Ann Inst Stat Math (2018) 70:789–806
https://doi.org/10.1007/s10463-017-0631-z

CrossMark

# Jackknife empirical likelihood for the difference of two volumes under ROC surfaces

**Yueheng An**[1] · **Yichuan Zhao**[1]

Received: 12 January 2016 / Revised: 28 November 2016 / Published online: 15 November 2017


**Abstract** The volume under a surface (VUS) is an effective measure for evaluating the discriminating power of a diagnostic test with three ordinal diagnostic groups. In this paper, we investigate the difference of two correlated VUS's to compare two treatments for discrimination of three-class classification data. A jackknife empirical likelihood (JEL) procedure is employed to avoid the variance estimation in the existing methods. We prove that the limiting distribution of the empirical log-likelihood ratio statistic follows a $\chi^2$ distribution. Extensive numerical studies show that the JEL confidence intervals outperform those based on the normal approximation method. The proposed method is also applied to the Alzheimer's disease data.

**Keywords** Jackknife empirical likelihood · Receiver operating characteristic (ROC) curve · Volume under an ROC surface

## 1 Introduction

The receiver operating characteristic (ROC) curve measures the performance of a binary classifier system by plotting the false-positive rate against the true-positive rate for various discrimination thresholds. The area under the ROC curve (AUC) provides a single value to summarize the performance of a diagnostic treatment. A multi-class classification procedure is necessary if the subjects are assigned to more than two groups simultaneously. Mossman (1999) evaluated a three-class classification treatment using the volume under the ROC surface (VUS). Thus, the VUS is proposed as an analogous measure to the AUC, which extends an ROC curve to an ROC surface in the three-class case. Tian et al. (2011) showed that the difference of two correlated

✉ Yichuan Zhao
yichuan@gsu.edu

1 Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30303, USA

VUS's is an efficient summary for the comparison of diagnostic accuracy with three ordinal diagnostic groups using parametric methods.

Wan (2012) developed an empirical likelihood (Owen 1988) inference procedure for the VUS, where it extended from AUC based on two-class data to VUS based on three-class data. The estimating equations for the difference of two correlated VUS's are much more complicated than the difference of two ROC curves or that of two AUC's. It is very difficult to construct an empirical likelihood (EL) confidence interval for such a difference as Owen (1988, 1990)'s EL method is too complicated to use. Since Jing et al. (2009) introduced the jackknife empirical likelihood (JEL) method, it has become possible and tractable for us to tackle with this challenging problem (cf. Gong et al. 2010; Yang and Zhao 2013, 2015). The JEL employs a $U$-statistic to avoid the nuisance parameters in the estimating equations. Therefore, it provides a reliable confidence interval by solving a simpler estimating equation of a pseudo-mean, which is based on $U$-statistic. Pan et al. (2013) made nonparametric inference for the VUS's using JEL, which is a univariate three-sample problem. Inspired by Pan et al. (2013), we extend it to a bivariate three-sample case and propose JEL methods for the difference of two correlated VUS's. Our simulation results demonstrate that proposed JEL confidence intervals outperform the normal approximation (NA) method for the difference of two correlated VUS's.

The rest of the paper is organized as follows. In Sect. 2, the JEL method is employed to construct the confidence intervals for the difference of two VUS's. Motivated by Pan et al. (2013), we prove that the limiting distribution of the empirical log-likelihood ratio statistic follows a $\chi^2$ distribution. In Sect. 3, we present the results of intensive simulation studies on the JEL confidence intervals, which have better performance than those based on the NA method in terms of coverage probability. In Sect. 4, the proposed method is illustrated by an Alzheimer's disease (AD) data set. All the proofs are provided in Appendix.

## 2 Inference procedure

Let $(X_1^T, X_2^T, \ldots, X_{n_1}^T)$, $(Y_1^T, Y_2^T, \ldots, Y_{n_2}^T)$ and $(Z_1^T, Z_2^T, \ldots, Z_{n_3}^T)$ represent i.i.d. samples of three independent populations, where $X_i = (X_{1i}, X_{2i})^T, i = 1, 2, \ldots, n_1$, $Y_j = (Y_{1j}, Y_{2j})^T$, $j = 1, 2, \ldots, n_2$, and $Z_k = (Z_{1k}, Z_{2k})^T$, $k = 1, 2, \ldots, n_3$. We adopt the same notations as Pan et al. (2013) did. Define the VUS with respect to the first component as $P(X_{11} < Y_{11} < Z_{11})$ and the VUS with respect to the second component as $P(X_{21} < Y_{21} < Z_{21})$, respectively. Therefore, the difference of two VUS's can be defined as

$$\begin{aligned}\theta &= P(X_{11} < Y_{11} < Z_{11}) - P(X_{21} < Y_{21} < Z_{21}) \\ &= E(I(X_{11} < Y_{11} < Z_{11})) - E(I(X_{21} < Y_{21} < Z_{21})) \\ &= E(I(X_{11} < Y_{11} < Z_{11}) - I(X_{21} < Y_{21} < Z_{21})),\end{aligned}$$

which can be estimated by

$$\hat{\theta} = \frac{1}{n_1 n_2 n_3} \sum_{\substack{i=1,\ldots,n_1,\\ j=1,\ldots,n_2,\\ k=1,\ldots,n_3}} [I(X_{1i} < Y_{1j} < Z_{1k}) - I(X_{2i} < Y_{2j} < Z_{2k})].$$

A $U$-statistic of degree (1, 1, 1) with a kernel $h(x; y; z)$ is defined as

$$U_n = \frac{1}{n_1 n_2 n_3} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} h(X_i; Y_j; Z_k),$$

which is an unbiased estimator of $\theta = Eh(X_i; Y_j; Z_k)$. In particular, if

$$h(X_i; Y_j; Z_k) = I(X_{1i} < Y_{1j} < Z_{1k}) - I(X_{2i} < Y_{2j} < Z_{2k}),$$

then $\theta = P(X_{11} < Y_{11} < Z_{11}) - P(X_{21} < Y_{21} < Z_{21})$. Therefore, we define the estimate of $\theta$ as a $U$-statistic.

$$U_n = \frac{1}{n_1 n_2 n_3} \sum_{i=1}^{n1} \sum_{j=1}^{n2} \sum_{k=1}^{n3} [I(X_{1i} < Y_{1j} < Z_{1k}) - I(X_{2i} < Y_{2j} < Z_{2k})].$$

For $i = 1, 2, \ldots, n_1$, $j = 1, 2, \ldots, n_2$, and $k = 1, 2, \ldots, n_3$, we denote

(1) the statistics for all observations as $U^0_{n_1,n_2,n_3} = U_n$;
(2) the statistic after removing $X_{i'}$ as

$$U^{-i',0,0}_{n_1-1,n_2,n_3} = ((n_1-1)n_2 n_3)^{-1} \sum_{\substack{i=1,\\ i\neq i'}}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} [I(X_{1i} < Y_{1j} < Z_{1k}) \\ -I(X_{2i} < Y_{2j} < Z_{2k})];$$

(3) the statistic after removing $Y_{j'}$ as

$$U^{0,-j',0}_{n_1,n_2-1,n_3} = (n_1(n_2-1)n_3)^{-1} \sum_{i=1}^{n_1} \sum_{\substack{j=1,\\ j\neq j'}}^{n_2} \sum_{k=1}^{n_3} [I(X_{1i} < Y_{1j} < Z_{1k}) \\ -I(X_{2i} < Y_{2j} < Z_{2k})];$$

