Spring 4-25-2011

A New Jackknife Empirical Likelihood Method for U-Statistics

Zhengbo Ma

A NEW JACKKNIFE EMPIRICAL LIKELIHOOD METHOD FOR U-STATISTICS

by

ZHENGBO MA

Under the Direction of Dr. Yichuan Zhao

ABSTRACT

U-statistics generalizes the concept of mean of independent identically distributed (*i.i.d.*) random variables and is widely utilized in many estimating and testing problems. The standard empirical likelihood (EL) for U-statistics is computationally expensive because of its nonlinear constraint. The jackknife empirical likelihood method largely relieves computation burden by circumventing the construction of the nonlinear constraint. In this thesis, we adopt a new jackknife empirical likelihood method to make inference for the general volume under the ROC surface (VUS), which is one typical kind of U-statistics. Monte Carlo simulations are conducted to show that the EL confidence intervals perform well in terms of the coverage probability and average length for various sample sizes.

INDEX WORDS: Confidence interval, U-statistics, Jackknife empirical likelihood

A NEW JACKKNIFE EMPIRICAL LIKELIHOOD METHOD FOR U-STATISTICS

by

ZHENGBO MA

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

in the College of Arts and Sciences

Georgia State University

2011

A NEW JACKKNIFE EMPIRICAL LIKELIHOOD METHOD FOR U-STATISTICS

by

ZHENGBO MA

Committee Chair:	Dr. Yichuan Zhao
Committee:	Dr. Yuanhui Xiao
	Dr. Ruiyan Luo
	Dr. Xu Zhang

Electronic Version Approved:

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
May 2011

ACKNOWLEDGEMENTS

First and foremost, I am most grateful to my advisor Professor Yichuan Zhao for sharing his wealth of knowledge and wisdom with me. I profitted greatly from his personality when facing troubles. I would like to gratefully and sincerely acknowledge those who have assisted me in my graduate study.

I would also like to thank my thesis committee, Dr. Yuanhui Xiao, Dr. Ruiyan Luo, and Dr. Xu Zhang for taking their precious time to read my thesis and providing me valuable suggestions.

Besides, I would like to acknowledge all the professors in the Department of Mathematics and Statistics at Georgia State University who taught me in my graduate study. They all assisted me to develop my statistical knowledge and skills.

I also wish to thank some student colleagues in Department of Statistics with whom I profitted a lot from useful discussions: Meng Zhao, Hanfang Yang, Ye Cui and Huayu Liu. Finally, I would thank my parents for their unconditional love and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1

INTRODUCTION

1.1 Empirical Likelihood

Empirical likelihood approach was first introduced by Owen (1988) to construct confidence regions for the mean of a random vector. It is an effective and flexible nonparametric method based on a data-driven likelihood ratio function, rather than a assumption that the whole data come from a known family of distributions. The empirical likelihood method enjoys the advantages of both nonparametric method and likelihood ratio function, such as producing confidence regions whose shape and orientation are determined completely and automatically by the data. It also has better asymptotic power properties and small sample performance.

After Owen's proposed his pioneering work, much attention has been attracted by the properties of the empirical likelihood approach. The empirical likelihood approach has been extended and applied in many different fields such as the work of Chen and Hall (1993), Qin and Lawless (1994) on estimating equations and the work of Ren (2008) and Keziou and Leoni (2008) on the two-sample problem. We refer to the bibliography of Owen (2001) for more extensive references.

In the following, we will give a brief description of the procedure of empirical likelihood for population mean. Let $F(x)$ be a common distribution function and $\{X_1, ..., X_n\}$ be independent and identically distributed (*i.i.d.*) sample from this distribution and θ is the population mean, which is also the parameter of interest. In the framework of empirical likelihood approach, we will replace the real distribution $F(x)$ by the weighted empirical distribution with such form of: $\sum_{i=1}^{n} p_i I(X_i \leq x)$, where $I(\cdot)$ is the indicator function and $\{p_i\}$ satisfies that $\sum_{i=1}^{n} p_i = 1$ and $p_i \geq 0 \quad (i = 1, ..., n)$. Then, the mean of this weighted empirical distribution is: $\theta_0 = \sum_{i=1}^{n} p_i X_i$.

The empirical likelihood function for θ, evaluated at $\theta = \theta_0$, is defined to be:

$$\prod_{i=1}^{n} p_i, \quad \sum_{i=1}^{n} p_i = 1, \quad p_i \geq 0, \quad \sum_{i=1}^{n} p_i X_i = \theta_0 \tag{1.1}$$

Define the empirical likelihood ratio basing on the above definition:

$$sup\{\prod_{i=1}^{n} np_i, \quad \sum_{i=1}^{n} p_i = 1, \quad p_i \geq 0, \quad \sum_{i=1}^{n} p_i X_i = \theta_0\}, \tag{1.2}$$

and its log-form, i.e. the log-empirical likelihood ratio is:

$$sup\{\sum_{i=1}^{n} log(np_i), \quad \sum_{i=1}^{n} p_i = 1, \quad p_i \geq 0, \quad \sum_{i=1}^{n} p_i X_i = \theta_0\}. \tag{1.3}$$

To optimize (1.3), the Lagrange multiplier method is applied. Let γ be the lagrangian multiplier. Then, we have:

$$p_i = \frac{1}{n} \frac{1}{1 + \gamma(X_i - \theta_0)} \quad i = 1, ..., n,$$

where the lagrangian multiplier γ satisfies:

$$\sum_{i=1}^{n} \frac{X_i - \theta_0}{1 + \gamma(X_i - \theta_0)} = 0$$

By the expression of $p_i \quad (i = 1, ..., n)$, the log-empirical likelihood ratio is:

$$-\sum_{i=1}^{n} log(1 + \gamma(X_i - \theta_0)). \tag{1.4}$$

Let $l(\theta_0)$ denote this log-empirical likelihood ratio. By the work of Owen, $-2l(\theta_0)$ converges to χ_1^2 in distribution by central limit theorem. This conclusion makes it possible to construct an $(1-\alpha)$ level confidence region for θ as:

$$\{\theta : -2l(\theta) \leq a\},$$

where a is chosen to satisfy $P\{\chi_1^2 \leq a\} = 1 - \alpha$.

1.2 U-statistics

The U-statistics, introduced by Halmos (1946) and Hoeffding (1948), is important in statistical practice. Considering K *i.i.d.* samples of random vectors, $y_{k,i}$ $(1 \leq i \leq n_k, 1 \leq k \leq K)$, the general U-statistics U_n with kernel function h of m_k arguments for the kth sample is defined as:

$$U_n = \left[\prod_{k=1}^{K} \binom{n_k}{m_k}\right]^{-1} \sum_{k=1}^{K} \sum_{(i_1,...,i_{m_k}) \in C_{m_k}^{n_k}} h(y_{1,i_1}, ...y_{1,i_{m_1}}; ...; y_{K,i_1}, ..., y_{K,i_{m_K}})$$

which is an unbiased estimation of $E(h)$.

