IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received October 26, 2016, accepted November 6, 2016, date of publication December 8, 2016, date of current version January 27, 2017.

Digital Object Identifier 10.1109/ACCESS.2016.2628102

A Fast Algorithm for Unbiased Estimation of Variance of AUC Based on Dynamic Programming

WEICHAO XU[1], (Member, IEEE), SHUN LIU[1], XU SUN[2], SIYANG LIU[1], AND YUN ZHANG[1]

[1]Department of Automatic Control, School of Automation, Guangdong University of Technology, Guangzhou 510006, China
[2]Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong

Corresponding author: W. Xu (wcxu@gdut.edu.cn)

This work was supported in part by National Natural Science Foundation of China under Project 61271380 and Project U1501251, in part by the Guangdong Natural Science Foundation under Grant 2014A030313515, and in part by the Guangzhou Science and Technology Plan under Project 201607010290.

ABSTRACT Receiver operating characteristic (ROC) curve is a plot traced out by pairs of false-positive rate and true-positive rate according to various decision threshold settings. The area under the ROC curve (AUC) is widely used as a figure of merit to summarize a diagnostic system's performance, a binary classifier's overall accuracy, or an energy detector's power. Exploiting the equivalent relationship between the sample version of AUC and Mann Whitney U statistic (MWUS), in this paper, we develop an efficient algorithm of linearithmic order, based on dynamic programming, for unbiased estimation of the mean and variance of MWUS. Monte Carlo simulations verify our algorithmic findings.

INDEX TERMS Area under the curve (AUC), dynamic programming, Mann-Whitney U statistic (MWUS), receiver operating characteristic (ROC).

I. INTRODUCTION

Receiver Operating Characteristic (ROC) analysis is a framework originating from the signal detection theory, with an emphasis for the analysis of radar images during World War II [1]–[3]. Since then, it has been applied in a wide spectrum of scientific and engineering areas, including data mining [4], computer vision [5], [6], signal processing [7]–[9], machine learning [10], [11], medicine [12]–[14], psychology [15] and biomedical informatics [16], among others. In essence, ROC analysis is a supervised method that requires prior knowledge of both the sample membership and underlying cumulative distribution functions (cdfs hereafter) which generate the two samples [17]. With this knowledge, a two-dimensional curve, called the ROC curve, can be plotted by pairs of false-positive rate and true-positive rate according to various decision threshold settings.

The area under the curve (AUC) is often considered as an overall summary of diagnostic accuracy in the literature [18], [19]. It is well known that the Mann-Whitney U statistic (MWUS) is an unbiased estimator of AUC [20]. To obtain the distributional information of this estimator, many researchers, such as DeLong *et al.* [21], Hanley *et al.* [22] and Brunner *et al.* [23]–[26], among others, have developed algorithms for computing the estimates of MWUS's variance. DeLong's algorithm is easy to implement, but is only asymptotically unbiased and in quadratic time. Hanley's algorithm, while having a linearithmic time complexity, is only unbiased when the samples follow negative exponential distributions [27], which is a rather ristrict assumption about the parent distribution of the samples. Brunner's rank-based method is also of linearithmic order, but, as demonstrated by the simulation results in this work, is only asymptotically unbiased. Recently, Xu *et al.* [28] proposed an efficient algorithm for estimating variance of the estimator, but only applicable to samples following continuous distributions.

To overcome the disadvantages of the existing methods, in this paper, we developed a linearithmic algorithm for unbiased estimation of the variance of MWUS, which possesses the following properties. Firstly, our algorithm is *unified*, that is, it can embrace samples from both continuous and non-continuous populations. Secondly, it is *unbiased*, that is, the mean of its output equals the population version of MWUS's variance, which is a desired feature in statistical analysis. Thirdly, it is in *linearithmic* time, that is, the time complexity is in the order of the product of sample size and its logarithm. Last but not least, the proposed algorithm is *nonparametric*, that is, it depends only on the samples, without making any parametric model assumptions concerning the functional forms of the cdfs.

The rest of this paper is organized as follows. Section II gives the basic definition of the unbiased estimator of AUC, i.e., the MWUS, as well as the associated general formulas concerning its variance. We also develop a linearithmic algorithm for computing MWUS and its variance based on dynamic programming. In section III, numerical experiments are undertaken to demonstrate the efficiency and unbiasedness of our algorithm. Finally, we draw our conclusion in section IV.

II. METHOD

For completeness and ease of later development, in this section we describe the definition of nonparametric estimator of AUC and the well known formulas concerning its expectation and variance. In addition, we develop an unbiased algorithm in nonparametric way, for estimating the variance of AUC without knowing the underlying parent distributions based on our efficient computing structure. For notational convenience, throughout we denote by $\mathbb{E}(\cdot)$ and $\mathbb{V}(\cdot)$ the expectation and variance of random variables, respectively.

A. UNBIASED ESTIMATIOR OF AUC

Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be two independent and identically distributed (i.i.d) samples drawn respectively from two populations (whose distributions can be either continuous or discrete). Then, from the work of Bamber [20], an empirical unbiased estimator of AUC can be computed over a kernel function $\mathcal{H}(\cdot)$ as:

$$\hat{\theta} \triangleq \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathcal{H}(X_i - Y_j). \tag{1}$$

where

$$\mathcal{H}(t) = \begin{cases} 1 & t > 0 \\ \frac{1}{2} & t = 0 \\ 0 & t < 0. \end{cases} \tag{2}$$

is the familiar Heaviside function. Note that (1) is also referred to as the MWUS in the statistics textbook.

B. EXPECTATION AND VARIANCE OF $\hat{\theta}$

Taking expectation on both sides of Eq. (2) leads readily to

$$\mathbb{E}(\hat{\theta}) = \underbrace{\Pr(X > Y) + \frac{1}{2}\Pr(X = Y)}_{\triangleq \theta}. \tag{3}$$

From (3), it follows that $\theta = 0.5$ when X and Y are identically distributed; whereas $\theta = 1$ if the values of Y are consistently larger than those of X. Therefore, AUC is often employed as an index to characterize the extent of separation of the distributions of X and Y.