(4) the statistic after removing $Z_{k'}$ as

$$U^{0,0,-k'}_{n_1,n_2,n_3-1} = (n_1 n_2(n_3-1))^{-1} \sum_{i=1}^{n_1} \sum_{j=1}^{n2} \sum_{\substack{k=1,\\ k\neq k'}}^{n_3} [I(X_{1i} < Y_{1j} < Z_{1k}) \\ -I(X_{2i} < Y_{2j} < Z_{2k})].$$

Then, define the jackknife pseudo-values by

$$\begin{aligned} V_{i,0,0} &= n_1 U^{0}_{n_1,n_2,n_3} - (n_1-1) U^{-i,0,0}_{n_1-1,n_2,n_3};\\ V_{0,j,0} &= n_2 U^{0}_{n_1,n_2,n_3} - (n_2-1) U^{0,-j,0}_{n_1,n_2-1,n_3};\\ V_{0,0,k} &= n_3 U^{0}_{n_1,n_2,n_3} - (n_3-1) U^{0,0,-k}_{n_1,n_2,n_3-1}. \end{aligned}$$

We obtain the following forms with some simple algebra,

$$\begin{aligned} V_{i,0,0} &= \frac{1}{n_2 n_3} \sum_{j_1=1}^{n_2} \sum_{k_1=1}^{n_3} [I(X_{1i} < Y_{1j} < Z_{1k}) - I(X_{2i} < Y_{2j} < Z_{2k})];\\ V_{0,j,0} &= \frac{1}{n_1 n_3} \sum_{i_1=1}^{n_1} \sum_{k_1=1}^{n_3} [I(X_{1i} < Y_{1j} < Z_{1k}) - I(X_{2i} < Y_{2j} < Z_{2k})];\\ V_{0,0,k} &= \frac{1}{n_1 n_2} \sum_{i_1=1}^{n_1} \sum_{j_1=1}^{n_2} [I(X_{1i} < Y_{1j} < Z_{1k}) - I(X_{2i} < Y_{2j} < Z_{2k})]; \end{aligned}$$

and

$$\bar{V}_{\cdot,0,0} = \bar{V}_{0,\cdot,0} = \bar{V}_{0,0,\cdot} = U_n,$$

where $\bar{V}_{\cdot,0,0}$, $\bar{V}_{0,\cdot,0}$, and $\bar{V}_{0,0,\cdot}$ are the averages of $V_{i,0,0}$, $V_{0,j,0}$, and $V_{0,0,k}$, respectively.

The following notations are needed throughout the paper (cf. Pan et al. 2013),

$$\begin{aligned} g_{1,0,0}(x) &= [P(x_{11} < Y_{11} < Z_{11}) - P(x_{21} < Y_{21} < Z_{21})] - \theta,\\ \sigma^2_{1,0,0} &= Var(g_{1,0,0}(X_1));\\ g_{0,1,0}(y) &= [P(X_{11} < y_{11} < Z_{11}) - P(X_{21} < y_{21} < Z_{21})] - \theta,\\ \sigma^2_{0,1,0} &= Var(g_{0,1,0}(Y_1));\\ g_{0,0,1}(z) &= [P(X_{11} < Y_{11} < z_{11}) - P(X_{21} < Y_{21} < z_{21})] - \theta,\\ \sigma^2_{0,0,1} &= Var(g_{0,0,1}(Z_1)); \end{aligned}$$

where $x = (x_1, x_2)^T$, $y = (y_1, y_2)^T$, and $z = (z_1, z_2)^T$.

Denote $(T_1, T_2, \ldots, T_n) = (T_1, T_2, \ldots, T_{n_1}, T_{n_1+1}, T_{n_1+2}, \ldots, T_{n_1+n_2}, T_{n_1+n_2+1}, \ldots, T_{n_1+n_2+n_3})$ $= (X_1^T, X_2^T, \ldots, X_{n_1}^T, Y_1^T, Y_2^T, \ldots, Y_{n_2}^T, Z_1^T, Z_2^T, \ldots, Z_{n_3}^T)$, where $n = n_1 + n_2 + n_3$. A one-sample $U$-statistic of degree three is defined as follows,

$$W_n = U_n(T_1, T_2, \ldots, T_n) = \binom{n}{3}^{-1} \sum_{1 \le i < j < k \le n} h(T_i, T_j, T_k),$$

where the kernel function

$$h(T_i, T_j, T_k) = \frac{\binom{n}{3}}{n_1 n_2 n_3}[I(X_{1i} < Y_{1,j-n_1} < Z_{1,k-n_1-n_2}) \\ -I(X_{2i} < Y_{2,j-n_1} < Z_{2,k-n_1-n_2})] \tag{1}$$

for $i = 1, 2, \ldots, n_1$, $j = n_1 + 1, n_1 + 2, \ldots, n_1 + n_2$, $k = n_1 + n_2 + 1, n_1 + n_2 + 2, \ldots, n$, and $1 \le i \le n_1 < j \le n_1 + n_2 < k \le n$, and $h(T_i, T_j, T_k) = 0$ otherwise. Note that $\theta = Eh(T_i, T_j, T_k)$, and $W_n = U_n$. Similar to Pan et al. (2013), we define the $U$-statistic with $T_l$ deleted as follows:

$$\begin{aligned}
& W_{n-1}^{(-l)} U_{n-1}(T_1, T_2, \ldots, T_{l-1}, T_{l+1}, \ldots, T_n) \\
&= \binom{n-1}{3}^{-1} \sum_{n-1,3}^{(-l)} h(T_i, T_j, T_k) \\
&= \binom{n-1}{3}^{-1} \left[ \sum_{i<j<k} h(T_i, T_j, T_k) - \sum_{j<k} h(T_l, T_j, T_k) \right. \\
& \quad \left. - \sum_{i<k} h(T_i, T_l, T_k) - \sum_{i<j} h(T_i, T_j, T_l) \right] \\
&= \binom{n-1}{3}^{-1} \left[ \binom{n}{3} W_n - \sum_{j<k} h(T_l, T_j, T_k) - \sum_{i<k} h(T_i, T_l, T_k) \right. \\
& \quad \left. - \sum_{i<j} h(T_i, T_j, T_l) \right],
\end{aligned}$$

where we denote the removal of $T_l$ as $(-l)$, $1 \le l \le n$.

