EMPIRICAL LIKELIHOOD WITH APPLICATIONS

WANG XIPING

(Master of Science, Northeast Normal University, P. R. China)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY

NATIONAL UNIVERSITY OF SINGAPORE

2010

Acknowledgements

I would like to express my deepest and most profound gratitude and thanks to my supervisors, Professor Bai Zhidong and Associate Professor Zhou Wang for their perspicacious guidance and continuous encouragement. Their insights and suggestions helped me improve my research skills. Their patience and encouragement carried me on through difficult times. Their strict attitude towards academic research, their kindness and understanding will always be remembered.

I wish to express my heartfelt gratitude to Assistant Professors Pan Guangming and Li Jialiang for their cooperation in my research projects, and to Dr Wang Xiaoying for discussions on various topics of the empirical likelihood method.

I would like to thank the university and the department for providing me with an NUS research scholarship which give me the valuable opportunity to study here. Assistance from the staff at the Department of Statistics and Applied Probability is gratefully appreciated.

I also wish to thank my friends, Ms. Papia Sultana, Ms. Zhao Jingyuan, Ms.

Wang Keyan, Ms. Zhao Wanting, Ms. Zhang Rongli, Mr. Li Mengxin, Mr. Hu Tao, Mr. Khang Tsung Fei, Mr. Wang Daqing, Mr. Loke Chok Kang and Mr. Jiang Binyan who have given me innumerous help in one way or another for their friendship and encouragement. All my friends whom I have forgotten to mention here are also greatly appreciated for their assistance and encouragement.

Finally, special appreciations are given to my wife, Li Yao, my parents and brother for their deep love, considerable understanding and continuous support in my life. I wish to dedicate this thesis to them.

Contents

Acknowledgements **i**

Summary **v**

List of Tables **vii**

1 Introduction **2**

1.1 Empirical likelihood 2

1.1.1 Empirical likelihood for mean functionals 4

1.2 U-statistics 6

1.2.1 Empirical likelihood for U-statistics 6

1.2.2 Jackknife empirical likelihood for U-statistics 8

1.3 Compound Poisson sum 9

1.4 Motivation and layout of the thesis 10

2 Interval Based Inference for $P(X < Y < Z)$ 15

2.1 Introduction 15

2.2 Methodology and main results 19

2.2.1 Asymptotic Normal approximations 19

2.2.2 JEL for the three-sample U-statistic U_n 23

2.3 Numerical study 28

2.4 Applications to real data 36

2.4.1 Chemical and overt diabetes data 37

2.4.2 Alzheimer's disease 39

2.4.3 Summary 40

2.5 Conclusion 41

2.6 Proof of Theorem 2.2.2 42

3 Interval Estimation of the Hypervolume under ROC Manifold 54

3.1 Introduction 54

3.2 Methodology and results 59

3.2.1 Asymptotic Normal approximations 59

3.2.2 JEL for the k-sample U-statistic U_n 61

3.3 Simulation study 64

3.4 Application to tissue biomarkers of synovitis 68

3.5 Discussion 72

3.6 Proof of Theorem 3.2.2 72

4 Empirical Likelihood for Compound Poisson Sum 76

4.1 Introduction 76

4.2 Methodology and results 80

4.3 Simulation study 83

4.4 Application to coal-mining disasters data 89

4.5 Proof of Theorem 4.2.1 90

5 Conclusions and Further Research 100

5.1 Conclusions 100

5.2 Further Research 102

Bibliography **104**

Summary

Empirical likelihood, first introduced by Thomas and Grunkemeier (1975) and later extended in Owen (1988, 1990), is an effective and flexible nonparametric method based on a data-driven likelihood ratio function. It enjoys many advantages over other nonparametric methods, such as automatic determination of the confidence region by the sample and transformation respecting, easy incorporation of side information, direct extension to biased sampling and censored data, good asymptotic power properties and Bartlete correctability. The empirical likelihood method can be used to find estimators, conduct hypothesis testing and construct small confidence intervals/regions. However, when treating with nonlinear statistics via the empirical likelihood method, the computation burden is quite heavy. The *Jackknife Empirical Likelihood* method, brought out by Jing et al. (2009), is surprisingly easy to cope with nonlinear statistics and largely relieves computation burden. In this thesis, we first apply the jackknife empirical likelihood method to make inference for the Volume Under the ROC Surface (VUS) and the Hypervolume Under the ROC Manifold (HUM) measures, which are straight extensions of the Area Under

the The Receiver Operating Characteristic (ROC) curve (AUC) for three-category and multi-category samples respectively. The popularity and importance of VUS and HUM are due to their capability of providing general measures of the differences amongst populations. Another problem in this thesis concerns the compound Poisson sum. Monte Carlo simulations are conducted to assess the performance of the proposed methods in finite samples. Some meaningful real datasets are analyzed.

List of Tables

2.1 $\theta_0 = 0.3407$, $F_1 = N(0,1)$, $F_2 = N(1,1)$ and $F_3 = N(1,2)$ 29

2.2 $\theta_0 = 0.6919$, $F_1 = Exp(8)$, $F_2 = Exp(1)$ and $F_3 = Exp(1/4)$ 31

2.3 $\theta_0 = 0.4019$, $F_1 = U(-1,1)$, $F_2 = Exp(2)$ and $F_3 = Cauchy(1,2)$ 32

2.4 $\theta_0 = 0.0454$, $F_1 = Cauchy(1,2)$, $F_2 = Exp(2)$ and $F_3 = U(-1,0.5)$ . . . 34

2.5 $\theta_0 = 0.9317$, $F_1 = N(-3,1)$, $F_2 = Exp(1)$ and $F_3 = Cauchy(6,1)$ 35

2.6 PLG, $\hat{\theta} = 0.7299$ 38

2.7 IR, $\hat{\theta} = 0.7161$ 38

2.8 MMSE, $\hat{\theta} = 0.3644$ 39

3.1 $F_1 = N(0,1)$, $F_2 = N(6,1)$, $F_3 = N(9,1)$, $F_4 = N(12,1)$ and $\theta_0 = 0.9662$ 65