The class of U-statistics includes many statistics in common use. Its consistency and asymptotic normality were proved in Hoeffding (1948). The distributional properties and its simple structure make them ideal for studying many estimating and testing problem. Thus, one useful application of U-statistics is to generate new statistics in practical cases. In the recent years, the research interest in this subject has been constantly increasing and led many academic works. One can refer to Lee (1990), and Koroljuk and Borovskich (1994) for detailed expositions of U-statistics.

1.3 Standard Empirical Likelihood for U-statistics

For doing inference for the expectations, one may attempt to apply the standard empirical likelihood method to U-statistics. Following the standard procedure, confidence intervals for the parameter of interest might be constructed by deriving asymptotic distribution for the empirical log-likelihood ratio of U-statistics. However, there will be heavy computation burdens since we need to solve several simultaneous nonlinear constraints. This question regarding the presence of nonlinear constraints was also explored by Wood et al. (1996). Let us take one-sample U-statistics with 2 degrees for example. Let $X_1, ..., X_n$ are independent and identically distributed ($i.i.d.$) random variables with common distribution function $F(x)$ and ϕ is the kernel function. Then the U-statistic is defined to be:

$$W_n = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} \phi(X_i, X_j),$$

where $\theta = E(\phi(X_i, X_j))$ is the parameter of interest.

By Wood's definition for W_n, the empirical likelihood and the log-empirical likelihood ratio should be:

$$\prod_{i=1}^{n} p_i, \quad \sum_{i=1}^{n} p_i = 1, \quad p_i \ge 0, \quad \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} n^2 p_i p_j \phi(X_i, X_j) = \theta_0 \tag{1.5}$$

$$sup \left\{ \sum_{i=1}^{n} log(np_i), \quad \sum_{i=1}^{n} p_i = 1, \quad p_i \ge 0, \quad \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} n^2 p_i p_j \phi(X_i, X_j) = \theta_0 \right\}. \tag{1.6}$$

We can obtain the empirical likelihood ratio for W_n By solving (1.6). However,there is no simple way available for such optimization problem involving n variables $p_1, ..., p_n$ with the nonlinear constraints. The numerical methods can give an approximate solution, but it has no help on deriving asymptotic distribution of the result from (1.6). One can refer to Jing et al. (2009) for excellent interpretations.

1.4 Jackknife Empirical Likelihood for U-statistics

To cope with the computational difficulty arising in the above section, an modified empirical likelihood approach is proposed by Jing et al. (2009) and Wang (2010), called as jackknife empirical likelihood.

Let's continue the above example of W_n to briefly describe the JEL procedure.

Applying the standard jackknife method to W_n, we obtain the jackknife pseudo-values :

$$\tilde{V}_s = nW_n - (n-1)W_{n-1}^{-s}, \quad (s = 1, ..., n)$$

where W_{n-1}^{-s} is the U-statistic after removing X_s. It is obvious that $E\tilde{V}_s = \theta_0$ due to the unbiasedness of U-statistics. For this pseudo sample, we can still construct one weighted empirical distribution with the probability vector $\mathbf{P} = \{p_1, ..., p_n\}$, and write its mean as: $\sum_{s=1}^{n} p_s\tilde{V}_s$. Applying the idea of standard empirical likelihood approach to the pseudo sample, we can define the jackknife empirical likelihood at θ_0 as:

$$\prod_{s=1}^{n} p_s, \quad \sum_{s=1}^{n} p_s = 1, \quad p_s \geq 0, \quad \sum_{s=1}^{n} p_s\tilde{V}_s = \theta_0$$

Hence, the corresponding jackknife empirical likelihood ratio and its log-form are:

$$sup\{\prod_{s=1}^{n}(np_s), \quad \sum_{s=1}^{n} p_s = 1, \quad p_s \geq 0, \quad \sum_{s=1}^{n} p_s\tilde{V}_s = \theta_0\}$$

$$sup\{\sum_{s=1}^{n} log(np_s), \quad \sum_{s=1}^{n} p_s = 1, \quad p_s \geq 0, \quad \sum_{s=1}^{n} p_s \tilde{V}_s = \theta_0\}$$

By the Lagrange multiplier method, the jackknife empirical log-likelihood ratio at θ_0 can be rewritten as:

$$l(\theta_0) = -\sum_{s=1}^{n} log(1 + \gamma(\tilde{V}_s - \theta_0)),$$

where γ satisfies the equation:

$$\sum_{s=1}^{n} \frac{\tilde{V}_s - \theta_0}{1 + \gamma(\tilde{V}_s - \theta_0)} = 0.$$

Jing et al. (2009) and Wang (2010) has proven the asymptotic distribution of $-2l(\theta_0)$ is χ_1^2. By this conclusion, the $(1-\alpha)$-level confidence interval for θ_0 can be constructed. Since circumvent the nonlinear constraint of optimization problem, the advantage of jackknife empirical likelihood on computation is apparent.

1.5 Motivation of the Thesis

Receiver operating characteristic (ROC) curve has been developed as an important tool to distinguish the quality of given classifier in diagnostic tests in decades. The area under the ROC curve (AUC) is a related topic for evaluating the accuracy of diagnostic tests of two-category classification data. Bamber (1975) shows that AUC is exactly $P(X < Y)$, the probability that a randomly selected observation from one population scores less than that from another population, which is the most commonly used measure of diagnostic accuracy for a continuous-scale diagnostic test. However, most of ROC analysis and AUC have been restricted to a classifier with just two classes. However, many real applications involve more than two classes and demand a methodology expansion. Mossman (1999) extended such three-class problems to the volume under the ROC surface (VUS). However, the multi-class problem obviously is more complex than the two-class one, because at least $d(d-1)$ dimensional variable for d classes are needed for obtaining their volumes, which is trivial when $d = 2$. The increase on dimensions can be seen as the costs for the various misclassifications. If we ignore the misclassifications, i.e., no specificities are concerned, then the VUS can be expressed simply as, see Nakas and Yiannoutsos (2004): $VUS = P(X < Y < Z)$. Most recently, Li (2009) also proposed a generalization of VUS for ordered multi-class problem, which is the linear combination of probabilities of the possible inequality relations between the three random variables X, Y, Z:

$$VUS' = \left(\frac{a_2}{a_2 + a_3} P(Y < Z) + \frac{a_3}{2(a_2 + a_3)}\right)\left(\frac{a_2}{a_2 + a_1} P(X < Y) + \frac{a_1}{2(a_2 + a_1)}\right) P(X < Z),$$

where $a_1 + a_2 + a_3 = 1$, $a_1 \geq 0$, $a_2 \geq 0$ and $a_3 \geq 0$.