From the work of Noether [29], the variance of $\hat{\theta}$ can be written as

$$\mathbb{V}(\hat{\theta}) = \frac{Q_0 + (n-1)Q_1 + (m-1)Q_2}{mn}. \tag{4}$$

where

$$Q_0 = \mathbb{E}\left[\mathcal{H}(X_i - Y_j)\mathcal{H}(X_i - Y_j)\right] - \theta^2, \tag{5}$$

$$Q_1 = \mathbb{E}\left[\mathcal{H}(X_i - Y_j)\mathcal{H}(X_i - Y_{j'})\right] - \theta^2 \quad j \neq j', \tag{6}$$

$$Q_2 = \mathbb{E}\left[\mathcal{H}(X_i - Y_j)\mathcal{H}(X_{i'} - Y_j)\right] - \theta^2 \quad i \neq i'. \tag{7}$$

C. UNBIASED ESTIMATOR OF $\mathbb{V}(\hat{\theta})$—SLOW VERSION

Theorem 1: Let $\hat{\theta}$ be defined as in (1) with respect to two i.i.d. samples $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ drawn from two populations, respectively. Let $\sigma^2_{\hat{\theta}}$ be a compact notation of $\mathbb{V}(\hat{\theta})$ in (4). Write

$$\hat{\sigma}^2_{\hat{\theta}} = \frac{\hat{Q}_0 + (n-1)\hat{Q}_1 + (m-1)\hat{Q}_2}{(m-1)(n-1)} \tag{8}$$

where

$$\hat{Q}_0 = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left[\mathcal{H}(X_i - Y_j)\mathcal{H}(X_i - Y_j)\right] - \hat{\theta}^2, \tag{9}$$

$$\hat{Q}_1 = \frac{1}{mn(n-1)}\sum_{i=1}^{m}\sum_{j\neq j'=1}^{n}\sum^{n}\left[\mathcal{H}(X_i - Y_j)\mathcal{H}(X_i - Y_{j'})\right] - \hat{\theta}^2, \tag{10}$$

$$\hat{Q}_2 = \frac{1}{m(m-1)n}\sum_{i\neq i'=1}^{m}\sum^{m}\sum_{j=1}^{n}\left[\mathcal{H}(X_i - Y_j)\mathcal{H}(X_{i'} - Y_j)\right] - \hat{\theta}^2. \tag{11}$$

Then $\hat{\sigma}^2_{\hat{\theta}}$ is an unbiased estimator of $\sigma^2_{\hat{\theta}}$, i.e., $\mathbb{E}(\hat{\sigma}^2_{\hat{\theta}}) = \sigma^2_{\hat{\theta}}$.

Proof: To show that $\mathbb{E}(\hat{\sigma}^2_{\hat{\theta}}) = \sigma^2_{\hat{\theta}}$, it suffices to evaluate the expectations of the $\hat{Q}$-terms in the numerator of (8). Applying the relationship of $\sigma^2_{\hat{\theta}} = \mathbb{E}(\hat{\theta}^2) - \theta^2$ yields

$$\mathbb{E}(\hat{\theta}^2) = \theta^2 + \sigma^2_{\hat{\theta}}. \tag{12}$$

Taking expectation of both sides of (9) and using (12), it follows that

$$\begin{aligned}\mathbb{E}(\hat{Q}_0) &= \mathbb{E}\left\{\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathcal{H}(X_i - Y_j)\mathcal{H}(X_i - Y_j) - \hat{\theta}^2\right\} \\ &= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbb{E}\left[\mathcal{H}(X_i - Y_j)\mathcal{H}(X_i - Y_j)\right] - \mathbb{E}(\hat{\theta}^2) \\ &= \mathbb{E}\left[\mathcal{H}(X_i - Y_j)\mathcal{H}(X_i - Y_j)\right] - \theta^2 - \sigma^2_{\hat{\theta}} \\ &= Q_0 - \sigma^2_{\hat{\theta}}.\end{aligned} \tag{13}$$

In a similar manner, we also have

$$\mathbb{E}(\hat{Q}_1) = Q_1 - \sigma^2_{\hat{\theta}}. \tag{14}$$

and

$$\mathbb{E}(\hat{Q}_2) = Q_2 - \sigma^2_{\hat{\theta}}. \tag{15}$$

Taking expectation of both sides of (8) and substituting(13)–(15) thereafter along with some straightforward algebra, we find $\mathbb{E}(\hat{\sigma}^2_{\hat{\theta}}) = \sigma^2_{\hat{\theta}}$, hence the result. □

TABLE 1. Quantities needed in the fast algorithm.

Quantities	# of events satisfying the relation inside ( )
S_1	$\mathcal{E}(X=Y)$
S_2	$\mathcal{E}(X>Y)$
S_3	$\mathcal{E}(X>Y>Y')$ or $\mathcal{E}(X>Y'>Y)$
S_4	$\mathcal{E}(X>Y=Y')$
S_5	$\mathcal{E}(X=Y'>Y)$ or $\mathcal{E}(X=Y>Y')$
S_6	$\mathcal{E}(X=Y=Y')$
S_7	$\mathcal{E}(X>X'>Y)$ or $\mathcal{E}(X'>X>Y)$
S_8	$\mathcal{E}(X=X'>Y)$
S_9	$\mathcal{E}(X'>X=Y)$ or $\mathcal{E}(X>X'=Y)$
S_{10}	$\mathcal{E}(X=X'=Y)$

D. UNBIASED ESTIMATOR OF $\mathbb{V}(\hat{\theta})$—FAST VERSION

It is noteworthy that, although unbiased, the algorithm based on Eq. (8)–(11) is very inefficient for large m and n, due to the cubic order of the time complexity of the Q_1 and Q_2 terms. Fortunately, however, a linearithmic algorithm can be constructed via rewriting Eqs. (1), (9)–(11) involved in Theorem 1 in terms of S_1 to S_{10} in Table 1, which stand for the number of events satisfying the relation inside the respective brackets. As will be shown later on, these S-terms can all be computed with dynamic programming. Note that in Table 1, X' and Y' stand for i.i.d copies of X and Y, respectively.

