Novel non-parametric point and interval estimation for competing probability

Dewi Rahardja\textsuperscript{a,∗}, Yongming Qu\textsuperscript{b} and Yan D. Zhao\textsuperscript{b}

\textsuperscript{a}Department of Statistical Science, Baylor University, Waco, TX, USA
\textsuperscript{b}Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN, USA

Abstract. The Wilcoxon-Mann-Whitney (WMW) test has been widely used as a nonparametric method to compare a continuous or ordinal variable between two groups. The WMW test essentially compares the competing probability \( \pi = \Pr(X > Y) + 0.5 \Pr(X = Y) \) with 0.5, where \( X \) and \( Y \) are independent random variables from two distributions. The competing probability is naturally meaningful and equal to the area under the Receiver Operating Characteristics (ROC) curve. To construct a confidence interval (CI) for \( \pi \), existing methods focused either only on continuous variables or only on ordinal variables. Furthermore, recently developed methods require complicated computation. In this paper, we propose a unified approach to construct a CI for \( \pi \) where the data can be continuous, ordinal, or a mixture of the two. The new approach gives a closed-form solution which is easy to compute. In addition, we propose a small sample modification which allows for constructing a CI even when the estimator for \( \pi \) is 0 or 1. Finally, simulation shows that the performance of the new method is comparable or superior to the existing methods.

Keywords: Area under the curve, competing probability, confidence intervals, ROC curve, small sample modification, Wilcoxon-Mann-Whitney test

1. Introduction

The Wilcoxon-Mann-Whitney (WMW) test, also known as the Wilcoxon rank sum test [19] or Mann-Whitney test [10], has been widely used as a nonparametric method to compare continuous or ordinal response variables between two groups. The WMW test first computes the competing probability \( \pi \) that one group (say, A) has larger responses than the other group (say, B), and then compares \( \pi \) with 0.5 which reflects no group difference. In other words, the WMW method tests the following null hypothesis:

\[
H_0 : \pi = 0.5,
\]

where the competing probability \( \pi \) is defined as

\[
\pi = \Pr(X > Y) + 0.5 \Pr(X = Y),
\]

and \( X \) and \( Y \) denote the response variables for group A and B, respectively. Here \( \pi \) is defined to incorporate ties because WMW test is frequently used on ordinal data where the probability of ties is not zero.

For the alternative hypothesis of \( H_0 \), commonly a location-shift parameterization is considered; i.e.,

\[
H_1 : F_X(\cdot - \Delta) = F_Y(\cdot),
\]

where \( F_X \) and \( F_Y \) are the cumulative distribution functions (CDF) of \( X \) and \( Y \), respectively, and \( \Delta \) is a location-shift parameter. In this case, a confidence interval (CI) for \( \Delta \) is usually constructed using the Hodges-Lehmann estimator [7]. However, O’Brien and Castelloe [12] pointed out that the assumption Eq. (3) for \( H_1 \) is rarely met in

\*Corresponding author. E-mail: Rahardja@gmail.com.
practice and it is a misconception that the WMW test compares two medians. Therefore, the Hodges-Lehmann CI for \( \Delta \) is rarely applicable and a relevant interval estimator associated with the WMW test is more useful.

The measure \( \pi \) is the probability that a randomly selected participant in Group A has a larger response than a randomly selected participant in Group B. Due to this natural interpretation, some researchers prefer to use the measure \( \pi \) than the mean-difference effect measure in the traditional linear models. For example [20], recommended using \( \pi \) to measure the differences between two distributions. In addition, \( \pi \) is closely related to the Receiver Operating Characteristics (ROC) curve methodology popular in the medical diagnostic testing. For example, let \( Y \) and \( X \) be continuous or ordinal test responses for a diseased and a nondiseased participant, respectively, and \( c \) be some cut-off point such that a participant is classified as diseased when the response is greater than \( c \). The ROC curve plots the true positive rates (sensitivity) \( \Pr(Y > c) \) versus the false positive rates (1-specificity) \( \Pr(X > c) \), as the cut-off point \( c \) runs through the range of possible test values. The area under the ROC curve (AUC) is the most commonly used summary measure of the diagnostic test. [3] showed that AUC = \( \pi \), which established a strong relationship between the ROC curve methodology and the WMW test.