In this thesis, we apply the JEL approach to make statistical inference for the simple form of VUS, $P(X < Y < Z)$ and its generalization respectively. The asymptotic distribution theory on the JEL statistics is also provided.

1.6 Structure

The rest of the thesis is organized as follows. In chapter 2, the jackknife empirical likelihood ratio statistic is constructed, the limiting distribution of the statistic is given, and the jackknife empirical likelihood based confidence interval for the U-statistics is constructed. In chapter 3, we report that the results of a simulation study on the finite sample performance of jackknife empirical likelihood based confidence interval on parameter of interest. The conclusion is given in chapter 4, and all the technical derivations are provided in the Appendix A.

Chapter 2

INFERENCE PROCEDURE

2.1 Multi-sample U-statistics and VUS

Consider independent samples, $X_1, X_2, ..., X_{n_1}$ from $F_1(x)$; $Y_1, Y_2, ..., Y_{n_2}$ from $F_2(y)$ and $Z_1, Z_2, ..., Z_{n_3}$ from $F_3(z)$. Let the indicator function $h(x, y, z) = I(x < y < z)$ be a kernel function, and denote θ_0 by the parameter of interest: $P(X < Y < Z)$, the most simple form of VUS. Then, it is trivial that $\theta_0 = P(X < Y < Z) = E\{I(x < y < z)\}$. By the definition in section 1.2, we can construct a U-statistic of degree (1,1,1) with the indicator kernel $I(x < y < z)$ in the form of:

$$U_n = \frac{1}{n_1 n_2 n_3} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} I(X_i < Y_j < Z_k), \tag{2.1}$$

which is a consistent and unbiased estimator of the parameter $\theta_0 = P(X < Y < Z)$. Similarly, for the generalized form of VUS, we can use the kernel function: $h(x, y, z) = \left(\frac{a_2}{a_2+a_3} I(y < z) + \frac{a_3}{2(a_2+a_3)}\right) \left(\frac{a_2}{a_2+a_1} I(x < y) + \frac{a_1}{2(a_2+a_1)}\right) I(x < z)$, where $a_1 + a_2 + a_3 = 1$, $a_1 \geq 0$, $a_2 \geq 0$ and $a_3 \geq 0$, to construct a U-statistic in the form of:

$$U_n = \frac{1}{n_1 n_2 n_3} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} h(X_i, Y_j, Z_k), \tag{2.2}$$

2.2 Jackknife Empirical Likelihood for VUS

The jackknife empirical likelihood (JEL) approach for multi-sample U-statistics is the extension of the JEL approach for one-sample U-statistics corresponding to the pooled sample of the multi-sample case. The pooled sample is $(T_1, T_2, ..., T_n) = (X_1, X_2, ..., X_{n_1}, Y_1, Y_2, ..., Y_{n_2}, Z_1, Z_2, ..., Z_{n_3})$, where $n = n_1 + n_2 + n_3$. For the simple form of VUS, $P(X < Y < Z)$, the corresponding U-statistics is:

$$\tilde{U}_n = \binom{n}{3}^{-1} \sum_{1 \le i < j < k \le n} \frac{n(n-1)(n-2)}{6n_1 n_2 n_3} I(X_i < Y_j < Z_k) I(1 \le i \le n_1 < j \le n_1 + n_2 < k \le n), \tag{2.3}$$

where the function $\frac{n(n-1)(n-2)}{6n_1n_2n_3} I(X_i < Y_j < Z_k) I(1 \le i \le n_1 < j \le n_1 + n_2 < k \le n)$ is also a function with respect to the sample sizes: n_1, n_2, n_3. Similar to the simple form, for the generalized form of VUS, $\left(\frac{a_2}{a_2+a_3} P(Y < Z) + \frac{a_3}{2(a_2+a_3)}\right) \left(\frac{a_2}{a_2+a_1} P(X < Y) + \frac{a_1}{2(a_2+a_1)}\right)$ $P(X < Z)$, $(a_1 + a_2 + a_3 = 1)$, the corresponding U-statistics is:

$$\tilde{U}_n = \binom{n}{3}^{-1} \sum_{1 \le i < j < k \le n} \frac{n(n-1)(n-2)}{6n_1 n_2 n_3} h(X_i, Y_j, Z_k) I(1 \le i \le n_1 < j \le n_1 + n_2 < k \le n), \tag{2.4}$$

where $h(X_i, Y_j, Z_k) = \left(\frac{a_2}{a_2+a_3} I(Y_j < Z_k) + \frac{a_3}{2(a_2+a_3)}\right) \left(\frac{a_2}{a_2+a_1} I(X_i < Y_j) + \frac{a_1}{2(a_2+a_1)}\right) I(X_i < Z_k)$, and the product of the above kernel and indicator is also a function with respect to the sample sizes: n_1, n_2, n_3.

By the idea of jackknife U-statistics, let $\tilde{U}_{n-1}^{-i}$ denote the U-statistic with respect to the sample deleting ith observation and the jackknife pseudo-value for such sample deleting ith observation be $\hat{V}_i = n\tilde{U}_n - (n-1)\tilde{U}_{n-1}^{-i}$. Note that $\tilde{U}_n = U_n$. It is trivial that $U_n = \frac{1}{n} \sum_{i=1}^{n} \hat{V}_i$.

Let $U_{n_1,n_2,n_3}^0 = U_n$ be the original U-statistics;
$U_{n_1-1,n_2,n_3}^{-i,0,0} = \frac{1}{(n_1-1)n_2n_3} \sum_{i'=1, i' \ne i}^{n_1} \sum_{j'=1}^{n_2} \sum_{k'=1}^{n_3} h(X_{i'} < Y_{j'} < Z_{k'})$, which is the U-statistics

after deleting X_i;

$U^{0,-j,0}_{n_1,n_2-1,n_3} = \frac{1}{n_1(n_2-1)n_3} \sum_{i'=1}^{n_1} \sum_{j'=1,j'\neq j}^{n_2} \sum_{k'=1}^{n_3} h(X_{i'} < Y_{j'} < Z_{k'})$, which is the U-statistics after deleting Y_j;

$U^{0,0,-k}_{n_1,n_2,n_3-1} = \frac{1}{n_1 n_2(n_3-1)} \sum_{i'=1}^{n_1} \sum_{j'=1}^{n_2} \sum_{k'=1,k'\neq k}^{n_3} h(X_{i'} < Y_{j'} < Z_{k'})$, which is the U-statistics after deleting Z_k.

Let $V_{i,0,0} = n_1 U^0_{n_1,n_2,n_3} - (n_1-1)U^{-i,0,0}_{n_1-1,n_2,n_3}$, $V_{0,j,0} = n_2 U^0_{n_1,n_2,n_3} - (n_2-1)U^{0,-j,0}_{n_1,n_2-1,n_3}$, and $V_{0,0,k} = n_3 U^0_{n_1,n_2,n_3} - (n_3-1)U^{0,0,-k}_{n_1,n_2,n_3-1}$.