Theorem 2: Let $\hat{\theta}$ be defined as in (1) with respect to two i.i.d samples, $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$, respectively. Then the estimator $\sigma^2_{\hat{\theta}}$ in Theorem 1 is equivalent to

$$\hat{\sigma}^2_{\hat{\theta}} = \hat{\varsigma}^2_{\hat{\theta}} = \frac{\hat{\mathcal{Q}}_0 + (n-1)\hat{\mathcal{Q}}_1 + (m-1)\hat{\mathcal{Q}}_2'}{(m-1)(n-1)}, \tag{16}$$

where

$$\hat{\theta} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathcal{H}(X_i - Y_j) = \frac{S_1/2 + S_2}{mn}, \tag{17}$$

$$\begin{aligned}\hat{\mathcal{Q}}_0 &= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}[\mathcal{H}(X_i - Y_j)\mathcal{H}(X_i - Y_j)] - \hat{\theta}^2\\ &= \frac{S_1/4 + S_2}{mn} - \hat{\theta}^2, \end{aligned}\tag{18}$$

$$\begin{aligned}\hat{\mathcal{Q}}_1 &= \frac{1}{mn(n-1)}\sum_{i=1}^{m}\sum_{j\neq j'=1}^{n}\sum^{n}[\mathcal{H}(X_i - Y_j)\mathcal{H}(X_i - Y_{j'})] - \hat{\theta}^2\\ &= \frac{2S_3 + S_4 + S_5 + (S_6 - S_1)/4 - S_2}{mn(n-1)} - \hat{\theta}^2, \end{aligned}\tag{19}$$

and

$$\begin{aligned}\hat{\mathcal{Q}}_2 &= \frac{1}{m(m-1)n}\sum_{j=1}^{n}\sum_{i\neq i'=1}^{m}\sum^{m}[\mathcal{H}(X_i - Y_j)\mathcal{H}(X_{i'} - Y_j)] - \hat{\theta}^2\\ &= \frac{2S_7 + S_8 + S_9 + (S_{10} - S_1)/4 - S_2}{m(m-1)n} - \hat{\theta}^2. \end{aligned}\tag{20}$$

Proof: Let $\mathcal{I}(\cdot)$ be the indicator function which equals unity (zero) if the statement inside the bracket is true (false). For compactness, write $n^{[2]} \triangleq n(n-1)$. Then it follows readily that

$$\mathcal{H}(X-Y) = \mathcal{I}(X>Y) + \frac{1}{2}\mathcal{I}(X=Y). \tag{21}$$

Substituting this result into (1), we have

$$\begin{aligned}\hat{\theta} &= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left[\mathcal{I}(X_i>Y_j) + \frac{1}{2}\mathcal{I}(X_i=Y_j)\right]\\ &= \frac{1}{mn}\underbrace{\sum_{i=1}^{m}\sum_{j=1}^{n}[\mathcal{I}(X_i>Y_j)]}_{\mathcal{E}(X>Y)=S_2} + \frac{1}{2mn}\underbrace{\sum_{i=1}^{m}\sum_{j=1}^{n}[\mathcal{I}(X_i=Y_j)]}_{\mathcal{E}(X=Y)=S_1}\\ &= \frac{S_1/2 + S_2}{mn}\end{aligned}$$

which is Eq. (17). From (9) and (10), we have, respectively,

$$\begin{aligned}\hat{\mathcal{Q}}_0 &= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left[\mathcal{I}(X_i>Y_j) + \frac{1}{2}\mathcal{I}(X_i=Y_j)\right]^2 - \hat{\theta}^2\\ &= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left[\mathcal{I}(X_i>Y_j) + \frac{1}{4}\mathcal{I}(X_i=Y_j)\right] - \hat{\theta}^2\\ &= \frac{1}{mn}\underbrace{\sum_{i=1}^{m}\sum_{j=1}^{n}\left[\mathcal{I}(X_i>Y_j)\right]}_{\mathcal{E}(X>Y)=S_2}\\ &\quad + \frac{1}{4mn}\underbrace{\sum_{i=1}^{m}\sum_{j=1}^{n}[\mathcal{I}(X_i=Y_j)]}_{\mathcal{E}(X=Y)=S_1} - \hat{\theta}^2\\ &= \frac{S_1/4 + S_2}{mn} - \hat{\theta}^2 = \hat{\mathcal{Q}}_0\end{aligned}$$

which verifies the statement in (18), and

$$\begin{aligned}\hat{\mathcal{Q}}_1 &= \frac{1}{mn^{[2]}}\sum_{i=1}^{m}\sum_{j\neq j'=1}^{n}\sum^{n}[\mathcal{H}(X_i - Y_j)\mathcal{H}(X_i - Y_{j'})] - \hat{\theta}^2\\ &= \frac{1}{mn^{[2]}}\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{j'=1}^{n}[\mathcal{H}(X_i - Y_j)\mathcal{H}(X_i - Y_{j'})]\\ &\quad - \frac{1}{mn^{[2]}}\sum_{i=1}^{m}\sum_{j=1}^{n}[\mathcal{H}(X_i - Y_j)\mathcal{H}(X_i - Y_j)] - \hat{\theta}^2\\ &= \frac{1}{mn^{[2]}}\underbrace{\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{j'=1}^{n}[\mathcal{I}(X_i>Y_j)\mathcal{I}(X_i>Y_{j'})]}_{\mathcal{E}(X>Y>Y')+\mathcal{E}(X>Y'>Y)+\mathcal{E}(X>Y=Y')\Rightarrow 2S_3+S_4}\\ &\quad + \frac{1}{2mn^{[2]}}\underbrace{\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{j'=1}^{n}[\mathcal{I}(X_i>Y_j)\mathcal{I}(X_i=Y_{j'})]}_{\mathcal{E}(X=Y'>Y)\Rightarrow S_5}\\ &\quad + \frac{1}{2mn^{[2]}}\underbrace{\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{j'=1}^{n}[\mathcal{I}(X_i>Y_{j'})\mathcal{I}(X_i=Y_j)]}_{\mathcal{E}(X=Y>Y')\Rightarrow S_5}\\ &\quad + \frac{1}{4mn^{[2]}}\underbrace{\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{j'=1}^{n}[\mathcal{I}(X_i=Y_j)\mathcal{I}(X_i=Y_{j'})]}_{\mathcal{E}(X=Y=Y')\Rightarrow S_6}\end{aligned}$$

$$- \frac{1}{mn^{[2]}} \underbrace{\sum_{i=1}^{m}\sum_{j=1}^{n}\left[\mathcal{I}(X_i > Y_j) + \frac{1}{4}\mathcal{I}(X_i = Y_j)\right]}_{S_2+\frac{1}{4}S_1} - \hat{\theta}^2$$

$$= \frac{2S_3 + S_4 + S_5 + \frac{1}{4}(S_6 - S_1) - S_2}{mn(n-1)} - \hat{\theta}^2$$

$$= \hat{\mathcal{Q}}_1$$

which confirms the result in Eq. (19). In a similar manner, we also have the result in Eq. (20). Hence the theorem follows. □