3.2 F_1=$Exp(8)$, F_2=$Exp(1)$, F_3=$Exp(1/4)$, F_4=$Exp(1/16)$, θ_0=0.5239 . . . 67

3.3 Sample sizes for synovitis data. 69

3.4 95% confidence intervals by JEL and Norm. 71

4.1 $F = Exp(1/2)$ and $\lambda_0 = 0.5$ 82

4.2 $F = N(1,1)$ and $\lambda_0 = 10$ 84

4.3 $F = U(0,1)$ and $\lambda_0 = 15$ 85

4.4 $F = Binomial(20,0.05)$ and $\lambda_0 = 20$ 87

4.5 CIs by EL, Normality, Edgeworth expansion and Kegler's method . 90

Chapter 1

Introduction

1.1 Empirical likelihood

Empirical likelihood (EL) is an effective and flexible nonparametric method based on a data-driven likelihood ratio function, which does not require us to assume the data coming from a known family of distributions. It was first introduced by Owen (1988, 1990) to construct confidence intervals/regions for population means, which extends the work in Thomas and Grunkemeier (1975) where a nonparametric likelihood ratio idea was used to construct confidence intervals for some survival function. The empirical likelihood method can be used to find estimators, conduct hypothesis testing and construct small confidence intervals/regions even when the data are incomplete. It enjoys many advantages over other nonparametric methods, such as automatic determination of the confidence region by the sample and transformation respecting, easy incorporation of side information, straight extension to biased sampling and censored data, better asymptotic power properties and Bartlete correctability (see Hall and LaScala (1992) for details).

Since Owen's pioneering work, much attention has been attracted by the beautiful properties of the EL method. See for example, Diciccio et al. (1991) for smooth functions of means, Qin (1993) and Chen and Sitter (1999) for biased sampling, Chen and Hall (1993), Qin and Lawless (1994) for estimation equations, Wang and Jing (1999, 2003) for partial linear models, and Zhang (1997a & 1997b) and Zhou and Jing (2003) for M-functionals and quantile, Chen and Qin (1993) and Zhong and Rao (2000) for random sampling. Some recent developments and applications of the empirical likelihood method include those for: additive risk models (Lu and Qi (2004)); longitudinal data and single-index models (You et al. (2006), Xue and Zhu (2006, 2007), Zhao and Jian (2007)); two-sample problems (Zhou and Liang (2005), Cao and Van Keilegom (2006), Ren (2008), Keziou and Leoni-Aubin (2008)); regression models (Zhao and Chen (2008), Zhao and Yang (2008)); time series models (Chan and Ling (2006), Nordman and Lahiri (2006), Otsu (2006), Chen and Gao (2007), Nordman et al. (2007), Guggenberger and Smith (2008)), copula (Chen et al. (2009)) and high dimensional data (Chen et al. (2009)). We refer to the bibliography of Owen (2001) for more extensive references.

1.1.1 Empirical likelihood for mean functionals

In this section, we provide a brief description of the elementary procedure of empirical likelihood for mean functionals. For simplicity, we consider the population mean. Suppose that $\mathbf{X_1}, \ldots, \mathbf{X_n} \in \mathbf{R}^q$ are independent and identically distributed (i.i.d.) random vectors with common distribution function (d.f.) $F(x)$. Let $\mathbf{p} = (p_1, \ldots, p_n)$ be a probability vector, i.e. $\sum_{i=1}^n p_i = 1$, $p_i \geq 0$ for $i = 1, \ldots, n$, and θ be the population mean. $F(x)$ assigns probability p_i to the ith atom $\mathbf{X_i}$. The empirical likelihood, evaluated at θ, is then given by

$$L(\theta) = \max\left\{\prod_{i=1}^n p_i : \sum_{i=1}^n p_i = 1, p_i \geq 0, \vartheta(F_{\mathbf{p}}) = \theta\right\},$$

where $\vartheta(F_{\mathbf{p}}) = \sum_{i=1}^n p_i \mathbf{X_i}$ is a mean functional, and $F_{\mathbf{p}}$ is the empirical d.f. of $\mathbf{X_1}$.

Since $\prod_{i=1}^n p_i$, subject to the restriction $\sum_{i=1}^n p_i = 1$, attains its maximum at $p_i = 1/n$, we can define the empirical likelihood ratio at θ by

$$R(\theta) = \max\left\{\prod_{i=1}^n (np_i) : \sum_{i=1}^n p_i = 1, p_i \geq 0, \vartheta(F_{\mathbf{p}}) = \theta\right\}. \tag{1.1}$$

To optimize (1.1), use Lagrange multiplier method and write

$$LH(\mathbf{p}) = \sum_{i=1}^n \log(p_i) - \lambda\left(\sum_{i=1}^n p_i - 1\right) - n\gamma^T\left(\sum_{i=1}^n p_i \mathbf{X}_i - \theta\right)$$

where A^T means the transpose of A. Now differentiating $LH(\mathbf{p})$ with respect to each p_i and setting all partial derivatives to zero, we have

$$p_i = \frac{1}{n} \cdot \frac{\mathbf{X_i} - \theta}{1 + \gamma^T(\mathbf{X_i} - \theta)} \qquad (i = 1, \ldots, n)$$

where the Lagrangian multiplier $\gamma = (\gamma_1, \ldots, \gamma_n)^T$ satisfies

$$\sum_{i=1}^{n} \frac{\mathbf{X_i} - \theta}{1 + \gamma^T(\mathbf{X_i} - \theta)} = 0. \tag{1.2}$$

Let

$$\mathbf{S} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{X_i} - \theta)(\mathbf{X_i} - \theta)^T$$

be a covariance matrix of $\mathbf{X}_1, \ldots, \mathbf{X}_n$ of full rank q and expand the left hand side of (1.2), we get

$$\gamma = \mathbf{S}^{-1}(\bar{\mathbf{X}} - \theta) + o_p(n^{-1/2})$$

where $\bar{\mathbf{X}}$ is the mean of $\mathbf{X}_1, \ldots, \mathbf{X}_n$ and $A_n = o_p(B_n)$ means A_n/B_n converges to 0 in probability.