Further, it can be shown easily that:

$$\begin{aligned}\hat{V}_i =& U_n + \frac{n-1}{n_1-1}(V_{i,0,0} - U_n)I(1 \le i \le n_1) + \frac{n-1}{n_2-1}(V_{0,j,0} - U_n)I(n_1+1 \le i \le n_1+n_2) \\ &+ \frac{n-1}{n_3-1}(V_{0,0,k} - U_n)I(n_1+n_2+1 \le i \le n),\end{aligned} \tag{2.5}$$

and all expectations of $\hat{V}_i$ are the same: $E\hat{V}_i = \theta_0$, $(1 \le i \le n)$.

Based on the pseudo sample of $\{\hat{V}_i\}_1^n$, which is asymptotically independent (Shao and Tu, 1995), let us apply usual empirical likelihood approach to estimate θ_0, the parameter of interest. Let $\{p_i\}_{i=1}^n$ satisfy that: $\sum_i^n p_i = 1$ and $p_i \le 0$, $(1 \le i \le n)$. The jackknife empirical likelihood (JEL) function for a estimated parameter θ_0, is evaluated as:

$$L(\theta_0) = sup\{\prod_{i=1}^n p_i : \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \hat{V}_i = \theta_0\}, \tag{2.6}$$

and the corresponding JEL ratio is:

$$R(\theta_0) = sup\{\prod_{i=1}^n np_i : \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \hat{V}_i = \theta_0\}. \tag{2.7}$$

If the condition: $\min_{1\le i\le n} \hat{V}_i < \theta_0 < \max_{1\le i\le n} \hat{V}_i$ holds, then the solution to the above optimal problem is:

$$p_i = \frac{1}{n}\frac{1}{1+\gamma(\hat{V}_i - \theta_0)}, \tag{2.8}$$

where γ satisfies

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{V}_i}{1+\gamma(\hat{V}_i - \theta_0)} = \theta_0. \tag{2.9}$$

By (2.8), we can rewrite the jackknife empirical likelihood ratio as:

$$R(\theta_0) = \prod_{i=1}^{n}\frac{1}{1+\gamma(\hat{V}_i - \theta_0)}, \tag{2.10}$$

and the jackknife empirical log-likelihood ratio is:

$$logR(\theta_0) = -\sum_{i=1}^{n} log(1+\gamma(\hat{V}_i - \theta_0)). \tag{2.11}$$

The following theorem 1 states that the the asymptotic distribution of the jackknife empirical log-likelihood ratio statistic still follows Wilks' theorem. Its proof is given in Appendix A.

Theorem 1. *Let* $\sigma^2_{1,0,0} = Var(E(h(X,Y,Z)|X))$, $\sigma^2_{0,1,0} = Var(E(h(X,Y,Z)|Y))$, $\sigma^2_{0,0,1} = Var(E(h(X,Y,Z)|Z))$. *Assume that:* $\sigma^2_{1,0,0} > 0$, $\sigma^2_{0,1,0} > 0$ *and* $\sigma^2_{0,0,1} > 0$*;* $0 < \underline{\lim}_{n\to\infty}\frac{n_1}{n_2} \leq \overline{\lim}_{n\to\infty}\frac{n_1}{n_2} < \infty$, $0 < \underline{\lim}_{n\to\infty}\frac{n_2}{n_3} \leq \overline{\lim}_{n\to\infty}\frac{n_2}{n_3} < \infty$*; and* $Eh^2(X,Y,Z) < \infty$*. Then, as* $\min(n_1,n_2,n_3) \to \infty$*, at the true value* θ_0*, we have:*

$$-2logR(\theta_0) \xrightarrow{d} \chi^2_1. \tag{2.12}$$

By theorem 1, an approximate $(1-\alpha)$ level confidence interval for θ_0 can be constructed as:

$$\boldsymbol{\Theta}_c = \{\theta : -2logR(\theta) \leq c\}, \tag{2.13}$$

where c is chosen to satisfy $P(\chi_1^2 \leq c) = 1 - \alpha$.

Chapter 3

NUMERICAL STUDIES

3.1 Monte Carlo Simulation

In this section, based on the theorems of the Jackknife Empirical likelihood (JEL), extensive simulation studies are conducted to explore the performance of the confidence intervals from this procedure for the generalized and simple VUS. We compare the coverage accuracy of the proposed modified jackknife empirical likelihood method (mJEL) with the original jackknife empirical likelihood method (JEl) in Wang (2010) and the empirical likelihood method (EL) in Li (2009). We consider four cases: the first two cases are the generalized forms of VUS and the last two case are the simple forms of VUS.

Firstly, in the simulation studies on generalized VUS, we choose: **case (1)**, $F_1 = N(1, 0.5)$, $F_2 = N(2, 0.5)$ and $F_3 = N(3, 0.5)$; **case (2)**, $F_1 = N(1, 3)$, $F_2 = N(2, 3)$ and $F_3 = N(3, 3)$. We generate 10,000 random samples from the above cases with sample sizes (16, 8, 16), (40, 20, 40) and (60, 30, 60). The coefficients are fixed as $a_1 : a_2 : a_3 = 2 : 1 : 2$.

Secondly, in the simulation studies on simple VUS, we choose: **case (3)**, $F_1 = N(0, 1)$, $F_2 = N(1, 1)$ and $F_3 = N(1, 2)$; **case (4)**, $F_1 = exp(8)$, $F_2 = exp(1)$ and $F_3 = exp(1/4)$.We generate 10,000 random samples from the above cases with sample sizes (15, 15, 15), (30, 30, 30) and (50, 50, 50).

In addition, one case about the small sample performance and calibration of the jackknife empirical likelihood is studied. An adjusted empirical likelihood (AEL) method proposed by Chen (2008) is applied to JEL procedure, trying to promote the coverage probability. we choose: **case (5)**, $F_1 = N(-3, 1)$, $F_2 = exp(1)$ and $F_3 = Cauchy(6, 1)$.We generate 5,000 random samples from the above cases with sample sizes (6, 6, 6).

Table 3.1. Confidence intervals for Normal distribution cases:

		mJEL		*JEL*		*EL*	
Sample size	*C.L.*	*C.P.*	*A.L.*	*C.P.*	*A.L.*	*C.P.*	*A.L.*
(16, 8, 16)	0.90	0.9019	0.0593	0.9019	0.0593	0.8712	0.0515
	0.95	0.9416	0.0721	0.9416	0.0721	0.9200	0.0626
	0.99	0.9871	0.0978	0.9871	0.0978	0.9554	0.0849
(40, 20 ,40)	0.90	0.8983	0.0399	0.8983	0.0399	0.8697	0.0346
	0.95	0.9470	0.0493	0.9470	0.0493	0.9252	0.0427
	0.99	0.9895	0.0711	0.9895	0.0711	0.9679	0.0617
(60, 30, 60)	0.90	0.9003	0.0283	0.9003	0.0283	0.8710	0.0244
	0.95	0.9506	0.0345	0.9506	0.0345	0.9279	0.0306
	0.99	0.9897	0.0458	0.9897	0.0458	0.9749	0.0407

NOTE:
The three distributions are $N(1, 0.5)$, $N(2, 0.5)$ and $N(3, 0.5)$.
$a_1 : a_2 : a_3 = 2 : 1 : 2$.
The true value of the parameter of interest is: 0.4124.
C.L. is confidence level,
C.P. is coverage probability,
A.L. is the average length of the interval.