E. EFFICIENT COMPUTATIONS OF S_1 TO S_{10} BASED ON DYNAMIC PROGRAMMING

Let $\mathcal{Z}_1, \ldots, \mathcal{Z}_{m+n}$ be a combined sequence of $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$. Sorting this sequence in ascending order yields a new sequence of order statistics [28], [30]–[34]

$$\underbrace{\mathcal{Z}_{(1)} = \cdots = \mathcal{Z}_{(1)}}_{\text{Block}_1} < \cdots < \underbrace{\mathcal{Z}_{(J)} = \cdots = \mathcal{Z}_{(J)} (= \mathcal{Z}_i)}_{\text{Block}_J}$$
$$< \cdots < \underbrace{\mathcal{Z}_{(K)} = \cdots = \mathcal{Z}_{(K)}}_{\text{Block}_K}. \quad (22)$$

Suppose that the elements of Block_J are all equal to $\mathcal{Z}_i$. Let a_i (c_i) be the number of X's (Y's) equals to $\mathcal{Z}_i$, for $i = 1, \ldots, K$. Then we can obtain two count vectors, say, $C_X \triangleq [a_1 \ \ldots \ a_K]$ and $C_Y \triangleq [c_1 \ \ldots \ c_K]$, based on the $\mathcal{Z}_{(i)}$-sequence in (22), which, as shown in Fig. 1, can be obtained in a linearithmic time, i.e., $\mathcal{O}[(m+n)\log(m+n)]$, by using the popular sorting algorithms in the textbook [35]. Given C_X and C_Y, we can compute all S-terms in linear time $\mathcal{O}(K)$, where $K \le m+n$. Below we only elaborate the procedure of computing the first three S terms. The algorithms for the rest terms can be constructed in a similar and straightforward manner, and thus omitted for brevity. Now let us explain the computing structure by looking into the definitions of S_1 to S_3 one by one.

1) PROCEDURE FOR COMPUTING S_1

From Table 1, the definition of S_1 is

$$S_1 = \mathcal{E}(X = Y) = \sum_{i=1}^{m}\sum_{j=1}^{n}\mathcal{I}(X_i = Y_j)$$
$$= \sum_{k=1}^{K}\left[\sum_{i=1}^{m}\mathcal{I}(X_i = Z_k)\right]\left[\sum_{j=1}^{n}\mathcal{I}(Y_j = Z_k)\right] = \sum_{k=1}^{K} a_k c_k. \quad (23)$$

If we construct a count matrix $\mathbf{C}1$ by stacking the two count vectors C_X (Row_1) and C_Y (Row_2) together, S_1 can then be calculated with the procedure depicted in Fig. 2(a). Specifically, the update rule is

$$\mathbf{C}_{[1,k]} = \begin{cases} \mathbf{C}_{[1,k]} \cdot \mathbf{C}_{[2,k]} & k = 1; \\ \mathbf{C}_{[1,k]} \cdot \mathbf{C}_{[2,k]} + \mathbf{C}_{[1,k-1]} & 2 \le k \le K. \end{cases}$$

Algorithm 1: Procedure of Computing Count Vectors

Data: $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$
Result: count vectors C_X and C_Y

```
1  begin
2      Z ⟵ X_1, ..., X_m, Y_1, ..., Y_n
3      W ⟵ sorted version of Z in ascending order
4      L ⟵ labels of elements in W
5      W ⟵ {W, W_{m+n} + 1}
6      i ⟵ 1
7      k ⟵ 1
8      CX_k ⟵ 0
9      CY_k ⟵ 0
10     while i ≤ m + n do
11         s ⟵ i
12         j ⟵ s
13         while W_j = W_s do
14             if L_j = 'X' then
15                 CX_k ⟵ CX_k + 1
16             else
17                 CY_k ⟵ CY_k + 1
18             end
19             j ⟵ j + 1
20         end
21         i ⟵ j
22         k ⟵ k + 1
23         CX_k ⟵ 0
24         CY_k ⟵ 0
25     end
26     K ⟵ k − 1
27     C_X ⟵ CX_1, ..., CX_K
28     C_Y ⟵ CY_1, ..., CY_K
29 end
```

FIGURE 1. **Fast algorithm for computing the two count vectors C_X and C_Y. In Line 3, $\mathcal{W}$ is the ordered $\mathcal{Z}$-sequence in (22); whereas in Line 4, $\mathcal{L}$ contains the labels of $\mathcal{W}_i$, $i = 1, \ldots, m+n$, i.e., $\mathcal{L}_i$ = 'X' if $\mathcal{W}_i$ comes from X-class, and $\mathcal{L}_i$ = 'Y' if $\mathcal{W}_i$ comes from Y-class. Line 5 appends an extra element $\mathcal{W}_{m+n} + 1$ (= $\max(\mathcal{Z}) + 1$) to $\mathcal{W}$ in order to prevent overflow in Line 13. After Lines 10 to 25, we obtain two lists, $\text{CX}_1, \ldots, \text{CX}_k$ and $\text{CY}_1, \ldots, \text{CY}_k$. Finally, in Lines 26 to 28, the extra (last) elements (due to Line 5) in CX-list and CY-list are removed and the rest are stored in C_X and C_Y, respectively. It is noteworthy that the most time consuming procedure is the sorting operation in Line 3, which can be accomplished by any efficient sorting algorithms, such as the familiar *quicksort* and *merge sort* that are available in the textbook [35].**

As the index k running from 1 to K, the final desired result $\sum a_k c_k$ in Eq. (23) is stored in the cell of $\mathbf{C}1_{[1,K]}$.

