

Fast Implementation of DeLong's Algorithm for Comparing the Areas Under Correlated Receiver Operating Characteristic Curves

Xu Sun, Weichao Xu*, *Member, IEEE*

Abstract

Among algorithms for comparing the areas under two or more correlated receiver operating characteristic (ROC) curves, DeLong's algorithm is perhaps the most widely used one due to its simplicity of implementation in practice. Unfortunately, however, the time complexity of DeLong's algorithm is of quadratic order (the product of sample sizes), thus making it time-consuming and impractical when the sample sizes are large. Based on an equivalent relationship between the Heaviside function and mid-ranks of samples, we improve DeLong's algorithm by reducing the order of time complexity from quadratic down to linearithmic (the product of sample size and its logarithm). Monte Carlo simulations verify the computational efficiency of our algorithmic findings in this work.

Index Terms

Area under the curve (AUC), DeLong's method, mid-rank, receiver operating characteristic (ROC).

Copyright (c) 2012 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

This work was supported in part by National Natural Science Foundation of China (Project 61271380), in part by Guangdong Natural Science Foundation (Project S2012010009870), in part by 100-Talents Scheme Funding from Guangdong University of Technology (Grant 112418006), in part by the Talent Introduction Special Funds from Guangdong Province (Grant 2050205), in part by a team project from Guangdong University of Technology (Grant GDUT2011-07), and in part by Project Program of Key Laboratory of Guangdong Higher Education Institutes of China (Grant 2013CXZDA015).

X. Sun and W. Xu are with the Department of Automatic Control, School of Automation, Guangdong University of Technology, Guangzhou, 510006, P. R. China (e-mail: xsun@mail2.gdut.edu.cn; wcxu@gdut.edu.cn).

*Corresponding Author.

Tel:+86-020-39322552 Fax:+86-020-39322469

EDICS:MLSAS-PATT; SAS-MALN; SAS-STAT

I. INTRODUCTION

Originated from detection theory developed during World War II [1]–[4], receiver operating characteristic (ROC) analysis has found a wide use in a number of fields, including medicine, psychology, bioinformatics, signal processing, and machine learning, just to name a few [5]–[9]. Geometrically, ROC curve is a two-dimensional curve traced out by pairs of false-positive rate and true-positive rate according to various decision threshold settings. Given the ROC curve, the area under the curve (AUC) can then be computed, either analytically or empirically, as a figure of merit to summarize a diagnostic system's performance [5], a binary classifier's overall accuracy [6], or a detector's power of detecting the presence of an unknown signal [7]–[9].

As revealed by Bamber [10], AUC can be estimated by the Mann-Whitney U statistic (MWUS) [11]. Based on such relationship, a plenty of nonparametric methods have been proposed in the literature [12]–[16], which formulate algorithms directly from the samples without making any parametric model assumptions on the forms of the parent populations. For samples drawn from continuous distributions, algorithms of linearithmic time complexity have been proposed by Xu *et al.* for estimating the mean and variance of AUC [17]. Unfortunately, however, sometimes in practice, the samples obey non-continuous distributions, that is, the probability of ties between samples are not zero. Under this circumstance, sub-quadratic estimators for the variance of and covariance between AUCs are still unavailable to the best of our knowledge. This lack of efficient algorithms makes AUC comparison computationally very expensive in scenarios (e.g. bioinformatics) involving massive data analysis.

Motivated by this unsatisfactory situation, in this work we improve the popular DeLong's algorithm [14] by reducing the time complexity from quadratic down to linearithmic order. This is accomplished through a relationship we find between the Heaviside function and the mid-ranks of samples.

The rest part of this paper is organized as follows. Section II gives the basic definition of AUC as well as some general notations employed throughout this work. Section III depicts our linearithmic algorithm after establishing a close relationship between the Heaviside function and mid-ranks associated with samples. In Section IV, we demonstrate the efficiency of our improved algorithm in terms of time-complexity by Monte Carlo experiments. Finally, we summarize our main finding and draw our conclusion in Section V.

II. DEFINITIONS

For completeness and ease of later development, this section describes the definition of nonparametric estimator of AUC as well as DeLong's formulas for estimating the variance of and covariance between correlated AUCs.