Like Pan et al. (2013), we define the jackknife pseudo-values by

$$\begin{aligned}
\hat{V}_l &= nW_n - (n-1)W_{n-1}^{(-l)} \\
&= nW_n - (n-1)\binom{n-1}{3}^{-1} \binom{n}{3} W_n \\
& \quad + (n-1)\binom{n-1}{3}^{-1} \left[ \sum_{l<j<k} h(T_l, T_j, T_k) + \sum_{i<l<k} h(T_i, T_l, T_k) \right. \\
& \quad \left. + \sum_{i<j<l} h(T_i, T_j, T_l) \right] \\
&= -\frac{2n}{n-3} U_n + \frac{6}{(n-2)(n-3)} \left[ \sum_{l<j<k} h(T_l, T_j, T_k) + \sum_{i<l<k} h(T_i, T_l, T_k) \right. \\
& \quad \left. + \sum_{i<j<l} h(T_i, T_j, T_l) \right].
\end{aligned}$$

Now plugging in Eq. (1), one has that

$$
\begin{aligned}
&\hat{V}_l - \frac{2n}{n-3}U_n + \frac{6}{(n-2)(n-3)}\frac{n(n-1)(n-2)}{6n_1n_2n_3}\\
&\left\{\sum_{j<k}\left[I(X_{1l} < Y_{1,j-n_1} < Z_{1,k-n_1-n_2}) - I(X_{2l} < Y_{2,j-n_1} < Z_{2,k-n_1-n_2})\right]\right.\\
&I(1 \le l \le n_1 < j \le n_1+n_2 < k \le n)\\
&+\sum_{i<k}\left[I(X_{1i} < Y_{1l} < Z_{1,k-n_1-n_2}) - I(X_{2i} < Y_{2l} < Z_{2,k-n_1-n_2})\right]\\
&I(1 \le i \le n_1 < l \le n_1+n_2 < k \le n)\\
&+\sum_{i<j}[I(X_{1i} < Y_{1,j-n_1} < Z_{1,l}) - I(X_{2i} < Y_{2,j-n_1} < Z_{2l})]\\
&\left. I(1 \le i \le n_1 < j \le n_1+n_2 < l \le n)\right\}\\
&= -\frac{2n}{n-3}U_n + \frac{n(n-1)}{(n-3)}\frac{1}{n_1n_2n_3}\\
&\left\{\sum_{j=n_1+1}^{n_1+n_2}\sum_{k=n_1+n_2+1}^{n}[I(X_{1l} < Y_{1,j-n_1} < Z_{1,k-n_1-n_2})\right.\\
&- I(X_{2l} < Y_{2,j-n_1} < Z_{2,k-n_1-n_2})]I(1 \le l \le n_1)\\
&+\sum_{i=1}^{n_1}\sum_{k=n_1+n_2+1}^{n}[I(X_{1i} < Y_{1l} < Z_{1,k-n_1-n_2})\\
&- I(X_{2i} < Y_{2l} < Z_{2,k-n_1-n_2})]I(n_1 < l \le n_1+n_2)\\
&+\sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n_1+n_2}[I(X_{1i} < Y_{1,j-n_1} < Z_{1,l})\\
&\left. -I(X_{2i} < Y_{2,j-n_1} < Z_{2l})]I(n_1+n_2 < l \le n)\right\}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
E(\hat{V}_l) = -\frac{2n}{n-3}\theta + \frac{n(n-1)}{(n-3)}&\left[\frac{\theta}{n_1}I(1 \le l \le n_1) + \frac{\theta}{n_2}I(n_1 < l \le n_1+n_2)\right.\\
&\left.+\frac{\theta}{n_3}I(n_1+n_2 < l \le n)\right].
\end{aligned}
$$

By adopting the idea of Jing et al. (2009) and Pan et al. (2013), we define the jackknife empirical likelihood ratio for $\theta$ as follows,

$$
R(\theta) = \sup_{p_1,\ldots,p_n,}\left\{\prod_{i=1}^{n}(np_i) : p_i > 0, \sum_{i=1}^{n}p_i = 1, \sum_{i=1}^{n}p_i\hat{V}_i - \sum_{i=1}^{n}p_iE\hat{V}_i = 0\right\}.
$$

Using Lagrange multiplier method, we have $\log R(\theta) = -\sum_{l=1}^{n} \log(1 + \gamma(\hat{V}_l - E\hat{V}_l))$, and $\gamma$ is the solution to the following equation

$$\frac{1}{n}\sum_{l=1}^{n} \frac{\hat{V}_l - E\hat{V}_l}{1+\gamma(\hat{V}_l - E\hat{V}_l)} = 0. \tag{2}$$

We establish the Wilk's theorem for the JEL at true value $\theta_0$.

**Theorem 1** *We assume the following conditions hold.*

*(a)* $\sigma^2_{1,0,0} > 0$, $\sigma^2_{0,1,0} > 0$, $\sigma^2_{0,0,1} > 0$, *and*
*(b)* $\frac{n}{n_t} \to c_t < \infty$, *where* $t = 1, 2, 3$ *and* $c_t$*'s are finite constants. As* $\min(n_1, n_2, n_3) \to \infty$, *the empirical log-likelihood ratio statistic at the true value* $\theta_0$

$$l(\theta_0) = -2\log R(\theta_0) \xrightarrow{d} \chi^2_1,$$

*where* $\chi^2_1$ *is a standard* $\chi^2$ *distribution with degree of freedom 1.*

Thus, the asymptotic $100(1-\alpha)\%$ JEL confidence interval for $\theta$ is given by

$$\{\theta : l(\theta) \le \chi^2_1(\alpha)\},$$

where $\chi^2_1(\alpha)$ is the upper $\alpha$-quantile of $\chi^2_1$.

## 3 Numerical studies

In this section, we carry out extensive simulations to study the finite sample performance of the proposed JEL for the difference of two VUS's. We also construct the confidence intervals based on the normal approximation (NA) method. The normal approximation method can be found in Lemma 1 of Appendix. Based on Lemma 1, the $100(1-\alpha)\%$ confidence intervals based on the normal approximation method can be constructed as

$$I = \left\{\theta : |U_n - \theta| \le Z_{\alpha/2}\hat{\sigma}\right\},$$

where $Z_{\alpha/2}$ is the upper $\alpha/2$ critical value for the standard normal distribution and $\hat{\sigma}$ is defined in Appendix. We compare the two methods in terms of average length and coverage probability of confidence intervals.

For Tables 1 and 2, the data follow the Marshall–Olkin bivariate exponential distribution (MOBVE), as in Marshall and Olkin (1967) and Balakrishnan (1996). $MOBVE(\lambda_1, \lambda_2, \lambda_3)$ has a CDF

$$F(w_1, w_2) = 1 - exp[-\lambda_1 w_1 - \lambda_2 w_2 - \lambda_3 max\{w_1, w_2\}],$$

**Table 1** 95% Confidence intervals for the difference of two VUS's. $(X_1^*, X_2^*) \sim MOBVE(\lambda_{x_1}, \lambda_{x_2}, \lambda_{x_3})$, $(Y_1^*, Y_2^*) \sim MOBVE(\lambda_{y_1}, \lambda_{y_2}, \lambda_{y_3})$, $(Z_1^*, Z_2^*) \sim MOBVE(\lambda_{z_1}, \lambda_{z_2}, \lambda_{z_3})$