Plugging the p_i's back into (1.1) and taking logarithm, we get the empirical log-likelihood ratio

$$-2\ell(\theta) = 2\sum_{i=1}^{n} \log\left(1 + \gamma^T(\mathbf{X}_i - \theta)\right).$$

Expanding $-2\ell(\theta)$, we have

$$-2\ell(\theta) = n(\bar{\mathbf{X}} - \theta)^T\mathbf{S}^{-1}(\bar{\mathbf{X}} - \theta) + o_p(1),$$

which converges in distribution to χ^2_q by central limit theorem. From this, an $(1-\alpha)$-level confidence region for θ can be constructed as

$$\Theta_c = \{\theta : -2\ell(\theta) \le c\}$$

where c is chosen to satisfy $P\{\chi^2_q \le c\} = 1 - \alpha$.

1.2 U-statistics

U-statistics were first introduced by Halmos (1946) as unbiased estimators of their expectations, and then were termed U-*statistics* by Hoeffding (1948). A U-statistic of degree k with kernel h is defined as

$$U_n = \binom{n}{k}^{-1} \sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} h(X_{i_1}, X_{i_2} \ldots, X_{i_k}).$$

The consistency and asymptotic normality of U-statistics were proved in Hoffding (1948). U-statistics are found to play a role in almost any statistical setting. From general Hoeffding-decomposition, we know that U-statistics are in fact successive generalization of sums of i.i.d random variables (r.v.'s), which has been the focus of probability theory for centuries. As many statistics occurring in estimation and testing problems behave asymptotically like independent r.v.'s, the study of U-statistics is of theoretical and practical importance, and limit theorems and certain asymptotic properties of U-statistics have been the subject of many academic articles. For comprehensive details of U-statistics, one may refer to Lee (1990), and Koroljuk and Borovskich (1994).

1.2.1 Empirical likelihood for U-statistics

Due to their wonderful properties, U-statistics have been widely used to do inference for their expectations. For example, one may attempt to apply Owen's empirical likelihood method to U-statistics, and derive asymptotic distribution for the empirical log-likelihood ratio, from which hypothesis testing could be done and confidence intervals might be constructed for the parameter one is interested in. However, the computation burden will be very heavy as we need to solve several simultaneous nonlinear equations.

To get a clear image of how heavy the computation burden is when dealing with nonlinear statistics, for simplicity, we take one-sample U-statistics for example. Suppose $X_1, \ldots, X_n$ are independent and identically distributed (i.i.d.) random variables with common distribution function $F(x)$. A one-sample U-statistic of degree 2 with symmetric kernel ψ can be defined to be

$$W_n = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} \psi(X_i, X_j), \tag{1.3}$$

and $\theta = E\psi(X_1, X_2)$ is the parameter of interest.

To apply the usual empirical likelihood method to W_n, let $\mathbf{p} = (p_1, \ldots, p_n)$ be a probability vector and write

$$\tilde{\theta}(F_{\mathbf{p}}) = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} n^2 p_i p_j \psi(X_i, X_j), \tag{1.4}$$

where $F_{\mathbf{p}}(x) = \sum_{i=1}^n p_i I_{\{X_i \le x\}}$. (1.3) and (1.4) coincide when $p_i = 1/n$ for $i = 1, \ldots, n$. Then the empirical likelihood can be defined by

$$\tilde{\ell} = \max\left\{ \prod_{i=1}^n p_i : \sum_{i=1}^n p_i = 1, \tilde{\theta}(F_{\mathbf{p}}) = \theta \right\}. \tag{1.5}$$

By solving (1.5), we obtain the empirical likelihood for W_n. However,the computational difficulty arises when one tries to do so: there is not simple methodology available for an optimization problem involving n variables $p_1, \ldots, p_n$ with $n+1$ nonlinear constraints. The situation becomes worse when n gets larger. One may also refer to Jing et al. (2009) for excellent interpretations.

1.2.2 Jackknife empirical likelihood for U-statistics

As we can see from Section 1.2.1, Owen's empirical likelihood encounters awkward computational difficulties when treating with nonlinear statistics. Fortunately, in 2009, Jing et al. brings out the so-called *Jackknife Empirical Likelihood* method, which can cope with nonlinear statistics promisingly.

Now as an illustration of the JEL procedure, we briefly describe it for W_n as follows.

Applying the standard jackknife method (Shao and Tu (1995)) to W_n (see Arvesen (1969) for jackknife to U-statistics), we obtain the jackknife pseudo-values $(s = 1, \ldots, n)$

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

and the jackknife estimator of θ: $n^{-1}\sum_{s=1}^{n}\tilde{V}_s$, where $W_{n-1}^{(-s)}$ is the U-statistic after removing X_s. If we write $\tilde{\theta}_p = \sum_{s=1}^{n} p_i E\tilde{V}_s$, it is obvious that $E\tilde{V}_s = \theta$ and $\tilde{\theta}_p = \theta$ due to the unbiasedness of U-statistics. Applying Owen's EL method to $\tilde{V}_s$, we get the empirical likelihood at θ:

$$\tilde{L}(\theta) = \max\left\{\prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i\tilde{V}_i = \tilde{\theta}_p\right\},$$

and we can define the JEL ratio by

$$\tilde{R}(\theta) = \max\left\{\prod_{i=1}^{n}(np_i) : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \tilde{V}_i = \tilde{\theta}_p\right\}.$$

The jackknife empirical log-likelihood ratio at θ then follows as

$$\log \tilde{R}(\theta) = -\sum_{i=1}^{n} \log\{1 + \gamma(\tilde{V}_i - \tilde{\theta}_p)\},$$

where γ satisfies the equation

$$\sum_{i=1}^{n} \frac{\tilde{V}_i - \tilde{\theta}_p}{1 + \gamma(\tilde{V}_i - \tilde{\theta}_p)} = 0.$$

The asymptotic distribution of $-2\log \tilde{R}(\theta)$ was proven to be χ_1^2 in Jing et al. (2009), from which $(1-\alpha)$-level confidence interval for θ can be constructed. The superiority of JEL over the usual empirical likelihood is apparent, since the optimization problem now involves only one nonlinear equation.