2) PROCEDURE FOR COMPUTING S_2

From Table 1, the definition of S_2 is

$$S_2 = \mathcal{E}(X > Y) = \sum_{i=1}^{m}\sum_{j=1}^{n}\mathcal{I}(X_i > Y_j)$$
$$= \sum_{k=2}^{K}\left[\sum_{i=1}^{m}\mathcal{I}(X_i = Z_k)\right]\left[\sum_{j=1}^{n}\mathcal{I}(Y_j < Z_k)\right]$$
$$= \sum_{k=2}^{K}\sum_{l=1}^{k-1} a_k c_l \quad (24)$$

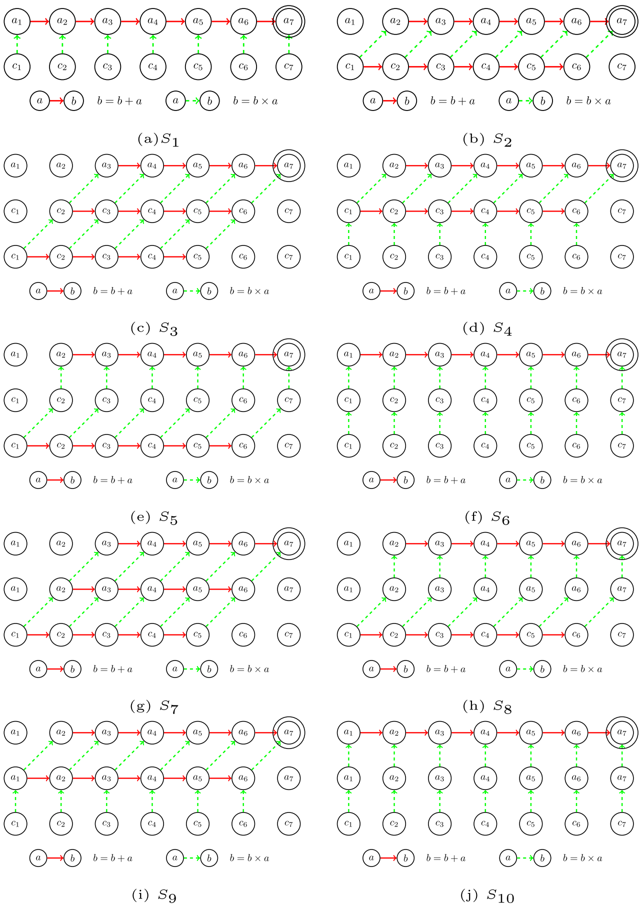

FIGURE 2. Diagrams for computing S_1 to S_{10} that listed in Table 1, where $K = 7$ is just for purpose of demonstration. (a) S_1. (b) S_2. (c) S_3. (d) S_4. (e) S_5. (f) S_6. (g) S_7. (h) S_8. (i) S_9. (j) S_{10}.

which can be computed, also based on the two count vectors C_X and C_Y, with the procedure shown in Fig. 2(b). Let $\mathbf{C}2$ be a $2 \times K$ matrix with first row being C_X and second row being C_Y. Then, with $\mathbf{C}2$, the update rule for S_2 is

$$\mathbf{C}_{[r,k]} = \begin{cases} 0 & r = 1, k = 1; \\ \mathbf{C}_{[r,k]} + \mathbf{C}_{[r,k-1]} & r = 2, 2 \le k \le K; \\ \mathbf{C}_{[r,k]} \cdot \mathbf{C}_{[r+1,k-1]} + \mathbf{C}_{[r,k-1]} & r = 1, 2 \le k \le K. \end{cases}$$

As k running from 2 to K, the final desired result (the rightmost term) in Eq. (24) is stored in the cell of $\mathbf{C}2_{[1,K]}$.

3) PROCEDURE FOR COMPUTING S_3 AND OTHERS

By its definition in Table 1, it follows that

$$S_3 = \mathcal{E}(X > Y > Y') = \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{l=1}^{n} \mathcal{I}(X_i > Y_j > Y_l) = \sum_{i=3}^{K}\sum_{j=2}^{i-1}\sum_{l=1}^{j-1} a_i c_j c_l. \tag{25}$$

To compute S_3, we first construct a $3 \times K$ count matrix $\mathbf{C}3$, with three rows being C_X, C_Y and C_Y, respectively. We further set $\mathbf{C}3_{[1,1]}$, $\mathbf{C}3_{[1,2]}$ and $\mathbf{C}3_{[2,1]}$ to be 0. Then, as shown in Fig. 2(c), the programming path goes from the southwest corner towards the northeast corner, with the corresponding update rule of

$$\mathbf{C}_{[r,k]} = \begin{cases} \mathbf{C}_{[r,k]} + \mathbf{C}_{[r,k-1]} & r = 3, 2 \le k \le K; \\ \mathbf{C}_{[r,k]} \cdot \mathbf{C}_{[r+1,k-1]} + \mathbf{C}_{[r,k-1]} & r = 2, 2 \le k \le K; \\ \mathbf{C}_{[r,k]} \cdot \mathbf{C}_{[r+1,k-1]} + \mathbf{C}_{[r,k-1]} & r = 1, 2 \le k \le K. \end{cases}$$

When k increases up to K, the desired value of S_3 is finally stored in the cell of $\mathbf{C}3_{[1,K]}$.

In a similar manner, it follows from Table 1 again that S_4–S_{10} can be expressed as

$$S_4 = \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{l=1}^{n} \mathcal{I}(X_i > Y_j = Y_l) = \sum_{i=2}^{K}\sum_{j=1}^{i-1} a_i c_j^2 \tag{26}$$

$$S_5 = \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{l=1}^{n} \mathcal{I}(X_i = Y_j > Y_l) = \sum_{i=2}^{K}\sum_{j=1}^{i-1} a_i c_i c_j \tag{27}$$

$$S_6 = \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{l=1}^{n} \mathcal{I}(X_i = Y_j = Y_l) = \sum_{i=1}^{K} a_i b_i^2 \tag{28}$$

$$S_7 = \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{l=1}^{n} \mathcal{I}(X_i > X_j > Y_l) = \sum_{i=3}^{K}\sum_{j=2}^{i-1}\sum_{l=1}^{j-1} a_i a_j c_l \tag{29}$$

$$S_8 = \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{l=1}^{n} \mathcal{I}(X_i = X_j > Y_l) = \sum_{i=2}^{K}\sum_{j=1}^{i-1} a_i^2 c_j \tag{30}$$

$$S_9 = \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{l=1}^{n} \mathcal{I}(X_i > X_j = Y_l) = \sum_{i=2}^{K}\sum_{j=1}^{i-1} a_i a_j c_j \tag{31}$$

$$S_{10} = \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{l=1}^{n} \mathcal{I}(X_i = X_j = Y_l) = \sum_{i=1}^{K} a_i^2 b_i \tag{32}$$

which can all be computed based on the diagrams illustrated in Fig. 2 (d)–(j), respectively.