Table 3.2. Confidence intervals for Normal distribution cases:

		mJEL		*JEL*		*EL*	
Sample size	*C.L.*	*C.P.*	*A.L.*	*C.P.*	*A.L.*	*C.P.*	*A.L.*
(16, 8, 16)	0.90	0.9019	0.0929	0.9019	0.0929	0.8692	0.0901
	0.95	0.9526	0.1211	0.9526	0.1211	0.9176	0.1103
	0.99	0.9871	0.1597	0.9871	0.1597	0.9554	0.1498
(40, 20 ,40)	0.90	0.9083	0.0899	0.9113	0.0908	0.8697	0.0726
	0.95	0.9577	0.1042	0.9591	0.1056	0.9279	0.0867
	0.99	0.9885	0.1305	0.9898	0.1324	0.9699	0.1041
(60, 30, 60)	0.90	0.8973	0.0833	0.8989	0.0836	0.8761	0.0694
	0.95	0.9502	0.0990	0.9511	0.0995	0.9339	0.0863
	0.99	0.9908	0.1258	0.9914	0.1270	0.9608	0.1173

NOTE:
The three distributions are $N(1,3)$, $N(2,3)$ and $N(3,3)$.
$a_1 : a_2 : a_3 = 2 : 1 : 2$.
The true value of the parameter of interest is: 0.2159.
C.L. is confidence level,
C.P. is coverage probability,
A.L. is the average length of the interval.

Table 3.3. Confidence intervals for Normal distribution cases:

		mJEL		*JEL*		*EL*	
Sample size	*C.L.*	*C.P.*	*A.L.*	*C.P.*	*A.L.*	*C.P.*	*A.L.*
(15, 15, 15)	0.90	0.9085	0.2968	0.9085	0.2968	0.8786	0.2675
	0.95	0.9532	0.3553	0.9532	0.3553	0.9317	0.3119
	0.99	0.9891	0.4693	0.9891	0.4693	0.9664	0.4128
(30, 30 ,30)	0.90	0.8949	0.2025	0.8949	0.2025	0.8743	0.1756
	0.95	0.9442	0.2417	0.9442	0.2417	0.9227	0.2108
	0.99	0.9895	0.3193	0.9895	0.3193	0.9669	0.277
(50, 50, 50)	0.90	0.8888	0.1548	0.8888	0.1548	0.8631	0.1336
	0.95	0.9402	0.1846	0.9402	0.1846	0.9193	0.1603
	0.99	0.9865	0.2458	0.9865	0.2458	0.9648	0.2113

NOTE:
The three distributions are $N(0,1)$, $N(1,1)$ and $N(1,2)$.
The true value of the parameter of interest is: 0.3407.
C.L. is confidence level,
C.P. is coverage probability,
A.L. is the average length of the interval.

Table 3.4. Confidence intervals for Exponential distribution cases:

		mJEL		*JEL*		*EL*	
Sample size	*C.l.*	*C.P.*	*A.L.*	*C.P.*	*A.L.*	*C.P.*	*A.L.*
(15, 15, 15)	0.90	0.9152	0.3052	0.9152	0.3052	0.8946	0.2652
	0.95	0.9563	0.3659	0.9563	0.3659	0.9351	0.3171
	0.99	0.9895	0.4871	0.9895	0.4871	0.9688	0.4238
(30, 30 ,30)	0.90	0.9077	0.2075	0.9077	0.2075	0.8874	0.1805
	0.95	0.9559	0.2491	0.9559	0.2491	0.9341	0.2162
	0.99	0.9918	0.3293	0.9918	0.3293	0.9694	0.2378
(50, 50, 50)	0.90	0.9075	0.1586	0.9075	0.1586	0.8863	0.1376
	0.95	0.9552	0.1894	0.9552	0.1894	0.9333	0.1643
	0.99	0.9911	0.2498	0.9911	0.2498	0.9684	0.2173

NOTE:
The three distributions are $Exp(8)$, $Exp(1)$ and $Exp(1/4)$.
The true value of the parameter of interest is: 0.6919.
C.L. is confidence level,
C.P. is coverage probability,
A.L. is the average length of the interval.

Table 3.5. Confidence intervals for three different distributions cases:

		mJEL		*JEL*		*AEL*	
Sample size	*C.l.*	*C.P.*	*A.L.*	*C.P.*	*A.L.*	*C.P.*	*A.L.*
(6, 6, 6)	0.90	0.4488	0.2198	0.4488	0.2198	0.4505	0.2421
	0.95	0.4489	0.2646	0.4489	0.2646	0.4492	0.2936
	0.99	0.4492	0.3591	0.4492	0.3591	0.4492	0.4117

NOTE:
The three distributions are $N(-3,1)$, $Exp(1)$ and $Cauchy(6,1)$.
The true value of the parameter of interest is: 0.9317.
C.L. is confidence level,
C.P. is coverage probability,
A.L. is the average length of the interval.

Chapter 4

SUMMARY AND FUTURE WORK

4.1 Summary

In this thesis, a new jackknife empirical likelihood method is proposed to construct the confidence intervals for VUS in the simple and generalized forms. The jackknife empirical likelihood ratio statistic can be proved to converge to the chi-square distribution asymptotically.

First, the proposed JEL method runs faster than the other traditional method as observed in simulation studies. And the simulation studies evaluate the finite sample numerical performance of the inference. The results from the new JEL and the original JEL are almost the same. All coverage probabilities of JEL are close to the corresponding nominal levels of 90%, 95% and 99%; and the larger sample sizes lead to more accurate coverage probabilities and smaller average length of the confidence intervals as well. However, with the accurate coverage probabilities, the average lengths of confidence intervals of JEL are larger than the average lengths from the traditional EL method.

For the small sample case, Chen's aEL is a useful method to promote the coverage probability for traditional empirical likelihood procedure. However, in the JEL case, this method seems not to work well. The coverage probabilities from the three methods are close to each other and the average length of the confidence intervals from Chen's method is significantly larger than the lengthes from the other naive methods.

4.2 Future Work

In the future, we can continue the study in more than one way.
First, real data sets can be applied to testify the performance of the proposed method. Second, to obtain a more efficient confidence interval, we can try to plug the bootstrap or other calibration methods into the procedure. Third, we will try this JEL procedure to deal with the missing data problem of VUS.