1.3 Compound Poisson sum

Let $\{X_j\}_{j=1}^{\infty}$ be a sequence of i.i.d. r.v.'s with common d.f. F. Define a renewal counting process $\{N(t), t > 0\}$ by $N(t) = \max\{k : T_k \leq t\}$, where T_k is the occurrence time of X_k. Then $N(t)$ can be interpreted as the number of occurrences X_k in $(0, t]$. Further, suppose that $\{N(t), t > 0\}$ is independent of the sequence $\{X_j\}_{j=1}^{\infty}$ and write

$$S_{N(t)} = \sum_{j=1}^{N(t)} X_j,$$

then the stochastic process $\{S_{N(t)}, t > 0\}$ is called a *renewal reward process* (for definiteness, we assume that $S_{N(t)} = 0$ if $N(t) = 0$). When $\{N(t), t > 0\}$ is a Poisson process, the renewal reward process $S_{N(t)}$ is termed as a *compound Poisson process* (CPP), which has various applications in the applied fields such as physics, industry, finance and risk management. See Helmers et al (2003) for some developments on compound Poisson sums and their relevance in finance. Excellent interpretations and more examples of CPPs may be found in Parzen (1967, p129-130), and Karlin and Taylor (1981, p426); see also Gnedenko and Korolev (1996) for the general theories of random sums.

1.4 Motivation and layout of the thesis

The Receiver Operating Characteristic (ROC) curve and the Area Under the ROC Curve (AUC) are standard statistical tools for evaluating the accuracy of diagnostic tests of two-category classification data. The ROC curve is a plot of sensitivity *versus* $1-$specificity as one changes the value of positivity. For a given threshold value c, the sensitivity and specificity of a test are respectively defined as

$$\text{Sensitivity} = P(Y > c) = 1 - F_2(c), \quad \text{Specificity} = P(X \le c) = F_1(c)$$

where F_1 and F_2 are the d.f.'s of X and Y respectively. The AUC is given by $\int_0^1 [1 - F_2(F_1^{-1}(t))]dt$, where F^{-1} is the inverse function of F. Bamber (1975) show that AUC is exactly $P(X < Y)$, the probability that a randomly selected observation from one population scores less than that from another population. AUC is the most commonly used measure of diagnostic accuracy for a continuous-scale diagnostic test. Because of its great importance, AUC has attracted much attention in the past decades. For example, one can refer to Swets and Pickett (1982), Johnson (1989), Hanley (1989), Newcombe (2006), Zhou (2008) and the monograph by Kotz et al. (2003) for some references and excellent reviews. Comprehensive descriptions of methods for diagnostic tests can be found in Zhou et al. (2002) and Pepe (2003).

In practice, however, many real applications involve more than two classes and demand a methodology expansion. The Volume Under the ROC Surface (VUS) and the Hypervolume Under the ROC Manifold (HUM) measures are direct extensions of AUC for three-category and multi-category samples, respectively. VUS and HUM have extensive applications in various areas since they provide global measures of the differences amongst populations.

The existing inference methods for such measures include the asymptotic normal approximation and the bootstrap resampling method. The normal approximation method may produce confidence intervals with unsatisfactory coverage when sample size is small while the bootstrap is computationally intensive.

In this thesis, on one hand, we develop JEL procedures to make statistical inference for VUS $P(X < Y < Z)$ and HUM $P(X_1 < X_2 < \cdots < X_k)$ respectively, and provide the corresponding asymptotic distribution theories. On the other hand, we employ Owen's empirical likelihood method to compound Poisson sum. Monte Carlo simulations are conducted to assess the performance of the proposed methods in finite samples. Some real datasets are also analyzed as applications of the proposed methods.

In Chapter 2, we make inference for $P(X < Y < Z)$ by applying two methods, normal approximation and JEL, to three-sample U-statistics. We propose the JEL method, because Owen's EL method for U-statistics is too complicated to apply in practice. The simulation results show that the two proposed methods work quite well and JEL always outperforms the normal approximation method. Practically, for simplicity purpose, we recommend the normal approximation method; for better statistical results, we suggest the reader to use the JEL method although it involves a bit more computation burden than the normal approximation one.

In Chapter 3, as the existing inference methods for $P(X_1 < X_2 < \cdots < X_k)$ are either imprecise or computationally intensive, we develop a JEL procedure and provide the corresponding distribution theories. As the results of simulation studies indicate, JEl performs reasonably well for small samples and can be implemented more efficiently than the bootstrap.

In Chapter 4, we apply Owen's EL method to do inference for the unit mean of compound Poisson sums. Compound Poisson sums have plenty of applications in physics, industry, finance, risk management and so on. They are frequently used to describe phenomena in applied probability when a single Poisson process fails to do so. It is well-known that for a renewal reward process $\{S_{N(t)} = \sum_{j=1}^{N(t)} X_j, t > 0\}$, if $N(t)/t$ converges in probability to a constant or, more generally, to a positive r.v., then $S_{N(t)}$ is asymptotically normally distributed. Especially, when $\{N(t), t > 0\}$ is a Poisson process with rate $\lambda > 0$, independent of the i.i.d. r.v.'s $X_1, X_2, ...$ with mean $\mu = EX_1$ and variance $\sigma^2 = \mathrm{Var}(X_1) > 0$, we can use this asymptotic normality to construct confidence intervals for $\lambda\mu$. But as pointed out by Helmers (2003), the usual normal approximation for compound Poisson sums usually performs very badly because, in real applications, the distribution of the X_i is often highly skewed to the right. This urges for better methods, e.g. the bootstrap or Edgeworth/saddlepoint approximations, to construct more accurate confidence intervals for $\lambda\mu$. One can also consider a studentized version of CPP to correct the skewness. Kegler (2007) uses

