IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received June 6, 2020, accepted June 21, 2020, date of publication July 22, 2020, date of current version August 5, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3011159

Fast and Unbiased Estimation of Volume Under Ordered Three-Class ROC Surface (VUS) With Continuous or Discrete Measurements

SHUN LIU[1], HONGBIN ZHU[1], KAI YI[2], XU SUN[3], (Member, IEEE), WEICHAO XU[1], (Member, IEEE), AND CHAO WANG[4], (Member, IEEE)

[1]School of Automation, Guangdong University of Technology, Guangzhou 510006, China
[2]MediaTek Inc., Shenzhen 518000, China
[3]Baidu, Inc., Beijing 100193, China
[4]School of Engineering and Digital Arts, University of Kent, Canterbury CT2 7NT, U.K.

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

This work was supported in part by the National Natural Science Foundation of China under Grant 61771148, Grant 61975248, and Grant 61875041, and in part by the Guangzhou Technology Project Plan under Grant 201902010045 and Grant 201802010037.

ABSTRACT Receiver Operating Characteristic (ROC) surfaces have been studied in the literature essentially during the last decade and are considered as a natural generalization of ROC curves in three-class problems. The volume under the surface (VUS) is useful for evaluating the performance of a trichotomous diagnostic system or a three-class classifier's overall accuracy when the possible disease condition or sample belongs to one of three ordered categories. In the areas of medical studies and machine learning, the VUS of a new statistical model is typically estimated through a sample of ordinal and continuous measurements obtained by some suitable specimens. However, discrete scales of the prediction are also frequently encountered in practice. To deal with such scenario, in this paper, we proposed a unified and efficient algorithm of linearithmic order, based on dynamic programming, for unbiased estimation of the mean and variance of VUS with unidimensional samples drawn from continuous or non-continuous distributions. Monte Carlo simulations verify our theoretical findings and developed algorithms.

INDEX TERMS Volume under the surface (VUS), variance, discrete measurements, dynamic programming, receiver operating characteristic (ROC).

I. INTRODUCTION

After decades of development since early 1950s, receiver operating characteristic (ROC) analysis has found abundant applications in a wide spectrum of scientific and engineering areas [1]–[3], including data mining [4], computer vision [5], [6], signal processing [7]–[10], machine learning [11]–[13], medical decision making [14], psychology [15], and biomedical informatics [16], among others. Traditionally, ROC analysis can only deal with the two-class problems. Given the prior knowledge of the sample membership (abnormal vs. normal), a two-dimensional ROC curve, which is a plot of false positive rate against true positive rate, can be traced out based on various decision threshold settings [17]. The area under the curve (AUC) can then be estimated, in either parametric [18], [19] or nonparametric ways [20]–[23], as a figure of merit for various purposes [24], [25].

The associate editor coordinating the review of this manuscript and approving it for publication was Yilun Shang.

However, three-class tasks are frequently encountered in practice, where the conventional ROC analysis falls short as two mutually exclusive outcomes is no longer applicable [26], [27]. In medicine, for example, the heart signal is sometimes categorized as Bradycardia (slower rhythm), normal, and Tachycardia (faster rhythm); and the blood pressure is classified as Hypertension (lower pressure), normal, and Hypertension (higher pressure). In communication, the amplitudes of transmitted signals fall into three categories, as negative (binary "0"), idle (baseline), and positive (binary "1"). Fortunately, the concept of ROC curve analysis has been extended to accommodate trichotomous problems thanks to the efforts of many researchers [28]–[36]. Similar to the dichotomous case, ROC analysis for three-class problems is also a supervised methodology requiring prior knowledge of the sample membership (abnormal, intermediate, normal, say). Given such knowledge, a three-dimensional surface can be traced out by a series of triplets associated with two various decision threshold choices [37]. The volume under the surface (VUS), in parallel to AUC in two-class problems, is usually employed for measuring the overall accuracy of the test. Following this direction, other researchers have proposed various methods to estimate the mean and variance of VUS in the case of ordered multi-class [33], [34], [38]–[40], a setting which is also referred to as ordinal regression [41]–[44]. Besides the above-mentioned methods focusing on one-dimensional ordered three-class measurements, other techniques for ROC analysis of high-dimensional data have also been proposed, including Mossman's three-way method [30], He's likelihood ratio based framework [35], Dreiseitl's nonparametric algorithms [31]. From the viewpoint of computation, these methods are unsatisfactory because the time complexity of them is polynomial, ranging from quintic order [30], [31], [33], [34], [38] to quadratic order [39]. To avoid such excessive computational load, recently, Liu *et al.* developed an unbiased and linearithmic algorithms for estimating the variance of VUS with continuous measurements.

It is worth noting that most of the existing nonparametric methodologies are proposed for samples drawn from continuous distributions. However, in practice, outcomes of the statistical models (their names vary with different fields, e.g. classifiers in machine learning, diagnostic in medicine, detector in signal processing), will sometimes be discrete, in other words, the parental distributions of outputs are determined by probability mass functions (pmfs) instead of probability density functions (pdfs). For example, the probability that a sample being categorized into three different classes via k nearest neighbor (KNN) will sometimes be the same, especially when the hyperparameter k is small [11]. In the case of decision making, observers who participate in a three-interval discrimination task are usually required to classify targets into a certain range rather than a specific value [28].

For two-class problem, methods to estimate area under the curve (AUC) and its variance by means of discrete measurements have been developed, for instance, in Kaufmann *et al.* [20], Neubert and Brunner [21], Konietschke *et al.* [22], Xu *et al.* [23]. Recently, the study of Duc *et al.* under discrete sample condition attracted researchers attention, but their work mainly focused on the nonignorable verification bias when achieving estimation. To the best of our knowledge, the point estimation of VUS established by discrete data has not been explicitly presented in literature, not to mention its variance. Issues of estimating the mean and variance of VUS based on discrete measurements are still scarcely considered in the statistical literature. Only Mossman's boostrap method [30] and the tie breaking used in Liu *et al.* [45] gave contribution in this direction, but the former algorithm is biased and time-consuming, while the latter combined with any existing algorithms based on continuous measurements, e.g. Waegeman *et al.* [39], also suffers sever bias (see Section V for more details).

Motivated by such unsatisfactory situation, in this paper, we developed a linearithmic algorithm for unbiased estimation of the mean and variance of VUS with three ordinal continuous or discrete measurements that are under the unidimensional assumption. Our algorithm possesses four advantages as follows. Firstly, our algorithm is unified, that is, it can simultaneously work for samples drawn from both continuous and discrete populations. Secondly, it is unbiased, in other words, the mean of its output is equivalent to the population version of VUS's variance, which is always a necessary feature in statistical inference. Thirdly, it is only in linearithmic time, that is, the time complexity is in the order of the product of sample size and its logarithm, which is a favorite property in big data trials [46], [47]. Last but not least, the derived algorithm is nonparametric since it depends only on samples, eliminating the need for considering any other parametric model assumptions regarding the functional form of cumulative distributional functions (cdfs hereafter).

The rest of this paper is constructed as follows. Section II gives the basic definition of an unbiased estimator of VUS and the related general formulation regarding its variance with discrete measurements based on the U-statistic theory. Section III derives the unbiased expression of the variance of the VUS's estimator with discrete measurements. A linearithmic algorithm for computing VUS and its corresponding variance based on dynamic programming is presented in Section IV. In Section V, simulation experiments are undertaken to demonstrate our theoretical and algorithmic findings. Finally, we arrive at conclusions in Section VI.

II. PRELIMINARIES

For completeness and ease of later discussion, this section presents some preliminaries concerning the probabilistic interpretation of VUS (population version), sample version of VUS and an unbiased estimator for the variance of the sample version. These results are mandatory for later development of fast algorithms.

A. PROBABILISTIC INTERPRETATION OF VUS

Let $\{X_{1i}\}_{i=1}^{n_1}$,$\{X_{2j}\}_{j=1}^{n_2}$,$\{X_{3k}\}_{k=1}^{n_3}$ be sample sets drawn from three independent distributions with discrete cdfs F_1, F_2 and F_3, respectively. As illustrated in [33], the following linear combination of four probabilities

$$\theta \triangleq \Pr(X_3 > X_2 > X_1) + \frac{1}{2}\Pr(X_3 = X_2 > X_1) + \frac{1}{2}\Pr(X_3 > X_2 = X_1) + \frac{1}{6}\Pr(X_3 = X_2 = X_1) \quad (1)$$

can be interpreted as the VUS corresponding to distributions F_1, F_2 and F_3. From (1), it follows that $\theta = 1/6$ when X_1, X_2 and X_3 are identically distributed, i.e., $F_1 = F_2 = F_3$; whereas $\theta = 1$ if, from left to right, X_1, X_2 and X_3 are perfectly separable, i.e., there is no overlap between any two of the three class distributions. Therefore, VUS can characterize the extend of separation of the three classes. For continuous measurements, that is, the distributions of X_1, X_2 and X_3 are all continuous, (1) will degenerate to the familiar form

$$\theta = \Pr(X_3 > X_2 > X_1). \quad (2)$$

B. SAMPLE VERSION OF VUS

Based on (1), a nonparametric estimator of VUS (sample version) can be constructed readily, as [33]

$$\hat{\theta} = \frac{1}{n_1 n_2 n_3} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} \mathcal{H}(X_{3k}, X_{2j}, X_{1i}), \tag{3}$$

where

$$\mathcal{H}(X_3, X_2, X_1) \triangleq \begin{cases} 1, & X_3 > X_2 > X_1 \\ 1/2, & X_3 = X_2 > X_1 \\ & \text{or } X_3 > X_2 = X_1 \\ 1/6, & X_3 = X_2 = X_1. \end{cases} \tag{4}$$

Note that the sample version (3) just defined is an unbiased estimator of θ, i.e., $\mathbb{E}(\hat{\theta}) = \theta$. See [28], [29], [31], [34] for the justification.

C. VARIANCE OF $\hat{\theta}$

Given the sample version $\hat{\theta}$ in (3), it is of crucial importance to estimate its variance if one needs to infer the confidence interval or perform a hypothesis test. According to the arguments of Nakas and Yiannoutsos [33] and Dreiseitl *et al.* [31], the variance of $\hat{\theta}$, denoted by $\mathbb{V}(\hat{\theta})$, can be expressed as in Lemma 1 below.

Lemma 1: The variance of $\hat{\theta}$ that is defined in (3) is

$$\begin{aligned} \mathbb{V}(\hat{\theta}) = \frac{1}{n_1 n_2 n_3}&[q_0 - \theta^2 + (n_3 - 1)(q_{12} - \theta^2) \\ &+(n_2 - 1)(q_{13} - \theta^2) + (n_1 - 1)(q_{23} - \theta^2) \\ &+(n_2 - 1)(n_3 - 1)(q_1 - \theta^2) \\ &+(n_1 - 1)(n_3 - 1)(q_2 - \theta^2) \\ &+(n_1 - 1)(n_2 - 1)(q_3 - \theta^2)], \end{aligned} \tag{5}$$

where

$$q_0 = \mathbb{E}\big[\mathcal{H}(X_3, X_2, X_1)^2\big], \tag{6}$$

$$q_{12} = \mathbb{E}\big[\mathcal{H}(X_3, X_2, X_1)\mathcal{H}(X_3', X_2, X_1)\big], \tag{7}$$

$$q_{13} = \mathbb{E}\big[\mathcal{H}(X_3, X_2, X_1)\mathcal{H}(X_3, X_2', X_1)\big], \tag{8}$$

$$q_{23} = \mathbb{E}\big[\mathcal{H}(X_3, X_2, X_1)\mathcal{H}(X_3, X_2, X_1')\big], \tag{9}$$

$$q_1 = \mathbb{E}\big[\mathcal{H}(X_3, X_2, X_1)\mathcal{H}(X_3', X_2', X_1)\big], \tag{10}$$

$$q_2 = \mathbb{E}\big[\mathcal{H}(X_3, X_2, X_1)\mathcal{H}(X_3', X_2, X_1')\big], \tag{11}$$

$$q_3 = \mathbb{E}\big[\mathcal{H}(X_3, X_2, X_1)\mathcal{H}(X_3, X_2', X_1')\big], \tag{12}$$

with X' being an i.i.d. copy of X.

Proof: Write $\mathcal{H}_{ijk} \triangleq \mathcal{H}(X_{3k}, X_{2j}, X_{1i})$. Let $\mathbb{C}(\cdot, \cdot)$ be the covariance of two variables. It follows that

$$\begin{aligned} \mathbb{V}(\hat{\theta}) &= \mathbb{C}(\hat{\theta}, \hat{\theta}) \\ &= \mathbb{C}\Big(\frac{1}{n_1 n_2 n_3} \sum_i \sum_j \sum_k \mathcal{H}_{ijk}, \\ &\qquad \frac{1}{n_1 n_2 n_3} \sum_i \sum_j \sum_k \mathcal{H}_{i'j'k'}\Big) \\ &= \frac{1}{n_1^2 n_2^2 n_3^2} \sum_i \sum_j \sum_k \sum_{i'} \sum_{j'} \sum_{k'} \mathbb{C}(\mathcal{H}_{ijk}, \mathcal{H}_{i'j'k'}). \end{aligned} \tag{13}$$

From (3), it follows that $\mathbb{E}(\mathcal{H}_{ijk}) = \theta$. Then

$$\begin{aligned} &\mathbb{C}(\mathcal{H}_{ijk}, \mathcal{H}_{i'j'k'}) \\ &\quad = \mathbb{E}(\mathcal{H}_{ijk} \cdot \mathcal{H}_{i'j'k'}) - \mathbb{E}(\mathcal{H}_{ijk}) \cdot \mathbb{E}(\mathcal{H}_{i'j'k'}) \\ &\quad = \mathbb{E}(\mathcal{H}_{ijk} \cdot \mathcal{H}_{i'j'k'}) - \theta^2. \end{aligned} \tag{14}$$

Since the triplets (X_{1i}, X_{2j}, X_{3k}) and $(X_{1i'}, X_{2j'}, X_{3k'})$ are independent, we have $\mathbb{C}(\mathcal{H}_{ijk}, \mathcal{H}_{i'j'k'}) = 0$ for $i \neq i', j \neq j', k \neq k'$. Then, the sixfold sum in Eq.(13), with the assistance of (14), can be rewritten into the summation of seven terms, as:

$$\begin{aligned} &\sum_i \sum_j \sum_k \sum_{i'} \sum_{j'} \sum_{k'} \mathbb{C}(\mathcal{H}_{ijk}, \mathcal{H}_{i'j'k'}) \\ &= \underbrace{\sum_i \sum_j \sum_k \Big[\mathbb{E}(\mathcal{H}_{ijk} \cdot \mathcal{H}_{ijk}) - \theta^2\Big]}_{=n_1 n_2 n_3 (q_0 - \theta^2)} \\ &\quad + \underbrace{\sum_i \sum_j \sum_k \sum_{k \neq k'} \Big[\mathbb{E}(\mathcal{H}_{ijk} \cdot \mathcal{H}_{ijk'}) - \theta^2\Big]}_{=n_1 n_2 n_3 (n_3 - 1)(q_{12} - \theta^2)} \\ &\quad + \underbrace{\sum_i \sum_j \sum_{j \neq j'} \sum_k \Big[\mathbb{E}(\mathcal{H}_{ijk} \cdot \mathcal{H}_{ij'k}) - \theta^2\Big]}_{=n_1 n_2 (n_2 - 1) n_3 (q_{13} - \theta^2)} \\ &\quad + \underbrace{\sum_i \sum_{i \neq i'} \sum_j \sum_k \Big[\mathbb{E}(\mathcal{H}_{ijk} \cdot \mathcal{H}_{i'jk}) - \theta^2\Big]}_{=n_1 (n_1 - 1) n_2 n_3 (q_{23} - \theta^2)} \\ &\quad + \underbrace{\sum_i \sum_j \sum_{j \neq j'} \sum_k \sum_{k \neq k'} \Big[\mathbb{E}(\mathcal{H}_{ijk} \cdot \mathcal{H}_{ij'k'}) - \theta^2\Big]}_{=n_1 n_2 (n_2 - 1) n_3 (n_3 - 1)(q_1 - \theta^2)} \\ &\quad + \underbrace{\sum_i \sum_{i \neq i'} \sum_j \sum_k \sum_{k \neq k'} \Big[\mathbb{E}(\mathcal{H}_{ijk} \cdot \mathcal{H}_{i'jk'}) - \theta^2\Big]}_{=n_1 (n_1 - 1) n_2 n_3 (n_3 - 1)(q_2 - \theta^2)} \\ &\quad + \underbrace{\sum_i \sum_{i \neq i'} \sum_j \sum_{j \neq j'} \sum_k \Big[\mathbb{E}(\mathcal{H}_{ijk} \cdot \mathcal{H}_{i'j'k}) - \theta^2\Big]}_{=n_1 (n_1 - 1) n_2 (n_2 - 1) n_3 (q_3 - \theta^2)}. \end{aligned} \tag{15}$$

The result (5) thus follows upon substitution of (15) into (13).