In summary, the research of VUS can be further investigated in many different aspects.

REFERENCES

[1] Bamber, D., The area above the ordinal dominance graph and the area below the receiver operating characteristic graph, *Journal of Mathematical Psychology*, Vol. 4, pp. 941-953, 1975.

[2] Chen, J. and Rao, J. N. K., Asymptotic normality under two-phase sampling designs, *Statistica Sinica*, Vol. 17, pp. 1047-1064, 2007.

[3] Chen, S. X. and Hall, P., Smoothed empirical likelihood confidence intervals for quantiles, *The Annals of Statistics*, Vol. 21, pp 1166-1181, 1993.

[4] Halmos, P. R., The theory of unbiased estimation, *Annals of Mathematical Statistics,*, Vol. 17, pp. 34-43, 1946.

[5] Hoeffding, W., A class of statistics with asymptotically normal distribution, *Annals of Mathematical Statistics,*, Vol. 19, pp. 293-325, 1948.

[6] Hsieh, F. and Turnbull, B. W., Nonparametric and semiparametric estimation of the receiver operating characteristic curve, *The Annals of Statistics*, Vol. 24, pp. 25-40, 1996.

[7] Jing, B. Y., Yuan, J. Q. and Zhou, W., Jackknife empirical likelihood, *Journal of the American Statistical Association*, Vol. 104, pp. 1224-1232, 2009.

[8] Keziou, A. and Leoni-Aubin, S., On empirical likelihood for semiparametric two-sample density ratio models, *J. Statist. Plann. Inference*, Vol. 138, pp. 915C928, 2008.

[9] Koroljuk, V. S. and Borovskich, Yu. V., *Theory of U-statistics, Mathematics and Its Applications*, Vol. 273, Kluwer Academic Publishers, 1994.

[10] Lehmann, E. L., Consistency and unbiasedness of certain nonparametric tests, *Ann. Math. Statistics*, Vol. 22, pp. 165C179, 1951.

[11] Lee, A. J., *U-statistics, Theory and Practice*, New York: Marcel Dekker,Inc., 1990.

[12] Li, Y., A generalization of AUC to an ordered multi-class diagnosis and application to longitudinal data analysis on intellectual outcome in pediatric braintumor patients, *Ph. D. Thesis*, Georgia State University, 2009.

[13] Mossman,D., Three-way ROCs, *Medical Decision Making*, Vol. 19, pp. 78-89, 1999.

[14] Nakas,C. T. and Yiannoutsos,C. T., Ordered multiple-class ROC analysis with continuous measurements, *Statistics in Medicine*, Vol. 23, pp. 3437-3449, 2004.

[15] Owen, A. B., Empirical likelihood ratio confidence intervals for a single functional, *Biometrika*, Vol. 75, pp. 237-249, 1988.

[16] Owen, A. B., Empirical likelihood ratio confidence regions. The Annals of Statistics, *Biometrika*, Vol. 18, pp. 90-120, 1990.

[17] Pepe, M. S., The Statistical Evaluation of Medical Tests for Classification and Prediction, *Oxford: Oxford University Press.*

[18] Qin, J. and Lawless, J., Empirical likelihood and general estimating equations, *The Annals of Statistics*, Vol. 22, pp. 300-325, 1994.

[19] Ren, J., Weighted empirical likelihood in some two-sample semiparametric models with various types of censored data, *Annals of Statistics*, Vol. 36, pp. 147-166, 2008.

[20] Shao, J. and Tu, D., *The Jackknife and Bootstrap*, Springer-Verlag, 1995.

[21] Su, H., Qin, Y. and Liang, H., Empirical Likelihood-Based Confidence Interval of ROC Curves, *Statistics in Biopharmaceutical Research*, Vol. 1, pp. 407-414, 2009.

[22] Wang, X., Empirical likelihoood with applications, *Ph. D. Thesis*, National University of Singapore, 2010.

[23] Wood, A.T.A., Do, K.A. and Broom, N.M., Sequential linearization of empirical likelihood constraints with application to U-statistics, *J. Comput. Graph. Stat.*, Vol. 5, pp. 365-385, 1996.

[24] Zou, K. H., Hall, W. J. and Shapiro, D. E. , Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests, *Statistics in Medicine*, Vol. 16, pp. 2143-2156, 1997.

APPENDICES

Appendix A: Lemmas and Proofs

The following two lemmas of Wang (2009) will be used later.

Lemma 1. *Suppose that* $\{X_i\}_{i=1}^{n_1}$, $\{Y_j\}_{j=1}^{n_2}$ *and* $\{Z_k\}_{k=1}^{n_3}$ *are three independent samples. Let* $\{\hat{V}_i\}_{i=1}^{n}$ *be the pseudo sample constructed by the procedure in section 2. Assume that* $\sigma^2_{1,0,0} > 0$, $\sigma^2_{0,1,0} > 0$ *and* $\sigma^2_{0,0,1} > 0$*; and* $\overline{\lim}_{n\to\infty} \frac{n_1}{n_2} > 0$, $\overline{\lim}_{n\to\infty} \frac{n_2}{n_3} > 0$. *Then as* $\min(n_1, n_2, n_3) \to \infty$, *we have:*

$$P\{\min_{1\le i\le n} \hat(V)_i < \theta_0 < \max_{1\le i\le n} \hat(V)_i\} \to 1$$

.

Let $\hat{\sigma}^2 = \frac{1}{n_1(n_1-1)} \sum_{i=1}^{n_1} \left(V_{i,0,0} - \frac{1}{n_1} \sum_{i=1}^{n_1} V_{i,0,0}\right)^2 + \frac{1}{n_2(n_2-1)} \sum_{j=1}^{n_2} \left(V_{0,j,0} - \frac{1}{n_2} \sum_{j=1}^{n_2} V_{0,j,0}\right)^2 + \frac{1}{n_3(n_3-1)} \sum_{k=1}^{n_3} \left(V_{0,0,k} - \frac{1}{n_3} \sum_{k=1}^{n_3} V_{0,0,k}\right)^2$ and $S^2_{n_1,n_2,n_3} = \frac{\sigma^2_{1,0,0}}{n_1} + \frac{\sigma^2_{0,1,0}}{n_2} + \frac{\sigma^2_{0,0,1}}{n_3}$.

By Lehmann (1951) and Koroljuk and Borovshich (1994), we have a central limit theorem for U-Statistics:

Lemma 2. *(a)* $U_n \xrightarrow{a.s.} \theta_0$ *as* $min(n_1, n_2, n_3) \to \infty$*;*

(b) Assume that $\sigma^2_{1,0,0} > 0$, $\sigma^2_{0,1,0} > 0$ *and* $\sigma^2_{0,0,1} > 0$*,then*

$$\frac{U_n - \theta}{S_{n_1,n_2,n_3}} \xrightarrow{d} N(0,1), as \min(n_1, n_2, n_3) \to \infty, \tag{1}$$

and

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

To prove the main theorem, we need some additional lemmas. Without loss of generality, we always suppose that $n_1 \leq n_2 \leq n_3$.