Let X_1, \dots, X_m and Y_1, \dots, Y_n be two independent and identically distributed (i.i.d) samples drawn from two populations (whose distributions can be either continuous or discrete). Then, based on the relationship between the MWUS and AUC, the sample version of AUC can be defined as

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

where

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

is the familiar Heaviside function.

Let $\mathbb{E}(\cdot)$, $\mathbb{V}(\cdot)$ and $\mathbb{C}(\cdot, \cdot)$ denote the mean, variance and covariance of (between) random variables, respectively. It is easily seen from (1) that $\hat{\theta}$ is an unbiased estimator of the corresponding population version θ , since

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

Let

$$\hat{\boldsymbol{\theta}} \triangleq \{\hat{\theta}^{(1)}, \dots, \hat{\theta}^{(k)}\} \quad (4)$$

be a vector of statistics representing the areas under the ROC curves derived from different readings $X_1^{(r)}, \dots, X_m^{(r)}$ and $Y_1^{(r)}, \dots, Y_n^{(r)}$ ($1 \leq r \leq k$) of k different experiments.

For the r th element of the vector, define the “structural components”

$$V_{10}(X_i^{(r)}) = \frac{1}{n} \sum_{j=1}^n \mathcal{H}(X_i^{(r)} - Y_j^{(r)}), i = 1, \dots, m \quad (5)$$

and

$$V_{01}(Y_j^{(r)}) = \frac{1}{m} \sum_{i=1}^m \mathcal{H}(X_i^{(r)} - Y_j^{(r)}), j = 1, \dots, n. \quad (6)$$

Also define two matrices $\mathbf{S}_{10} \triangleq [s_{10}^{(r,s)}]_{k \times k}$ and $\mathbf{S}_{01} \triangleq [s_{01}^{(r,s)}]_{k \times k}$ such that

$$s_{10}^{(r,s)} = \frac{1}{m-1} \sum_{i=1}^m [V_{10}(X_i^{(r)}) - \hat{\theta}^{(r)}] [V_{10}(X_i^{(s)}) - \hat{\theta}^{(s)}] \quad (7)$$

and

$$s_{01}^{(r,s)} = \frac{1}{n-1} \sum_{j=1}^n [V_{01}(Y_j^{(r)}) - \hat{\theta}^{(r)}] [V_{01}(Y_j^{(s)}) - \hat{\theta}^{(s)}]. \quad (8)$$

Then, DeLong *et al.* [14] proposed a variance-covariance matrix estimator for the vector $\hat{\boldsymbol{\theta}}$ in (4), as

$$\mathbf{S} = \frac{1}{m} \mathbf{S}_{10} + \frac{1}{n} \mathbf{S}_{01}. \quad (9)$$

When the vector $\hat{\theta}$ contains only one element, that is, $r = s = 1$ in (7) and (8), the covariance estimator in (9) reduces to a variance estimator $\mathbb{V}(\hat{\theta})$.

III. LINEARITHMIC ALGORITHMS

Possessing the time complexities of orders $\mathcal{O}(mn)$ and $\mathcal{O}[kmn + k^2(m+n)]$, respectively, the algorithm of $\hat{\theta}$ in (1) and the algorithm of \mathbf{S} in (9) are both computationally inefficient, especially when the sample sizes m and n are large. However, by the relationship of the Heaviside function and mid-ranks shown in Lemma 1 below, linearithmic algorithms can be formulated based on DeLong's formulas (5)–(9).

A. Relationship between mid-ranks and $\mathcal{H}(\cdot)$

Let $\mathcal{Z}_1, \dots, \mathcal{Z}_M$ be a sequence of real numbers. Sorting the sequence in ascending order yields a new sequence, termed the order statistics [18]–[23], as

$$\underbrace{\mathcal{Z}_{(1)} = \dots = \mathcal{Z}_{(1)}}_{\text{Block}_1} < \dots < \underbrace{\mathcal{Z}_{(J)} = \dots = \mathcal{Z}_{(J)} (= \mathcal{Z}_i)}_{\text{Block}_J} < \dots < \underbrace{\mathcal{Z}_{(K)} = \dots = \mathcal{Z}_{(K)}}_{\text{Block}_K}. \quad (10)$$

Suppose that Block_J , whose elements are all equal to \mathcal{Z}_i , starts at position a and ends at position b in the sorted sequence (10). Then the mid-ranks of \mathcal{Z}_i 's in the original sequence is defined as [24]

$$\mathbb{T}_{\mathcal{Z}}(\mathcal{Z}_i) = \frac{1}{b - a + 1} \sum_{k=a}^b k = \frac{a + b}{2}. \quad (11)$$

Given (11), it then follows that mid-ranks are closed related to the Heaviside function $\mathcal{H}(\cdot)$, as shown in the following lemma.