III. NUMERICAL RESULTS

In this section, we compare our dynamic programming based algorithm (Theorem 2) with the rank-based state-of-the-art one [23]–[26], in terms of the unbiasedness as well as the computational efficiency, for estimating the variance of the MWUS defined in Eq. (1). Throughout this section, Monte Carlo experiments are undertaken for sample sizes within [5, 100], with the number of trials being 10^6 for reason of accuracy. All simulations are undertaken by functions in Matlab `Statistics Toolbox`™.

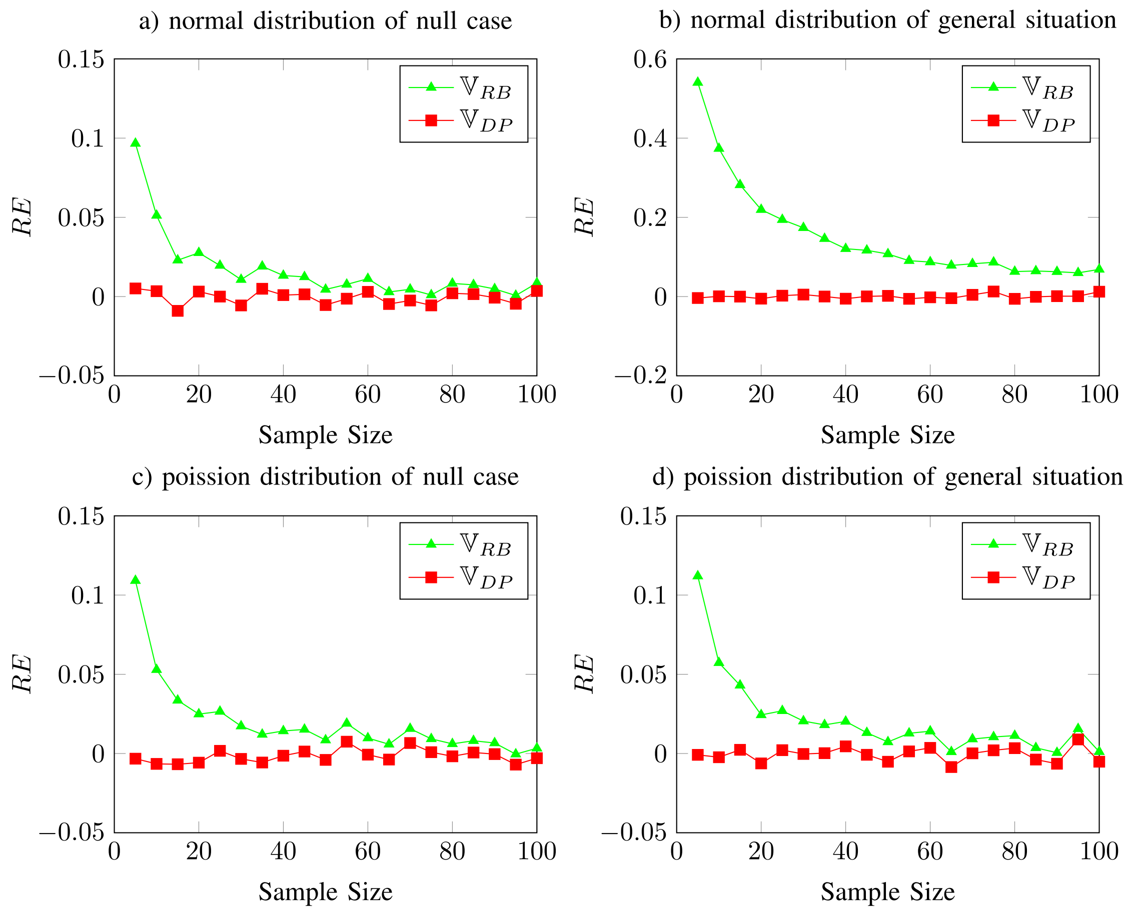

FIGURE 3. **Comparison of unbiasedness, in terms of RE, between the estimator in Eq. (16) and the rank-based method described in [23]–[26]. a) Null case under Normal distribution, where both X and Y follow $\mathcal{N}(0, 1)$. b) Non-null case under Normal distribution, where X follows $\mathcal{N}(0, 1)$ and Y follows $\mathcal{N}(4, 1)$. c) Null case under Poisson distribution, where both X and Y follow Poisson distributions with pmfs $\Pr(X = k) = \Pr(Y = k) = 2^k e^{-2}/k!$ for $k = 0, 1, 2, \ldots, \infty$. d) Non-null case under Poisson distribution, with pmfs of $\Pr(X = k) = 2^k e^{-2}/k!$ and $\Pr(Y = k) = 4^k e^{-4}/k!$ for $k = 0, 1, 2, \ldots, \infty$.**

A. VERICICATION OF UNBIASEDNESS

We first illustrate that, in contrast to the rank-based method [23]–[26], our method in (16) is an unbiased estimator of the variance of MWUS. In the following, we generate the two samples X and Y based on normal distributions (denoted by $\mathcal{N}(\mu, \sigma^2)$ with mean μ and variance σ^2) and Poisson distributions, respectively. Specifically, the following four scenarios are included in this study:

1) X and Y both follow the standard normal distribution $\mathcal{N}(0, 1)$,
2) X and Y follow $\mathcal{N}(0, 1)$ and $\mathcal{N}(4, 1)$, respectively,
3) X and Y both follow a Poisson distribution with pmfs $\Pr(X = k) = \Pr(Y = k) = 2^k e^{-2}/k!$ for $k = 0, 1, 2, \ldots, \infty$,
4) X follows a Poisson distribution with pmf $\Pr(X = k) = 2^k e^{-2}/k!$ and Y follows a Poisson distribution with pmf $\Pr(Y = k) = 4^k e^{-4}/k!$, where $k = 0, 1, 2, \ldots, \infty$.

Under the four scenarios mentioned above, we compare the two methods in terms of the *relative error* (RE), which is defined by

$$RE_\delta \triangleq \frac{\mathbb{E}(\hat{\mathbb{V}}_\delta - \mathbb{V}_E)}{\mathbb{V}_E}$$

where the suffix $\delta \in \{DP, RB\}$ indicates one of the two method mentioned above, $\hat{\mathbb{V}}_{DP}$ stands for our method based on dynamic programming in (16), $\hat{\mathbb{V}}_{RB}$ for the rank-based method [23]–[26], and $\mathbb{V}_E$ for the empirical variance calculated based on Monte Carlo experiments.