□

III. UNBIASED ESTIMATORS

A. UNBIASED ESTIMATOR OF θ – SLOW VERSION

As remarked before, (3) is an unbiased estimator of the VUS θ defined by (1). However, a direct implementation of (3) is computationally rather inefficient, since the time complexity is cubic. Nevertheless, as illustrated later on, a linearithmic algorithm is available by means of dynamic programming.

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

Based on Lemma 1, an unbiased estimator of $\mathbb{V}(\hat{\theta})$ can be established, as stated in the following Lemma 2.

Lemma 2: Let $\hat{\theta}$ be defined as in (3) associated with three i.i.d. samples $\{X_{1i}\}_{i=1}^{n_1}$, $\{X_{2j}\}_{j=1}^{n_2}$ and $\{X_{3k}\}_{k=1}^{n_3}$ drawn from three discrete distributions respectively. Let $\hat{\sigma}_{\hat{\theta}}^2$ be the estimator of $\mathbb{V}(\hat{\theta})$. Denoted by $n_i^{[2]} = n_i(n_i-1)$, $i = 1, 2, 3$. Then an unbiased estimator of $\mathbb{V}(\hat{\theta})$ can be formulated as

$$\begin{aligned}\hat{\sigma}_{\hat{\theta}}^2 = &\frac{1}{(n_1-1)(n_2-1)(n_3-1)}\\ &\times[\hat{q}_0 - \hat{\theta}^2 + (n_3-1)(\hat{q}_{12}-\hat{\theta}^2)\\ &+(n_2-1)(\hat{q}_{13}-\hat{\theta}^2) + (n_1-1)(\hat{q}_{23}-\hat{\theta}^2)\\ &+(n_2-1)(n_3-1)(\hat{q}_1-\hat{\theta}^2)\\ &+(n_1-1)(n_3-1)(\hat{q}_2-\hat{\theta}^2)\\ &+(n_1-1)(n_2-1)(\hat{q}_3-\hat{\theta}^2)]\end{aligned} \tag{16}$$

where

$$\hat{\theta} = \frac{1}{n_1n_2n_3}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}\mathcal{H}(X_{3k}, X_{2j}, X_{1i}), \tag{17}$$

$$\hat{q}_0 = \frac{1}{n_1n_2n_3}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}\left[\mathcal{H}(X_{3k}, X_{2j}, X_{1i})\right]^2, \tag{18}$$

$$\begin{aligned}\hat{q}_{12} = &\frac{1}{n_1n_2n_3^{[2]}}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k\neq k'=1}^{n_3}\sum^{n_3}[\mathcal{H}(X_{3k}, X_{2j}, X_{1i})\\ &\times\mathcal{H}(X_{3k'}, X_{2j}, X_{1i})],\end{aligned} \tag{19}$$

$$\begin{aligned}\hat{q}_{13} = &\frac{1}{n_1n_2^{[2]}n_3}\sum_{i=1}^{n_1}\sum_{j\neq j'=1}^{n_2}\sum^{n_2}\sum_{k=1}^{n_3}[\mathcal{H}(X_{3k}, X_{2j}, X_{1i})\\ &\times\mathcal{H}(X_{3k}, X_{2j'}, X_{1i})],\end{aligned} \tag{20}$$

$$\begin{aligned}\hat{q}_{23} = &\frac{1}{n_1^{[2]}n_2n_3}\sum_{i\neq i'=1}^{n_1}\sum^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}[\mathcal{H}(X_{1k}, X_{2j}, X_{1i})\\ &\cdot\mathcal{H}(X_{3k}, X_{2j}, X_{1i'})],\end{aligned} \tag{21}$$

$$\begin{aligned}\hat{q}_1 = &\frac{1}{n_1n_2^{[2]}n_3^{[2]}}\sum_{i=1}^{n_1}\sum_{j\neq j'=1}^{n_2}\sum^{n_2}\sum_{k\neq k'=1}^{n_3}\sum^{n_3}[\mathcal{H}(X_{3k}, X_{2j}, X_{1i})\\ &\times\mathcal{H}(X_{3k'}, X_{2j'}, X_{1i})],\end{aligned} \tag{22}$$

$$\begin{aligned}\hat{q}_2 = &\frac{1}{n_1^{[2]}n_2n_3^{[2]}}\sum_{i\neq i'=1}^{n_1}\sum^{n_1}\sum_{j=1}^{n_2}\sum_{k\neq k'=1}^{n_3}\sum^{n_3}[\mathcal{H}(X_{3k}, X_{2j}, X_{1i})\\ &\times\mathcal{H}(X_{3k'}, X_{2j}, X_{1i'})],\end{aligned} \tag{23}$$

$$\begin{aligned}\hat{q}_3 = &\frac{1}{n_1^{[2]}n_2^{[2]}n_3}\sum_{i\neq i'=1}^{n_1}\sum^{n_1}\sum_{j\neq j'=1}^{n_2}\sum^{n_2}\sum_{k=1}^{n_3}[\mathcal{H}(X_{3k}, X_{2j}, X_{1i})\\ &\times\mathcal{H}(X_{3k}, X_{2j'}, X_{1i'})].\end{aligned} \tag{24}$$

Proof:

To demonstrate that $\mathbb{E}(\hat{\sigma}_{\hat{\theta}}^2) = \sigma_{\hat{\theta}}^2$, it suffices to verify the unbiasedness of the $\hat{q}$-terms with the corresponding q-terms in the numerator of (16). Based on the definitions of q-terms, it follows readily that

$$\mathbb{E}(\hat{q}_\zeta) = q_\zeta, \tag{25}$$

where $\zeta \in \{0, 12, 13, 23, 1, 2, 3\}$ stands for the subscript of q-terms. Using the relationship $\sigma_{\hat{\theta}}^2 = \mathbb{E}(\hat{\theta}^2) - \theta^2$, we have

$$\mathbb{E}(\hat{\theta}^2) = \theta^2 + \sigma_{\hat{\theta}}^2. \tag{26}$$

Taking expectation of both sides of (16) and (26), it follows that

$$\mathbb{E}(\hat{q}_0) = \theta(1-\theta) - \sigma_{\hat{\theta}}^2 \tag{27}$$

and

$$\mathbb{E}(\hat{q}_\zeta - \hat{\theta}^2) = q_\zeta - \theta^2 - \sigma_{\hat{\theta}}^2. \tag{28}$$

A substitution of (27)–(28) into the expectation of (16) along with some straightforward algebra, we have

$$\mathbb{E}(\hat{\sigma}_{\hat{\theta}}^2) = \sigma_{\hat{\theta}}^2,$$

and the theorem thus follows. □

IV. FAST ALGORITHMS

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

It is noticeable that a direct implementation based on (16)–(24) in Theorem 2 is very time-consuming for large n_1, n_2, and n_3, due to the quintic summations involved in the $\hat{q}$-terms. Fortunately, however, following our previous work [23], [40], a linearirthmic algorithm can be conveniently applied via rewriting (17)–(24) in terms of the quantities S illustrated in Table 3, which stand for the number of events satisfying the relation inside the corresponding brackets. As shown below, these S-terms can all be computed with dynamic programming. Note that in Table 3, X_1', X_2' and X_3' stand for i.i.d copies of X_1, X_2 and X_3, respectively.

Theorem 1: Let $\hat{\theta}$ be defined as in (3) with respect to three i.i.d. samples, $X_{11}, \ldots, X_{1n_1}$, $X_{21}, \ldots, X_{2n_2}$, and $X_{31}, \ldots, X_{3n_3}$ respectively. Denoted by $n^{[2]} = n_i(n_i-1)$, $i = 1, 2, 3$. Then the estimator $\hat{\sigma}_{\hat{\theta}}^2$ in Lemma 2 is equivalent to

$$\begin{aligned}\hat{\sigma}_{\hat{\theta}}^2 = \hat{\varsigma}_{\hat{\theta}}^2 = &\frac{1}{(n_1-1)(n_2-1)(n_3-1)}[\hat{\mathcal{Q}}_0 - \hat{\theta}^2\\ &+(n_3-1)(\hat{\mathcal{Q}}_{12}-\hat{\theta}^2) + (n_2-1)(\hat{\mathcal{Q}}_{13}-\hat{\theta}^2)\\ &+(n_1-1)(\hat{\mathcal{Q}}_{23}-\hat{\theta}^2)\\ &+(n_2-1)(n_3-1)(\hat{\mathcal{Q}}_1-\hat{\theta}^2)\\ &+(n_1-1)(n_3-1)(\hat{\mathcal{Q}}_2-\hat{\theta}^2)\\ &+(n_1-1)(n_2-1)(\hat{\mathcal{Q}}_3-\hat{\theta}^2)],\end{aligned} \tag{29}$$

where

$$\begin{aligned}\hat{\theta} &= \frac{1}{n_1n_2n_3}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}\mathcal{H}(X_{3k}, X_{2j}, X_{1i})\\ &= \frac{1}{n_1n_2n_3}(S_0^{(1)} + \frac{1}{2}S_0^{(2)} + \frac{1}{2}S_0^{(3)} + \frac{1}{6}S_0^{(4)}),\end{aligned} \tag{30}$$

$$\begin{aligned}\hat{\mathcal{Q}}_0 &= \frac{1}{n_1n_2n_3}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}\left[\mathcal{H}(X_{3k}, X_{2j}, X_{1i})\right]^2\\ &= \frac{1}{n_1n_2n_3}(S_0^{(1)} + \frac{1}{4}S_0^{(2)} + \frac{1}{4}S_0^{(3)} + \frac{1}{36}S_0^{(4)}),\end{aligned} \tag{31}$$

$$\hat{\mathcal{Q}}_{12} = \hat{q}_{12} = \frac{1}{n_1 n_2 n_3^{[2]}}(2S_{12}^{(1)} + S_{12}^{(2)} + S_{12}^{(3)} + \frac{1}{2}S_{12}^{(4)} + \frac{1}{4}S_{12}^{(5)} + \frac{1}{4}S_{12}^{(6)} + \frac{1}{6}S_{12}^{(7)} + \frac{1}{36}S_{12}^{(8)} - n_1 n_2 n_3 \hat{\mathcal{Q}}_0), \tag{32}$$

$$\hat{\mathcal{Q}}_{13} = \hat{q}_{13} = \frac{1}{n_1 n_2^{[2]} n_3}(2S_{13}^{(1)} + S_{13}^{(2)} + S_{13}^{(3)} + S_{13}^{(4)} + \frac{1}{2}S_{13}^{(5)} + \frac{1}{4}S_{13}^{(6)} + \frac{1}{4}S_{13}^{(7)} + \frac{1}{36}S_{13}^{(8)} - n_1 n_2 n_3 \hat{\mathcal{Q}}_0), \tag{33}$$

$$\hat{\mathcal{Q}}_{23} = \hat{q}_{23} = \frac{1}{n_1^{[2]} n_2 n_3}(2S_{23}^{(1)} + S_{23}^{(2)} + S_{23}^{(3)} + \frac{1}{2}S_{23}^{(4)} + \frac{1}{4}S_{23}^{(5)} + \frac{1}{4}S_{23}^{(6)} + \frac{1}{6}S_{23}^{(7)} + \frac{1}{36}S_{23}^{(8)} - n_1 n_2 n_3 \hat{\mathcal{Q}}_0), \tag{34}$$

$$\hat{\mathcal{Q}}_1 = \hat{q}_1 = \frac{1}{n_1 n_2^{[2]} n_3^{[2]}}(4S_1^{(1)} + 3S_1^{(2)} + 2S_1^{(3)} + 2S_1^{(4)} + 2S_1^{(5)} + 2S_1^{(6)} + \frac{3}{2}S_1^{(7)} + S_1^{(8)} + S_1^{(9)} + S_1^{(10)} + S_1^{(11)} + S_1^{(12)} + S_1^{(13)} + S_1^{(14)} + \frac{1}{2}S_1^{(15)} + \frac{1}{2}S_1^{(16)} + \frac{1}{2}S_1^{(17)} + \frac{1}{2}S_1^{(18)} + \frac{1}{3}S_1^{(19)} + \frac{1}{4}S_1^{(20)} + \frac{1}{4}S_1^{(21)} + \frac{1}{6}S_1^{(22)} + \frac{1}{6}S_1^{(23)} + \frac{1}{36}S_1^{(24)} - n_1 n_2 n_3^{[2]} \hat{\mathcal{Q}}_{12} - n_1 n_2^{[2]} n_3 \hat{\mathcal{Q}}_{13} - n_1 n_2 n_3 \hat{\mathcal{Q}}_0), \tag{35}$$

$$\hat{\mathcal{Q}}_2 = \hat{q}_2 = \frac{1}{n_1^{[2]} n_2 n_3^{[2]}}(4S_2^{(1)} + 2S_2^{(2)} + 2S_2^{(3)} + 2S_2^{(4)} + 2S_2^{(5)} + S_2^{(6)} + S_2^{(7)} + S_2^{(8)} + \frac{5}{6}S_2^{(9)} + \frac{1}{2}S_2^{(10)} + \frac{1}{2}S_2^{(11)} + \frac{1}{4}S_2^{(12)} + \frac{1}{4}S_2^{(13)} + \frac{1}{6}S_2^{(14)} + \frac{1}{6}S_2^{(15)} + \frac{1}{36}S_2^{(16)} - n_1 n_2 n_3^{[2]} \hat{\mathcal{Q}}_{12} - n_1^{[2]} n_2 n_3 \hat{\mathcal{Q}}_{23} - n_1 n_2 n_3 \hat{\mathcal{Q}}_0), \tag{36}$$

$$\hat{\mathcal{Q}}_3 = \hat{q}_3 = \frac{1}{n_1^{[2]} n_2^{[2]} n_3}(4S_3^{(1)} + 3S_3^{(2)} + 2S_3^{(3)} + 2S_3^{(4)} + 2S_3^{(5)} + 2S_3^{(6)} + \frac{3}{2}S_3^{(7)} + S_3^{(8)} + S_3^{(9)} + S_3^{(10)} + S_3^{(11)} + S_3^{(12)} + S_3^{(13)} + S_3^{(14)} + \frac{1}{2}S_3^{(15)} + \frac{1}{2}S_3^{(16)} + \frac{1}{2}S_3^{(17)} + \frac{1}{2}S_3^{(18)} + \frac{1}{3}S_3^{(19)} + \frac{1}{4}S_3^{(20)} + \frac{1}{4}S_3^{(21)} + \frac{1}{6}S_3^{(22)} + \frac{1}{6}S_3^{(23)} + \frac{1}{36}S_3^{(24)} - n_1 n_2^{[2]} n_3 \hat{\mathcal{Q}}_{13} - n_1^{[2]} n_2 n_3 \hat{\mathcal{Q}}_{23} - n_1 n_2 n_3 \hat{\mathcal{Q}}_0). \tag{37}$$