Lemma 1: The mid-ranks of \mathcal{Z}_i 's in $\mathcal{Z}_1, \dots, \mathcal{Z}_M$ can be computed by

$$\mathbb{T}_{\mathcal{Z}}(\mathcal{Z}_i) = \sum_{j=1}^M \mathcal{H}(\mathcal{Z}_i - \mathcal{Z}_j) + \frac{1}{2}. \quad (12)$$

Proof: It suffices to show that the right side of (12) equals the rightmost term in (11), as

$$\sum_{j=1}^M \mathcal{H}(\mathcal{Z}_i - \mathcal{Z}_j) + \frac{1}{2} = \sum_{k=1}^{a-1} 1 + \frac{1}{2} \sum_{k=a}^b 1 + \frac{1}{2} = \frac{a + b}{2}. \quad (13)$$

■

Note that from (10) and (11), $\mathbb{T}_{\mathcal{Z}}(\mathcal{Z}_i)$, $i = 1, \dots, M$, can be obtained in linearithmic time, i.e., $\mathcal{O}(M \log M)$, by using the popular quick sort algorithm [25] (See Fig. 1). The relationship of (12), while useful for developing efficient algorithms later on, is not employed to calculate $\mathbb{T}_{\mathcal{Z}}(\mathcal{Z}_i)$, $i = 1, \dots, M$, due to its quadratic time complexity.

Algorithm 1: Procedure of Calculating Mid-ranks

Data: a sequence $\mathcal{Z}_1, \dots, \mathcal{Z}_M$
Result: the associated mid-ranks $\mathbb{T}_{\mathcal{Z}}(\mathcal{Z}_1), \dots, \mathbb{T}_{\mathcal{Z}}(\mathcal{Z}_M)$

```

1 begin
2    $M \leftarrow$  length of  $\mathcal{Z}$ 
3    $\mathcal{W} \leftarrow$  sorted version of  $\mathcal{Z}$  in ascending order
4    $\mathcal{I} \leftarrow$  position indices of  $\mathcal{Z}$  in  $\mathcal{W}$ 
5    $\mathcal{W} \leftarrow \{\mathcal{W}, \mathcal{W}_M + 1\}$ 
6    $\mathbb{T} \leftarrow$  a list of  $M$  zeros
7    $i \leftarrow 1$ 
8   while  $i \leq M$  do
9      $a \leftarrow i$ 
10     $j \leftarrow a$ 
11    while  $\mathcal{W}_j = \mathcal{W}_a$  do
12       $j \leftarrow j + 1$ 
13    end
14     $b \leftarrow j - 1$ 
15    for  $k = a$  to  $b$  do
16       $\mathbb{T}_k = (a + b)/2$ 
17    end
18     $i \leftarrow b + 1$ 
19  end
20  for  $i = 1$  to  $M$  do
21     $k \leftarrow \mathcal{I}_i$ 
22     $\mathbb{T}_{\mathcal{Z}}(\mathcal{Z}_k) \leftarrow \mathbb{T}_i$ 
23  end
24 end

```

Fig. 1. Fast algorithm for calculating the mid-ranks of a sequence. Note that in Line 3, \mathcal{W} is the ordered \mathcal{Z} -sequence in (10); whereas in Line 4, \mathcal{I} contains the position indices of \mathcal{Z}_i , $i = 1, \dots, M$, in \mathcal{W} , i.e., $\mathcal{I}_i = k$ if $\mathcal{W}_k = \mathcal{Z}_i$. Line 5 is to append an extra element $\mathcal{W}_M + 1$ ($= \max(\mathcal{Z}) + 1$) to \mathcal{W} in order to prevent overflow in Line 12. After Lines 8 to 19, the list \mathbb{T} contains the mid-ranks of \mathcal{W} already. At last, the rest lines complete the rearrangement of the mid-ranks with respect to the original sequence $\mathcal{Z}_1, \dots, \mathcal{Z}_M$.