Lemma 3. *Let* $S_n = \frac{1}{n}\sum_{i=1}^{n}(\hat{V}_i - \theta_0)^2$. *Under the conditions of lemma*1, *as* $\min(n_1, n_2, n_3) \to \infty$, $S_n = nS^2_{n_1,n_2,n_3} + o(1)$ *a.s.*.

Proof of Lemma 3. Let $\xi_{n_i} = \psi(\hat{V}_i - \theta_0)$, where $\psi(x)$ is nondecreasing, twice differentiable with bounded first and second derivatives such that:

$$\psi(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ a(x) & \text{if } 0 < x < \delta \\ 1 & \text{if } x \geq \delta \end{cases}$$

with $0 < a(x) < 1$ for $0 < x < \delta$.

For $1 \leq i \leq n_1$, by the definition of pseudo value $\hat{V}_i$, we have:

$$\begin{aligned}\hat{V}_i - \theta_0 = (U_n - \theta_0) + \frac{n-1}{n_1 - 1}(V_{i,0,0} - U_n)I(1 \leq i \leq n_1) + \frac{n-1}{n_2 - 1}(V_{0,j,0} - U_n)I(n_1 + 1 \leq i \leq n_1 + n_2) \\ + \frac{n-1}{n_3 - 1}(V_{0,0,k} - U_n)I(n_1 + n_2 + 1 \leq i \leq n),\end{aligned}$$

Then by lemma 2, we have:

$$
\begin{aligned}
S_n &= \frac{1}{n}\sum_{i=1}^{n}(\hat{V}_i - \theta_0)^2 \\
&= \frac{(n-1)^2}{n}[\frac{1}{(n_1-1)^2}\sum_{i=1}^{n_1}(\hat{V}_{i,0,0} - U_n)^2 + \frac{1}{(n_2-1)^2}\sum_{j=n_1+1}^{n_1+n_2}(\hat{V}_{0,j,0} - U_n)^2 \\
&+ \frac{1}{(n_3-1)^2}\sum_{i=n_1+n_2+1}^{n}(\hat{V}_{0,0,k} - U_n)^2] + (U_n - \theta_0)^2 \\
&= n\hat{\sigma}^2 + o(1) \qquad a.s. \\
&= nS^2_{n_1,n_2,n_3} + o(1) \qquad a.s.
\end{aligned}
$$

□

Lemma 4. *Let* $H_n = \max_{1\le i\le n_1<j\le n_1+n_2<k\le n} |h(X_i, Y_j, Z_k)|$. *Under the conditions of lemma 1, we have* $H_n = o(n^{1/2})$ *a.s.*

Proof of Lemma 4.

$$
\begin{aligned}
H_n &= \max_{1\le i\le n_1<j\le n_1+n_2<k\le n} |h(X_i, Y_j, Z_k)| \\
&= n^{-1/2} \max_{1\le i<j<k\le n} |h(X_i, Y_{j-n_1}, Z_{k-n_1-n_2})I(1 \le i \le n_1 < j \le n_1 + n_2 < k \le n)|.
\end{aligned}
$$

where $I(1 \le i \le n_1 < j \le n_1 + n_2 < k \le n)$ is an indicator function. By the above equation, we can consider of another form of the maximum:

$$
H_n = \max_{1\le i<j<k\le n} |\tilde{h}(X_i, Y_j, Z_k)|,
$$

where $\tilde{h}(X_i, Y_j, Z_k) = h(X_i, Y_{j-n_1}, Z_{k-n_1-n_2})I(1 \le i \le n_1 < j \le n_1 + n_2 < k \le n)$.

Then, for any two integers n' and n'' satisfying $1 < n' < n'' \le n$, by Markov's inequality and Borel-Cantelli Lemma,

$$n'^{-1/2} \max_{1 \le i < n'} |\tilde{h}(X_i, Y_{n'}, Z_{n''})| \to 0 \quad a.s..$$

Since

$$\begin{aligned} & n^{-1/2} \max_{1 \le i < j < k \le n} |\tilde{h}(X_i, Y_j, Z_k)| \\ =& n^{-1/2} \max_{1 < k \le n} \{\max_{j<k}\{\max_{i<j} |\tilde{h}(X_i, Y_j, Z_k)|\}\} \\ \le & \max_{1 < k \le n} \{k^{-1/2} \max_{j<k}\{\max_{i<j} |\tilde{h}(X_i, Y_j, Z_k)|\}\} \\ \le & \max_{1 < k \le n} \{\max_{j<k}\{j^{-1/2} \max_{i<j} |\tilde{h}(X_i, Y_j, Z_k)|\}\}. \end{aligned}$$

Then, the two above statement implies that $n^{-1/2} \max_{1 \le i < j < k \le n} |\tilde{h}(X_i, Y_j, Z_k)| \to 0$ *a.s.*, which completes the proof. □

Lemma 5. *Let $K_n = \max_{1 \le i \le n} |\hat{V}_i - \theta_0|$. Under the conditions of lemma 1, $K_n = o(n^{1/2})$ a.s. and $\frac{1}{n} \sum_{i=1}^{n} |\hat{V}_i - \theta_0|^3 = o(n^{1/2})$ a.s..*

Proof of Lemma 5. By the definition of the pseudo sample $\{\hat{V}_i\}_{i=1}^n$,

$$|\hat{V}_i - \theta_0| \le (2C-1)H_n + |\theta_0|, \quad (1 \le i \le n)$$

where C satisfies that $\max(\frac{n-1}{n_1-1}, \frac{n-1}{n_2-1}, \frac{n-1}{n_3-1}) \le C < \infty$. Since $H_n = o(n^{1/2})$, then

$$K_n = o(n^{1/2}) a.s. \tag{3}$$

By lemma 3 and under the conditions in this lemma:

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}|\hat{V}_i-\theta_0|^3 &\le \frac{1}{n}\sum_{i=1}^{n}|\hat{V}_i-\theta_0|^2 K_n \\
&= S_n K_n \\
&\le C'(\sigma^2_{1,0,0}+\sigma^2_{0,1,0}+\sigma^2_{0,0,1})o(n^{1/2}) \\
&= o(n^{1/2}).
\end{aligned}
$$

where C' satisfies that $\max(\frac{n}{n_1},\frac{n}{n_2},\frac{n}{n_3}) \le C' < \infty$. Then, the proof is completed. $\square$