$$\left(e^{\log(S_{N(t)}/t) - z_{\alpha/2}\sqrt{\Delta_{N(t)}/S^2_{N(t)}}}, e^{\log(S_{N(t)}/t) + z_{\alpha/2}\sqrt{\Delta_{N(t)}/S^2_{N(t)}}} \right) \tag{1.6}$$

as confidence interval for $\lambda\mu$, where

$$S_{N(t)} = \sum_{j=1}^{N(t)} X_j,\ \Delta_{N(t)} = \sum_{j=1}^{N(t)} X_j^2,\ \Phi(z_{\alpha/2}) = 1 - \alpha/2.$$

However, this method is applicable only when $S_{N(t)} > 0$.

Therefore, we propose Owen's empirical likelihood to meet the demand for better inference methods. The idea of applying Owen's EL for compound Poisson sum is as follows.

From the viewpoint of conditional expectation, since

$$\lambda\mu t = E\left(\sum_{j=1}^{N(t)} X_j\right),$$

we argue that

$$E\left(\sum_{j=1}^{N(t)} X_j \middle| N(t) = n\right) \approx \lambda\mu t.$$

This leads us to consider the following EL

$$L(\theta|N(t) = n) = \max\left\{\prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1, p_i \geq 0, \vartheta(G_p) = \theta t/n\right\},$$

where $\vartheta(G_p) = \sum_{i=1}^{n} p_i X_i$ and $\theta = \lambda\mu$. Owen's EL method is then applied to the mean functional $\sum_{i=1}^{n} p_i X_i$ and an asymptotic theory for the *adjusted* empirical log-likelihood ratio is developed.

Chapter 2

Interval Based Inference for $P(X < Y < Z)$

2.1 Introduction

Let X, Y and Z be three r.v.'s. The "stress-strength" models of the types $P(X < Y)$, $P(X < Y < Z)$ have extensive applications in various subareas of engineering (often in reliability theory), psychology, genetics, clinical trials and so on, since these models provide general measures of the differences amongst populations. For more detailed descriptions on stress-strength models, one is referred to the monograph by Kotz et al. (2003) and references therein.

One such important case is $P(X < Y)$. In context of medicine and genetics, a popular topic is the analysis of the discriminatory accuracy of a diagnostic test or marker in distinguishing between diseased and non-diseased individuals, through the receiver-operating characteristic (ROC) curves. The ROC curve is a plot of sensitivity *versus* 1-specificity as one changes the value of positivity. The area under the ROC curve (AUC), is exactly $P(X < Y)$ (see, Bamber 1975), which is a general index of diagnostic accuracy. An individual is diagnosed as diseased or non-diseased according to whether the marker value is greater than or less than or equal to a specified threshold value.

Recently, lots of efforts have been devoted to the extension of ROC methodology to three-class diagnostic problems. Mossman (1999) showed that the volume under the ROC surface (VUS) equals $\theta = P(X < Y < Z)$, the probability that three measurements will be classified in the correct order $X < Y < Z$, where the ROC surface is a direct generalization of the two-sample ROC curve to the three-category classification problems. A motivation to study θ is from cancer diagnosis and treatment, where an important practical issue is to determine a set of genes which can optimally classify tumors, and diagnostic procedures need to assign individuals to one of the outcome tumor types. Generally speaking, ROC curves are not applicable to the situations where there are more than two tumor types. In such cases, one may convert the tumor types into pairs and evaluate all pairs of classes using two-class ROC analysis (Obuchowshi et al., 2001), but the problem is that this method does not provide an assessment of overall accuracy (Nakas et al., 2007). There are many other methods that, for assessing the overall accuracy of classification when there are more than two diseased classes, have been proposed and one can refer to the paper of Li et al. (2008) and Sampat et al. (2009) for excellent reviews of such related work and references. One can also find many interesting practical examples in Kotz et al. (2003)

Here are some other examples.

1. Many devices can not function at high temperatures, neither can do at very low temperatures. Extreme outer environmental conditions could result in failure of the devices.

2. One's normal blood pressure must lie within the systolic and diastolic pressures limits, as one will be identified as hypertensive if the blood pressure is abnormally high and hypotensive when it is abnormally low.

3. For a healthy individual, his/her level of blood sugar should lie within some range since hypoglycemia is a major cause of chronic fatigue while glycemia is most directly associated to chronic increase of diabetes mellitus.

4. To cure some disease, one must take a moderate dose of drug , because too much drug will result in side-effect and be harmful, but a relatively small dose of drug might fail to cure the disease.

It is clear from these examples that this stress-strength relation $P(X < Y < Z)$ reflects a number of real-world phenomena and one may also find many other applications of it.

In the literature, there are also some papers concerning the point estimation of θ. Hlawka (1975) suggests to estimate θ by three-sample U-statistics, Chandra and Owen (1975) construct MLEs and UMVUEs for $P(X_1 < Y, ..., X_l < Y)$ and $P(X < Y_1, ..., X < Y_l)$ in some special cases, which is related to θ by a formula provided in Singh (1980) where normal populations are considered, Dutta and Sriwastav (1986) deal with the estimation of θ when X, Y and Z are exponentially distributed, and Ivshin (1988) investigates the *Maximum Likelihood Estimate* (MLE) and *Uniformly Minimum Variance Unbiased Estimate* (UMVUE) of θ when X, Y and Z are either uniform or exponential r.v.'s. with unknown location parameters.