B. Fast algorithm for computing $\hat{\theta}$

Lemma 2: Denote by Z_1, \dots, Z_N , $N = m + n$, the concatenated sequence of X_1, \dots, X_m and Y_1, \dots, Y_n . Then

$$\sum_{j=1}^n \mathcal{H}(X_i - Y_j) = \mathbb{T}_Z(X_i) - \mathbb{T}_X(X_i) \quad (14)$$

$$\sum_{i=1}^m \mathcal{H}(Y_j - X_i) = \mathbb{T}_Z(Y_j) - \mathbb{T}_Y(Y_j) \quad (15)$$

Proof: It follows from (12) along with the definition of the Z -sequence that

$$\begin{aligned}\mathbb{T}_Z(X_i) &= \sum_{j=1}^N \mathcal{H}(X_i - Z_j) + \frac{1}{2} \\ &= \underbrace{\sum_{j=1}^m \mathcal{H}(X_i - X_j)}_{\mathbb{T}_X(X_i)} + \frac{1}{2} + \sum_{j=1}^n \mathcal{H}(X_i - Y_j) \\ \mathbb{T}_Z(Y_j) &= \sum_{i=1}^N \mathcal{H}(Y_j - Z_i) + \frac{1}{2} \\ &= \underbrace{\sum_{i=1}^n \mathcal{H}(Y_j - Y_i)}_{\mathbb{T}_Y(Y_j)} + \frac{1}{2} + \sum_{i=1}^m \mathcal{H}(Y_j - X_i)\end{aligned}$$

which leads to (14) and (15), respectively. ■

With Lemma 2, it follows readily that $\hat{\theta}$ can be computed by a linerithmic algorithm, as

$$\begin{aligned}\hat{\theta} &= \frac{1}{mn} \sum_{i=1}^m [\mathbb{T}_Z(X_i) - \mathbb{T}_X(X_i)] \\ &= \frac{1}{mn} \sum_{i=1}^m \mathbb{T}_Z(X_i) - \frac{m+1}{2n}.\end{aligned}\tag{16}$$

Remark 1: Note that (16) has been mentioned in passing in Lehmann's book [26]. We present it here just for completeness as well as ease of reference in the later part of this paper.

C. Fast Implementation of DeLong's Algorithm

Theorem 1: The two terms (5) and (6) in DeLong's algorithm (9) are numerically equivalent with

$$V_{10}(X_i^{(r)}) = \frac{\mathbb{T}_{Z^{(r)}}(X_i^{(r)}) - \mathbb{T}_{X^{(r)}}(X_i^{(r)})}{n}, i = 1, \dots, m\tag{17}$$

and

$$V_{01}(Y_j^{(r)}) = 1 - \frac{\mathbb{T}_{Z^{(r)}}(Y_j^{(r)}) - \mathbb{T}_{Y^{(r)}}(Y_j^{(r)})}{m}, j = 1, \dots, n\tag{18}$$

respectively.

Proof: Using the relationship $\mathcal{H}(X_i - Y_j) = 1 - \mathcal{H}(Y_j - X_i)$ and (15), we have

$$\sum_{i=1}^m \mathcal{H}(X_i - Y_j) = m - [\mathbb{T}_Z(Y_j) - \mathbb{T}_Y(Y_j)].\tag{19}$$

Then the results follow directly by substituting (14) and (19) into (5) and (6). ■

Remark 2: Although Theorem 1 is established in the context of AUC comparison, other methods using similar statistics as $\hat{\theta}$ can also gain benefit from this algorithmic finding, e.g., the localization ROC analysis [27].

Fig. 2 summarizes the fast implementation of DeLong's algorithm based on Theorem 1. Specifically, Lines 4 to 20 purport to obtain θ , V_{10} - and V_{01} -terms based on (16), (17) and (18), respectively; whereas Lines 21 to 33 are to obtain the matrix \mathbf{S} based on (7)–(9). It is easily seen that the improved algorithm is dominated by the mid-ranks algorithm in Fig. 1, which can be accomplished by the popular quick sort algorithm. Therefore, the overall time complexity of our algorithm is of order $\mathcal{O}[k(m+n)\log(m+n) + k^2(m+n)]$, much lower than the original order of $\mathcal{O}[kmn + k^2(m+n)]$ when m and n are large.