Proof: See Appendix. □

B. FAST COMPUTING STRUCTURE FOR QUANTITIES S

Let $\mathcal{Z}_1, \cdots, \mathcal{Z}_N$, $N = n_1 + n_2 + n_3$ be a combined sequence consist of $X_{11}, \cdots, X_{1n_1}$, $X_{21}, \cdots, X_{2n_2}$, $X_{31}, \cdots, X_{3n_3}$. We yield the sequence of order statistic by

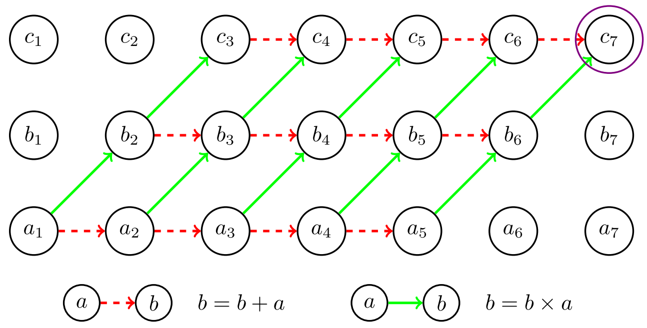


FIGURE 1. Counter matrix $\mathbf{C}_0^{(1)}$ and its diagram for computing $S_0^{(1)} = \sum_{k=3}^{K}\sum_{j=2}^{k-1}\sum_{i=1}^{j-1} c_k b_j a_i$, where $K = 7$ is for purpose of visualizing.

sorting the $\mathcal{Z}$-sequence in ascending order [48]–[50]

$$\underbrace{\mathcal{Z}_{(1)} = \cdots = \mathcal{Z}_{(1)}}_{\text{Block}_1} < \cdots < \underbrace{\mathcal{Z}_{(J)} = \cdots = \mathcal{Z}_{(J)}(= Z_i)}_{\text{Block}_J} < \cdots < \underbrace{\mathcal{Z}_{(K)} = \cdots = \mathcal{Z}_{(K)}}_{\text{Block}_K}. \tag{38}$$

Suppose that the elements of Block$_J$ are all equal to Z_i. Let a_i, b_i, c_i be the number of X_1's, X_2's and X_3's equaling to Z_i, respectively, for $i = 1, \ldots, K$. Then three count vectors corresponding to X_1, X_2 and X_3 can be attained, each is based on $\mathcal{Z}$-sequence (38), denoted by $\mathcal{C}_{X_1} \triangleq [a_1 \ \ldots \ a_K]$, $\mathcal{C}_{X_2} \triangleq [b_1 \ \ldots \ b_K]$ and $\mathcal{C}_{X_3} \triangleq [c_1 \ \ldots \ c_K]$. As shown in literature [40], $\mathcal{Z}$-sequence can be attained in a linearithmic time, i.e., $\mathcal{O}(N \log N)$, by means of some well-known and efficient sorting algorithms in the textbook [51]. Having the three count vectors, all the S-terms of Table 3 can be computed up to linear time $\mathcal{O}(5K)$, where $K \leq N$. In the following, we intend to comprehensively explain the computing structure by investigating the definitions of $S_0^{(1)}$, $S_0^{(2)}$, $S_{12}^{(5)}$ and $S_1^{(7)}$, respectively.

We first start from analysis of $S_0^{(1)}$. According to Table 3, the definition of $S_0^{(1)}$ is

$$S_0^{(1)} = \mathcal{E}(X_3 > X_2 > X_1) = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3} \mathcal{I}(X_{3k} > X_{2j} > X_{1i}) = \sum_{k=3}^{K}\sum_{j=2}^{k-1}\sum_{i=1}^{j-1} c_k b_j a_i, \tag{39}$$

where $\mathcal{I}(\cdot)$ is an indicator function that equals to 1(0) when the condition in bracket is *True*(*False*). It follows that (39) can be readily implemented via dynamic programing. Put more specially, we first stack $\mathcal{C}_{X_3}(Row_1)$, $\mathcal{C}_{X_2}(Row_2)$ and $\mathcal{C}_{X_1}(Row_3)$ aforementioned to construct a $3 \times K$ matrix $\mathbf{C}_0^{(1)}$, then further set $\mathbf{C}_0^{(1)}{}_{[2,1]}$ and $\mathbf{C}_0^{(1)}{}_{[1,2]}$ to be 0. As depicted in Figure. 1, the programming path goes from the lower left corner towards the upper right one in a linear time $\mathcal{O}(3K)$,

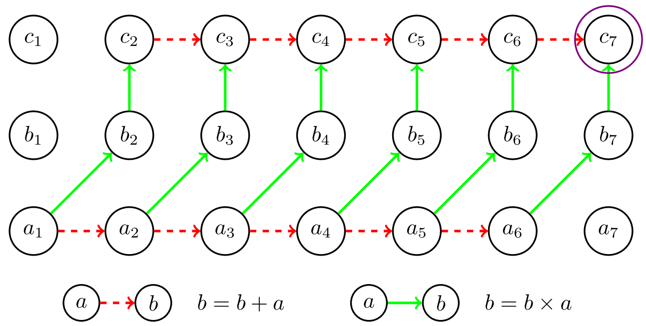


FIGURE 2. Counter matrix $\mathbf{C}_0^{(2)}$ and its diagram for computing $S_0^{(2)} = \sum_{j=2}^{K}\sum_{i=1}^{j-1} c_j b_j a_i$, where $K = 7$ is for purpose of visualizing.

with the update rule:

$$\mathbf{C}_{[I,J]} = \begin{cases} \mathbf{C}_{[I,J]}+\mathbf{C}_{[I,J-1]} & I=3,\ 2\le J\le K-2 \\ \mathbf{C}_{[I,J]}\cdot\mathbf{C}_{[I+1,J-1]}+\mathbf{C}_{[I,J-1]} & I=2,\ 2\le J\le K-1 \\ \mathbf{C}_{[I,J]}\cdot\mathbf{C}_{[I+1,J-1]}+\mathbf{C}_{[I,J-1]} & I=1,\ 3\le J\le K. \end{cases} \tag{40}$$

Finally, the desired value of $S_0^{(1)}$ is stored in the cell $\mathbf{C}_{\mathbf{0}\,[1,K]}^{(\mathbf{1})}$ as the index I, J iterating from 3 to 1 and 2 to K, respectively.

Regard to $S_0^{(2)}$ described in Table 3, it follows that

$$\begin{aligned} S_0^{(2)} &= \mathcal{E}(X_3 = X_2 > X_1) \\ &= \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}\mathcal{I}(X_{3k}=X_{2j}>X_{1i}) \\ &= \sum_{j=2}^{K}\sum_{i=1}^{j-1} c_j b_j a_i. \end{aligned} \tag{41}$$

We need to construct another counter matrix same as $\mathbf{C}_0^{(1)}$, denoted by $\mathbf{C}_0^{(2)}$, and set $\mathbf{C}_{\mathbf{0}\,[1,1]}^{(\mathbf{2})}$ to be 0. As illustrated in Figure 2, the programming path goes from the southwest corner towards the northeast corner in a linear time $\mathcal{O}(3K)$ with the update rule

$$\mathbf{C}_{[I,J]} = \begin{cases} \mathbf{C}_{[I,J]}+\mathbf{C}_{[I,J-1]} & I=3,\ 2\le J\le K-1 \\ \mathbf{C}_{[I,J]}\cdot\mathbf{C}_{[I+1,J-1]} & I=2,\ 2\le J\le K \\ \mathbf{C}_{[I,J]}\cdot\mathbf{C}_{[I+1,J]}+\mathbf{C}_{[I,J-1]} & I=1,\ 2\le J\le K. \end{cases} \tag{42}$$

Eventually, when the updating finished, the value of $S_0^{(2)}$ that we desired is saved in cell $\mathbf{C}_{\mathbf{0}\,[1,K]}^{(\mathbf{2})}$.

As to $S_{12}^{(5)}$, the definition of which follows that

$$\begin{aligned} S_{12}^{(5)} &= \mathcal{E}(X_3 = X_3' > X_2 = X_1) \\ &= \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}\sum_{l=1}^{n_3}\mathcal{I}(X_{3l} = X_{3k} > X_{2j} = X_{1i}) \\ &= \sum_{l=2}^{K}\sum_{j=1}^{l-1} c_l^2 b_j a_j, \end{aligned} \tag{43}$$

in this case, to begin with, we stack two $\mathcal{C}_{X_3}$ (Row_1 and Row_2 respectively), $\mathcal{C}_{X_2}$(Row_3) and $\mathcal{C}_{X_1}$(Row_4) to form a brand new counter matrix, denoted by $\mathbf{C}_{12}^{(5)}$. We further appoint 0 into cell $\mathbf{C}_{\mathbf{12}\,[1,1]}^{(\mathbf{5})}$. Then the desired output of $S_{12}^{(5)}$ is saved in cell $\mathbf{C}_{\mathbf{12}\,[1,K]}^{(\mathbf{5})}$ after the indexes I and J run from 1 to 3 and 1 to K respectively in a linear time ($\mathcal{O}(4K)$) with update rule

$$\mathbf{C}_{[I,J]} = \begin{cases} \mathbf{C}_{[I,J]}\cdot\mathbf{C}_{[I+1,J]} & I=3,\ 1\le J\le K-1 \\ \mathbf{C}_{[I,J]}+\mathbf{C}_{[I,J-1]} & I=3,\ 2\le J\le K-1 \\ \mathbf{C}_{[I,J]}\cdot\mathbf{C}_{[I+1,J-1]} & I=2,\ 2\le J\le K \\ \mathbf{C}_{[I,J]}\cdot\mathbf{C}_{[I+1,J]}+\mathbf{C}_{[I,J-1]} & I=1,\ 2\le J\le K. \end{cases} \tag{44}$$

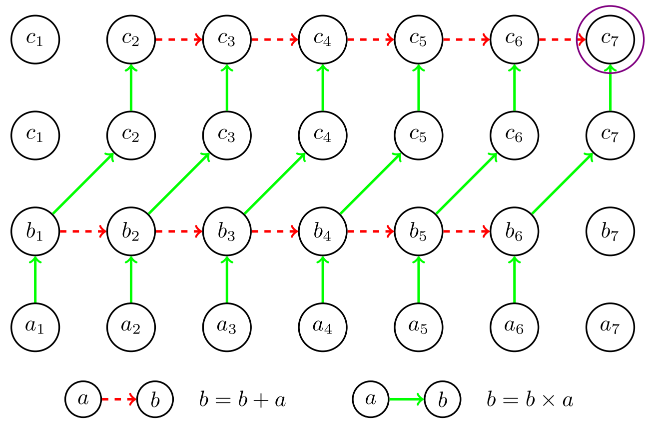


FIGURE 3. Counter matrix $\mathbf{C}_{12}^{5}$ and its diagram for computing $S_{12}^{5} = \sum_{l=2}^{K}\sum_{j=1}^{l-1} c_l^2 b_j a_j$, where $K = 7$ is for purpose of visualizing.

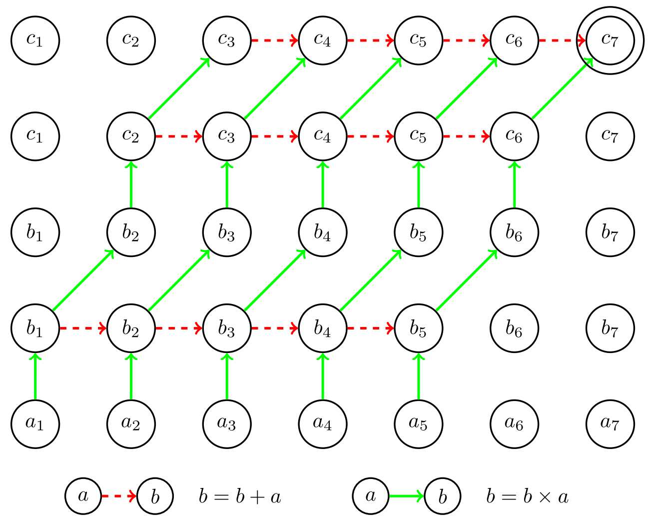


FIGURE 4. Counter matrix $\mathbf{C}_1^{7}$ and its diagram for computing $S_1^{7} = \sum_{k=3}^{K}\sum_{j=2}^{k-1}\sum_{i=1}^{j-1} c_k c_j b_j b_i a_i$, where $K = 7$ is for purpose of visualizing.