Proof of Theorem 1. By (2.9), we have:

$$
\begin{aligned}
0 &= \left|\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{V}_i-\theta_0}{1+\gamma(\hat{V}_i-\theta_0)}\right| \\
&= \frac{1}{n}\left|\sum_{i=1}^{n}(\hat{V}_i-\theta_0)-\gamma\frac{(\hat{V}_i-\theta_0)^2}{1+\gamma(\hat{V}_i-\theta_0)}\right| \\
&\ge \frac{|\gamma|S_n}{1+|\gamma|K_n}-|U_n-\theta_0|
\end{aligned}
$$

By lemma 2, $|U_n-\theta_0| = O_p(n^{-1/2})$. By lemma 3 and lemma 5, it follows that

$$
\frac{|\gamma|}{1+|\gamma|K_n} = O_p(n^{-1/2}),
$$

$$
|\gamma| = O_p(n^{-1/2}).
$$

For convenience, if let $\eta_i=\gamma(\hat{V}_i-\theta_0)$, then

$$
\begin{aligned}
\max_{1\le i\le n}|\eta_i| &= |\gamma|K_n \\
&= O_p(n^{-1/2})o(n^{1/2}) \qquad (4)\\
&= o_p(1).
\end{aligned}
$$

Plugging this estimator of η_i back into (2.9), we have

$$0 = U_n - \theta_0 - \gamma S_n + \frac{1}{n}\sum_{i=1}^{n}\frac{(\hat{V}_i - \theta_0)\eta_i}{1+\eta_i}.$$

Since

$$\frac{1}{n}\sum_{i=1}^{n}\frac{(\hat{V}_i - \theta_0)\eta_i}{1+\eta_i} = o(n^{1/2})O_p(n^{-1})O_p(1) = o_p(n^{-1/2}),$$

then

$$\gamma = \frac{U_n - \theta_0}{S_n} + \beta, \tag{5}$$

where $\beta = o_p(n^{-1/2})$. By Taylor expansion, we have $\log(1+\eta_i) = \eta_i - \frac{\eta_i^2}{2} + \zeta_i$, where as $n \to \infty$ $P\{|\zeta_i| \le C|\eta_i|^3, 1 \le i \le n\} \to 1$ for a number C $(0 < C < \infty)$.

Similar to the proof of Theorem 1 in Owen (1990), it can be shown that

$$\begin{aligned} -2\log R(\theta_0) &= 2\sum_{i=1}^{n}\log(1+\eta_i) \\ &= \frac{n(U_n-\theta_0)^2}{S_n} - nS_n\beta^2 + 2\sum_{i=1}^{n}\zeta_i. \end{aligned}$$

Since

$$|nS_n\beta^2| = n(nS^2_{n_1,n_2,n_3} + o(1))o_p(n_{-1}) = o_p(1),$$

$$|2\sum_{i=1}^{n}\zeta_i| \le C|\gamma|^3\sum_{i=1}^{n}|\hat{V}_i - \theta|^3 = O_p(n^{-3/2})o(n^{3/2}) = o_p(1),$$

and by lemma 2 and lemma 3, as $\min(n_1, n_2, n_3) \to \infty$,

$$\frac{n(U_n-\theta_0)^2}{S_n} \xrightarrow{d} \chi^2_1.$$

Then, by Slutsky's theorem, we have $-2\log R(\theta_0) \xrightarrow{d} \chi^2_1$, which completes the proof.

$\square$

Appendix B: Matlab Code for the Simulation Study

```
function [fres rres]=jel_conf_o(x,y,z,alpha);

v_hat=psuedo(x,y,z);
[fres rres]=confint(v_hat,alpha);

function [lend rend]=confint(V,alpha);

 epsilon=10^(-4);
 tAlpha=chi2inv(alpha,1);
 theta=mean(V);

 delta=1;
        deltaR=theta;

        while pandingzhi(V,deltaR)<tAlpha
            deltaR=deltaR+abs(delta);
        end

        deltaRL=theta;
        deltaRR=deltaR;

        while abs(deltaRL-deltaRR)>2*epsilon
            deltaRM=(deltaRL+deltaRR)/2;

            if (pandingzhi(V,deltaRR)-tAlpha)*(pandingzhi(V,deltaRM)-tAlpha)<0
                deltaRL=deltaRM;
            elseif (pandingzhi(V,deltaRL)-tAlpha)*(pandingzhi(V,deltaRM)-tAlpha)<0
                deltaRR=deltaRM;
            else
                break
            end
        end

        deltaRM=(deltaRL+deltaRR)/2;
        deltaL=theta;

        while pandingzhi(V,deltaL)<tAlpha
            deltaL=deltaL-abs(delta);
        end
```

```
        deltaLR=theta;
        deltaLL=deltaL;

        while abs(deltaLR-deltaLL)>2*epsilon
            deltaLM=(deltaLL+deltaLR)/2;
            if (pandingzhi(V,deltaLL)-tAlpha)*(pandingzhi(V,deltaLM)-tAlpha)<0
                deltaLR=deltaLM;
            elseif (pandingzhi(V,deltaLR)-tAlpha)*(pandingzhi(V,deltaLM)-tAlpha)<0
                deltaLL=deltaLM;
            else
                break
            end
        end

        deltaLM=(deltaLL+deltaLR)/2;
        lend=deltaLM;
        rend=deltaRM;

function t=pandingzhi(V,theta)

t=-2*elm(V', theta);
function pseudovalue=psuedo(x,y,z)
nx=length(x);
ny=length(y);
nz=length(z);
n=nx+ny+nz;
totalsum=0;

for i=1:nx;
    for j=1:ny;
        for k=1:nz;
            totalsum=totalsum+kernel(x(i),y(j),z(k));
        end;
    end;
end;

for index=1:n;
    if index<nx+1;
        partialsum=0;
        i=index;
        for j=1:ny;
```

```
            for k=1:nz;
                partialsum=partialsum+kernel(x(i),y(j),z(k));
            end;
        end;

        pseudovalue(index)=(1-(n-1)/(nx-1))*(totalsum/(nx*ny*nz))+((n-1)/(nx-1))*
        (partialsum/(ny*nz));
    elseif index>nx & index<nx+ny+1;
        partialsum=0;
        j=index-nx;
        for i=1:nx;
            for k=1:nz;
                partialsum=partialsum+kernel(x(i),y(j),z(k));
            end;
        end;

        pseudovalue(index)=(1-(n-1)/(ny-1))*(totalsum/(nx*ny*nz))+((n-1)/(ny-1))*
        (partialsum/(nx*nz));
    elseif index>nx+ny;
        partialsum=0;
        k=index-nx-ny;
        for i=1:nx;
            for j=1:ny;
                partialsum=partialsum+kernel(x(i),y(j),z(k));
            end;
        end;

        pseudovalue(index)=(1-(n-1)/(nz-1))*(totalsum/(nx*ny*nz))+((n-1)/(nz-1))*
        (partialsum/(ny*nx));
    end;
end;

function result=kernel(x,y,z)
if x<y &y<z;
    kernel1=1;
else kernel1=0;
end;
result=kernel1;
```