Because the competing probability \( \pi \) is naturally meaningful and is equal to the AUC, in this paper we consider a more general null hypothesis

\[
H_0 : \pi = \pi_0, \tag{4}
\]

where \( \pi_0 \in (0, 1) \), and we aim to develop an interval estimator for \( \pi \).

Constructing confidence interval (CI) for \( \pi \) has been studied in the literature, and the existing methods are applicable either to continuous responses or to ordinal responses. When the responses \( Y \) and \( X \), are continuous [15], proposed a confidence interval for \( \pi \); based on the asymptotic normality of the Mann-Whitney statistic. Since \( \pi \) is restricted to \([0, 1]\) [13], improved Sen’s CI by first obtaining CI for the logit of \( \pi \) and then constructing CI for \( \pi \) using inverse-logit transformation. However, Pepe’s approach does not work when the estimator for \( \pi \) may be 0 or 1 for small samples. Recently [14] used an Empirical Likelihood method to construct CI for \( \pi \) and examined the performance of their method with Sen’s, Pepe’s, and some other methods. Based on their simulation results, Qin and Zhou recommended the use of Empirical Likelihood (EL) CI for \( \pi \) when the underlying distributions for \( Y \) and \( X \) are unknown, which is usually the case in practice.

When the responses \( Y \) and \( X \) are ordinal, there exist two main strategies for constructing CI for \( \pi \). The first strategy is parametric and makes distributional assumptions for an underlying (latent) variable. This approach uses ordinal regression methods to construct CI [16]. However, this approach suffers from the difficulty of verifying the assumption of the underlying latent variable. The second strategy does not make distributional assumptions on the response variables. For example [5], proposed a method by estimating the variance of \( \pi \) using a purely nonparametric method. Nevertheless, the method breaks down when sample sizes for either group are small and/or the observed \( \pi \) is close to 1. To overcome this difficulty [17], proposed a Profile-Likelihood (PL) inference for constructing CI for \( \pi \). They showed that their CIs have acceptable coverage probabilities, even when sample sizes are small and the diagnostic tests have high accuracies. Recently [12] suggested a new method which first artificially altered the data and then applied the generalized odds ratio idea [1] to obtain point and interval estimators. However, they did not assess the type I error level of their method and the alteration of data seems arbitrary.

Although both EL and PL were claimed to be superior to previous existing methods, they are only applicable for either continuous or ordinal responses and are computationally difficult. There may be some situations where the data have both continuous and ordinal responses. For example, a continuous variable is truncated at the very low and/or very high values. For such situations, neither of the above two approaches can be directly applied. In addition, if there are too many categories for the ordinal data, the profile-likelihood method does not work because it will be computationally infeasible. In this paper, we propose a method that is computationally simple and works for both continuous and ordinal responses. We first consider a nonparametric unbiased point estimator that is a generalization of the Mann-Whitney \( U \) statistic [10] to the case with ties. Then, we compute its variance using an approach similar to [5] with additional consideration for the small sample situation. Finally, a Wald type CI for \( \pi \) is constructed.

The rest of the paper is organized as follows. In Section 2 we describe a nonparametric unbiased estimator for the competing probability \( \pi \) and compute its variance. Based on this estimator together with its variance estimator, we then develop CI for the competing probabilities for the WMW test. In Section 3 we study the performance of these methods through simulation. We illustrate our methods using an example in Section 4. Finally, some discussion can be found in Section 5.
2. Methods

Suppose that we observe two independent random samples $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ with CDF $F_X$ and $F_Y$, respectively. As discussed in Section 1, WMW statistic detects the group difference by testing the null hypothesis (1). To facilitate development, we express the competing probability $\pi$ as $\pi = E[(\delta(X - Y) + 1)/2]$, where

$$
\delta(t) = \begin{cases} 
1 & \text{if } t > 0 \\
0 & \text{if } t = 0 \\
-1 & \text{if } t < 0 
\end{cases}
$$