Fig. 3 shows the comparison results, in terms of RE, with respect to the two methods under the four scenarios. The top two panels are results corresponding to normal distributions (continuous case); whereas the bottom two panels are results corresponding to Poisson distributions (discrete case). It is clear that $\hat{\mathbb{V}}_{DP}$ outperfoms $\hat{\mathbb{V}}_{RB}$, in the sense that the former's REs are all approximately zero, which confirms the unbiasedness of our method. On the other hand, all the curves of RE_{RB} deviate from zero, especially when the sample sizes are small. However, with increase of sample size, RE_{RB} decreases gradually towards zero. In other words, $\hat{\mathbb{V}}_{RB}$ is only an asymptotically unbiased estimator of the variance of MWUS.

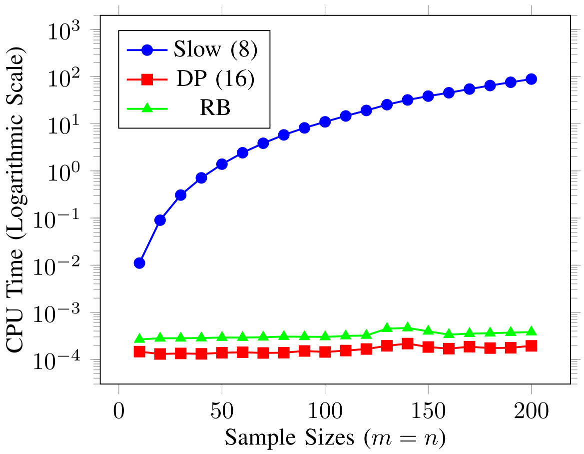

FIGURE 4. Comparative results of CPU time between the algorithms based on (8), (16) and the rank-based method. For simplicity, the sample sizes of X and Y are set to be equal. A log scale is used for better visual effect.

B. COMPARISON OF COMPUTATIONAL LOADS

To demonstrate the computational efficiency of our new algorithm, we generate two normal samples, each being i.i.d, i.e., $\{X_i\}_{i=1}^m \sim \mathcal{N}(0, 0.4)$ and $\{Y_j\}_{j=1}^n \sim \mathcal{N}(1, 0.6)$. Since the parameters have little if no effect on the computational speed comparison, they are chosen arbitrarily. Fig. 4 compares the computational loads between the algorithms based on Eq. (8), Eq. (16), and the rank-based approach [23]–[26] over the sample sizes $m = n = 10(10)200$, where 10(10)200 stands for a list starting from 10 to 200 with an increment of 10. Each of the algorithms is run for 100 times for stability. As shown in Fig. 4, it is observed that, although both in linearithmic time $\mathcal{O}[(m+n)\log(m+n)]$, our algorithm in (16) runs a little faster than the rank-based approach. And it is of no suprise that the version established in (8) is the slowest one, since its time complexity is in a cubic order.

IV. CONCLUSION

In this paper, we proposed an efficient algorithm for computing the variance of MWUS, an unbiased estimator of AUC, based on dynamic programming. Theoretical derivations suggest that (a) it can act as an unbiased estimator for the variance of MWUS, and (b) its time complexity is of linearithmic order, much lower than a conventional one and comparable with the state-of-the-art rank-based method. Besides these advantages, the structure of this algorithm can be easily extended to the three- or multi-class cases [36]. In addition, a current method for cell detection can be improved based on our algorithm [37]. The methodology established in this work might shed new light on the topic of ROC analysis, which is an indispensable tool in many scientific and engineering areas.

REFERENCES

[1] W. Peterson, T. Birdsall, and W. Fox, "The theory of signal detectability," *Trans. IRE Prof. Group Inf. Theory*, vol. 4, no. 4, pp. 171–212, Sep. 1954.
[2] D. V. Meter and D. Middleton, "Modern statistical approaches to reception in communication theory," *Trans. IRE Prof. Group Inf. Theory*, vol. 4, no. 4, pp. 119–145, Sep. 1954.
[3] A. Wald, "Statistical decision functions," *Ann. Math. Statist.*, vol. 20, no. 2, pp. 165–205, 1949.
[4] L. I. Kuncheva, "Change detection in streaming multivariate data using likelihood detectors," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 5, pp. 1175–1180, Feb. 2013.
[5] K. Bowyer, C. Kranenburg, and S. Dougherty, "Edge detector evaluation using empirical ROC curves," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1. Dec. 1999, p. 359
[6] Y. Yitzhaky and E. Peli, "A method for objective edge detection evaluation and detector parameter selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 8, pp. 1027–1033, Aug. 2003.
[7] A. Oukaci, J.-P. Delmas, and P. Chevalier, "Performance analysis of LRT/GLRT-based array receivers for the detection of a known real-valued signal corrupted by noncircular interferences," *Signal Process.*, vol. 91, no. 10, pp. 2323–2331, 2011.
[8] G. Guo, M. Mandal, and Y. Jing, "A robust detector of known signal in non-Gaussian noise using threshold systems," *Signal Process.*, vol. 92, no. 11, pp. 2676–2688, 2012.
[9] V. Hari, G. Anand, A. B. Premkumar, and A. Madhukumar, "Design and performance analysis of a signal detector based on suprathreshold stochastic resonance," *Signal Process.*, vol. 92, no. 7, pp. 1745–1757, 2012.
[10] W. A. Yousef, R. F. Wagner, and M. H. Loew, "Assessing classifiers from two independent data sets using ROC analysis: A nonparametric approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1809–1817, Nov. 2006.
[11] K. R. Varshney, R. J. Prenger, T. L. Marlatt, B. Y. Chen, and W. G. Hanley, "Practical ensemble classification error bounds for different operating points," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 11, pp. 2590–2601, Nov. 2013.
[12] S. G. Baker, "The central role of receiver operating characteristic (ROC) curves in evaluating tests for the early detection of cancer," *J. Nat. Cancer Inst.*, vol. 95, no. 7, pp. 511–515, 2003.
[13] K. H. Zou, A. J. O'Malley, and L. Mauri, "Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models," *Circulation*, vol. 115, no. 5, pp. 654–657, 2007.
[14] R. Kumar and A. Indrayan, "Receiver operating characteristic (ROC) curve for medical researchers," *Indian Pediatrics*, vol. 48, no. 4, pp. 277–287, 2011.
[15] S. D. Gronlund, J. T. Wixted, and L. Mickes, "Evaluating eyewitness identification procedures using receiver operating characteristic analysis," *Current Directions Psychol. Sci.*, vol. 23, no. 1, pp. 3–10, 2014.
[16] T. A. Lasko, J. G. Bhagwat, K. H. Zou, and L. Ohno-Machado, "The use of receiver operating characteristic curves in biomedical informatics," *J. Biomed. Informat.*, vol. 38, no. 5, pp. 404–415, Oct. 2005.
[17] J. D. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference*. Berlin, Germany: Springer, 2011.
[18] M. A. Cleves *et al.*, "Comparative assessment of three common algorithms for estimating the variance of the area under the nonparametric receiver operating characteristic curve," *Stata J.*, vol. 3, no. 3, pp. 280–289, 2002.
[19] F. E. Harrell, K. L. Lee, R. M. Califf, D. B. Pryor, and R. A. Rosati, "Regression modelling strategies for improved prognostic prediction," *Statist. Med.*, vol. 3, no. 2, pp. 143–152, 1984.
[20] D. Bamber, "The area above the ordinal dominance graph and the area below the receiver operating characteristic graph," *J. Math. Psychol.*, vol. 12, no. 4, pp. 387–415, 1975.
[21] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach," *Biometrics*, vol. 44, no. 3, pp. 837–845, 1988.
[22] J. A. Hanley *et al*, "Receiver operating characteristic (ROC) methodology: The state of the art," *Critical Rev. Diagnostic Imag.*, vol. 29, no. 3, pp. 307–335, 1989.
[23] J. Kaufmann, C. Werner, and E. Brunner, "Nonparametric methods for analysing the accuracy of diagnostic tests with multiple readers," *Statist. Methods Med. Res.*, vol. 14, no. 2, pp. 129–146, 2005.
[24] K. Neubert and E. Brunner, "A studentized permutation test for the non-parametric Behrens–Fisher problem," *Comput. Statist. Data Anal.*, vol. 51, no. 10, pp. 5192–5204, 2007.
[25] F. Konietschke and M. Pauly, "A studentized permutation test for the non-parametric Behrens–Fisher problem in paired data," *Electron. J. Statist.*, vol. 6, pp. 1358–1372, Jan. 2012.
[26] M. Pauly, T. Asendorf, and F. Konietschke, "Permutation tests and confidence intervals for the area under the roc-curve," ulm univ., Germany, Tech. Rep., 2014.