IV. NUMERICAL RESULTS

This section demonstrates the computational efficiency of the mid-ranks-based algorithm established in Theorems 1, by a Monte Carlo experiment for $k = 2$, a scenario which is not only the most fundamental, but also the motivating example for the research topic of this letter [13]. All samples are generated by functions in MATLAB Statistics Toolbox™.

Consider the following hypothetical scenario. We have a two-class p -dimensional data set $\{\mathbf{X}_1, \dots, \mathbf{X}_m, \mathbf{Y}_1, \dots, \mathbf{Y}_n\}$ based on which two linear classifiers are designed, that is, we have trained two $1 \times p$ row vectors \mathbf{W}_1 (by Fisher's LDA, say) and \mathbf{W}_2 (by SVM, say) which will project the original p -dimensional dataset onto two one-dimensional spaces, respectively. Write $X_i^{(1)} \triangleq \mathbf{W}_1 \mathbf{X}_i$, $Y_j^{(1)} \triangleq \mathbf{W}_1 \mathbf{Y}_j$, $X_i^{(2)} \triangleq \mathbf{W}_2 \mathbf{X}_i$, $Y_j^{(2)} \triangleq \mathbf{W}_2 \mathbf{Y}_j$, $i = 1 \dots, m$ and $j = 1 \dots, n$. Based on these two one-dimensional data set, two estimators of the corresponding AUCs, $\hat{\theta}^{(1)}$ and $\hat{\theta}^{(2)}$, can be computed from (16). The question is, which classifier is better, or, in terms of AUC, which one of $\hat{\theta}^{(1)}$ and $\hat{\theta}^{(2)}$ is greater. To answer this question, it is natural to investigate the magnitude of the statistic [13]

$$z \triangleq \frac{\hat{\theta}^{(1)} - \hat{\theta}^{(2)}}{\sqrt{\mathbb{V}[\hat{\theta}^{(1)} - \hat{\theta}^{(2)}]}} = \frac{\hat{\theta}^{(1)} - \hat{\theta}^{(2)}}{\sqrt{\mathbb{V}[\hat{\theta}^{(1)}] + \mathbb{V}[\hat{\theta}^{(2)}] - 2\mathbb{C}[\hat{\theta}^{(1)}, \hat{\theta}^{(2)}]}}$$

where the denominator involves all terms in the matrix \mathbf{S} defined by (9). Under the null hypothesis, z can be well approximated by the standard normal distribution [13]. Therefore, if the value of z deviates too much from zero, e.g., $z > 1.96$, it is thus reasonable to consider that $\hat{\theta}^{(1)} > \hat{\theta}^{(2)}$ with the significance level $p < 0.05$. When the sample sizes m and n are large, and/or one employs the resampling technique to test the statistical significance of the z -statistic, it is desirable to use a faster algorithm, such as the one proposed in this work, other than the original version of DeLong's algorithm.