| $c$ | $(\lambda_{x_1}, \lambda_{x_2}, \lambda_{x_3}; \lambda_{y_1}, \lambda_{y_2}, \lambda_{y_3}; \lambda_{z_1}, \lambda_{z_2}, \lambda_{z_3})$ | $(n_1, n_2, n_3)$ | JEL | | NA | |
|---|---|---|---|---|---|---|
| | | | CP (%) | AL | CP (%) | AL |
| 0 | (1, 1, 0; 2, 2, 0; 3, 3, 0) | (10, 10, 10) | 90.8 | .203 | 90.0 | .190 |
| | | (20, 25, 30) | 94.1 | .126 | 92.4 | .120 |
| | | (30, 30, 30) | 94.8 | .108 | 93.6 | .104 |
| | | (60, 60, 60) | 94.6 | .074 | 94.5 | .072 |
| | | (80, 80, 80) | 94.5 | .064 | 94.7 | .063 |
| | | (100, 100, 100) | 95.5 | .057 | 95.3 | .056 |
| 0.25 | $(\frac{3}{5}, \frac{3}{5}, \frac{2}{5}; \frac{6}{5}, \frac{6}{5}, \frac{4}{5}; \frac{9}{5}, \frac{9}{5}, \frac{6}{5})$ | (10, 10, 10) | 91.1 | .188 | 89.6 | .176 |
| | | (20, 25, 30) | 94.4 | .112 | 93.5 | .106 |
| | | (30, 30, 30) | 94.9 | .096 | 93.7 | .093 |
| | | (60, 60, 60) | 95.9 | .066 | 95.2 | .064 |
| | | (80, 80, 80) | 94.2 | .056 | 93.5 | .055 |
| | | (100, 100, 100) | 94.2 | .050 | 93.6 | .049 |
| 0.5 | $(\frac{1}{3}, \frac{1}{3}, \frac{2}{3}; \frac{2}{3}, \frac{2}{3}, \frac{4}{3}; 1, 1, 2)$ | (10, 10, 10) | 89.6 | .167 | 86.9 | .158 |
| | | (20, 25, 30) | 94.4 | .100 | 93.0 | .094 |
| | | (30, 30, 30) | 94.2 | .087 | 93.0 | .083 |
| | | (60, 60, 60) | 92.8 | .058 | 90.8 | .056 |
| | | (80, 80, 80) | 95.3 | .049 | 94.6 | .049 |
| | | (100, 100, 100) | 95.7 | .044 | 95.1 | .044 |
| 0.75 | $(\frac{1}{7}, \frac{1}{7}, \frac{6}{7}; \frac{2}{7}, \frac{2}{7}, \frac{12}{7}; \frac{3}{7}, \frac{3}{7}, \frac{18}{7})$ | (10, 10, 10) | 90.1 | .151 | 87.1 | .144 |
| | | (20, 25, 30) | 95.1 | .088 | 92.4 | .084 |
| | | (30, 30, 30) | 95.9 | .075 | 93.1 | .073 |
| | | (60, 60, 60) | 94.8 | .051 | 93.6 | .050 |
| | | (80, 80, 80) | 95.9 | .044 | 94.7 | .043 |
| | | (100, 100, 100) | 93.5 | .039 | 92.2 | .038 |
| 0.9 | $(\frac{1}{19}, \frac{1}{19}, \frac{18}{19}; \frac{2}{19}, \frac{2}{19}, \frac{36}{19}; \frac{3}{19}, \frac{3}{19}, \frac{54}{19})$ | (10, 10, 10) | 93.5 | .143 | 90.1 | .136 |
| | | (20, 25, 30) | 94.6 | .080 | 92.3 | .078 |
| | | (30, 30, 30) | 94.7 | .069 | 93.7 | .067 |
| | | (60, 60, 60) | 95.4 | .046 | 94.9 | .046 |
| | | (80, 80, 80) | 94.9 | .040 | 94.6 | .039 |
| | | (100, 100, 100) | 95.4 | .035 | 94.6 | .035 |

The correlations $c_1 = c_2 = c_3 = c$, and sample sizes $n_{x_1} = n_{x_2} = n_1$, $n_{y_1} = n_{y_2} = n_2$, $n_{z_1} = n_{z_2} = n_3$
*JEL* jackknife empirical likelihood, *NA* normal approximation, *CP* (%) coverage probability, *AL* average length

where $w_1, w_2 > 0$, $\lambda_t \geq 0$ and at least one $\lambda_t$ is positive, $t = 1, 2, 3$. The marginal distributions of $(W_1, W_2)$ are exponential with expectations $(\lambda_1 + \lambda_3)$ and $(\lambda_2 + \lambda_3)$, respectively. Their correlation $c$ is $\lambda_3/(\lambda_1 + \lambda_2 + \lambda_3)$. In this simulation study, the first population $X = (X_1, X_2) = (\rho_x X_1^*, X_2^*)$, where $(X_1^*, X_2^*) \sim$

**Table 2** 90% Confidence intervals for the difference of two VUS's. $(X_1^*, X_2^*) \sim MOBVE(\lambda_{x_1}, \lambda_{x_2}, \lambda_{x_3})$, $(Y_1^*, Y_2^*) \sim MOBVE(\lambda_{y_1}, \lambda_{y_2}, \lambda_{y_3})$, $(Z_1^*, Z_2^*) \sim MOBVE(\lambda_{z_1}, \lambda_{z_2}, \lambda_{z_3})$

| $c$ | $(\lambda_{x_1}, \lambda_{x_2}, \lambda_{x_3}; \lambda_{y_1}, \lambda_{y_2}, \lambda_{y_3}; \lambda_{z_1}, \lambda_{z_2}, \lambda_{z_3})$ | $(n_1, n_2, n_3)$ | JEL | | NA | |
|---|---|---|---|---|---|---|
| | | | CP (%) | AL | CP (%) | AL |
| 0 | (1, 1, 0; 2, 2, 0; 3, 3, 0) | (10, 10, 10) | 87.2 | .168 | 85.0 | .190 |
| | | (20, 25, 30) | 88.7 | .105 | 87.8 | .101 |
| | | (30, 30, 30) | 89.8 | .090 | 88.3 | .087 |
| | | (60, 60, 60) | 89.7 | .062 | 88.9 | .061 |
| | | (80, 80, 80) | 90.2 | .053 | 89.4 | .053 |
| | | (100, 100, 100) | 88.8 | .047 | 88.6 | .047 |
| 0.25 | $(\frac{3}{5}, \frac{3}{5}, \frac{2}{5}; \frac{6}{5}, \frac{6}{5}, \frac{4}{5}; \frac{9}{5}, \frac{9}{5}, \frac{6}{5})$ | (10, 10, 10) | 86.9 | .156 | 85.8 | .148 |
| | | (20, 25, 30) | 90.6 | .093 | 88.9 | .089 |
| | | (30, 30, 30) | 90.4 | .080 | 89.5 | .078 |
| | | (60, 60, 60) | 90.5 | .055 | 90.3 | .054 |
| | | (80, 80, 80) | 90.0 | .047 | 89.8 | .046 |
| | | (100, 100, 100) | 90.3 | .042 | 88.7 | .041 |
| 0.5 | $(\frac{1}{3}, \frac{1}{3}, \frac{2}{3}; \frac{2}{3}, \frac{2}{3}, \frac{4}{3}; 1, 1, 2)$ | (10, 10, 10) | 85.7 | .139 | 83.7 | .133 |
| | | (20, 25, 30) | 90.0 | .083 | 87.5 | .080 |
| | | (30, 30, 30) | 90.5 | .072 | 89.5 | .071 |
| | | (60, 60, 60) | 87.3 | .048 | 86.4 | .047 |
| | | (80, 80, 80) | 91.2 | .041 | 90.8 | .041 |
| | | (100, 100, 100) | 90.0 | .037 | 89.9 | .037 |
| 0.75 | $(\frac{1}{7}, \frac{1}{7}, \frac{6}{7}; \frac{2}{7}, \frac{2}{7}, \frac{12}{7}; \frac{3}{7}, \frac{3}{7}, \frac{18}{7})$ | (10, 10, 10) | 86.3 | .126 | 83.3 | .121 |
| | | (20, 25, 30) | 90.4 | .073 | 89.1 | .070 |
| | | (30, 30, 30) | 91.4 | .062 | 89.8 | .061 |
| | | (60, 60, 60) | 88.5 | .042 | 87.6 | .042 |
| | | (80, 80, 80) | 91.0 | .037 | 89.2 | .036 |
| | | (100, 100, 100) | 87.7 | .032 | 87.3 | .032 |
| 0.9 | $(\frac{1}{19}, \frac{1}{19}, \frac{18}{19}; \frac{2}{19}, \frac{2}{19}, \frac{36}{19}; \frac{3}{19}, \frac{3}{19}, \frac{54}{19})$ | (10, 10, 10) | 90.0 | .119 | 87.1 | .114 |
| | | (20, 25, 30) | 91.0 | .067 | 88.7 | .065 |
| | | (30, 30, 30) | 91.3 | .057 | 89.1 | .056 |
| | | (60, 60, 60) | 90.1 | .039 | 89.6 | .038 |
| | | (80, 80, 80) | 91.9 | .033 | 91.0 | .033 |
| | | (100, 100, 100) | 90.2 | .029 | 90.0 | .029 |

