, $\lambda \to 0$, $\sigma_3^2/\lambda + \sigma_4^2/(1-\lambda)$ is close to σ_3^2/λ, and the estimate of $\sigma_4^2/(1-\lambda)$ from the sample quantile $G_n(y)$ becomes negligible. In this case, both methods adequately estimate the variance, which explains the good performance of the He and Escobar estimator in these settings. Conversely, in settings where m is much larger than n where the variance is close to $\sigma_4^2/(1-\lambda)$ and balanced samples (e.g.,$(100, 100)$ and $(200, 200)$) where $\sigma_4^2/(1-\lambda)$ is non-negligible, the He and

Escobar estimator performs poorly. In contrast to the He and Escobar estimator, our estimator which fully accounts for σ_3^2/λ and $\sigma_4^2/(1-\lambda)$ performs well in almost all settings.

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

In this simulation study, we evaluate the performance of our three proposed methods: the NA method, jackknife method (JKN), and JEL for pAUC and pODC. We generated data sets consisting of two samples, $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$. For data sets (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 $j = 1, \ldots, n$. For data sets (B), $X_i \sim Exp(1)$ and $Y_j \sim N(1, 0.5^2)$. For data sets (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 1000 data sets.

Simulation results for the pAUC in Table 2 show that coverage probabilities are

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

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

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

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 for the largest sample size across all three estimators and for all distributions (i.e., 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. Compared with the other methods, the proposed JEL method has the advantage of a narrower CI and similar coverage compared with the NA and jackknife methods. 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.

In addition to the above two simulation studies, we use simulations to verify the consistency of jackknife variance estimators (Lemma 3 and Corollary 3). We generate data from normal and exponential distributions as described above for data sets A, B, and C. For each setting, we generated 50 repetitions and computed the mean squared error (MSE) of jackknife variance estimators of $\widetilde{pAUC}(P_0)$ and $\widetilde{pODC}(P_0)$ at $P_0 = 0.4$ or 0.6. All the plots in Figures 3 and 4 clearly show a decrease in MSE as the sample

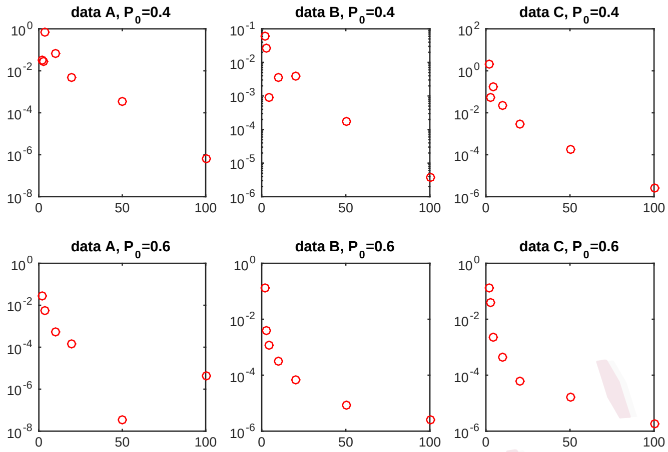


Figure 3: MSE for the jackknife variance estimator for the partial AUC.

size increases and it supports the consistency of the jackknife variance estimator. The Matlab code for these simulations is available from the authors upon request.

4 Real Application

In this section, we illustrate the proposed approaches for the partial AUC using data from the Pancreatic Cancer Serum Biomarkers study. We calculate 95% NA and JEL confidence intervals for the pAUC at varying levels P_0 from 0 to 1 for two biomarkers CA-125 (V1) and CA-19-9 (V2), respectively. 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 can not 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.

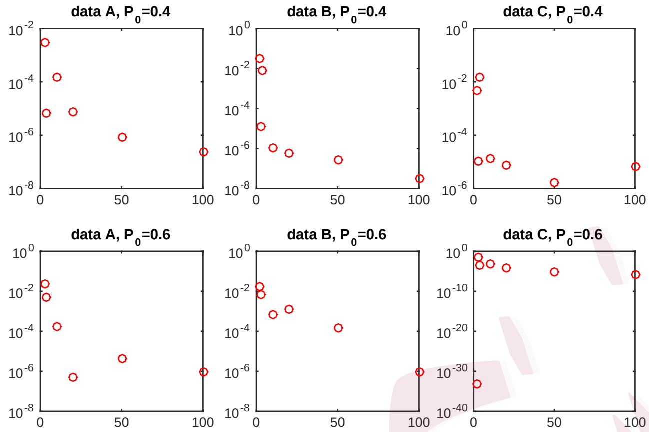


Figure 4: MSE for the jackknife variance estimator for the partial area under ODC.