[27] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.
[28] W. Xu, J. Dai, Y. Hung, and Q. Wang, "Estimating the area under a receiver operating characteristic (ROC) curve: Parametric and nonparametric ways," *Signal Process.*, vol. 93, no. 11, pp. 3111–3123, 2013.
[29] G. E. Noether, *Elements of Nonparametric Statistics*. New York, NY, USA: Wiley, 1967.
[30] W. Xu, C. Chang, Y. Hung, S. Kwan, and P. C. W. Fung, "Order statistic correlation coefficient and its application to association measurement of biosignals," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, vol. 2. May 2006, p. 2.
[31] W. Xu, C. Chang, Y. S. Hung, S. K. Kwan, and P. C. W. Fung, "Order statistics correlation coefficient as a novel association measurement with applications to biosignal analysis," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5552–5563, Dec. 2007.
[32] W. Xu, Y. Hung, M. Niranjan, and M. Shen, "Asymptotic mean and variance of gini correlation for bivariate normal samples," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 522–534, Jun. 2010.
[33] M. Kendall, J. D. Gibbons, *Rank Correlation Methods*, 5th ed. New York, NY, USA: Oxford Univ. Press, 1990.
[34] E. L. Lehmann and H. J. D'Abrera, *Nonparametrics: Statistical Methods Based on Ranks*. NewYork, NY, USA: Springer, 2006.
[35] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms*, 2rd ed. Cambridge, MA, USA: MIT Press, 2001.
[36] C. T. Nakas and C. T. Yiannoutsos, "Ordered multiple-class ROC analysis with continuous measurements," *Statist. Med.*, vol. 23, no. 22, pp. 3437–3449, 2004.
[37] G. E. D'Errico, "Adaptive cell detection by application of ROC methodology," in *Proc. 21st IMEKO World Congr. Meas. Res. Ind.*, Prague, Czech Republic, Aug. 2015, pp. 1–5.

WEICHAO XU (M'06) received the B.Eng. and M.Eng. degrees in electrical engineering from the University of Science and Technology of China, Hefei, in 1993 and 1996, respectively, and the Ph.D. degree in biomedical engineering from The University of Hong Kong, Hong Kong, in 2002. He was a Research Associate with the Department of Electrical and Electronic Engineering, The University of Hong Kong, from 2002 to 2010. In 2011, he entered the 100-Talent Scheme in the Guangdong University of Technology, where he is currently a Professor. His research interests are in the areas of mathematical statistics, computational statistics, information theory, machine learning, optimization theory, and digital signal processing and applications.

SHUN LIU received the B.Eng. degree in control theory and engineering from the Department of Automatic Control, School of Automation, Guangdong University of Technology, Guangzhou, China, in 2015, where he is currently pursuing the master's degree in control theory and engineering. His research interests include statistical signal processing, pattern recognition, and machine learning.

XU SUN received the B.Eng. and M.Eng. degrees in control theory and control engineering from the Department of Automatic Control, School of Automation, Guangdong University of Technology, Guangzhou, China, in 2013 and 2016, respectively. He is currently a Research Assistant with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong. His research interests include statistical signal processing, algorithm design, statistical learning theory, and intelligent systems.

SIYANG LIU received the B.Eng. degree in control theory and engineering from the Department of Automatic Control, School of Automation, Guangdong University of Technology, Guangzhou, China, in 2016, where she is currently pursuing the M.Eng. degree in control theory and engineering. Her research interests include statistical signal processing, pattern recognition, and machine learning.

YUN ZHANG received the B.S. and M.S. degrees in automatic engineering from Hunan University, Changsha, China, in 1982 and 1986, respectively, and the Ph.D. degree in automatic engineering from the South China University of Science and Technology, Guangzhou, China, in 1998. He is currently a Professor with the School of Automation, Guangdong University of Technology, Guangzhou. His research interests include intelligent control systems, network systems, and signal processing.