Although Dreiseitl et al. (2000) derive variance estimators for VUS using U-statistic theory, the variance becomes complicated as the number of categories increases and is difficult to apply. Nakas et al. (2004) used bootstrap method, but this is also computationally intensive. Further, a glance at the literature reveals that there is not simple method available for constructing confidence intervals (CIs) for θ via three-sample U-statistics; however, our proposed methods provide easier and better alternative tools to deal with such problems.

In this chapter, we employ normal approximation and the JEL method to make statistical inference for θ, assuming that the three samples are independent, without ties among them. In Section 2.2, we present our two methods. Simulation results are presented in Section 2.3 to illustrate and compare the performance of these methods. Real data sets are analyzed in Section 2.4. Proofs are deferred to Section

2.6.

2.2 Methodology and main results

2.2.1 Asymptotic Normal approximations

Let $(X_1, ..., X_{n_1})$, $(Y_1, ..., Y_{n_2})$ and $(Z_1, ..., Z_{n_3})$ be samples from three different populations with d.f.'s F_1, F_2 and F_3, respectively. Assume that the three samples are independent. A U-statistic of degree $(1, 1, 1)$ with a kernel $h(x; y; z)$ is defined as

$$U = \frac{1}{n_1 n_2 n_3} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} h(X_i; Y_j; Z_k), \tag{2.1}$$

which is a consistent and unbiased estimator of our parameter of interest $\theta = Eh(X_1; Y_1; Z_1)$. Particularly, if $h(x; y; z)$ is equal to the indicator function $I_{\{x<y<z\}}$, then $\theta = P(X_1 < Y_1 < Z_1)$, the probability that three measurements, one from each population, will be in correct order. Hence we can make inference on θ by means of the statistic

$$U_n = \frac{1}{n_1 n_2 n_3} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} I_{\{X_i<Y_j<Z_k\}}.$$

Write $\sigma^2 = E(U_n - \theta)^2$. Citing a result in Koroljuk and Borovshich (1994), we have a central limit theorem (CLT) for U_n, i.e., $(U_n - \theta)/\sigma \to_d N(0, 1)$ as $\min(n_1, n_2, n_3) \to \infty$, where "$\to_d$" means convergence in distribution. But we can not directly use this asymptotic normality to make statistical inference on θ because

σ^2 is usually unknown. So we must replace σ^2 by its estimator. One consistent estimator $\hat{\sigma}^2$ of σ^2 can be constructed as follows.

For $i = 1, ..., n_1$, $j = 1, ..., n_2$ and $k = 1, ..., n_3$, denote:

(1) $U^0_{n_1,n_2,n_3}$=U_n, the original statistics based on all observations;

(2) $U^{-i,0,0}_{n_1-1,n_2,n_3}$, the statistics after deleting X_i, given by

$$((n_1-1)n_2n_3)^{-1}\sum_{i_1=1,i_1\neq i}^{n_1}\sum_{j_1=1}^{n_2}\sum_{k_1=1}^{n_3} I_{\{X_{i_1}<Y_{j_1}<Z_{k_1}\}};$$

(3) $U^{0,-j,0}_{n_1,n_2-1,n_3}$, the statistics after deleting Y_j, given by

$$(n_1(n_2-1)n_3)^{-1}\sum_{i_1=1}^{n_1}\sum_{j_1=1,j_1\neq j}^{n_2}\sum_{k_1=1}^{n_3} I_{\{X_{i_1}<Y_{j_1}<Z_{k_1}\}};$$

(4) $U^{0,0,-k}_{n_1,n_2,n_3-1}$, the statistics after deleting Z_k, given by

$$(n_1n_2(n_3-1))^{-1}\sum_{i_1=1}^{n_1}\sum_{j_1=1}^{n_2}\sum_{k_1=1,k_1\neq k}^{n_3} I_{\{X_{i_1}<Y_{j_1}<Z_{k_1}\}},$$

and

$$V_{i,0,0} = n_1U^0_{n_1,n_2,n_3} - (n_1-1)U^{-i,0,0}_{n_1-1,n_2,n_3}; \tag{2.2}$$

$$V_{0,j,0} = n_2U^0_{n_1,n_2,n_3} - (n_2-1)U^{0,-j,0}_{n_1,n_2-1,n_3};$$

$$V_{0,0,k} = n_3U^0_{n_1,n_2,n_3} - (n_3-1)U^{0,0,-k}_{n_1,n_2,n_3-1}.$$

Some simple calculations show that

$$V_{i,0,0} = \frac{1}{n_2 n_3} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} I_{\{X_i<Y_j<Z_k\}}; \tag{2.3}$$
$$V_{0,j,0} = \frac{1}{n_1 n_3} \sum_{i=1}^{n_1} \sum_{k=1}^{n_3} I_{\{X_i<Y_j<Z_k\}};$$
$$V_{0,0,k} = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} I_{\{X_i<Y_j<Z_k\}},$$

and

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

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

Similar to Arversen (1969) and Sen (1960), we propose a consistent estimator of $\mathrm{Var}(U_n)$ given by

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n_1(n_1-1)} \sum_{i=1}^{n_1} (V_{i,0,0} - \overline{V}_{\cdot,0,0})^2 \\ &\quad + \frac{1}{n_2(n_2-1)} \sum_{j=1}^{n_2} (V_{0,j,0} - \overline{V}_{0,\cdot,0})^2 \\ &\quad + \frac{1}{n_3(n_3-1)} \sum_{k=1}^{n_3} (V_{0,0,k} - \overline{V}_{0,0,\cdot})^2. \end{aligned} \tag{2.4}$$

Further, to state the results, define

$$g_{1,0,0}(x) = P(x < Y_1 < Z_1) - \theta,\ \sigma^2_{1,0,0} = \mathrm{Var}(g_{1,0,0}(X_1));$$

$$g_{0,1,0}(y) = P(X_1 < y < Z_1) - \theta,\ \sigma^2_{0,1,0} = \mathrm{Var}(g_{0,1,0}(Y_1));$$

$$g_{0,0,1}(z) = P(X_1 < Y_1 < z) - \theta,\ \sigma^2_{0,0,1} = \mathrm{Var}(g_{0,0,1}(Z_1)).$$