At last we concentrate on $S_1^{(7)}$, the definition of which is

$$\begin{aligned} S_1^{(7)} &= \mathcal{E}(X_3 > X_3' = X_2 > X_2' = X_1) \\ &= \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_2}\sum_{l=1}^{n_3}\sum_{m=1}^{n_3}\mathcal{I}(X_{3m} > X_{3l} = X_{2k} > X_{2j} = X_{1i}) \\ &= \sum_{k=3}^{K}\sum_{j=2}^{k-1}\sum_{i=1}^{j-1} c_k c_j b_j b_i a_i. \end{aligned} \tag{45}$$

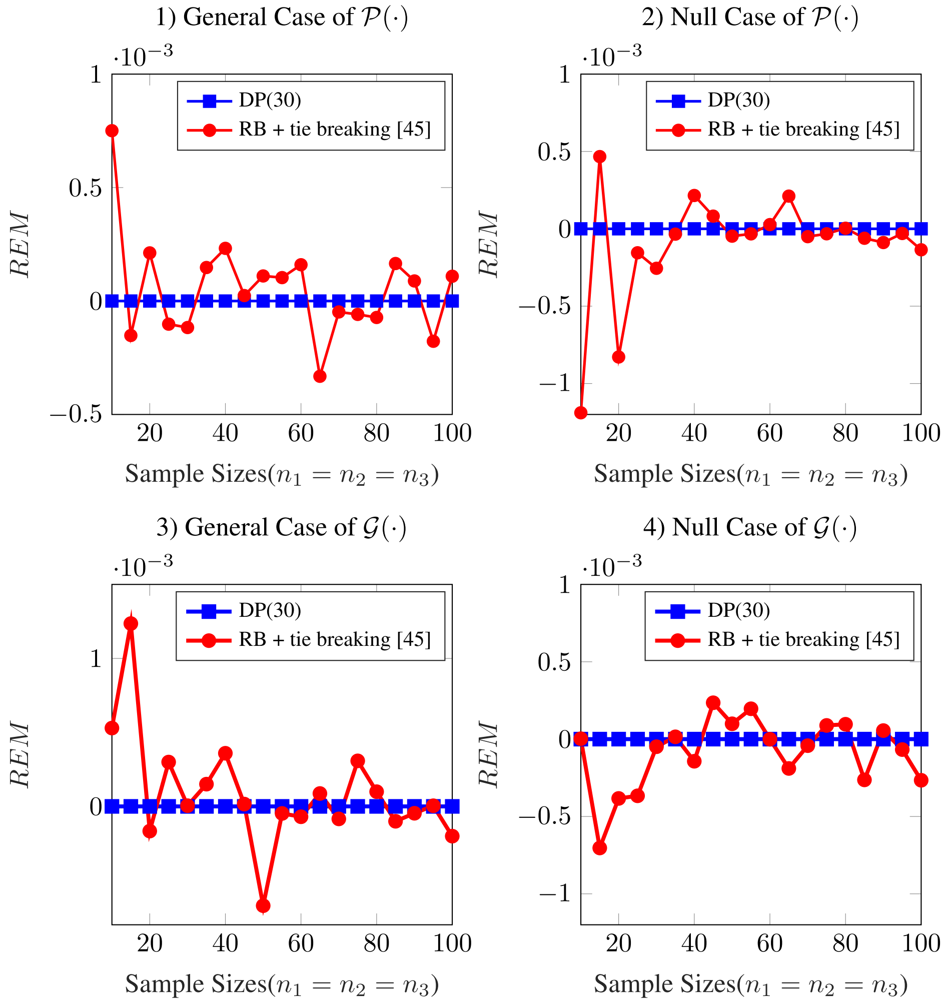


FIGURE 5. **Comparison of unbiasedness, in terms of *REM*, between the estimator in (30) and the rank-based method proposed in [45]. 1) Non-null case under Poisson distribution, where X_1, X_2 and X_3 follow $\mathcal{P}(10)$, $\mathcal{P}(12)$ and $\mathcal{P}(14)$, respectively. 2) Null case under Poisson distribution, where X_1, X_2 and X_3 all follow $\mathcal{P}(15)$. 3) Non-null case under Geometric distribution, with pmfs of $\Pr(X_1 = k) = (1-0.3)^{k-1}0.3, k = 1, 2, 3, \ldots, \infty$, $\Pr(X_2 = k) = (1-0.2)^{k-1}0.2, k = 1, 2, 3, \ldots, \infty$ and $\Pr(X_3 = k) = (1-0.1)^{k-1}0.1, k = 1, 2, 3, \ldots, \infty$. 4) Null case under Geometric distribution, where X_1, X_2 and X_3 all follow Poisson distribution with pmfs of $\Pr(X_3 = k) = (1-0.1)^{k-1}0.1$, $k = 1, 2, 3, \ldots, \infty$.**

In order to form the paths for programming $S_1^{(7)}$, a $5 \times K$ counter matrix, denoted by $\mathbf{C}_1^{(7)}$, should be created. In this matrix, the first two rows are both $\mathcal{C}_{X_3}$, the next two lines, i.e. Row_2 and Row_3, are $\mathcal{C}_{X_2}$, and the last row is equal to $\mathcal{C}_{X_1}$. We further put 0 into cells $\mathbf{C}_1^{(7)}{}_{[1,2]}$, $\mathbf{C}_1^{(7)}{}_{[2,1]}$ and $\mathbf{C}_1^{(7)}{}_{[3,1]}$, respectively. Then the desired value of $S_1^{(7)}$ will be stored in cell $\mathbf{C}_1^{(7)}{}_{[1,K]}$ after operating the indexes I and J with update rule

$$\mathbf{C}_{[I,J]} = \begin{cases} \mathbf{C}_{[I,J]} \cdot \mathbf{C}_{[I+1,J]} & I{=}4,\ 1{\le}J{\le}K-2 \\ \mathbf{C}_{[I,J]} + \mathbf{C}_{[I,J-1]} & I{=}4,\ 2{\le}J{\le}K-2 \\ \mathbf{C}_{[I,J]} \cdot \mathbf{C}_{[I+1,J-1]} & I{=}3,\ 2{\le}J{\le}K-1 \\ \mathbf{C}_{[I,J]} \cdot \mathbf{C}_{[I+1,J]} + \mathbf{C}_{[I,J-1]} & I{=}2,\ 2{\le}J{\le}K-1 \\ \mathbf{C}_{[I,J]} \cdot \mathbf{C}_{[I+1,J-1]} + \mathbf{C}_{[I,J-1]} & I{=}1,\ 3{\le}J{\le}K. \end{cases} \tag{46}$$

The algorithms for the rest S-terms can be constructed in a similar and straightforward manner according to Algorithm 1, and thus omitted for brevity.

V. NUMERIC RESULT

To evaluate our theoretical and algorithmic findings, in this section, we compare our dynamic programming based algorithm (Theorem 1) with the state-of-the-art algorithms for calculating the mean and variance of VUS [33], [39], [45], in terms of unbiasedness and computational efficiency. Without losing generality, throughout this section, the Monte Carlo simulation is undertaken for sample sizes from 10 to 100 with an increment of 10. Moreover, we set the sample sizes of each class to be the same, i.e. $n_1 = n_2 = n_3$, for simplicity. The number of trials is specified to be 10^5 for inhibiting experimental fluctuation if there is no additional statement. All samples of random variables following

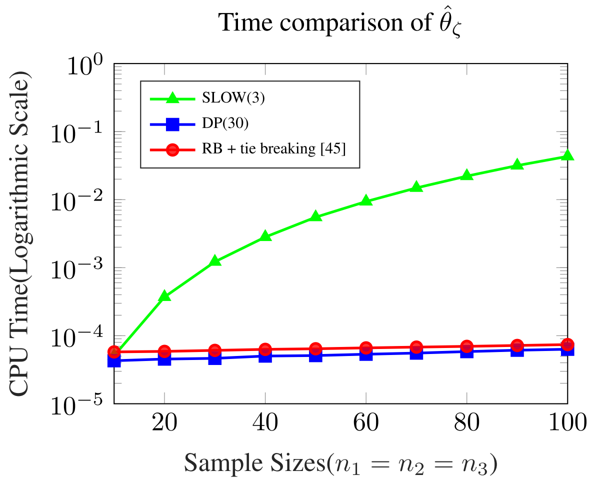


FIGURE 6. **Comparative results of CPU time between the algorithms based on (3), (30) and the rank-based method in [45]. For simplicity, the sample sizes of X_1, X_2 and X_3 are set to be equal. A log scale is used for better visualization.**

various distributions are generated by functions in Matlab `Statistics Toolbox`™.

A. COMPARISON OF ALGORITHMS FOR ESTIMATION OF MEAN OF VUS

We first compare the capacity of our algorithm for calculating the mean of VUS (30), denoted by $\hat{\theta}_{\text{DP}}$, with its slow version (3) and the state-of-the-art rank-based algorithm [45], denoted by $\hat{\theta}_{\text{SLOW}}$ and $\hat{\theta}_{\text{RB}}$ respectively. Ties between classes or within class lead to $\hat{\theta}_{\text{RB}}$ cannot be applied directly. To settle this problem, i.e. to break the tie, we add a zero-mean Gaussian noise with a tiny variance (10^{-5} in this work) while the parental distribution are discrete, which is a tie-breaking technique suggested in [17].

1) VERIFICATION OF UNBIASEDNESS

It suffices to analysis the unbiasedness of $\hat{\theta}_{\text{DP}}$ and $\hat{\theta}_{\text{RB}}$ with baseline $\hat{\theta}_{\text{SLOW}}$ only if there exist ties between classes or within the same class, because the expectation of $\hat{\theta}_{\text{SLOW}}$ and $\hat{\theta}_{\text{DP}}$ will be naturally degenerated from (1) to (2). In the following, we produce the three samples X_1, X_2 and X_3 based on Poisson distribution with parameter λ and Geometric distribution with parameter P, denoted by $\mathcal{P}(\lambda)$ and $\mathcal{G}(P)$ respectively, in four scenarios, which is listed as follows:

Under the four scenarios mentioned above, we evaluate the two methods by exploiting *Relative Error of Mean* (*REM*), which is defined by

$$REM \triangleq \frac{\mathbb{E}(\hat{\theta}_\zeta - \hat{\theta}_{\text{SLOW}})}{\theta_{\text{SLOW}}} \tag{47}$$

where the subscript $\zeta \in \{\text{DP}, \text{RB}\}$ indicates one of the two methods mentioned above and $\hat{\theta}_{\text{SLOW}}$ is an unbiased estimator of VUS, which is used for verifying the correctness of derivation in (30).

Algorithm 1 Procedure of Computing S in General Case

Input : Counter matrix $\mathbf{C}$ of size $m \times K$ and relation vector $\mathcal{R}$ of size $1 \times (m-1)$ holding either '>' or '='

Output: S corresponding to $\mathbf{C}$ and $\mathcal{R}$

```
begin
    m ⟵ the number of rows in C;
    K ⟵ the number of columns in C;
    CountL ⟵ the number of '>' in R;
    StartInd ⟵ 0;
    ComputeLen ⟵ K − CountL;
    L ⟵ a zero vector of length ComputeLen;
    for i = 1, 2, ..., ComputeLen do
        L_i ⟵ C_{m,i};
    end
    for i = m − 1, m − 2, ..., 1 do
        if R_{m−i} = '=' then
            for j = 1, 2, ..., ComputeLen do
                L_j ⟵ L_j × C_{i,StartInd+j};
            end
        else
            StartInd ⟵ StartInd + 1;
            for k = 2, 3, ..., ComputeLen do
                L_k ⟵ L_k + L_{k−1};
            end
            for l = 1, 2, ..., ComputeLen do
                L_l = L_l × C_{i,StartInd+l};
            end
        end
    end
    S ⟵ sum of all elements in L ;
end
```

Figure 5 shows the comparison results, in terms of *REM* over sample sizes, with respect to the two methods under four scenarios illustrated in Table 1. The top two panels are results corresponding to Poisson distribution, whereas the bottom panels are belonging to Geometric distribution. It is clear that $\hat{\theta}_{\text{DP}}$ outperforms $\hat{\theta}_{\text{RB}}$, in the sense that *REM*s of $\hat{\theta}_{\text{DP}}$ completely fits the line $REM = 0$, which confirms the unbiasedness of (30). On the other hand, it is of no surprise that all the curves of $\hat{\theta}_{\text{RB}}$ fluctuate around zero as the sample size increases since we have added a tiny value to break the tie of each measurement, which indicates that the combination of $\hat{\theta}_{\text{RB}}$ and tie breaking technique is biased if input data follows discrete distribution.

TABLE 1. **Parameters of the distributions followed by X_1, X_2 and X_3 for computing the mean of VUS under four scenarios.**

Scenario	X_1	X_2	X_3
1	$\mathcal{P}(10)$	$\mathcal{P}(12)$	$\mathcal{P}(14)$
2	$\mathcal{P}(15)$	$\mathcal{P}(15)$	$\mathcal{P}(15)$
3	$\mathcal{G}(0.3)$	$\mathcal{G}(0.2)$	$\mathcal{G}(0.1)$
4	$\mathcal{G}(0.1)$	$\mathcal{G}(0.1)$	$\mathcal{G}(0.1)$

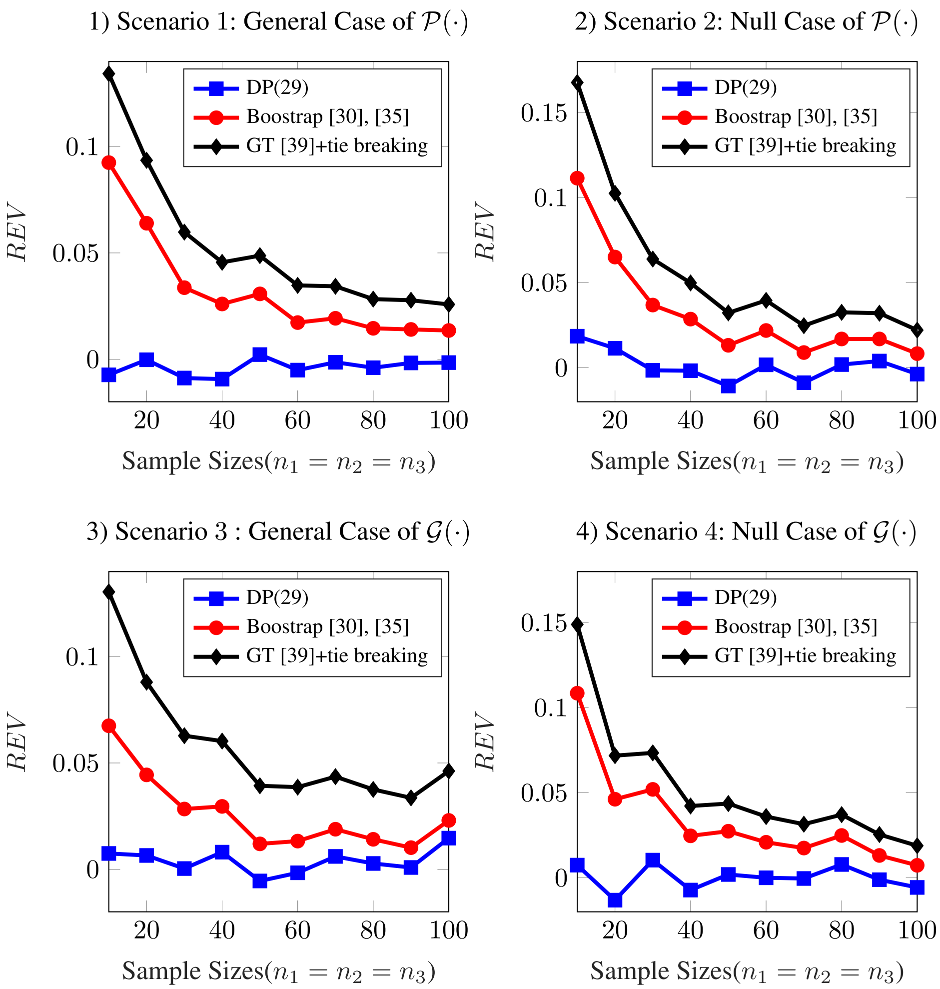


FIGURE 7. Comparison of unbiasedness, in terms of *REV*, among the estimator in (29), the boostrap technique and the combination of work in [39] and tie breaking. 1) Non-null case under Poisson distribution, where X_1, X_2 and X_3 follow $\mathcal{P}(10)$, $\mathcal{P}(20)$ and $\mathcal{P}(30)$, respectively. 2) Null case under Poisson distribution, where X_1, X_2 and X_3 all follow $\mathcal{P}(15)$. 3) Non-null case under Geometric distribution, with pmfs of $\Pr(X_1 = k) = (1-0.2)^{k-1}0.2, k = 1, 2, 3, \ldots, \infty$, $\Pr(X_2 = k) = (1-0.15)^{k-1}0.15, k = 1, 2, 3, \ldots, \infty$ and $\Pr(X_3 = k) = (1-0.1)^{k-1}0.1, k = 1, 2, 3, \ldots, \infty$. 4) Null case under Geometric distribution, where X_1, X_2 and X_3 all follow Geometric distribution with pmfs of $\Pr(X_3 = k) = (1-0.1)^{k-1}0.1, k = 1, 2, 3, \ldots, \infty$.