Similar to the Mann-Whitney U-statistic, a straightforward nonparametric unbiased estimator for $\pi$ that incorporates ties is

$$
\hat{\pi} = (mn)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n}[\delta(X_i - Y_j) + 1]/2.
$$

To test the null hypothesis Eq. (1), Wilcoxon used $\hat{\pi}$ and the variance of $\hat{\pi}$ under the null hypothesis Eq. (1) to construct a $z$-statistic:

$$
z_0 = \frac{\hat{\pi} - 0.5}{\sqrt{V_0[\hat{\pi}]}} \sim N(0, 1),
$$

(5)

where

$$
V_0[\hat{\pi}] = (12mn)^{-1}(m + n + 1)
$$

(6)

is the variance of $\hat{\pi}$ under the null hypothesis Eq. (1) when ties are ignored (see, e.g., [9]).

Although Wilcoxon statistic can be used for testing Eq. (1), it cannot be directly applied to construct CI for $\pi$. In order to do so, the variance of $\hat{\pi}$ and an estimator of the variance are needed. With tedious algebra, we express the variance of $\hat{\pi}$ as

$$
V[\hat{\pi}] = (4mn)^{-1}[(m - 1)(p_2 - p_1^2) + (n - 1)(p_3 - p_1^2) + p_4 - p_1^2],
$$

(7)

where

$$
p_1 = E[\delta(X - Y)],
$$

$$
p_2 = E[\delta(X - Y)\delta(X^* - Y)],
$$

$$
p_3 = E[\delta(X - Y)\delta(X - Y^*)],
$$

$$
p_4 = E[(\delta(X - Y))^2],
$$

$X$ and $X^*$ are independently and identically distributed (i.i.d.) from $F_X$, and $Y$ and $Y^*$ are i.i.d. from $F_Y$. An estimator $\hat{V}^*[\hat{\pi}]$ of $V[\hat{\pi}]$ is immediately available when replacing $p_1, p_2, p_3, p_4$ by their sample estimates; i.e.,

$$
\hat{V}^*[\hat{\pi}] = (4mn)^{-1}[(m - 1)(\hat{p}_2 - \hat{p}_1^2) + (n - 1)(\hat{p}_3 - \hat{p}_1^2) + \hat{p}_4 - \hat{p}_1^2],
$$

(8)

where

$$
\hat{p}_1 = (mn)^{-1}\sum_{1 \leq i \leq m, 1 \leq j \leq n}\delta(X_i - Y_j),
$$

$$
\hat{p}_2 = [mn(m - 1)]^{-1}\sum_{1 \leq j \leq n, 1 \leq i \neq k \leq m}\delta(X_i - Y_j)\delta(X_k - Y_j),
$$

$$
\hat{p}_3 = [mn(n - 1)]^{-1}\sum_{1 \leq i \leq m, 1 \leq j \neq l \leq n}\delta(X_i - Y_j)\delta(X_i - Y_l),
$$

and

$$
\hat{p}_4 = [mn(m - 1)(n - 1)]^{-1}\sum_{1 \leq i \leq m, 1 \leq j \neq k \neq l \leq n}\delta(X_i - Y_j)\delta(X_k - Y_l),
$$

(9)
Simulation controls the conservativeness of the CI. The smaller it is, the more biased and may not have good small sample properties. Therefore, we construct an unbiased estimator of \( \pi \) from \((8)\) where ordinal data are generated. The parameters used in the three simulation scenarios are described as follows. The simulation settings are adopted from \((14)\) where continuous data are generated; the third simulation setting is adopted from \((17)\) where ordinal data are generated. The Appendix for details. Although the variance estimator \( \hat{\pi}_m \) given in \((8)\) is a consistent estimator for \( \pi \), it is biased and may not have good small sample properties. Therefore, we construct an unbiased estimator of \( \hat{\pi}_m \) by

\[
\hat{\pi} = \frac{1}{(m+n-1)/(mn)} \sum_{1 \leq i \leq m, 1 \leq j \leq n} [\delta(X_i - Y_j)]^2
\]

It can be verified that \( \hat{\pi}_1, \hat{\pi}_2, \hat{\pi}_3 \) and \( \hat{\pi}_4 \) are unbiased and consistent estimators for \( p_1, p_2, p_3 \) and \( p_4 \), respectively. See the Appendix for details.