Theorem 2.2.1 *(a)* $U_n \xrightarrow{a.s.} \theta$ *as* $\min(n_1, n_2, n_3) \to \infty$*;*

(b) Assume that $\sigma^2_{1,0,0} > 0$, $\sigma^2_{0,1,0} > 0$, $\sigma^2_{0,0,1} > 0$, *and let* $S^2_{n_1,n_2,n_3} = \sigma^2_{1,0,0}/n_1 + \sigma^2_{0,1,0}/n_2 + \sigma^2_{0,0,1}/n_3$. *Then,*

$$\frac{U_n - \theta}{S_{n_1,n_2,n_3}} \xrightarrow{d} N(0,1), \text{ as } \min(n_1,n_2,n_3) \to \infty \tag{2.5}$$

and

$$\hat{\sigma}^2 - S^2_{n_1,n_2,n_3} = o_p((\min(n_1,n_2,n_3))^{-1}). \tag{2.6}$$

Proof. For the proof of part (a) and (2.5), refer to p151-153 of Koroljuk and Borovskich (1994). The proof of (2.6) is trivial and hence omitted.

Now by Theorem 2.2.1, we have CLT for the Studentized U_n, i.e.,

$$(U_n - \theta)/\hat{\sigma} \to_d N(0,1)$$

as $\min(n_1,n_2,n_3) \to \infty$, which provides an approach to construct CIs for θ. A two-sided $(1-\alpha)$ level CI based on the asymptotic normality is

$$(U_n - z_{\alpha/2}\hat{\sigma}, U_n + z_{\alpha/2}\hat{\sigma}). \tag{2.7}$$

From Dreiseitl (2003), one can derive the variance estimator of U_n as

$$\begin{aligned}\widehat{\text{Var}}(U_n) &= \frac{1}{n_1n_2n_3}[\theta(1-\theta) + (n_3-1)(\hat{q}_{12}-\theta^2) + (n_2-1)(\hat{q}_{13}-\theta^2) \\ &\quad +(n_1-1)(\hat{q}_{23}-\theta^2) + (n_2-1)(n_3-1)(\hat{q}_1-\theta^2) \\ &\quad +(n_1-1)(n_3-1)(\hat{q}_2-\theta^2) + (n_1-1)(n_2-1)(\hat{q}_3-\theta^2)],\end{aligned} \tag{2.8}$$

where

$$\hat{q}_{12} = \frac{1}{n_1 n_2 n_3 (n_3 - 1)} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} \sum_{\substack{K=1 \\ K \neq k}}^{n_3} I_{\{X_i < Y_j < Z_k\}} I_{\{X_i < Y_j < Z_K\}}$$

$$\hat{q}_{13} = \frac{1}{n_1 n_2 n_3 (n_2 - 1)} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} \sum_{\substack{J=1 \\ J \neq j}}^{n_2} I_{\{X_i < Y_j < Z_k\}} I_{\{X_i < Y_J < Z_k\}}$$

$$\hat{q}_{23} = \frac{1}{n_1 n_2 n_3 (n_1 - 1)} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} \sum_{\substack{I=1 \\ I \neq i}}^{n_1} I_{\{X_i < Y_j < Z_k\}} I_{\{X_I < Y_j < Z_k\}}$$

$$\hat{q}_{1} = \frac{1}{n_1 n_2 n_3 (n_2 - 1)(n_3 - 1)} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} \sum_{\substack{J=1 \\ J \neq j}}^{n_2} \sum_{\substack{K=1 \\ K \neq k}}^{n_3} I_{\{X_i < Y_j < Z_k\}} I_{\{X_i < Y_J < Z_K\}}$$

$$\hat{q}_{2} = \frac{1}{n_1 n_2 n_3 (n_1 - 1)(n_3 - 1)} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} \sum_{\substack{I=1 \\ I \neq i}}^{n_1} \sum_{\substack{J=1 \\ J \neq j}}^{n_2} I_{\{X_i < Y_j < Z_k\}} I_{\{X_I < Y_J < Z_k\}}$$

$$\hat{q}_{3} = \frac{1}{n_1 n_2 n_3 (n_1 - 1)(n_2 - 1)} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} \sum_{\substack{I=1 \\ I \neq i}}^{n_1} \sum_{\substack{K=1 \\ K \neq k}}^{n_3} I_{\{X_i < Y_j < Z_k\}} I_{\{X_I < Y_j < Z_K\}}.$$

Comparing (2.4) with (2.8), we can conclude that these two estimators of the variance of U_n do not necessarily equal and (2.8) is unbiased for $\mathrm{Var}(U_n)$ but computationally intensive. More interestingly, in our simulation studies, we find that the value (2.8) is always smaller than that of (2.4). Further, as sample sizes increase, the computation burden of (2.8) become strikingly heavy.

2.2.2 JEL for the three-sample U-statistic U_n

JEL introduced by Jing et al. (2008) is a marriage of two popular nonparametric approaches, jackknife and Owen's empirical likelihood method. For the reader's convenience, we briefly describe JEL for general one-sample U-statistics as follows.