2) COMPARISON OF COMPUTATIONAL LOADS

To demonstrate the computational efficiency of our new method, we generate three sample sets following Poisson distribution, each being i.i.d., i.e. $\{X_{1i}\}_{i=1}^{n_1} \sim \mathcal{P}(1)$, $\{X_{2j}\}_{j=1}^{n_2} \sim \mathcal{P}(2)$ and $\{X_{1k}\}_{k=1}^{n_3} \sim \mathcal{P}(3)$. Since the parameter λ has little if no effect on the computational speed comparison, they are chosen arbitrary. Figure 6 compares the computational loads among three algorithms based on (3), (30) and the rank-based approach in [45] respectively over the sample sizes $n_1 = n_2 = n_3$ starting from 10 to 100 with an increment of 10. Each of the algorithms runs 10^3 times for stability. According to Figure 6, our method $\hat{\theta}_{\mathrm{DP}}$ runs a little bit faster than $\hat{\theta}_{\mathrm{RB}}$. However, since the difference is too tiny, we consider both linearithmic algorithms are comparable in terms of computational efficiency. And it is no doubt that the version established in (3) is the slowest one since its time complexity is in a cubic order.

B. COMPARISON OF ALGORITHMS FOR ESTIMATION OF VARIANCE OF VUS

In the following we investigate the capacity of our algorithm for computing the variance of VUS in (29), denoted by $\hat{\mathbb{V}}_{\mathrm{DP}}$, with the algorithm based on graph theory (state-of-the-art) proposed by Waegeman *et al.* [39] and boostrap, a widely used technique in many literature [30], [33], [35], denoted by $\hat{\mathbb{V}}_{\mathrm{GT}}$ and $\hat{\mathbb{V}}_{\mathrm{Boostrap}}$ respectively, in the aspects of unbiasedness and computational efficiency. Tie-breaking technique is also employed in $\hat{\mathbb{V}}_{\mathrm{GT}}$, because its original version is only established for continuous inputs. Regarding to the prevalent boostrap technique $\hat{\mathbb{V}}_{\mathrm{Boostrap}}$, we set the number of replications to be 200, which was suggested in [30], [35]. Furthermore, for purpose of fairness and efficiency of the experiment, the fast implementation of (30) is used in each replication of $\hat{\mathbb{V}}_{\mathrm{Boostrap}}$.

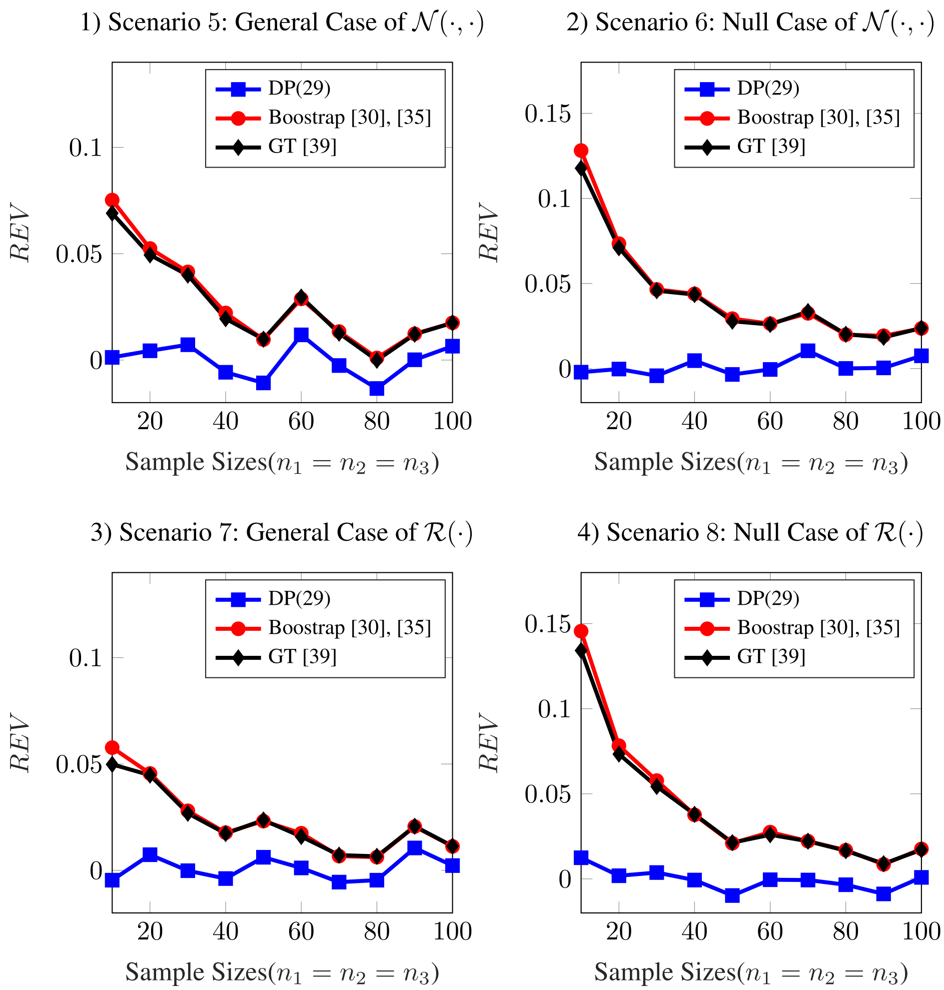


FIGURE 8. Comparison of unbiasedness, in terms of *REV*, among the estimator in (29), the boostrap technique and a method proposed in [39]. 1) Non-null case under Normal distribution, where X_1, X_2 and X_3 follow $\mathcal{N}(0,1)$, $\mathcal{N}(1,1)$ and $\mathcal{N}(2,1)$, respectively. 2) Null case under Normal distribution, where X_1, X_2 and X_3 all follow distribution $\mathcal{N}(0,1)$. 3) Non-null case under Rayleigh distribution. X_1 follows Rayleigh distribution with pdfs $\Pr(X_1 = x) = x\exp(-\frac{x^2}{2})$, X_2 follows Rayleigh distribution with pdfs $\Pr(X_2 = x) = \frac{x}{4}\exp(-\frac{x^2}{8})$ and X_3 follows $\Pr(X_3 = x) = \frac{x}{9}\exp(-\frac{x^2}{18})$, where $x \geq 0$. 4) Null case under Rayleigh distribution, where X_1, X_2 and X_3 all follow Rayleigh distribution with pdfs of $\Pr(X_1 = x) = \Pr(X_2 = x) = \Pr(X_3 = x) = x\exp(-\frac{x^2}{2})$, where $x \geq 0$.

TABLE 2. Parameters of the distributions followed by X_1, X_2 and X_3 for computing the variance of VUS under four scenarios.

Scenario	X_1	X_2	X_3
1	$\mathcal{P}(10)$	$\mathcal{P}(20)$	$\mathcal{P}(30)$
2	$\mathcal{P}(15)$	$\mathcal{P}(15)$	$\mathcal{P}(15)$
3	$\mathcal{G}(0.2)$	$\mathcal{G}(0.15)$	$\mathcal{G}(0.1)$
4	$\mathcal{G}(0.1)$	$\mathcal{G}(0.1)$	$\mathcal{G}(0.1)$
5	$\mathcal{N}(0,1)$	$\mathcal{N}(1,1)$	$\mathcal{N}(2,1)$
6	$\mathcal{N}(0,1)$	$\mathcal{N}(0,1)$	$\mathcal{N}(0,1)$
7	$\mathcal{R}(1)$	$\mathcal{R}(2)$	$\mathcal{R}(3)$
8	$\mathcal{R}(1)$	$\mathcal{R}(1)$	$\mathcal{R}(1)$

1) VERIFICATION OF UNBIASEDNESS

In order to proof that, in contrast to $\hat{\mathbb{V}}_{\mathrm{GT}}$ and $\hat{\mathbb{V}}_{\mathrm{Boostrap}}$, our method in (29) is an unified and unbiased estimator of the variance of VUS, we also generate data that follow continuous distributions, e.g. normal distribution (denoted by $\mathcal{N}(\mu,\sigma^2)$ with mean μ and variance σ^2) and Rayleigh distribution (denoted by $\mathcal{R}(\sigma)$ with parameter σ), for demonstration. The eight scenarios we considered in the experiment are summarized as follows,

Similar to the procedure aforementioned, we defined *Relative Error of Variance* (*REV*) of each method for computing the variance of VUS as

$$REV \triangleq \frac{\mathbb{E}(\hat{\mathbb{V}}_{\xi} - \hat{\mathbb{V}}_{\mathrm{E}})}{\mathbb{V}_{\mathrm{E}}}, \tag{48}$$

where suffix $\xi \in \{\mathrm{DP}, \mathrm{GT}, \mathrm{Boostrap}\}$ and $\hat{\mathbb{V}}_{\mathrm{E}}$ stands for the empirical variance calculated based on Monte Carlo simulation.

Figure 7 depicts the comparison results, in terms of *REV*, with respect to the three methods under four discrete distribution scenarios. The top two graphs are attained when the three

TABLE 3. Quantities needed in the fast algorithm (29).