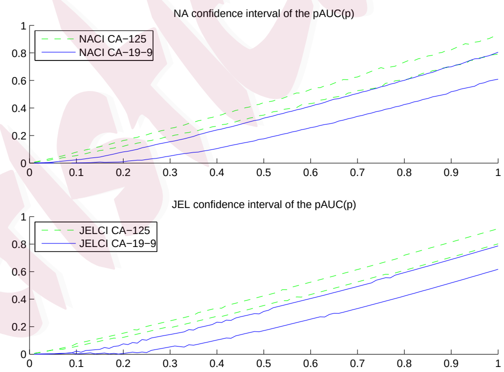


Figure 5: 95% point-wise JEL and NA confidence intervals for partial AUC's with Pancreatic Cancer Serum Biomarkers data.

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 prove related theorems about normal approximation method and jackknife empirical likelihood method. The proposed consistent jackknife variance estimator is straightforward to implement. Using our approach, the theoretical results can be extended to partial areas with two boundary points from P_1 to P_2, $0 < P_1 < P_2 < 1$, i.e., $pAUC(P_2) - pAUC(P_1)$ and $pODC(P_2) - pODC(P_1)$.

From our simulation studies, all of our proposed interval estimation methods, including the normal approximation, the jackknife, and jackknife empirical likelihood are robust and relatively simple to carry out. The jackknife empirical likelihood method provides data-driven and asymmetric confidence interval, but the jackknifing process may have a large computational burden for large data sets, which we will address in future studies.

The variance estimation by He and Escobar's method is represented as equation (3) of p. 5294 in He and Escobar (2008). Their estimation of variance has the origin from Sen (1960) and Bamber (1975) according to U-statistics properties. However, the estimation equation of pAUC, which is equation (2) in He and Escobar (2008), is a trimmed U-statistics instead of a typical two-sample U-statistics. Because the method of U-statistics is incorrectly applied to trimmed U-statistics, i.e., pAUC, He and Escobar's variance estimator is not consistent, and fails to include the trimmed effect for estimating the sample quantiles r_0 and r_1 at equation (3) in He and Escobar (2008). Although the

method of He and Escobar (2008) has some drawbacks, it is innovative in the application of two-sample trimmed U-statistic (Janssen et al., 1987) to the 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 et al. (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. It is worthwhile to study the jackknife empirical likelihood approach for the difference in two correlated pAUC's which is a natural extension of the JEL approach in this paper.

Acknowledgments

Hanfang Yang is supported by the National Natural Science Foundation of China (No. 11501567). Yichuan Zhao is supported by the NSF Grant, the NSA Grant and the RIG Grant, Georgia State University. The authors would like to express the sincere thanks to the two referees, Associate Editor, and Co-Editor for their suggestions leading to an improvement of the paper. The authors are grateful to 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, *Statistica Sinica*, Vol. 22, pp. 1457-1477.

Arvesen, J. (1969), Jackknifing U-Statistics, *Ann. Math. Statist.*, Vol. 40(6), pp. 2076-2100.

Bamber, D. (1975), The area above the ordinal dominance graph and the area below the receiver operating characteristic graph, *Journal of Mathematical Psychology*, Vol. 12, pp. 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*, Vol 44, pp. 837-845.

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

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.*, Vol. 27, pp. 5291-5308.

Hoeffding, W. (1948), A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics*, Vol. 19, pp. 293-325.

Hsieh, F. and Turnbull, B.W. (1996), Non-parametric and semi-parametric estimation of the receiver operating characteristic curve, *The Annals of Statistics*, Vol. 24, pp. 25-40.

Janssen, P., Serfling, R., Veraverbeke, N. (1987), Asymptotic normality of U-statistics based on trimmed samples, *Journal of Statistical Planning and Inference*, Vol. 16, pp. 63-74.

Jing, B., Yuan, J., and Zhou, W. (2009), Jackknife empirical likelihood, *Journal of the American Statistical Association*, Vol. 104, pp. 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.*

Miller, R.G. (1974), The jackknife: a review, *Biometrika*, Vol. 61, pp. 1-15.

Sen, P. K. (1960), On some convergence properties of U-statistics, *Calcutta Statistical Association Bulletin*, Vol. 10, pp. 1-18.

Shao, J. and Wu, C. F. J. (1989), A general theory for jackknife variance estimation, *Annals of Statistics*, Vol. 17, pp. 1176-1197.

Wang, Z. and Chang, Y. (2011), Marker selection via maximizing the partial area under the ROC curve of linear risk scores, *Biostatistics*, Vol. 12, pp. 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*, Vol. 76, pp. 585-592

Yang, H. and Zhao, Y. (2013), Jackknife empirical likelihood for the difference of ROC curve, *Journal of Multivariate Analysis*, Vol.115, pp. 270-284.

Zhang D. Z., Zhou X. H., Freeman D. H., 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, *Statistics in Medicine*, Vol. 21, pp. 701-715.