The correlations $c_1 = c_2 = c_3 = c$, and sample sizes $n_{x_1} = n_{x_2} = n_1, n_{y_1} = n_{y_2} = n_2, n_{z_1} = n_{z_2} = n_3$
*JEL* jackknife empirical likelihood, *NA* normal approximation, *CP* (%) coverage probability, *AL* average length

$MOBVE(\lambda_{x_1}, \lambda_{x_2}, \lambda_{x_3})$, and $\rho_x = 3$. The second population $Y = (Y_1, Y_2) = (\rho_y Y_1^*, Y_2^*)$, where $(Y_1^*, Y_2^*) \sim MOBVE(\lambda_{y_1}, \lambda_{y_2}, \lambda_{y_3})$, and $\rho_y = 2$. The third population $Z = (Z_1, Z_2) = (\rho_z Z_1^*, Z_2^*)$, where $(Z_1^*, Z_2^*) \sim MOBVE(\lambda_{z_1}, \lambda_{z_2}, \lambda_{z_3})$, and $\rho_z = 1$. The $\lambda_{x_t}, \lambda_{y_t}, \lambda_{z_t}$'s differ for various correlations, where the correlations $c_1, c_2$, and $c_3$ are chosen as 0, 0.25, 0.5, 0.75, and 0.9. We also guarantee the marginal

**Table 3** 95% Confidence intervals for the difference of two VUS's. $X_1 \sim N(\mu_{x_1}, 1)$, $X_2 \sim N(\mu_{x_2}, 1)$, $Y_1 \sim N(\mu_{y_1}, 1)$, $Y_2 \sim N(\mu_{y_2}, 1)$, $Z_1 \sim N(\mu_{z_1}, 1)$, $Z_2 \sim N(\mu_{z_2}, 1)$

| $c$ | $(\mu_{x_1}, \mu_{x_2}, \mu_{y_1}, \mu_{y_2}, \mu_{z_1}, \mu_{z_2})$ | $(n_1, n_2, n_3)$ | JEL | | NA | |
|---|---|---|---|---|---|---|
| | | | CP (%) | AL | CP (%) | AL |
| 0 | (5, 3, 4, 2, 4, 1) | (10, 10, 10) | 90.1 | .182 | 89.2 | .171 |
| | | (20, 25, 30) | 93.6 | .112 | 92.7 | .106 |
| | | (30, 30, 30) | 94.8 | .097 | 93.3 | .093 |
| | | (60, 60, 60) | 95.7 | .067 | 94.0 | .065 |
| | | (80, 80, 80) | 95.4 | .057 | 94.8 | .056 |
| | | (100, 100, 100) | 94.8 | .051 | 94.5 | .050 |
| 0.25 | (5, 3, 4, 2, 4, 1) | (10, 10, 10) | 91.3 | .176 | 90.0 | .165 |
| | | (20, 25, 30) | 94.1 | .107 | 93.1 | .101 |
| | | (30, 30, 30) | 93.9 | .092 | 93.5 | .088 |
| | | (60, 60, 60) | 94.6 | .063 | 93.7 | .062 |
| | | (80, 80, 80) | 95.1 | .054 | 94.3 | .053 |
| | | (100, 100, 100) | 94.6 | .048 | 92.3 | .047 |
| 0.5 | (5, 3, 4, 2, 4, 1) | (10, 10, 10) | 89.2 | .163 | 88.0 | .153 |
| | | (20, 25, 30) | 93.6 | .101 | 92.1 | .096 |
| | | (30, 30, 30) | 93.9 | .084 | 91.4 | .081 |
| | | (60, 60, 60) | 94.7 | .058 | 93.9 | .056 |
| | | (80, 80, 80) | 95.1 | .050 | 94.6 | .049 |
| | | (100, 100, 100) | 95.2 | .044 | 93.8 | .043 |
| 0.75 | (5, 3, 4, 2, 4, 1) | (10, 10, 10) | 89.6 | .148 | 87.6 | .140 |
| | | (20, 25, 30) | 93.9 | .088 | 91.4 | .084 |
| | | (30, 30, 30) | 94.7 | .075 | 91.5 | .072 |
| | | (60, 60, 60) | 94.9 | .051 | 93.3 | .050 |
| | | (80, 80, 80) | 95.3 | .043 | 94.0 | .042 |
| | | (100, 100, 100) | 95.5 | .039 | 95.0 | .038 |
| 0.9 | (5, 3, 4, 2, 4, 1) | (10, 10, 10) | 91.3 | .134 | 88.4 | .127 |
| | | (20, 25, 30) | 93.8 | .079 | 91.3 | .075 |
| | | (30, 30, 30) | 95.6 | .067 | 92.3 | .065 |
| | | (60, 60, 60) | 95.6 | .046 | 94.8 | .045 |
| | | (80, 80, 80) | 95.6 | .039 | 95.1 | .038 |
| | | (100, 100, 100) | 95.1 | .035 | 94.7 | .034 |

The correlations $c_1 = c_2 = c_3 = c$, and sample sizes $n_{x_1} = n_{x_2} = n_1$, $n_{y_1} = n_{y_2} = n_2$, $n_{z_1} = n_{z_2} = n_3$
*JEL* jackknife empirical likelihood, *NA* normal approximation, *CP* (%) coverage probability, *AL* average length

distributions $X_1^* \sim exp(1)$, $X_2^* \sim exp(1)$, $Y_1^* \sim exp(2)$, $Y_2^* \sim exp(2)$, $Z_1^* \sim exp(3)$, and $Z_2^* \sim exp(3)$.

In Tables 3 and 4, the data are generated from the bivariate normal distributions. The distributions are: $(X_1, X_2) \sim N(\mu_x, \Sigma_x)$, $(Y_1, Y_2) \sim N(\mu_y, \Sigma_y)$, $(Z_1, Z_2) \sim$