Quantities	Numbers of events satisfying the relation inside ()	Quantities	Numbers of events satisfying the relation inside ()
$S_0^{(1)}$	$\mathcal{E}(X_3 > X_2 > X_1)$	$S_0^{(2)}$	$\mathcal{E}(X_3 = X_2 > X_1)$
$S_0^{(3)}$	$\mathcal{E}(X_3 > X_2 = X_1)$	$S_0^{(4)}$	$\mathcal{E}(X_3 = X_2 = X_1)$
$S_{12}^{(1)}$	$\mathcal{E}(X_3 > X_3' > X_2 > X_1)$ or $\mathcal{E}(X_3' > X_3 > X_2 > X_1)$	$S_{12}^{(2)}$	$\mathcal{E}(X_3 = X_3' > X_2 > X_1)$
$S_{12}^{(3)}$	$\mathcal{E}(X_3 > X_3' = X_2 > X_1)$ or $\mathcal{E}(X_3' > X_3 = X_2 > X_1)$	$S_{12}^{(4)}$	$\mathcal{E}(X_3 > X_3' > X_2 = X_1)$ or $\mathcal{E}(X_3' > X_3 > X_2 = X_1)$
$S_{12}^{(5)}$	$\mathcal{E}(X_3 = X_3' > X_2 = X_1)$	$S_{12}^{(6)}$	$\mathcal{E}(X_3 = X_3' = X_2 > X_1)$
$S_{12}^{(7)}$	$\mathcal{E}(X_3 > X_3' = X_2 = X_1)$ or $\mathcal{E}(X_3' > X_3 = X_2 = X_1)$	$S_{12}^{(8)}$	$\mathcal{E}(X_3 = X_3' = X_2 = X_1)$
$S_{13}^{(1)}$	$\mathcal{E}(X_3 > X_2 > X_2' > X_1)$ or $\mathcal{E}(X_3 > X_2' > X_2 > X_1)$	$S_{13}^{(2)}$	$\mathcal{E}(X_3 > X_2 = X_2' > X_1)$
$S_{13}^{(3)}$	$\mathcal{E}(X_3 = X_2 > X_2' > X_1)$ or $\mathcal{E}(X_3 = X_2' > X_2 > X_1)$	$S_{13}^{(4)}$	$\mathcal{E}(X_3 > X_2 > X_2' = X_1)$ or $\mathcal{E}(X_3 > X_2' > X_2 = X_1)$
$S_{13}^{(5)}$	$\mathcal{E}(X_3 = X_2 > X_2' = X_1)$ or $\mathcal{E}(X_3 = X_2' > X_2 = X_1)$	$S_{13}^{(6)}$	$\mathcal{E}(X_3 > X_2 = X_2' = X_1)$
$S_{13}^{(7)}$	$\mathcal{E}(X_3 = X_2 = X_2' > X_1)$	$S_{13}^{(8)}$	$\mathcal{E}(X_3 = X_2 = X_2' = X_1)$
$S_{23}^{(1)}$	$\mathcal{E}(X_3 > X_2 > X_1 > X_1')$ or $\mathcal{E}(X_3 > X_2' > X_1' > X_1)$	$S_{23}^{(2)}$	$\mathcal{E}(X_3 > X_2 > X_1 = X_1')$
$S_{23}^{(3)}$	$\mathcal{E}(X_3 > X_2 = X_1 > X_1')$ or $\mathcal{E}(X_3 > X_2 = X_1' > X_1)$	$S_{23}^{(4)}$	$\mathcal{E}(X_3 = X_2 > X_1 > X_1')$ or $\mathcal{E}(X_3 = X_2 > X_1' > X_1)$
$S_{23}^{(5)}$	$\mathcal{E}(X_3 > X_2 = X_1 = X_1')$	$S_{23}^{(6)}$	$\mathcal{E}(X_3 = X_2 > X_1 = X_1')$
$S_{23}^{(7)}$	$\mathcal{E}(X_3 = X_2 = X_1 > X_1')$ or $\mathcal{E}(X_3 = X_2 = X_1' > X_1)$	$S_{23}^{(8)}$	$\mathcal{E}(X_3 = X_2 = X_1 = X_1')$
$S_1^{(1)}$	$\mathcal{E}(X_3 > X_3' > X_2 > X_2' > X_1)$ or $\mathcal{E}(X_3' > X_3 > X_2 > X_2' > X_1)$ $\mathcal{E}(X_3 > X_3' > X_2' > X_2 > X_1)$ or $\mathcal{E}(X_3' > X_3 > X_2' > X_2 > X_1)$	$S_1^{(2)}$	$\mathcal{E}(X_3 > X_3' = X_2 > X_2' > X_1)$ or $\mathcal{E}(X_3' > X_3 = X_2' > X_2 > X_1)$ $\mathcal{E}(X_3 > X_3' = X_2' > X_2 > X_1)$ or $\mathcal{E}(X_3' > X_3 = X_2 > X_2' > X_1)$
$S_1^{(3)}$	$\mathcal{E}(X_3 > X_2 > X_3' > X_2' > X_1)$ or $\mathcal{E}(X_3' > X_2' > X_3 > X_2 > X_1)$	$S_1^{(4)}$	$\mathcal{E}(X_3 = X_3' > X_2 > X_2' > X_1)$ or $\mathcal{E}(X_3 = X_3' > X_2' > X_2 > X_1)$
$S_1^{(5)}$	$\mathcal{E}(X_3 > X_3' > X_2 = X_2' > X_1)$ or $\mathcal{E}(X_3' > X_3 > X_2' = X_2 > X_1)$	$S_1^{(6)}$	$\mathcal{E}(X_3 > X_3' > X_2 > X_2' = X_1)$ or $\mathcal{E}(X_3' > X_3 > X_2 > X_2' = X_1)$ $\mathcal{E}(X_3 > X_3' > X_2' > X_2 = X_1)$ or $\mathcal{E}(X_3' > X_3 > X_2' > X_2 = X_1)$
$S_1^{(7)}$	$\mathcal{E}(X_3 > X_3' = X_2 > X_2' = X_1)$ or $\mathcal{E}(X_3' > X_3 = X_2' > X_2 = X_1)$ $\mathcal{E}(X_3 > X_3' = X_2' > X_2 = X_1)$ or $\mathcal{E}(X_3' > X_3 = X_2 > X_2' = X_1)$	$S_1^{(8)}$	$\mathcal{E}(X_3 = X_3' > X_2 = X_2' > X_1)$
$S_1^{(9)}$	$\mathcal{E}(X_3 = X_2 > X_3' > X_2' > X_1)$ or $\mathcal{E}(X_3' = X_2' > X_3 > X_2 > X_1)$	$S_1^{(10)}$	$\mathcal{E}(X_3 > X_2 > X_3' = X_2' > X_1)$ or $\mathcal{E}(X_3' > X_2' > X_3 = X_2 > X_1)$
$S_1^{(11)}$	$\mathcal{E}(X_3 = X_3' = X_2 > X_2' > X_1)$ or $\mathcal{E}(X_3 = X_3' = X_2' > X_2 > X_1)$	$S_1^{(12)}$	$\mathcal{E}(X_3 > X_3' = X_2 = X_2' > X_1)$ or $\mathcal{E}(X_3' > X_3' = X_2 = X_2' > X_1)$
$S_1^{(13)}$	$\mathcal{E}(X_3 > X_2 > X_3' > X_2' = X_1)$ or $\mathcal{E}(X_3' > X_2' > X_3 > X_2 = X_1)$	$S_1^{(14)}$	$\mathcal{E}(X_3 = X_3' > X_2 > X_2' = X_1)$ or $\mathcal{E}(X_3 = X_3' > X_2' > X_2 = X_1)$
$S_1^{(15)}$	$\mathcal{E}(X_3 = X_2 > X_3' = X_2' > X_1)$ or $\mathcal{E}(X_3' = X_2' > X_3 = X_2 > X_1)$	$S_1^{(16)}$	$\mathcal{E}(X_3 = X_2 > X_3' > X_2' = X_1)$ or $\mathcal{E}(X_3' = X_2' > X_3 > X_2 = X_1)$
$S_1^{(17)}$	$\mathcal{E}(X_3 = X_3' = X_2 > X_2' = X_1)$ or $\mathcal{E}(X_3 = X_3' = X_2' > X_2 = X_1)$	$S_1^{(18)}$	$\mathcal{E}(X_3 > X_3' > X_2 = X_2' = X_1)$ or $\mathcal{E}(X_3' > X_3 > X_2 = X_2' = X_1)$
$S_1^{(19)}$	$\mathcal{E}(X_3 > X_2 > X_3' = X_2' = X_1)$ or $\mathcal{E}(X_3' > X_2' > X_3 = X_2 = X_1)$	$S_1^{(20)}$	$\mathcal{E}(X_3 = X_2 = X_3' = X_2' > X_1)$
$S_1^{(21)}$	$\mathcal{E}(X_3 = X_3' > X_2 = X_2' = X_1)$	$S_1^{(22)}$	$\mathcal{E}(X_3 = X_2 > X_3' = X_2' = X_1)$ or $\mathcal{E}(X_3' = X_2' > X_3 = X_2 = X_1)$
$S_1^{(23)}$	$\mathcal{E}(X_3 > X_3' = X_2 = X_2' = X_1)$ or $\mathcal{E}(X_3' > X_3 = X_2 = X_2' = X_1)$	$S_1^{(24)}$	$\mathcal{E}(X_3 = X_3' = X_2 = X_2' = X_1)$
$S_2^{(1)}$	$\mathcal{E}(X_3 > X_3' > X_2 > X_1 > X_1')$ or $\mathcal{E}(X_3' > X_3 > X_2 > X_1 > X_1')$ $\mathcal{E}(X_3 > X_3' > X_2 > X_1' > X_1)$ or $\mathcal{E}(X_3' > X_3 > X_2 > X_1' > X_1)$	$S_2^{(2)}$	$\mathcal{E}(X_3 = X_3' > X_2 > X_1 > X_1')$ or $\mathcal{E}(X_3 = X_3' > X_2 > X_1' > X_1)$
$S_2^{(3)}$	$\mathcal{E}(X_3 > X_3' > X_2 > X_1 = X_1')$ or $\mathcal{E}(X_3' > X_3 > X_2 > X_1 = X_1')$	$S_2^{(4)}$	$\mathcal{E}(X_3 > X_3' = X_2 > X_1 > X_1')$ or $\mathcal{E}(X_3' > X_3 = X_2 > X_1 > X_1')$ $\mathcal{E}(X_3 > X_3' = X_2 > X_1' > X_1)$ or $\mathcal{E}(X_3' > X_3 = X_2 > X_1' > X_1)$
$S_2^{(5)}$	$\mathcal{E}(X_3 > X_3' > X_2 = X_1 > X_1')$ or $\mathcal{E}(X_3' > X_3 > X_2 = X_1 > X_1')$ $\mathcal{E}(X_3 > X_3' > X_2 = X_1' > X_1)$ or $\mathcal{E}(X_3' > X_3 > X_2 = X_1' > X_1)$	$S_2^{(6)}$	$\mathcal{E}(X_3 = X_3' > X_2 > X_1 = X_1')$
$S_2^{(7)}$	$\mathcal{E}(X_3 > X_3' = X_2 > X_1 = X_1')$ or $\mathcal{E}(X_3' > X_3 = X_2 > X_1 = X_1')$	$S_2^{(8)}$	$\mathcal{E}(X_3 = X_3' > X_2 = X_1 > X_1')$ or $\mathcal{E}(X_3 = X_3' > X_2 = X_1' > X_1)$
$S_2^{(9)}$	$\mathcal{E}(X_3 > X_3' = X_2 = X_1 > X_1')$ or $\mathcal{E}(X_3' > X_3 = X_2 = X_1 > X_1')$ $\mathcal{E}(X_3 > X_3' = X_2 = X_1' > X_1)$ or $\mathcal{E}(X_3' > X_3 = X_2 = X_1' > X_1)$	$S_2^{(10)}$	$\mathcal{E}(X_3 = X_3' = X_2 > X_1 > X_1')$ or $\mathcal{E}(X_3 = X_3' = X_2 > X_1' > X_1)$
$S_2^{(11)}$	$\mathcal{E}(X_3 > X_3' > X_2 = X_1 = X_1')$ or $\mathcal{E}(X_3' > X_3 > X_2 = X_1 = X_1')$	$S_2^{(12)}$	$\mathcal{E}(X_3 = X_3' = X_2 > X_1 = X_1')$
$S_2^{(13)}$	$\mathcal{E}(X_3 = X_3' > X_2 = X_1 = X_1')$	$S_2^{(14)}$	$\mathcal{E}(X_3 = X_3' = X_2 = X_1 > X_1')$ or $\mathcal{E}(X_3 = X_3' = X_2 = X_1' > X_1)$
$S_2^{(15)}$	$\mathcal{E}(X_3 > X_3' = X_2 = X_1 = X_1')$ or $\mathcal{E}(X_3' > X_3 = X_2 = X_1 > X_1')$	$S_2^{(16)}$	$\mathcal{E}(X_3 = X_3' = X_2 = X_1 = X_1')$
$S_3^{(1)}$	$\mathcal{E}(X_3 > X_2 > X_2' > X_1 > X_1')$ or $\mathcal{E}(X_3 > X_2' > X_2 > X_1 > X_1')$ $\mathcal{E}(X_3 > X_2 > X_2' > X_1' > X_1)$ or $\mathcal{E}(X_3 > X_2' > X_2 > X_1' > X_1)$	$S_3^{(2)}$	$\mathcal{E}(X_3 > X_2 > X_2' = X_1 > X_1')$ or $\mathcal{E}(X_3 > X_2' > X_2 = X_1 > X_1')$ $\mathcal{E}(X_3 > X_2 > X_2' = X_1' > X_1)$ or $\mathcal{E}(X_3 > X_2' > X_2 = X_1' > X_1)$
$S_3^{(3)}$	$\mathcal{E}(X_3 > X_2 > X_1 > X_2' > X_1')$ or $\mathcal{E}(X_3 > X_2' > X_1' > X_2 > X_1)$	$S_3^{(4)}$	$\mathcal{E}(X_3 > X_2 > X_2' > X_1 = X_1')$ or $\mathcal{E}(X_3 > X_2' > X_2 > X_1 = X_1')$
$S_3^{(5)}$	$\mathcal{E}(X_3 > X_2 = X_2' > X_1 > X_1')$ or $\mathcal{E}(X_3 > X_2 = X_2' > X_1' > X_1)$	$S_3^{(6)}$	$\mathcal{E}(X_3 = X_2 > X_2' > X_1 > X_1')$ or $\mathcal{E}(X_3 = X_2' > X_2 > X_1 > X_1')$ $\mathcal{E}(X_3 = X_2 > X_2' > X_1' > X_1)$ or $\mathcal{E}(X_3 = X_2' > X_2 > X_1' > X_1)$
$S_3^{(7)}$	$\mathcal{E}(X_3 = X_2 > X_2' = X_1 > X_1')$ or $\mathcal{E}(X_3 = X_2' > X_2 = X_1 > X_1')$ $\mathcal{E}(X_3 = X_2 > X_2' = X_1' > X_1)$ or $\mathcal{E}(X_3 = X_2' > X_2 = X_1' > X_1)$	$S_3^{(8)}$	$\mathcal{E}(X_3 > X_2 = X_2' > X_1 = X_1')$
$S_3^{(9)}$	$\mathcal{E}(X_3 = X_2 > X_1 > X_2' > X_1')$ or $\mathcal{E}(X_3 = X_2' > X_1' > X_2 > X_1)$	$S_3^{(10)}$	$\mathcal{E}(X_3 = X_2 > X_2' > X_1 = X_1')$ or $\mathcal{E}(X_3 = X_2' > X_2 > X_1 = X_1')$
$S_3^{(11)}$	$\mathcal{E}(X_3 > X_2 = X_1 > X_2' > X_1')$ or $\mathcal{E}(X_3 > X_2' = X_1' > X_2 > X_{)}$	$S_3^{(12)}$	$\mathcal{E}(X_3 > X_2 > X_1 > X_2' = X_1')$ or $\mathcal{E}(X_3 > X_2' > X_1' > X_2 = X_1)$
$S_3^{(13)}$	$\mathcal{E}(X_3 > X_2 = X_2' = X_1 > X_1')$ or $\mathcal{E}(X_3 > X_2 = X_2' = X_1' > X_1)$	$S_3^{(14)}$	$\mathcal{E}(X_3 > X_2 > X_2' = X_1 = X_1')$ or $\mathcal{E}(X_3 > X_2' > X_2 = X_1 = X_1')$
$S_3^{(15)}$	$\mathcal{E}(X_3 = X_2 = X_2' > X_1 > X_1')$ or $\mathcal{E}(X_3 = X_2 = X_2' > X_1' > X_1)$	$S_3^{(16)}$	$\mathcal{E}(X_3 = X_2 > X_1 > X_2' = X_1')$ or $\mathcal{E}(X_3 = X_2' > X_1' > X_2 = X_1)$
$S_3^{(17)}$	$\mathcal{E}(X_3 = X_2 > X_2' = X_1 = X_1')$ or $\mathcal{E}(X_3 = X_2' > X_2 = X_1 = X_1')$	$S_3^{(18)}$	$\mathcal{E}(X_3 > X_2 = X_1 > X_2' = X_1')$ or $\mathcal{E}(X_3 > X_2 = X_1' > X_2 = X_1)$
$S_3^{(19)}$	$\mathcal{E}(X_3 = X_2 = X_1 > X_2' > X_1')$ or $\mathcal{E}(X_3 = X_2' = X_1' > X_2 > X_1)$	$S_3^{(20)}$	$\mathcal{E}(X_3 = X_2 = X_2' > X_1 = X_1')$
$S_3^{(21)}$	$\mathcal{E}(X_3 > X_2 = X_2' = X_1 = X_1')$	$S_3^{(22)}$	$\mathcal{E}(X_3 = X_2 = X_2' = X_1 > X_1')$ or $\mathcal{E}(X_3 = X_2 = X_2' = X_1' > X_1)$
$S_3^{(23)}$	$\mathcal{E}(X_3 = X_2 = X_1 > X_2' = X_1')$ or $\mathcal{E}(X_3 = X_2' = X_1' > X_2 = X_1)$	$S_3^{(24)}$	$\mathcal{E}(X_3 = X_2 = X_2' = X_1 = X_1')$

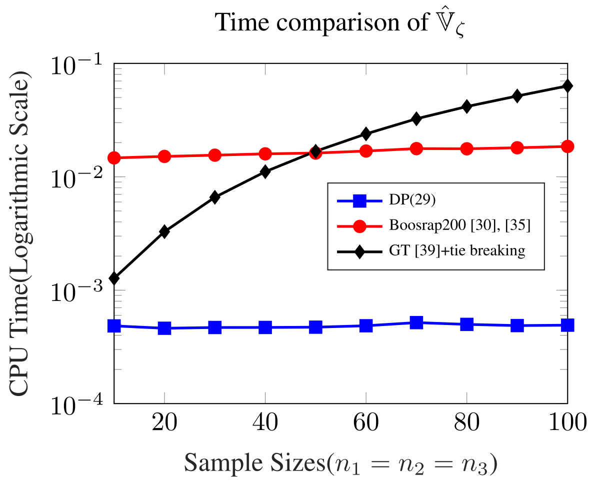


FIGURE 9. Comparative results of CPU time among three algorithms based on (29),the boostrap involves 200 times of replications and the method proposed by Waegeman *et al.* in [39], respectively. For simplicity, the sample sizes of X_1, X_2 and X_3 are set to be equal. A log scale is used for better visualization.

sample sets follow Poisson distribution, and the remaining graphs are results corresponding to Geometric distribution. It is obvious that our dynamic programming based method $\hat{\mathbb{V}}_{\text{DP}}$ outperforms the other two, in the sense that $\hat{\mathbb{V}}_{\text{DP}}$'s *REV*s are all approximately zero, which confirms the unbiasedness of our algorithm. Besides, all the curves of $\hat{\mathbb{V}}_{\text{GT}}$ and $\hat{\mathbb{V}}_{\text{Boostrap}}$ deviate from zero, especially when the sample sizes are small. As the sample sizes increase, the curves of both $\hat{\mathbb{V}}_{\text{GT}}$ and $\hat{\mathbb{V}}_{\text{Boostrap}}$ tend to convergence. The latter's *REV* gradually decreases towards zero, which validates that $\hat{\mathbb{V}}_{\text{Boostrap}}$ is only an asymptotically unbiased estimator of the variance of VUS; while though convergence, the large gap between *REV*s of $\hat{\mathbb{V}}_{\text{GT}}$ and line $REV = 0$ announces that $\hat{\mathbb{V}}_{\text{GT}}$ with tie breaking is not a rigorous solution for computing the variance of VUS when there exist ties, since the accuracy of variance must be explicitly controlled when estimating confident interval or performing hypothesis test.