Then, \( \hat{\pi} \) is slightly larger than \( \hat{\pi}_4 \) and should have better properties. Therefore, we will use \( \hat{\pi} \) as the estimator for \( \hat{\pi}_4 \) for the rest of the paper.

Following Pepe's approach, we propose to construct a \( z \)-statistic using a logit transformation:

\[
z = \frac{\logit(\hat{\pi}) - \logit(\pi)}{\sqrt{V[\hat{\pi}](\hat{\pi}(1-\hat{\pi}))^2}} \sim N(0, 1).
\]

Then, a \((1 - \alpha)100\%\) CI can be constructed as

\[
(\exp(L)/(1 + \exp(L)), \exp(U)/(1 + \exp(U)))
\]

where 

\[
L = \logit(\hat{\pi}) - Z_{1-\alpha/2} \sqrt{V[\hat{\pi}](\hat{\pi}(1-\hat{\pi}))^2}, \quad U = \logit(\hat{\pi}) + Z_{1-\alpha/2} \sqrt{V[\hat{\pi}](\hat{\pi}(1-\hat{\pi}))^2},
\]

and \( Z_{1-\alpha/2} \) is the \((1 - \alpha/2)\) quantile of the standard normal distribution. However, as pointed out earlier, for small sample size and very small or large \( \pi \) (e.g., \( \pi = 0.05 \) or \( 0.95 \)), often \( \hat{\pi} \) is close to or even equal to 0 or 1. In such a situation, the CI given in Eq. \((11)\) either has poor coverage probability or is impossible to construct. To handle such special cases, we propose a small sample modification to Eq. \((11)\) for the \((1 - \alpha)100\%\) confidence interval of \( \pi \):

\[
CI = \begin{cases} 
\left( \frac{\hat{\pi} - Z_{1-\alpha} \sqrt{V_0[\hat{\pi}]}}{1-a}, 1 \right) & \text{if } \hat{\pi} \geq 1-a \\
\left( 0, \frac{\hat{\pi} + Z_{1-\alpha} \sqrt{V_0[\hat{\pi}]}}{a} \right) & \text{if } \hat{\pi} \leq a \\
\text{Equation (11)} & \text{otherwise,}
\end{cases}
\]

where \( V_0[\hat{\pi}] \) is given by Eq. \((6)\), \( a = [(mn)/(m + n + 1)]^{-0.5+\theta} \), and \( \theta \) can be any positive value. The large sample properties of the new confidence interval do not change because the estimator \( \hat{\pi} \) is consistent with a rate of approximately \([(mn)/(m + n + 1)]^{-1} \). Because \( V[\hat{\pi}] \) decreases as \( \pi \) moves away from 0.5, \( V_0[\hat{\pi}] \) is generally larger than \( V[\hat{\pi}] \). Therefore, when \( \hat{\pi} \) is close to 0 or 1, the CI given above in Eq. \((12)\) is generally more conservative than that constructed directly from Eq. \((10)\). For data with small sample sizes, the empirical constant \( \theta \) used in computing \( \alpha \) controls the conservativeness of the CI. The smaller \( \theta \), the more conservative the CI is.