**Table 4** 90% confidence intervals for the difference of two VUS's. $X_1 \sim N(\mu_{x_1}, 1)$, $X_2 \sim N(\mu_{x_2}, 1)$, $Y_1 \sim N(\mu_{y_1}, 1), Y_2 \sim N(\mu_{y_2}, 1), Z_1 \sim N(\mu_{z_1}, 1), Z_2 \sim N(\mu_{z_2}, 1)$. The correlations $c_1 = c_2 = c_3 = c$, and sample sizes $n_{x_1} = n_{x_2} = n_1, n_{y_1} = n_{y_2} = n_2, n_{z_1} = n_{z_2} = n_3$

| $c$ | $(\lambda_{x_1}, \lambda_{x_2}, \lambda_{y_1}, \lambda_{y_2}, \lambda_{z_1}, \lambda_{z_2})$ | $(n_1, n_2, n_3)$ | JEL | | NA | |
|---|---|---|---|---|---|---|
| | | | CP (%) | AL | CP (%) | AL |
| 0 | (5, 3, 4, 2, 4, 1) | (10, 10, 10) | 86.1 | .151 | 85.3 | .143 |
| | | (20, 25, 30) | 88.9 | .093 | 88.1 | .089 |
| | | (30, 30, 30) | 90.9 | .080 | 89.2 | .078 |
| | | (60, 60, 60) | 91.0 | .056 | 90.3 | .055 |
| | | (80, 80, 80) | 90.3 | .048 | 89.8 | .047 |
| | | (100, 100, 100) | 89.5 | .043 | 89.5 | .042 |
| 0.25 | (5, 3, 4, 2, 4, 1) | (10, 10, 10) | 86.9 | .146 | 85.0 | .138 |
| | | (20, 25, 30) | 90.1 | .089 | 88.7 | .085 |
| | | (30, 30, 30) | 90.9 | .076 | 89.5 | .074 |
| | | (60, 60, 60) | 89.9 | .053 | 88.5 | .052 |
| | | (80, 80, 80) | 90.9 | .045 | 90.0 | .045 |
| | | (100, 100, 100) | 88.5 | .040 | 87.4 | .039 |
| 0.5 | (5, 3, 4, 2, 4, 1) | (10, 10, 10) | 86.5 | .135 | 83.8 | .128 |
| | | (20, 25, 30) | 89.7 | .084 | 88.0 | .081 |
| | | (30, 30, 30) | 88.5 | .070 | 86.5 | .068 |
| | | (60, 60, 60) | 91.4 | .048 | 89.6 | .047 |
| | | (80, 80, 80) | 92.3 | .041 | 91.7 | .041 |
| | | (100, 100, 100) | 88.2 | .037 | 88.3 | .036 |
| 0.75 | (5, 3, 4, 2, 4, 1) | (10, 10, 10) | 86.8 | .123 | 85.0 | .117 |
| | | (20, 25, 30) | 90.7 | .073 | 87.4 | .070 |
| | | (30, 30, 30) | 89.6 | .062 | 87.2 | .060 |
| | | (60, 60, 60) | 90.6 | .043 | 89.1 | .042 |
| | | (80, 80, 80) | 90.8 | .036 | 90.0 | .036 |
| | | (100, 100, 100) | 90.4 | .032 | 89.7 | .032 |
| 0.9 | (5, 3, 4, 2, 4, 1) | (10, 10, 10) | 87.8 | .111 | 85.0 | .107 |
| | | (20, 25, 30) | 90.5 | .065 | 87.3 | .063 |
| | | (30, 30, 30) | 90.9 | .055 | 88.1 | .054 |
| | | (60, 60, 60) | 92.9 | .038 | 91.2 | .038 |
| | | (80, 80, 80) | 91.6 | .032 | 90.8 | .032 |
| | | (100, 100, 100) | 90.6 | .029 | 90.0 | .029 |

*JEL* Jackknife empirical likelihood, *NA* normal approximation, *CP* (%) coverage probability, *AL* average length

$N(\mu_z, \Sigma_z)$, where $\mu_x = (5, 3)$, $\mu_y = (4, 2)$, and $\mu_z = (4, 2)$, and the covariance matrices are

$$\Sigma_x = \Sigma_y = \Sigma_z = \begin{pmatrix} 1 & c \\ c & 1 \end{pmatrix}$$

as the correlation $c$ varies.

The sample sizes for $x$, $y$ and $z$ of $(n_{x_1}, n_{x_2}, n_{y_1}, n_{y_2}, n_{z_1}, n_{z_2})$ are (10, 10, 10, 10, 10, 10), (20, 20, 25, 25, 30, 30), (30, 30, 30, 30, 30, 30), (60, 60, 60, 60, 60, 60), (80, 80, 80, 80, 80, 80), and (100, 100, 100, 100, 100, 100). The nominal levels of the confidence intervals are 95 and 90%; 1000 iterations are repeated to generate the data.

From Tables 1, 2, 3 and 4, we make the following conclusions.

1. For different correlations, sample sizes, and parameters of the distributions, the coverage probabilities of the confidence intervals based on JEL methods and NA methods are close to nominal levels.
2. In almost all the scenarios, as the sample sizes increase, the coverage probabilities of the confidence intervals for the two methods get closer to the nominal level, and the average lengths of the intervals decrease. This is reasonable since larger sample sizes provide more information of the data.
3. For the same sample sizes, as the correlations increase, the coverage probabilities of the confidence intervals for the two methods are closer to the nominal level, and the average lengths of the intervals decrease. JEL interval estimates outperform the normal approximation interval estimates for various sample sizes and correlation coefficients.

## 4 Real data analysis

In this section, the proposed confidence intervals of the difference of two VUS's are illustrated using a data set of the diagnosis for early-stage Alzheimer's disease (AD) from the Alzheimer's disease Research Center (ADRC) at Washington University (see Xiong et al. 2006). The severity of dementia of Alzheimer type can be staged by the clinical dementia rating (CDR), a score based on several clinical evaluations and neuropsychometric measurements. We concentrate on the following three diagnostic groups: non-demented group (CDR 0), very mildly demented group (CDR 0.5), and mildly demented group (CDR 1). The data set includes 14 neuropsychometric markers from 118 cases aged 75 falling into the three diagnostic categories mentioned above. Out of the 14 measures, we compare the diagnostic accuracies between the scores from two neuropsychometric tests. One of them is a measure of semantic memory, named as the information subset of the Wechsler Adult Intelligence Scale (WAIS), see Wechsler (1955). The other is an untimed visuospatial measure called Visual Retention Test (Form D, copy), as in Storandt and Hill (1989).

By deleting the individuals with results of missing values, we have 22 patients from mildly demented group (CDR 1), 44 patients from very mildly demented group (CDR 0.5), and 45 participants from non-demented group (CDR 0).

For CDR 1 group, the sample mean is $(-2.125, -1.769)$, the sample covariance matrix is

$$\begin{pmatrix} 1.298 & 0.786 \\ 0.786 & 5.751 \end{pmatrix},$$

and the correlation of the two attributes is 0.288.

For CDR 0.5 group, the sample mean is $(-0.607, -0.551)$, the sample covariance matrix is

$$\begin{pmatrix} 1.167 & 1.302 \\ 1.302 & 3.476 \end{pmatrix},$$

and the correlation of the two attributes is 0.647.

For CDR 0 group, the sample mean is $(0.631, 0.202)$, the sample covariance matrix is

$$\begin{pmatrix} 0.712 & 0.164 \\ 0.164 & 0.445 \end{pmatrix},$$

and the correlation of the two attributes is 0.292.

The interval estimate of the difference of the two VUS's based on the JEL method is $(0.350, 0.634)$ at 90% confidence level and $(0.324, 0.662)$ at 95% confidence level. The NA confidence interval is $(0.375, 0.604)$ at 90% confidence level and $(0.353, 0.627)$ at 95% confidence level. Therefore, the information subset of the WAIS possesses a stronger discrimination power than that of Visual Retention Test (Form D, copy).