Algorithm 2: Improved DeLong's Algorithm

Data: $X^{(r)} \triangleq \{X_i^{(r)}\}_{i=1}^m$, $Y^{(r)} \triangleq \{Y_j^{(r)}\}_{j=1}^n$, $1 \leq r \leq k$
Result: $S_{k \times k}$ in (9) and $\theta_{1 \times k}$ in (4)

```

1 begin
2    $m \leftarrow \text{length of } X_1^{(1)}, \dots, X_m^{(1)}$ 
3    $n \leftarrow \text{length of } Y_1^{(1)}, \dots, Y_n^{(1)}$ 
4   for  $r = 1$  to  $k$  do
5      $\theta^{(r)} \leftarrow 0$ 
6      $Z^{(r)} \leftarrow \text{concatenation of } X^{(r)} \text{ and } Y^{(r)}$ 
7      $\mathbb{T}_{Z^{(r)}} \leftarrow \text{mid-ranks of } Z^{(r)} \text{ by Algorithm 1}$ 
8      $\mathbb{T}_{X^{(r)}} \leftarrow \text{mid-ranks of } X^{(r)} \text{ by Algorithm 1}$ 
9      $\mathbb{T}_{Y^{(r)}} \leftarrow \text{mid-ranks of } Y^{(r)} \text{ by Algorithm 1}$ 
10    for  $i = 1$  to  $m$  do
11       $\mathbb{T}_{Z^{(r)}}(X_i^{(r)}) \leftarrow \mathbb{T}_{Z^{(r)}}(Z_i^{(r)})$ 
12       $V_{10}^{(r)}(X_i) \leftarrow [\mathbb{T}_{Z^{(r)}}(X_i^{(r)}) - \mathbb{T}_{X^{(r)}}(X_i^{(r)})]/n$ 
13       $\theta^{(r)} \leftarrow \theta^{(r)} + \mathbb{T}_{Z^{(r)}}(X_i^{(r)})$ 
14    end
15     $\theta^{(r)} \leftarrow \theta^{(r)} / (m \times n) - (m + 1) / (2 \times n)$ 
16    for  $j = 1$  to  $n$  do
17       $\mathbb{T}_{Z^{(r)}}(Y_j^{(r)}) \leftarrow \mathbb{T}_{Z^{(r)}}(Z_{m+j}^{(r)})$ 
18       $V_{01}^{(r)}(Y_j) \leftarrow$ 
19         $1 - [\mathbb{T}_{Z^{(r)}}(Y_j^{(r)}) - \mathbb{T}_{Y^{(r)}}(Y_j^{(r)})]/m$ 
20    end
21  end
22  for  $r = 1$  to  $k$  do
23    for  $s = 1$  to  $k$  do
24       $s_{10}^{(r,s)} \leftarrow 0$ 
25       $s_{01}^{(r,s)} \leftarrow 0$ 
26      for  $i = 1$  to  $m$  do
27         $s_{10}^{(r,s)} \leftarrow s_{10}^{(r,s)} + [V_{10}(X_i^{(r)}) - \theta^{(r)}] \times$ 
28           $[V_{10}(X_i^{(s)}) - \theta^{(s)}]$ 
29      end
30      for  $j = 1$  to  $n$  do
31         $s_{01}^{(r,s)} \leftarrow s_{10}^{(r,s)} + [V_{01}(Y_j^{(r)}) - \theta^{(r)}] \times$ 
32           $[V_{01}(Y_j^{(s)}) - \theta^{(s)}]$ 
33      end
34       $S_{r,s} \leftarrow s_{10}^{(r,s)} / (m - 1) + s_{01}^{(r,s)} / (n - 1)$ 
35    end
36  end

```

Fig. 2. Fast implementation of DeLong's algorithm based on Theorem 1.

To simulate the scenario mentioned above, we generate two one-dimensional normal samples, representing those projected from the high-dimensional space. Specifically, $\{X_i^{(1)}\}_{i=1}^m$ and $\{X_i^{(2)}\}_{i=1}^m$ are i.i.d. samples follow a bivariate normal distribution $\mathcal{N}(1, 2, 1, 1, 0.8)$; whereas $\{Y_j^{(1)}\}_{j=1}^n$ and $\{Y_j^{(2)}\}_{j=1}^n$ are i.i.d. follow a bivariate normal distribution $\mathcal{N}(2, 3, 1, 1, 0.8)$. The notation $\mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ stands for a bivariate norm distribution with means μ_1 and μ_2 , variances σ_1^2 and σ_2^2 , and correlation ρ . Note that the parameters are chosen rather arbitrarily, since they have little if not no effect on the computational speed comparison.

Fig. 3 contrasts the computational speeds over $m = n = 10(10)200$ between the original quadratic versions and the newly developed linearithmic versions for computing \mathcal{S} , where m and n denote the sample sizes of X - and Y -class, respectively. The notation 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. 3, the linearithmic algorithm proposed in this work does reduce the time complexity significantly.

V. SUMMARY

In this paper we investigated the problems of comparing the areas under two or more correlated receiver operating characteristic curves from nonparametric viewpoints. Our contribution is an improvement of the popular DeLong's algorithm in the sense of reducing the time complexity from the original quadratic order down to the present linearithmic order, based on an equivalent relationship we find between the Heaviside function and mid-ranks of samples. These algorithmic findings might shed new light on the topic of ROC analysis involving many scientific areas including signal processing.