Figure 8 is an experimental result when the class distributions are continuous. With respect to the comparison of unbiasedness, a conclusion similar to the discrete case in Figure 7 above can also be drawn. Better than the other two methods, the proposed algorithm based on dynamic programming remains unbiased. Interestingly, the performance of $\mathbb{V}_{\text{GT}}$ is improved and close to that of $\mathbb{V}_{\text{Boostrap}}$, due to the removal of tie breaking. However, both of them are only asymptotic unbiased estimators.

2) COMPARISON OF COMPUTATIONAL LOADS

Finally, we examine the computational speed of $\hat{\mathbb{V}}_\xi$. Since the generation of sample has little effect on the calculation time, we remain using the data produced in section V-A2 for comparison. As shown in Figure 9, it is observed that, our method $\hat{\mathbb{V}}_{\text{DP}}$ based on (29) always outperforms the other two approaches, running around 100 times faster than the boostrap approach $\hat{\mathbb{V}}_{\text{Boostrap}}$ that is also of linearithmic time order $\mathcal{O}[(n_1+n_2+n_3)\log(n_1+n_2+n_3)]$ as well. When the sample sizes are small (less than 50), $\hat{\mathbb{V}}_{\text{Boostrap}}$ is more time consuming than $\hat{\mathbb{V}}_{\text{GT}}$ with tie breaking. However, as described in [39], *REV*s of $\hat{\mathbb{V}}_{\text{GT}}$ with tie breaking soars up dramatically with the increase of sample sizes, since its time complexity is in a quadratic order.

VI. CONCLUSION

In this paper, we proposed an efficient and unified algorithm, based on dynamic programming for computing the mean and variance of VUS with continuous or non-continuous measurements. Theoretical and experimental derivations suggest that (a) it can act as an unbiased estimator for the mean and variance of VUS, (b) it is simultaneously applicable for continuous and discrete inputs, and (c) its time complexity is of linearithmic order, comparable with the state-of-the-art methods for computing the mean of VUS, and much lower than the state-of-the-art methods proposed by Waegement *et al.* in the case of computing the variance of VUS. Besides these advantages, the structure of this algorithm can be easily extended to the multi-class cases [47]. 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.

APPENDIX. TABLE OF *S*

The basic events employed in Theorem 1 are summarized in Table 3.

APPENDIX. PROOF OF THEOREM 1

Let $\mathcal{I}(\cdot)$ be a indicator function that equals to unity(zeros) if the statement in bracket is *True*(*False*) and $\mathcal{E}(\cdot)$ defined in Table 3 be the number of times that events in the bracket are established, respectively. Based on (1)(3)(4) and the indicator function, we get

$$\begin{aligned}
\hat{\theta} &= \frac{1}{n_1n_2n_3}\Bigg[\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}\mathcal{I}(X_{3k} > X_{2j} > X_{1i}) \\
&\quad +\frac{1}{2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}\mathcal{I}(X_{3k} = X_{2j} > X_{1i}) \\
&\quad +\frac{1}{2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}\mathcal{I}(X_{3k} > X_{2j} = X_{1i}) \\
&\quad +\frac{1}{6}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}\mathcal{I}(X_{3k} = X_{2j} = X_{1i})\Bigg] \\
&= \frac{1}{n_1n_2n_3}\Bigg[\mathcal{E}(X_3 > X_2 > X_1) + \frac{1}{2}\mathcal{E}(X_3 = X_2 > X_1) \\
&\quad +\frac{1}{2}\mathcal{E}(X_3 > X_2 = X_1) + \frac{1}{6}\mathcal{E}(X_3 = X_2 = X_1)\Bigg] \\
&= \frac{1}{n_1n_2n_3}(S_0^{(1)} + \frac{1}{2}S_0^{(2)} + \frac{1}{2}S_0^{(3)} + \frac{1}{6}S_0^{(4)}),
\end{aligned}$$

which is (30). From (21) and (22), we have $\hat{Q}_{23}$, as shown at the bottom of this page and $\hat{Q}_1$, as shown at the bottom of the next page which confirm the results of (34) and (35), respectively. In a similar manner, the rest $\hat{Q}$-terms in (29) can also be obtained by referring to Table 3. This completes the proof.

$$
\begin{aligned}
\hat{Q}_{23} =& \frac{1}{n_1^{[2]} n_2 n_3} \Bigg[\sum_{i=1}^{n_1} \sum_{i'=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} \mathcal{H}(X_{3k}, X_{2j}, X_{1i}) \cdot \mathcal{H}(X_{3k}, X_{2j}, X_{1i'}) \\
& - \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} \mathcal{H}(X_{3k}, X_{2j}, X_{1i})^2 \Bigg] \\
=& \frac{1}{n_1^{[2]} n_2 n_3} \Bigg[\underbrace{\sum_{i=1}^{n_1} \sum_{i'=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} \mathcal{I}(X_{3k} > X_{2j} > X_{1i}) \cdot \mathcal{I}(X_{3k} > X_{2j} > X_{1i'})}_{\mathcal{E}(X_3>X_2>X_1>X_1')+\mathcal{E}(X_3>X_2>X_1'>X_1)+\mathcal{E}(X_3>X_2>X_1=X_1') \Rightarrow 2S_{23}^{(1)}+S_{23}^{(2)}} \\
& + \frac{1}{2} \underbrace{\sum_{i=1}^{n_1} \sum_{i'=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} \mathcal{I}(X_{3k} > X_{2j} > X_{1i}) \cdot \mathcal{I}(X_{3k} > X_{2j} = X_{1i'})}_{\mathcal{E}(X_3>X_2=X_1'>X_1) \Rightarrow S_{23}^{(3)}} \\
& + \frac{1}{4} \underbrace{\sum_{i=1}^{n_1} \sum_{i'=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} \mathcal{I}(X_{3k} = X_{2j} > X_{1i}) \cdot \mathcal{I}(X_{3k} = X_{2j} > X_{1i'})}_{\mathcal{E}(X_3=X_2>X_1>X_1')+\mathcal{E}(X_3=X_2>X_1'>X_1)+\mathcal{E}(X_3=X_2>X_1=X_1') \Rightarrow 2S_{23}^{(4)}+S_{23}^{(6)}} \\
& + \frac{1}{12} \underbrace{\sum_{i=1}^{n_1} \sum_{i'=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} \mathcal{I}(X_{3k} = X_{2j} > X_{1i}) \cdot \mathcal{I}(X_{3k} = X_{2j} = X_{1i'})}_{\mathcal{E}(X_3=X_2=X_1'>X_1) \Rightarrow S_{23}^{(7)}} \\
& + \frac{1}{2} \underbrace{\sum_{i=1}^{n_1} \sum_{i'=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} \mathcal{I}(X_{3k} > X_{2j} = X_{1i}) \cdot \mathcal{I}(X_{3k} > X_{2j} > X_{1i'})}_{\mathcal{E}(X_3>X_2=X_1>X_1') \Rightarrow S_{23}^{(3)}} \\
& + \frac{1}{4} \underbrace{\sum_{i=1}^{n_1} \sum_{i'=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} \mathcal{I}(X_{3k} > X_{2j} = X_{1i}) \cdot \mathcal{I}(X_{3k} > X_{2j} = X_{1i'})}_{\mathcal{E}(X_3>X_2=X_1=X_1') \Rightarrow S_{23}^{(5)}} \\
& + \frac{1}{12} \underbrace{\sum_{i=1}^{n_1} \sum_{i'=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} \mathcal{I}(X_{3k} = X_{2j} = X_{1i}) \cdot \mathcal{I}(X_{3k} = X_{2j} > X_{1i'})}_{\mathcal{E}(X_3=X_2=X_1>X_1') \Rightarrow S_{23}^{(7)}} \\
& + \frac{1}{36} \underbrace{\sum_{i=1}^{n_1} \sum_{i'=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} \mathcal{I}(X_{3k} = X_{2j} = X_{1i}) \cdot \mathcal{I}(X_{3k} = X_{2j} = X_{1i'})}_{\mathcal{E}(X_3=X_2=X_1=X_1') \Rightarrow S_{23}^{(8)}} \\
& - n_1 n_2 n_3 \hat{Q}_0 \Bigg] \\
=& \frac{1}{n_1^{[2]} n_2 n_3} \Big(2S_{23}^{(1)} + S_{23}^{(2)} + S_{23}^{(3)} + \frac{1}{2} S_{23}^{(4)} + \frac{1}{4} S_{23}^{(5)} + \frac{1}{4} S_{23}^{(6)} + \frac{1}{6} S_{23}^{(7)} + \frac{1}{36} S_{23}^{(8)} \\
& - S_0^{(1)} - \frac{1}{4} S_0^{(2)} - \frac{1}{4} S_0^{(3)} - \frac{1}{36} S_0^{(4)} \Big)
\end{aligned}
$$

$$
\begin{aligned}
\hat{Q}_1 &= \frac{1}{n_1 n_2^{[2]} n_3^{[2]}} \Bigg[\sum_{i=1}^{n_1} \sum_{j \neq j'=1}^{n_2} \sum^{n_2} \sum_{k \neq k'=1}^{n_3} \sum^{n_3} \mathcal{H}(X_{3k}, X_{2j}, X_{1i}) \cdot \mathcal{H}(X_{3k'}, X_{2j'}, X_{1i}) \Bigg] \\
&= \frac{1}{n_1 n_2^{[2]} n_3^{[2]}} \Bigg[\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{j'=1}^{n_2} \sum_{k=1}^{n_3} \sum_{k'=1}^{n_3} \mathcal{H}(X_{3k}, X_{2j}, X_{1i}) \cdot \mathcal{H}(X_{3k'}, X_{2j'}, X_{1i}) \\
&\quad - \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k \neq k'=1}^{n_3} \sum^{n_3} \mathcal{H}(X_{3k}, X_{2j}, X_{1i}) \cdot \mathcal{H}(X_{3k'}, X_{2j}, X_{1i}) \\
&\quad - \sum_{i=1}^{n_1} \sum_{j \neq j'=1}^{n_2} \sum^{n_2} \sum_{k=1}^{n_3} \mathcal{H}(X_{3k}, X_{2j}, X_{1i}) \cdot \mathcal{H}(X_{3k}, X_{2j'}, X_{1i}) \\
&\quad - \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} \mathcal{H}(X_{3k}, X_{2j}, X_{1i})^2 \Bigg] \\
&= \frac{1}{n_1 n_2^{[2]} n_3^{[2]}} \Bigg[\underbrace{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{j'=1}^{n_2} \sum_{k=1}^{n_3} \sum_{k'=1}^{n_3} \mathcal{I}(X_{3k} > X_{2j} > X_{1i}) \cdot \mathcal{I}(X_{3k'} > X_{2j'} > X_{1i})}_{4S_1^{(1)} + 2S_1^{(2)} + 2S_1^{(3)} + 2S_1^{(4)} + 2S_1^{(5)} + S_1^{(8)}} \\
&\quad + \frac{1}{2} \underbrace{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{j'=1}^{n_2} \sum_{k=1}^{n_3} \sum_{k'=1}^{n_3} \mathcal{I}(X_{3k} > X_{2j} > X_{1i}) \cdot \mathcal{I}(X_{3k'} = X_{2j'} > X_{1i})}_{S_1^{(2)} + S_1^{(8)} + S_1^{(10)} + S_1^{(11)} + S_1^{(12)}} \\
&\quad + \frac{1}{2} \underbrace{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{j'=1}^{n_2} \sum_{k=1}^{n_3} \sum_{k'=1}^{n_3} \mathcal{I}(X_{3k} > X_{2j} > X_{1i}) \cdot \mathcal{I}(X_{3k'} > X_{2j'} = X_{1i})}_{2S_1^{(6)} + S_1^{(7)} + S_1^{(13)} + S_1^{(14)}} \\
&\quad + \frac{1}{6} \underbrace{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{j'=1}^{n_2} \sum_{k=1}^{n_3} \sum_{k'=1}^{n_3} \mathcal{I}(X_{3k} > X_{2j} > X_{1i}) \cdot \mathcal{I}(X_{3k'} = X_{2j'} = X_{1i})}_{S_1^{(19)}} \\
&\quad + \frac{1}{2} \underbrace{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{j'=1}^{n_2} \sum_{k=1}^{n_3} \sum_{k'=1}^{n_3} \mathcal{I}(X_{3k} = X_{2j} > X_{1i}) \cdot \mathcal{I}(X_{3k'} > X_{2j'} > X_{1i})}_{S_1^{(2)} + S_1^{(9)} + S_1^{(10)} + S_1^{(11)} + S_1^{(12)}} \\
&\quad + \frac{1}{4} \underbrace{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{j'=1}^{n_2} \sum_{k=1}^{n_3} \sum_{k'=1}^{n_3} \mathcal{I}(X_{3k} = X_{2j} > X_{1i}) \cdot \mathcal{I}(X_{3k'} = X_{2j'} > X_{1i})}_{2S_1^{(15)} + S_1^{(20)}} \\
&\quad + \frac{1}{4} \underbrace{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{j'=1}^{n_2} \sum_{k=1}^{n_3} \sum_{k'=1}^{n_3} \mathcal{I}(X_{3k} = X_{2j} > X_{1i}) \cdot \mathcal{I}(X_{3k'} > X_{2j'} = X_{1i})}_{S_1^{(7)} + S_1^{(16)} + S_1^{(17)}} \\
&\quad + \frac{1}{12} \underbrace{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{j'=1}^{n_2} \sum_{k=1}^{n_3} \sum_{k'=1}^{n_3} \mathcal{I}(X_{3k} = X_{2j} > X_{1i}) \cdot \mathcal{I}(X_{3k'} = X_{2j'} = X_{1i})}_{S_1^{(22)}} \\
&\quad + \frac{1}{2} \underbrace{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{j'=1}^{n_2} \sum_{k=1}^{n_3} \sum_{k'=1}^{n_3} \mathcal{I}(X_{3k} > X_{2j} = X_{1i}) \cdot \mathcal{I}(X_{3k'} > X_{2j'} > X_{1i})}_{2S_1^{(6)} + S_1^{(7)} + S_1^{(13)} + S_1^{(14)}}
\end{aligned}
$$