3. Simulation

In this section, we present a simulation study with 3 settings to examine the coverage probabilities and the mean widths of the newly proposed confidence interval estimator for the competing probability \( \pi \). The first two simulation settings are adopted from \((14)\) where continuous data are generated; the third simulation setting is adopted from \((17)\) where ordinal data are generated. The parameters used in the three simulation scenarios are described as follows. In the first simulation setting, we generate \( X \sim N(\sqrt{\theta}^{-1} \Phi^{-1}(\pi), 2) \) and \( Y \sim N(0, 1) \), where \( \Phi(\cdot) \) is the CDF of the standard normal distribution and \( \pi \) is taken to be 0.80 and 0.95. In the second simulation setting, we generate \( X \sim LN(401^{1/2} \Phi^{-1}(\pi), 200) \) and \( Y \sim LN(0, 1) \), where \( LN(\mu, \sigma^2) \) denotes the lognormal distribution with parameters \( \mu \) and \( \sigma^2 \), and \( \pi \) is taken to be 0.80 and 0.95. In the third simulation, in order to achieve \( \pi = 0.95 \), we generate \( X \) and \( Y \) with values of 1, 2, 3 and 4 with probabilities \( P_x = (0.53203, 0.3, 0.1, 0.0634, 0.000634) \) and \( P_y = (0.00634, 0.0634, 0.1, 0.3, 0.53203) \); respectively; Similarly, we choose \( P_x = (0.43626, 0.3, 0.15, 0.1034, 0.01034) \) and \( P_y = (0.01034, 0.1034, 0.15, 0.3, 0.43626) \) to achieve \( \pi = 0.90 \).

Tables 1, 2 and 3 show the empirical coverage probabilities (c.p.) and the mean widths (m.w.) of the 90% and 95% CIs based on 10,000 simulations. In all the simulations we use \((12)\) to calculate CIs with \( \theta \) chosen to be 1.0.
Table 1
Coverage probabilities (c.p.) and mean widths (m.w.) of confidence intervals for \( \pi \) based on 10,000 simulations (\( X \sim N(5^{1/2} \Phi^{-1}(\pi), 2) \) and \( Y \sim N(0, 1) \))

<table>
<thead>
<tr>
<th>( \pi )</th>
<th>( (m, n) )</th>
<th>90% CI</th>
<th>95% CI</th>
<th>EL*</th>
<th>NEW</th>
<th>EL*</th>
<th>NEW</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8</td>
<td>(50, 50)</td>
<td>0.8938</td>
<td>0.1499</td>
<td>0.9044</td>
<td>0.1513</td>
<td>0.9407</td>
<td>0.1783</td>
</tr>
<tr>
<td></td>
<td>(80, 80)</td>
<td>0.8996</td>
<td>0.1188</td>
<td>0.9019</td>
<td>0.1196</td>
<td>0.9431</td>
<td>0.1419</td>
</tr>
<tr>
<td></td>
<td>(100, 100)</td>
<td>0.8988</td>
<td>0.1064</td>
<td>0.8991</td>
<td>0.1070</td>
<td>0.9489</td>
<td>0.1269</td>
</tr>
<tr>
<td></td>
<td>(30, 30)</td>
<td>0.8817</td>
<td>0.1246</td>
<td>0.9054</td>
<td>0.1256</td>
<td>0.9468</td>
<td>0.1487</td>
</tr>
</tbody>
</table>

*Simulation results for EL method were from [14].

Table 2
Coverage probabilities (c.p.) and mean widths (m.w.) of confidence intervals for \( \pi \) based on 10,000 simulations (\( X \sim LN(401^{1/2} \Phi^{-1}(\pi), 20) \) and \( Y \sim LN(0, 1) \))

<table>
<thead>
<tr>
<th>( \pi )</th>
<th>( (m, n) )</th>
<th>90% CI</th>
<th>95% CI</th>
<th>EL*</th>
<th>NEW</th>
<th>EL*</th>
<th>NEW</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8</td>
<td>(50, 50)</td>
<td>0.8941</td>
<td>0.1771</td>
<td>0.9055</td>
<td>0.1811</td>
<td>0.9354</td>
<td>0.2102</td>
</tr>
<tr>
<td></td>
<td>(80, 80)</td>
<td>0.8969</td>
<td>0.1413</td>
<td>0.8974</td>
<td>0.1433</td>
<td>0.9437</td>
<td>0.1680</td>
</tr>
<tr>
<td></td>
<td>(100, 100)</td>
<td>0.8988</td>
<td>0.1268</td>
<td>0.9052</td>
<td>0.1282</td>
<td>0.9442</td>
<td>0.1507</td>
</tr>
<tr>
<td></td>
<td>(30, 30)</td>
<td>0.8862</td>
<td>0.1415</td>
<td>0.9045</td>
<td>0.1434</td>
<td>0.9421</td>
<td>0.1678</td>
</tr>
<tr>
<td></td>
<td>(50, 50)</td>
<td>0.8990</td>
<td>0.1269</td>
<td>0.8976</td>
<td>0.1282</td>
<td>0.9516</td>
<td>0.1508</td>
</tr>
</tbody>
</table>

*Simulation results for EL method were from [14].