## 5 Discussion

In this paper, we make elaborate efforts to provide an alternative method in evaluating diagnostic tests through the jackknife empirical likelihood procedure. A new inference technique is constructed to compare the diagnostic treatments in discriminating three-class data. We apply bivariate three-sample $U$-statistic to obtain interval estimates for the difference of VUS's and establish the Wilk's theorem for the $U$-statistic rigorously. The corresponding coverage probability and average length of the confidence intervals are calculated based on the Wilk's theorem. Our JEL method for the bivariate three-sample $U$-statistic is an extension of the existing JEL methods for the univariate multi-sample $U$-statistic (see Jing et al. 2009 and Pan et al. 2013). We also presented the normal approximation method to make inference for the difference of two correlated VUS's. The extensive simulation studies show the advantages of the JEL method over the normal approximation method in terms of coverage probability for the confidence intervals.

In the future, we will investigate the adjusted JEL confidence intervals for the difference of two correlated VUS's to improve the coverage probability. On the other hand, we will also study the partial volume under surface (PVUS), which is another important and powerful measure for the evaluation of the diagnostic tests. Finally, we will explore the JEL confidence intervals for VUS and PVUS with incomplete data.

**Acknowledgements** The authors would like to thank two reviewers and the associate editor for their helpful comments, which lead to a significant improvement in the paper. Yichuan Zhao appreciates the support from the NSA Grant (H98230-12-1-0209) and NSF Grants (DMS-1613176).

## Appendix: Proof of Theorem 1

The variance $Var(U_n)$ can be estimated by a consistent estimator $\hat{\sigma}^2$ as in Sen (1960) and Arvesen (1969),

$$\hat{\sigma}^2 = \frac{1}{n_1(n_1-1)}\sum_{i=1}^{n_1}(V_{i,0,0}-\bar{V}_{\cdot,0,0})^2 + \frac{1}{n_2(n_2-1)}\sum_{j=1}^{n_2}(V_{0,j,0}-\bar{V}_{0,\cdot,0})^2 + \frac{1}{n_3(n_3-1)}\sum_{k=1}^{n_3}(V_{0,0,k}-\bar{V}_{0,0,\cdot})^2.$$

**Lemma 1** *We have the following conclusions.*

*(a) The U-statistic* $U_n \xrightarrow{a.s.} \theta_0$ *as* $min(n_1, n_2, n_3) \to \infty$;
*(b) Suppose that* $\sigma^2_{1,0,0} > 0$, $\sigma^2_{0,1,0} > 0$, $\sigma^2_{0,0,1} > 0$, *and denote* $S^2_{n_1,n_2,n_3} = \sigma^2_{1,0,0}/n_1 + \sigma^2_{0,1,0}/n_2 + \sigma^2_{0,0,1}/n_3$. *As* $min(n_1, n_2, n_3) \to \infty$,

$$\frac{U_n-\theta_0}{S_{n_1,n_2,n_3}} \xrightarrow{d} N(0,1), \tag{3}$$

*and*

$$\hat{\sigma}^2 - S^2_{n_1,n_2,n_3} = o_p((min(n_1,n_2,n_3))^{-1}). \tag{4}$$

The proof of part (a) and Eqs. (3) and (4) can be found in Arvesen (1969) and Kowalski and Tu (2007).

**Lemma 2** *Let* $S_n = n^{-1}\sum_{l=1}^{n}(\hat{V}_l - E\hat{V}_l)^2$. *We assume the same conditions as (a) and (b) in Theorem 1. Then as* $n_1 \to \infty$,

$$S_n = nS^2_{n_1,n_2,n_3} + o_p(1).$$

*Proof of Lemma 2* For $1 \le l \le n_1$, it is clear that

$$\hat{V}_l - E\hat{V}_l = \frac{n(n-1)}{(n-3)n_1}(V_{l,0,0}-U_n) + \frac{n(n-2n_1-1)}{(n-3)n_1}(U_n-\theta_0),$$

and

$$\begin{aligned}
&\frac{1}{n_1}\sum_{l=1}^{n_1}(V_{l,0,0}-U_n)(U_n-\theta_0)\\
&=(U_n-\theta_0)\left\{\frac{1}{n_1n_2n_3}\sum_{i=1}^{n1}\sum_{j=1}^{n2}\sum_{k=1}^{n3}[I(X_{1i}<Y_{1j}<Z_{1k})-I(X_{2i}<Y_{2j}<Z_{2k})]-U_n\right\}\\
&=0.
\end{aligned}$$

As Pan et al. (2013) and Wang (2010) did, we have that

$$\sum_{l=1}^{n_1}(\hat{V}_l-E\hat{V}_l)^2=\left[\frac{n(n-1)}{(n-3)n_1}\right]^2\sum_{l=1}^{n_1}(V_{l,0,0}-U_n)^2+\left[\frac{n(n-2n_1-1)}{(n-3)n_1}\right]^2 n_1(U_n-\theta_0)^2.$$

For $n_1 < l \leq n_1+n_2$, one has that

$$\sum_{l=n_1+1}^{n_1+n_2}(\hat{V}_l-E\hat{V}_l)^2=\left[\frac{n(n-1)}{(n-3)n_2}\right]^2\sum_{l=n_1}^{n_1+n_2}(V_{0,l,0}-U_n)^2 +\left[\frac{n(n-2n_2-1)}{(n-3)n_2}\right]^2 n_2(U_n-\theta_0)^2.$$

For $n_1+n_2 < l \leq n$, we have that (see Pan et al. 2013)

$$\sum_{l=n_1+n_2+1}^{n}(\hat{V}_l-E\hat{V}_l)^2=\left[\frac{n(n-1)}{(n-3)n_3}\right]^2\sum_{l=n_1+n_2}^{n}(V_{0,0,l}-U_n)^2 +\left[\frac{n(n-2n_3-1)}{(n-3)n_3}\right]^2 n_3(U_n-\theta_0)^2.$$

Therefore,

$$\begin{aligned} S_n=&\frac{1}{n}\left[\frac{n(n-1)}{(n-3)}\right]^2\left[\frac{1}{n_1^2}\sum_{l=1}^{n_1}(V_{l,0,0}-\bar{V}_{\cdot,0,0})^2+\frac{1}{n_2^2}\sum_{l=n_1+1}^{n_1+n_2}(V_{0,l,0}-\bar{V}_{0,\cdot,0})^2\right.\\ &\left.+\frac{1}{n_3^2}\sum_{l=n_1+n_2+1}^{n}(V_{0,0,l}-\bar{V}_{0,0,\cdot})^2\right]\\ &+\frac{1}{n}\left[\frac{n}{(n-3)}\right]^2\left[\frac{(n-2n_1-1)^2}{n_1}+\frac{(n-2n_2-1)^2}{n_2}+\frac{(n-2n_3-1)^2}{n_3}\right](U_n-\theta_0)^2. \end{aligned} \tag{5}$$

From the LLN of $U$-statistic, we have the conclusion $U_n-\theta_0 = O_p(n_1^{-1/2})$. The second term in Eq. (5) is

$$\frac{n}{(n-3)^2}\left[\frac{(n-2n_1-1)^2}{n_1}+\frac{(n-2n_2-1)^2}{n_2}+\frac{(n-2n_3-1)^2}{n_3}\right](U_n-\theta_0)^2 = O_p(n^{-1}).$$