REFERENCES

- [1] W. Peterson, T. Birdsall, and W. Fox, "The theory of signal detectability," *IRE Professional Group on Information Theory*, vol. 4, no. 4, pp. 171–212, 1954.
- [2] D. Van Meter and D. Middleton, "Modern statistical approaches to reception in communication theory," *IRE Professional Group on Information Theory*, vol. 4, no. 4, pp. 119–145, 1954.
- [3] A. Wald, "Statistical decision functions," *The Annals of Mathematical Statistics*, pp. 165–205, 1949.
- [4] J. A. Hanley, "Receiver operating characteristic (ROC) methodology: the state of the art," *Critical Reviews in Diagnostic Imaging*, vol. 29, no. 3, p. 307, 1989.
- [5] D. Mossman, "Three-way ROCs," *Medical Decision Making*, vol. 19, no. 1, pp. 78–89, 1999.
- [6] W. Yousef, R. Wagner, and M. Loew, "Assessing classifiers from two independent data sets using ROC analysis: A nonparametric approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1809–1817, nov. 2006.
- [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 Processing*, vol. 91, no. 10, pp. 2323–2331, 2011.

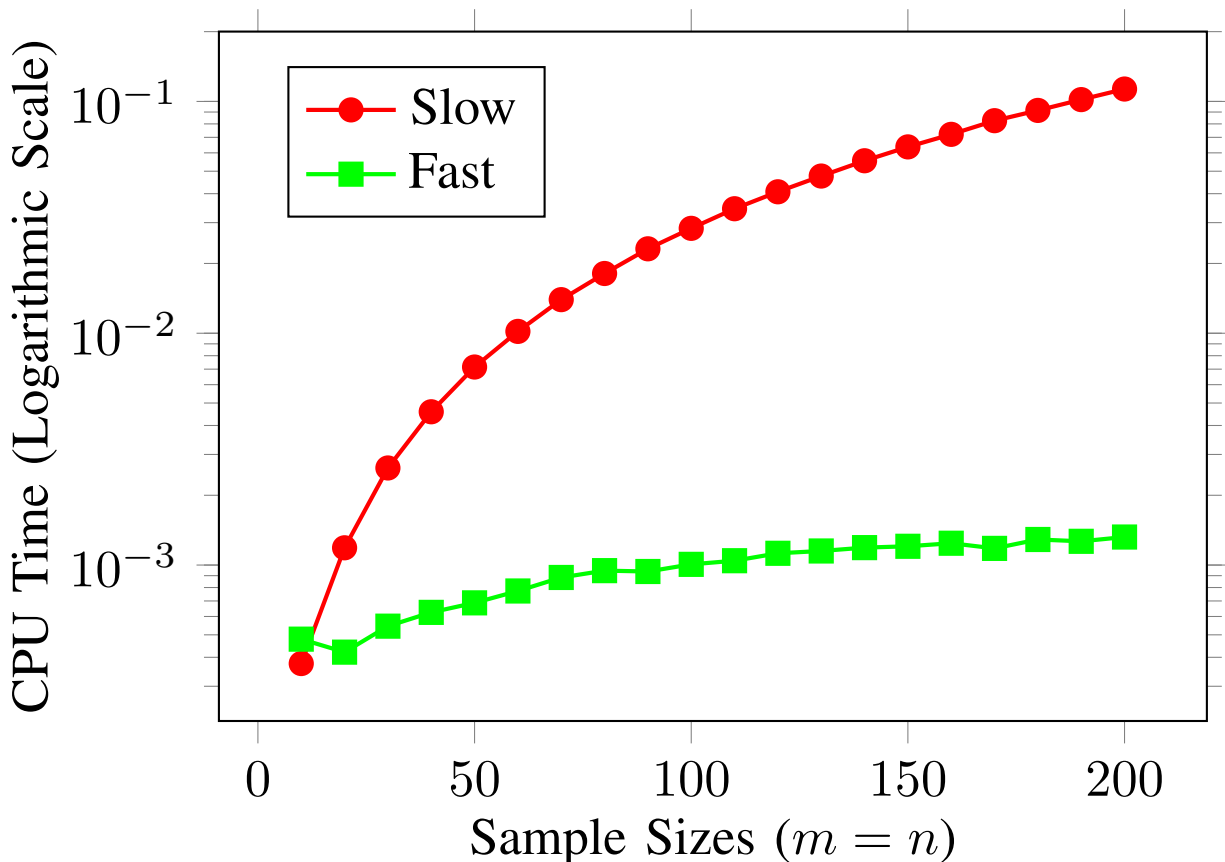


Fig. 3. Contrast of computational speeds between the conventional covariance estimator proposed by DeLong et al. using (1), (5) and (6) and our efficient version proposed in Theorem 1 using (16), (17) and (18). For simplicity, the sample sizes of X - and Y -class for each experiments are set to be equal, namely, $m = n = 10(10)200$. All samples are drawn from bivariate normal distributions. A logarithmic scale is used for a better visual effect.

- [8] G. Guo, M. Mandal, and Y. Jing, "A robust detector of known signal in non-gaussian noise using threshold systems," *Signal Processing*, vol. 92, no. 11, pp. 2676–2688, 2012.
- [9] V. Hari, G. Anand, A. Premkumar, and A. Madhukumar, "Design and performance analysis of a signal detector based on suprathreshold stochastic resonance," *Signal Processing*, vol. 92, no. 7, pp. 1745–1757, 2012.
- [10] D. Bamber, "The area above the ordinal dominance graph and the area below the receiver operating characteristic graph," *Journal of Mathematical Psychology*, vol. 12, no. 4, pp. 387–415, 1975.
- [11] H. Mann and D. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *The Annals of Mathematical Statistics*, vol. 18, no. 1, pp. 50–60, 1947.
- [12] J. Hanley, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 743, pp. 29–36, 1982.