Table 3
Coverage probabilities (c.p.) and mean widths (m.w.) of 95% confidence intervals for \( \pi \) based on 10,000 simulations

<table>
<thead>
<tr>
<th>( \pi )</th>
<th>( (m, n) )</th>
<th>PL*</th>
<th>NEW</th>
<th>PL*</th>
<th>NEW</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>(5, 25)</td>
<td>0.9772</td>
<td>0.314</td>
<td>0.9723</td>
<td>0.289</td>
</tr>
<tr>
<td></td>
<td>(15, 15)</td>
<td>0.9744</td>
<td>0.202</td>
<td>0.9763</td>
<td>0.226</td>
</tr>
<tr>
<td></td>
<td>(10, 50)</td>
<td>0.9778</td>
<td>0.206</td>
<td>0.9715</td>
<td>0.207</td>
</tr>
<tr>
<td></td>
<td>(30, 30)</td>
<td>0.9423</td>
<td>0.144</td>
<td>0.9479</td>
<td>0.150</td>
</tr>
<tr>
<td></td>
<td>(5, 25)</td>
<td>0.9772</td>
<td>0.270</td>
<td>0.9753</td>
<td>0.253</td>
</tr>
<tr>
<td></td>
<td>(15, 15)</td>
<td>0.9825</td>
<td>0.149</td>
<td>0.9758</td>
<td>0.191</td>
</tr>
<tr>
<td></td>
<td>(10, 50)</td>
<td>0.9777</td>
<td>0.167</td>
<td>0.9769</td>
<td>0.176</td>
</tr>
<tr>
<td></td>
<td>(30, 30)</td>
<td>0.9762</td>
<td>0.100</td>
<td>0.9752</td>
<td>0.109</td>
</tr>
</tbody>
</table>

*Simulation results for PL method were from [17].

Results from Tables 1 and 2 show that the coverage probability and the mean width of the 90% and 95% CI from the new proposed method are comparable with or superior to those using the empirical likelihood (EL) method. For example, in Table 1, for \( (m, n) = (50, 50) \) and \( \pi = 0.95 \), the coverage probability of the 95% CI using EL was 0.8964 while the coverage probability using the proposed method is 0.9332. In the meantime, the mean widths of the CI were similar for the two methods. The CI constructed by the proposed method is generally conservative. In some situations where sample sizes are between 50 and 100, the coverage probability is less than the nominal level.
Table 4  
AUCs and the 95% confidence interval based on CT ratings for cancer in the abdomen and beyond

<table>
<thead>
<tr>
<th>Institution</th>
<th>Diseased?</th>
<th>CT Rating</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>( \hat{\pi} )</th>
<th>PL CI</th>
<th>New CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 NO</td>
<td></td>
<td></td>
<td>9</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0.969</td>
<td>(0.701, 0.998)</td>
<td>(0.697, 1.000)</td>
</tr>
<tr>
<td>YES</td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 NO</td>
<td></td>
<td></td>
<td>7</td>
<td>21</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>0.872</td>
<td>(0.696, 0.962)</td>
<td>(0.671, 0.959)</td>
</tr>
<tr>
<td>YES</td>
<td></td>
<td></td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 NO</td>
<td></td>
<td></td>
<td>3</td>
<td>31</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>0.946</td>
<td>(0.837, 0.987)</td>
<td>(0.823, 0.985)</td>
</tr>
<tr>
<td>YES</td>
<td></td>
<td></td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 NO</td>
<td></td>
<td></td>
<td>11</td>
<td>7</td>
<td>5</td>
<td>0</td>
<td>4</td>
<td>0.926</td>
<td>(0.807, 0.976)</td>
<td>(0.822, 0.971)</td>
</tr>
<tr>
<td>YES</td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 NO</td>
<td></td>
<td></td>
<td>15</td>
<td>8</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1.000</td>
<td>(0.804, 1.000)</td>
<td>(0.782, 1.000)</td>
</tr>
<tr>
<td>YES</td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.929</td>
<td>(0.877, 0.963)</td>
<td>(0.884, 0.966)</td>
</tr>
</tbody>
</table>

However, in these cases, the coverage probability for the EL method was even worse. The results in Table 3 indicate that for ordinal data, the coverage probability and the mean width of the CI for the proposed method are comparable to those using the PL method.