Moreover, the 1st term of Eq. (5) is (cf. Wang 2010)

$$
\begin{aligned}
& n\left(\frac{n-1}{n-3}\right)^2\left[\frac{1}{n_1^2}\sum_{l=1}^{n_1}\left(V_{l,0,0}-\bar{V}_{\cdot,0,0}\right)^2+\frac{1}{n_2^2}\sum_{l=n_1+1}^{n_1+n_2}\left(V_{0,l,0}-\bar{V}_{0,\cdot,0}\right)^2\right.\\
& \left.+\frac{1}{n_3^2}\sum_{l=n_1+n_2+1}^{n}(V_{0,0,l}-\bar{V}_{0,0,\cdot})^2\right]\\
&= n\hat{\sigma}^2+o_p(1).
\end{aligned}
$$

Using Eq. (4), we prove Lemma 2. □

**Lemma 3** *Let* $Q_n = \max_{1\le l\le n} |\hat{V}_l - \theta_0|$*. Under the same conditions as in Lemma 2, we have* $Q_n = o_p(n^{1/2})$ *and* $n^{-1}\sum_{l=1}^{n} |\hat{V}_l - \theta_0|^3 = o_p(n^{1/2})$*.*

*Proof of Lemma 3* For $1 \le l \le n_1$, we have (see Wang 2010)

$$
|\hat{V}_l - E\hat{V}_l| \le \left|\frac{n}{n_1}\frac{n-1}{n-3}V_{l,0,0}\right| + \left|\frac{n}{n_1}\frac{n-1}{n-3}U_n\right| + \left|\frac{n(n-2n_1-1)}{(n-3)n_1}(U_n-\theta_0)\right|.
$$

Note that $|V_{l,0,0}| \le \tilde{H}_n$, and $|U_n| \le \tilde{H}_n$, where

$$
\tilde{H}_n = \max_{1\le i\le n_1<l\le n_1+n_2<k\le n} |h(X_i, Y_j, Z_k)|.
$$

Therefore, $|\hat{V}_l - E\hat{V}_l| \le c^*\tilde{H}_n + c^*\tilde{H}_n + c^*|U_n - \theta|$, where $c^*$ is a constant. Similar to Pan et al. (2013) and Wang (2010), we have $\tilde{H}_n = o_p(n^{1/2})$ and $U_n - \theta_0 = O_p(n^{-1/2})$. Therefore, $|\hat{V}_l - E\hat{V}_l| = o_p(n^{1/2})$ for $1 \le l \le n_1$. For $n_1 < l \le n_1 + n_2$, $|\hat{V}_l - E\hat{V}_l| \le 2c^*\tilde{H}_n + c^*|U_n - \theta_0|$. Thus, $|\hat{V}_l - E\hat{V}_l| = o_p(n^{1/2})$ for $n_1 < l \le n_1 + n_2$. For $n_1 + n_2 < l \le n$, we have $|\hat{V}_l - E\hat{V}_l| = o_p(n^{1/2})$ similarly as before. Thus $Q_n = o_p(n^{1/2})$. Hence,

$$
n^{-1}\sum_{l=1}^{n} |\hat{V}_l - E\hat{V}_l|^3 = o_p(n^{1/2})(nS^2_{n_1,n_2,n_3} + o_p(1)) = o_p(n^{1/2}).
$$

□

*Proof of Theorem 1* The proof follows the same lines of Pan et al. (2013). Recall $U_n = 1/n\sum_{l=1}^{n}\hat{V}_l$ and $\theta_0 = 1/n\sum_{l=1}^{n} E\hat{V}_l$. Then

$$
|U_n - \theta_0| \ge |\gamma|\frac{1}{1+|\gamma|\max|\hat{V}_l - E\hat{V}_l|}\frac{1}{n}\sum_{l=1}^{n}(\hat{V}_l - E\hat{V}_l)^2 \ge |\gamma|\frac{S_n}{1+|\gamma|Q_n}.
$$

We have that $|\gamma| = O_p(n^{-1/2})$. Using Taylor's expansion, one has that

$$-2\log R(\theta_0) = 2\sum_{l=1}^{n}\left\{\gamma(\hat{V}_l - E\hat{V}_l) - \frac{1}{2}\left[\gamma(\hat{V}_l - E\hat{V}_l)\right]^2\right\} + o_p(1). \tag{6}$$

Let $F_0 = 1/n\sum_{l=1}^{n}\frac{\gamma^2(\hat{V}_l - E\hat{V}_l)^3}{1+\gamma(\hat{V}_l - E\hat{V}_l)}$. By Eq. (2), we have that

$$2n\gamma(U_n - \theta_0) - nS_n\gamma^2 = n\frac{(U_n - \theta_0)^2}{S_n} - \frac{nF_0^2}{S_n}.$$

One can obtain $F_0 = o_p(n^{-1/2})$. Thus Eq. (6) can be re-expressed as follows,

$$-2\log R(\theta_0) = n\frac{(U_n - \theta_0)^2}{S_n} + o_p(1).$$

Combining Lemmas 1, 2 and 3, we finish the proof. □

## References


Arvesen, J. (1969). Jackknifing U-statistics. *The Annals of Mathematical Statistics*, *40*(6), 2096–2100.

Balakrishnan, K. (1996). *Exponential distribution: Theory, methods and applications*. Amsterdam: CRC Press. illustrated edition.

Gong, Y., Peng, L., Qi, Y. (2010). Smoothed jackknife empirical likelihood method for ROC curve. *Journal of Multivariate Analysis*, *101*, 1520–1531.

Jing, B. Y., Yuan, J., Zhou, W. (2009). Jackknife empirical likelihood. *Journal of the American Statistical Association*, *104*, 1224–1232.

Kowalski, J., Tu, X. M. (2007). *Modern applied U-statistics*. Hoboken, New Jersey: Wiley.

Marshall, A. W., Olkin, I. (1967). A multivariate exponential distribution. *Journal of American Statistical Association*, *62*, 30–41.

Mossman, D. (1999). Three-way ROC's. *Medical Decision Making*, *19*(1), 78–89.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, *75*(2), 237–249.

Owen, A. B. (1990). Empirical likelihood confidence regions. *The Annals of Statistics*, *18*, 90–120.

Pan, G., Wang, X., Zhou, W. (2013). Nonparametric statistical inference for P(X < Y < Z). *The Indian Journal of Statistics*, *75-A*, 118–138.

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

Storandt, M., Hill, R. D. (1989). Very mild senile dementia of the Alzheimer type. II. Psychometric test performance. *Archives of Neurology*, *46*(4), 383–386.

Tian, L., Xiong, C., Lai, C. Y., Vexler, A. (2011). Exact confidence interval estimation for the difference in diagnostic accuracy with three ordinal diagnostic groups. *Journal of Statistical Planning and Inference*, *141*(1), 549–558.

Wan, S. (2012). An empirical likelihood confidence interval for the volume under ROC surface. *Statistics and Probability Letters*, *82*(7), 1463–1467.

Wang, X. (2010). *Empirical likelihood with applications*. Thesis, National University of Singapore.

Wechsler, D. (1955). *Manual for the Wechsler adult intelligence scale* (Vol. vi). Oxford: Psychological Corporation.

Xiong, C., van Belle, G., Miller, J. P., Morris, J. C. (2006). Measuring and estimating diagnostic accuracy when there are three ordinal diagnostic groups. *Statistics in Medicine*, *2006*(25), 1251–1273.

Yang, H., Zhao, Y. (2013). Smoothed jackknife empirical likelihood inference for the difference of ROC curves. *Journal of Multivariate Analysis*, *115*, 270–284.

Yang, H., Zhao, Y. (2015). Smoothed jackknife empirical likelihood inference for ROC curves with missing data. *Journal of Multivariate Analysis*, *140*, 123–138.