- [13] J. A. Hanley and B. J. McNeil, "A method of comparing the areas under receiver operating characteristic curves derived from the same cases," *Radiology*, vol. 148, no. 3, pp. 839–843, 1983.

- [14] E. DeLong, D. DeLong, and D. Clarke-Pearson, "Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach," *Biometrics*, pp. 837–845, 1988.
- [15] L. Popescu, "Nonparametric ROC and LROC analysis," *Medical Physics*, vol. 34, p. 1556, 2007.
- [16] K. Zou, W. Hall, and D. Shapiro, "Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests," *Statistics in Medicine*, vol. 16, no. 19, pp. 2143–2156, 1998.
- [17] W. Xu, J. Dai, Y. Hung, and Q. Wang, "Estimating the area under a receiver operating characteristic (roc) curve: Parametric and nonparametric ways," *Signal Processing*, vol. 93, no. 11, pp. 3111–3123, 2013.
- [18] W. Xu, C. Chang, Y. S. Hung, S. K. Kwan, and P. C. W. Fung, "Order statistic correlation coefficient and its application to association measurement of biosignals," *Proc. Int. Conf. Acoustics, Speech, Signal Process. (ICASSP) 2006*, vol. 2, pp. II–1068–II–1071, May 2006.
- [19] —, "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.
- [20] W. Xu, C. Chang, Y. S. Hung, and P. C. W. Fung, "Asymptotic properties of order statistics correlation coefficient in the normal cases," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2239–2248, Jun. 2008.
- [21] W. Xu, Y. S. Hung, M. Niranjana, 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, Feb. 2010.
- [22] W. Xu, Y. Hou, Y. Hung, and Y. Zou, "A comparative analysis of Spearman's rho and Kendall's tau in normal and contaminated normal models," *Signal Processing*, vol. 93, pp. 261–276, Jan. 2013.
- [23] R. Ma, W. Xu, Q. Wang, and W. Chen, "Robustness analysis of three classical correlation coefficients under contaminated gaussian model," *Signal Processing*, vol. 104, pp. 51–58, 2014.
- [24] M. Kendall and J. D. Gibbons, *Rank Correlation Methods*, 5th ed. New York: Oxford University Press, 1990.
- [25] T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein *et al.*, *Introduction to algorithms*. MIT press Cambridge, 2001, vol. 2.
- [26] E. L. Lehmann and H. J. D'Abrera, *Nonparametrics: statistical methods based on ranks*. Springer New York, 2006.
- [27] A. Wunderlich and F. Noo, "A nonparametric procedure for comparing the areas under correlated lroc curves," *IEEE Transactions on Medical Imaging*, vol. 31, no. 11, pp. 2050–2061, 2012.