$$+\underbrace{\frac{1}{4}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{j'=1}^{n_2}\sum_{k=1}^{n_3}\sum_{k'=1}^{n_3}\mathcal{I}(X_{3k} > X_{2j} = X_{1i})\cdot\mathcal{I}(X_{3k'} = X_{2j'} > X_{1i})}_{S_1^{(7)}+S_1^{(16)}+S_1^{(17)}}$$

$$+\underbrace{\frac{1}{4}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{j'=1}^{n_2}\sum_{k=1}^{n_3}\sum_{k'=1}^{n_3}\mathcal{I}(X_{3k} > X_{2j} = X_{1i})\cdot\mathcal{I}(X_{3k'} = X_{2j'} > X_{1i})}_{S_1^{(7)}+S_1^{(16)}+S_1^{(17)}}$$

$$+\underbrace{\frac{1}{4}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{j'=1}^{n_2}\sum_{k=1}^{n_3}\sum_{k'=1}^{n_3}\mathcal{I}(X_{3k} > X_{2j} = X_{1i})\cdot\mathcal{I}(X_{3k'} > X_{2j'} = X_{1i})}_{2S_1^{(18)}+S_1^{(21)}}$$

$$+\underbrace{\frac{1}{12}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{j'=1}^{n_2}\sum_{k=1}^{n_3}\sum_{k'=1}^{n_3}\mathcal{I}(X_{3k} > X_{2j} = X_{1i})\cdot\mathcal{I}(X_{3k'} = X_{2j'} = X_{1i})}_{S_1^{(23)}}$$

$$+\underbrace{\frac{1}{6}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{j'=1}^{n_2}\sum_{k=1}^{n_3}\sum_{k'=1}^{n_3}\mathcal{I}(X_{3k} = X_{2j} = X_{1i})\cdot\mathcal{I}(X_{3k'} > X_{2j'} > X_{1i})}_{S_1^{(19)}}$$

$$+\underbrace{\frac{1}{12}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{j'=1}^{n_2}\sum_{k=1}^{n_3}\sum_{k'=1}^{n_3}\mathcal{I}(X_{3k} = X_{2j} = X_{1i})\cdot\mathcal{I}(X_{3k'} = X_{2j'} > X_{1i})}_{S_1^{(22)}}$$

$$+\underbrace{\frac{1}{12}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{j'=1}^{n_2}\sum_{k=1}^{n_3}\sum_{k'=1}^{n_3}\mathcal{I}(X_{3k} = X_{2j} = X_{1i})\cdot\mathcal{I}(X_{3k'} > X_{2j'} = X_{1i})}_{S_1^{(23)}}$$

$$+\underbrace{\frac{1}{36}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{j'=1}^{n_2}\sum_{k=1}^{n_3}\sum_{k'=1}^{n_3}\mathcal{I}(X_{3k} = X_{2j} = X_{1i})\cdot\mathcal{I}(X_{3k'} = X_{2j'} = X_{1i})}_{S_1^{(24)}}$$

$$\left.-n_1n_2n_3^{[2]}\hat{\mathcal{Q}}_{12} - n_1n_2^{[2]}n_3\hat{\mathcal{Q}}_{13} - n_1n_2n_3\hat{\mathcal{Q}}_0\right]$$

$$= \frac{1}{n_1n_2^{[2]}n_3^{[2]}}(4S_1^{(1)} + 3S_1^{(2)} + 2S_1^{(3)} + 2S_1^{(4)} + 2S_1^{(5)} + 2S_1^{(6)} + \frac{3}{2}S_1^{(7)} + S_1^{(8)} + S_1^{(9)} + S_1^{(10)}$$

$$+S_1^{(11)} + S_1^{(12)} + S_1^{(13)} + S_1^{(14)} + \frac{1}{2}S_1^{(15)} + \frac{1}{2}S_1^{(16)} + \frac{1}{2}S_1^{(17)} + \frac{1}{2}S_1^{(18)} + \frac{1}{3}S_1^{(19)}$$

$$+\frac{1}{4}S_1^{(20)} + \frac{1}{4}S_1^{(21)} + \frac{1}{6}S_1^{(22)} + \frac{1}{6}S_1^{(23)} + \frac{1}{36}S_1^{(24)} - n_1n_2n_3^{[2]}\hat{\mathcal{Q}}_{12} - n_1n_2^{[2]}n_3\hat{\mathcal{Q}}_{13} - n_1n_2n_3\hat{\mathcal{Q}}_0),$$

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. van 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," *J. Roy. Stat. Soc.*, vol. 46, no. 253, pp. 165–205, 1950.

[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, May 2013.

[5] 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.

[6] X. Sun, Y. Xu, M. Tan, H. Fu, W. Zhao, T. You, and J. Liu, "Localizing optic disc and cup for glaucoma screening via deep object detection networks," in *Computational Pathology and Ophthalmic Medical Image Analysis*. Cham, Switzerland: Springer, 2018, pp. 236–244.
[7] 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, Nov. 2012.
[8] 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, Oct. 2011.
[9] V. N. Hari, G. V. Anand, A. B. Premkumar, and A. S. Madhukumar, "Design and performance analysis of a signal detector based on suprathreshold stochastic resonance," *Signal Process.*, vol. 92, no. 7, pp. 1745–1757, Jul. 2012.
[10] C. Catal, O. Alan, and K. Balkan, "Class noise detection based on software metrics and ROC curves," *Inf. Sci.*, vol. 181, no. 21, pp. 4867–4877, Nov. 2011.
[11] 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.
[12] R. C. Prati, G. E. A. P. A. Batista, and M. C. Monard, "Evaluating classifiers using ROC curves," *IEEE Latin Amer. Trans.*, vol. 6, no. 2, pp. 215–222, Jun. 2008.
[13] M. Zolghadri and E. Mansoori, "Weighting fuzzy classification rules using receiver operating characteristics (ROC) analysis," *Inf. Sci.*, vol. 177, no. 11, pp. 2296–2307, Jun. 2007.
[14] M. H. Zweig and G. Campbell, "Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine," *Clin. Chem.*, vol. 39, no. 4, pp. 561–577, Apr. 1993.
[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, Feb. 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: Revised and Expanded*. New York, NY, USA: CRC Press, 2014.
[18] K. O. Hajian-Tilaki, J. A. Hanley, L. Joseph, and J.-P. Collet, "A comparison of parametric and nonparametric approaches to ROC analysis of quantitative diagnostic tests," *Med. Decis. Making*, vol. 17, no. 1, pp. 94–102, Feb. 1997.
[19] K. Molodianovitch, D. Faraggi, and B. Reiser, "Comparing the areas under two correlated ROC curves: Parametric and non-parametric approaches," *Biometrical J.*, vol. 48, no. 5, pp. 745–757, Oct. 2006.
[20] J. Kaufmann, C. Werner, and E. Brunner, "Nonparametric methods for analysing the accuracy of diagnostic tests with multiple readers," *Stat. Methods Med. Res.*, vol. 14, no. 2, pp. 129–146, Apr. 2005.
[21] 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, Jun. 2007.
[22] F. Konietschke and M. Pauly, "A studentized permutation test for the non-parametric Behrens–Fisher problem in paired data," *Electron. J. Statist.*, vol. 6, no. 0, pp. 1358–1372, 2012.
[23] W. Xu, S. Liu, X. Sun, S. Liu, and Y. Zhang, "A fast algorithm for unbiased estimation of variance of AUC based on dynamic programming," *IEEE Access*, vol. 4, pp. 9553–9560, 2016.
[24] M. A. Cleves, "Comparative assessment of three common algorithms for estimating the variance of the area under the nonparametric receiver operating characteristic curve," *Stata J., Promoting Commun. Statist. Stata*, vol. 2, no. 3, pp. 280–289, Sep. 2002.
[25] 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, Apr. 1984.
[26] A. C. Patel and M. K. Markey, "Comparison of three-class classification performance metrics: A case study in breast cancer cad," in *Proc. Med. Imag., Image Perception, Observer Perform., Technol. Assessment*, vol. 5749, 2005, pp. 581–590.
[27] N. Bhooshan, M. Giger, D. Edwards, Y. Yuan, S. Jansen, H. Li, L. Lan, H. Sattar, and G. Newstead, "Computerized three-class classification of MRI-based prognostic markers for breast cancer," *Phys. Med. Biol.*, vol. 56, no. 18, p. 5995, 2011.
[28] B. K. Scurfield, "Multiple-event forced-choice tasks in the theory of signal detectability," *J. Math. Psychol.*, vol. 40, no. 3, pp. 253–269, Sep. 1996.
[29] B. K. Scurfield, "Generalization of the theory of signal detectability *ton-eventm*-dimensional forced-choice tasks," *J. Math. Psychol.*, vol. 42, no. 1, pp. 5–31, Mar. 1998.
[30] D. Mossman, "Three-way ROCs," *Med. Decis. Making*, vol. 19, no. 1, pp. 78–89, Jan. 1999.
[31] S. Dreiseitl, L. Ohno-Machado, and M. Binder, "Comparing three-class diagnostic tests by three-way ROC analysis," *Med. Decis. Making*, vol. 20, no. 3, pp. 323–331, Jul. 2000.
[32] P. S. Heckerling, "Parametric three-way receiver operating characteristic surface analysis using mathematica," *Med. Decis. Making*, vol. 21, no. 5, pp. 409–417, Oct. 2001.
[33] 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.
[34] L. Kang and L. Tian, "Estimation of the volume under the ROC surface with three ordinal diagnostic categories," *Comput. Statist. Data Anal.*, vol. 62, pp. 39–51, Jun. 2013.
[35] X. He and E. C. Frey, "The meaning and use of the volume under a three-class ROC surface (VUS)," *IEEE Trans. Med. Imag.*, vol. 27, no. 5, pp. 577–588, May 2008.
[36] X. He, B. D. Gallas, and E. C. Frey, "Three-class ROC analysis—Toward a general decision theoretic solution," *IEEE Trans. Med. Imag.*, vol. 29, no. 1, pp. 206–215, Jan. 2010.
[37] C. T. Nakas, "Developments in roc surface analysis and assessment of diagnostic markers in three-class classification problems," *REVSTAT-Stat. J.*, vol. 12, no. 1, pp. 43–65, 2014.
[38] M. P. Sampat, A. C. Patel, Y. Wang, S. Gupta, C.-W. Kan, A. C. Bovik, and M. K. Markey, "Indexes for three-class classification performance assessment—An empirical comparison," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 3, pp. 300–312, May 2009.
[39] W. Waegeman, B. De Baets, and L. Boullart, "On the scalability of ordered multi-class ROC analysis," *Comput. Statist. Data Anal.*, vol. 52, no. 7, pp. 3371–3388, Mar. 2008.
[40] S. Liu, W. Xu, X. Sun, and Y. Zhang, "Fast and unbiased estimation of volume under ordered three-class ROC surface (VUS) based on dynamic programming," *IEEE Access*, vol. 7, pp. 63972–63982, 2019.
[41] V. Torra, J. Domingoferrer, J. Mateosanz, and M. Ng, "Regression for ordinal variables without underlying continuous variables," *Inf. Sci.*, vol. 176, no. 4, pp. 465–474, Feb. 2006.
[42] M. Kadziński, S. Greco, and R. Słowiński, "Robust ordinal regression for dominance-based rough set approach to multiple criteria sorting," *Inf. Sci.*, vol. 283, pp. 211–228, Nov. 2014.
[43] R. H. B. Christensen, "Ordinal—Regression models for ordinal data," *R Package Version*, vol. 28, p. 2015, Jun. 2015.
[44] F. E. Harrell, Jr., *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. New York, NY, USA: Springer, 2015.
[45] S. Liu, X. Sun, W. Xu, Y. Zhang, and J. Dai, "Null distribution of volume under ordered three-class ROC surface (VUS) with continuous measurements," *IEEE Signal Process. Lett.*, vol. 25, no. 12, pp. 1855–1859, Dec. 2018.
[46] V. Cevher, S. Becker, and M. Schmidt, "Convex optimization for big data: Scalable, randomized, and parallel algorithms for big data analytics," *IEEE Signal Process. Mag.*, vol. 31, no. 5, pp. 32–43, Sep. 2014.
[47] A. Fahad, N. Alshatri, Z. Tari, A. Alamri, I. Khalil, A. Y. Zomaya, S. Foufou, and A. Bouras, "A survey of clustering algorithms for big data: Taxonomy and empirical analysis," *IEEE Trans. Emerg. Topics Comput.*, vol. 2, no. 3, pp. 267–279, Sep. 2014.
[48] W. Xu, C. Chang, Y. S. Hung, S. K. Kwan, and P. Chin Wan 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.
[49] W. Xu, Y. S. 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, Feb. 2010.
[50] W. Xu, J. Dai, Y. S. Hung, and Q. Wang, "Estimating the area under a receiver operating characteristic (ROC) curve: Parametric and non-parametric ways," *Signal Process.*, vol. 93, no. 11, pp. 3111–3123, Nov. 2013.
[51] U. Manber, *Introduction to Algorithms: A Creative Approach*, vol. 4. Reading, MA, USA: Addison-Wesley, 1989.

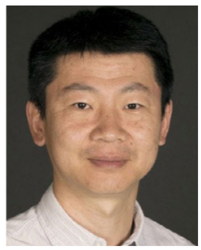
SHUN LIU received the B.Eng. degree from the School of Automation, Guangdong University of Technology, Guangzhou, China, in 2015, where he is currently pursuing the Ph.D. degree. His research interests include statistical signal processing, pattern recognition, and machine learning.

HONGBIN ZHU was born in Guangdong, China. He is currently pursuing the graduate's degree with the Department of Automatic Control, School of Automation, Guangdong University of Technology, China. He is a good Programmer in MATLAB language. His current research interests include statistical signal processing and machine learning.

KAI YI received the B.Eng. degree from the School of Automation, Guangdong University of Technology, Guangzhou, China, in 2017. He is currently a Hardware Engineer at Mediatek Inc. His research interests include statistical signal processing and machine learning.

XU SUN (Member, IEEE) received the B.Eng. and M.Eng. degrees from the School of Automation, Guangdong University of Technology, Guangzhou, China, in 2013 and 2016, respectively. He is currently a Senior Engineer with Baidu, Inc. His research interests include statistical signal processing, statistical learning theory, and intelligent systems.

WEICHAO XU (Member, IEEE) 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 include mathematical statistics, computational statistics, information theory, machine learning, optimization theory, and digital signal processing and its applications.



CHAO WANG (Member, IEEE) received the B.Eng. degree in opto-electrical engineering from Tianjin University, China, in 2002, the M.Sc. degree in optics from Nankai University, China, in 2005, and the Ph.D. degree in electrical and computer engineering from the University of Ottawa, Canada, in 2011. From 2011 to 2012, he was an NSERC Postdoctoral Fellow with the Photonics Laboratory, University of California at Los Angeles (UCLA), USA. He is currently a Senior Lecturer with the School of Engineering and Digital Arts, University of Kent, where he first joined as a Lecturer, in 2013.