4. An example

We apply the method in Section 2 to data from an ovarian cancer study [17,8]. The results of a CT rating for cancer in the abdomen and beyond are ordinal values: 1 = definitely benign; 2 = benign; 3 = indeterminate; 4 = malignant; 5 = definitely malignant. Here we calculate the AUC as well as the 95% confidence interval based on the CT rating. Table 4 shows the data, the estimate of AUC and the 95% confidence intervals based on the profile likelihood (PL) method and the method described in Section 2. It can be seen that the confidence intervals calculated by PL and the new proposed method were similar.

5. Discussion

In this paper we proposed a new method for constructing CI for the competing probabilities \( \pi \) associated with the WMW test. The newly proposed method has three major improvements over the existing methods. First, the new method is a unified approach which works for continuous data, ordinal data, and a mixture of the two. Second, the new method includes a small sample modification which enables constructing a CI even when the estimate \( \hat{\pi} \) is close to 0 or 1. Lastly, the new method has closed-form formula and is very easy to compute. We note that the computation of the new method involves choosing a constant \( \theta \). Although the constant \( \theta = 1.0 \) used in our simulation seems to work well, choosing the optimal \( \theta \) deserves further research. For a particular problem, some simulations may be needed to determine what \( \theta \) should be used. It should be noted that no theoretical work has been done for small sample properties for empirical likelihood and profile likelihood methods.

Acknowledgement

We would like to thank Dr. Ilya Lipkovich and an anonymous referee for useful comments.

Appendix

In the appendix we show that \( \hat{p}_1, \hat{p}_2, \hat{p}_3 \) and \( \hat{p}_4 \) are unbiased and consistent estimators for \( p_1, p_2, p_3 \) and \( p_4 \), respectively. We will use \( \hat{p}_1 \) as an example and the proofs for \( \hat{p}_2, \hat{p}_3 \) and \( \hat{p}_4 \) are similar.
To show $\hat{p}_1$ is unbiased for $p_1$, we first note that $E(\delta(X_i - Y_j)) = p_1$. Therefore, we have
\[
E(\hat{p}_1) = (mn)^{-1} \sum_{1 \leq i \leq m, 1 \leq j \leq n} E(\delta(X_i - Y_j)) = p_1
\]
and hence $\hat{p}_1$ is unbiased for $p_1$.

To show $\hat{p}_1$ is consistent for $p_1$, we first compute
\[
V(\hat{p}_1) = (mn)^{-2} \sum_{1 \leq i \leq m, 1 \leq j \leq n} V(\delta(X_i - Y_j)) + (mn)^{-2} \sum_{1 \leq j \leq n, 1 \leq i \neq j \leq m} \text{Cov}(\delta(X_i - Y_j), \delta(X_k - Y_j)) + (mn)^{-2} \sum_{1 \leq i \leq m, 1 \leq j \neq l \leq n} \text{Cov}(\delta(X_i - Y_j), \delta(X_k - Y_l))
\]
\[
= p_1 - \frac{p_1^2}{mn} + \frac{(m-1)(p_2 - p_2^2)}{mn} + \frac{(n-1)(p_3 - p_3^2)}{mn} + 0.
\]
Clearly $V(\hat{p}_1)$ goes to zero when $m$ and $n$ go to infinity; therefore, by Chebyshev Theorem, $\hat{p}_1$ is a consistent estimator of $p_1$.

References

[10] H.B. Mann and D.R. Whitney, On a test of whether one of two random variables is stochastically larger than the other, Annals of Mathematical Statistics 18 (1947), 50–60.