Let $\mathbf{Z}_1, ..., \mathbf{Z}_n$ be independent (not necessarily identically distributed) r.v's and

$$T_n = T(\mathbf{Z}_1, ..., \mathbf{Z}_n) = \binom{n}{m}^{-1} \sum_{1 \leq i_1 < ... < i_m \leq n} h(\mathbf{Z}_{i_1}, ..., \mathbf{Z}_{i_m})$$

be a one-sample U-statistic of degree m as an unbiased estimator of the parameter θ, that is $\theta = Eh(\mathbf{Z}_1, \ldots, \mathbf{Z}_m)$. Define the jackknife pseudo-values by

$$\widehat{V}_i = nT_n - (n-1)T_{n-1}^{(-i)},$$

where $T_{n-1}^{(-i)} = T(\mathbf{Z}_1, ..., \mathbf{Z}_{i-1}, \mathbf{Z}_{i+1}, ..., \mathbf{Z}_n)$ is the statistic T_{n-1} computed on the sample of $n-1$ r.v.'s from the original data set by deleting the ith data value. Its expression is as follows,

$$T_{n-1}^{(-i)} = \binom{n-1}{m}^{-1} \sum_{(n-1,m)}^{(-i)} h(\mathbf{Z}_{j_1}, ..., \mathbf{Z}_{j_m}),$$

here and after, $\sum_{(n-1,m)}^{(-i)}$ denotes the summation over all possible indices $(j_1, ..., j_m)$ chosen from $(1, ..., i-1, i+1, ..., n)$, subject to the restriction $1 \leq j_1 < ... < j_m \leq n$.

The jackknife estimator of θ is simply the average of the pseudo-values:

$$\widehat{T}_n(jack) \cong \frac{1}{n} \sum_{i=1}^{n} \widehat{V}_i.$$

One advantage of $\widehat{T}_n(jack)$ over T_n is its smaller bias (see Quenouille (1956) and Tukey (1958)). Another one is that $\widehat{V}_i$'s are asymptotically independent (Shi (1984)).

Let $\mathbf{p} = (p_1, ..., p_n)$ be a probability vector, i.e., $\sum_{i=1}^{n} p_i = 1$ and $p_i \geq 0$ for $1 \leq i \leq n$. Let $G_\mathbf{p}$ be the d.f. which assigns probability p_i to the ith pseudo-value

$\widehat{V}_i$ and consider the mean functional $\vartheta(G_{\mathbf{p}}) = \sum_{i=1}^{n} p_i \widehat{V}_i$. The JEL, evaluated at θ, is

$$L(\theta) = \max \left\{ \prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1,\ p_i \geq 0,\ \vartheta(G_{\mathbf{p}}) = \theta_{\mathbf{p}} \right\}$$

with $\theta_{\mathbf{p}} = \sum_{i=1}^{n} p_i E\widehat{V}_i$.

Since $\prod_{i=1}^{n} p_i$, subject to the constraint $\sum_{i=1}^{n} p_i = 1$, attains its maximum n^{-n} at $p_i = n^{-1}$, we can define the JEL ratio at θ by

$$R(\theta) = \max \left\{ \prod_{i=1}^{n} (np_i) : \ \sum_{i=1}^{n} p_i (\widehat{V}_i - E\widehat{V}_i) = 0 \right\}. \tag{2.9}$$

Using Lagrange multiplier methods, when

$$\min_{1 \leq i \leq n} (\widehat{V}_i - E\widehat{V}_i) < 0 < \max_{1 \leq i \leq n} (\widehat{V}_i - E\widehat{V}_i),$$

the above maximum is attained at

$$p_i = \frac{1}{n} \cdot \frac{1}{1 + \gamma(\widehat{V}_i - E\widehat{V}_i)}, \tag{2.10}$$

where γ satisfies

$$f(\gamma) \equiv \frac{1}{n} \sum_{i=1}^{n} \frac{\widehat{V}_i - E\widehat{V}_i}{1 + \gamma(\widehat{V}_i - E\widehat{V}_i)} = 0. \tag{2.11}$$

After substituting the p_i's into (2.9) by those obtained in (2.10) and taking the logarithm of $R(\theta)$, we get the nonparametric jackknife empirical log-likelihood ratio

$$\log R(\theta) = -\sum_{i=1}^{n} \log\{1 + \gamma(\widehat{V}_i - E\widehat{V}_i)\}.$$

One might attempt to apply the usual EL (Owen, 1988&1990) method to this type of problems. However, there is computational difficulty caused by the presence of nonlinear constraints, since we need to solve several nonlinear equations simultaneously, which will be more difficult as the sample size n gets larger. Fortunately, the JEL method can efficiently overcome this difficulty.

To apply the JEL to the three-sample U-statistic U_n, let

$$\begin{aligned} & n = n_1 + n_2 + n_3, \\ & (\mathbf{Z}_1, \cdots, \mathbf{Z}_n) = (X_1, \cdots, X_{n_1}, Y_1, \cdots, Y_{n_2}, Z_1, \cdots, Z_{n_3}), \end{aligned} \tag{2.12}$$

and

$$T_n = U_n = \binom{n}{3}^{-1} \sum_{1 \le i < j < k \le n} \tilde{h}(\mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_k),$$

where

$$\tilde{h}(\mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_k) = \frac{n(n-1)(n-2)}{6n_1 n_2 n_3} I_{\{X_i < Y_{j-n_1} < Z_{k-n_1-n_2}\}}$$

for $1 \le i \le n_1 < j \le n_1 + n_2 < k \le n$, and 0 otherwise.

Similar to the one-sample U-statistics case, for $1 \le i \le n$, we have

$$\begin{aligned} U_{n-1}^{(-i)} &= U(\mathbf{Z}_1, ..., \mathbf{Z}_{i-1}, \mathbf{Z}_{i+1}, ..., \mathbf{Z}_n) \\ &= \binom{n-1}{3}^{-1} \sum_{(n-1,3)}^{(-i)} \tilde{h}(\mathbf{Z}_{i_1}, \mathbf{Z}_{j_1}, \mathbf{Z}_{k_1}) \\ &= \binom{n-1}{3}^{-1} \Bigg[\binom{n}{3} U_n - \sum_{i<j<k} \widetilde{h}(\mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_k) \\ &\quad - \sum_{j<i<k} \widetilde{h}(\mathbf{Z}_j, \mathbf{Z}_i, \mathbf{Z}_k) - \sum_{j<k<i} \widetilde{h}(\mathbf{Z}_j, \mathbf{Z}_k, \mathbf{Z}_i) \Bigg]. \end{aligned}$$

It follows that the jackknife pseudo-values are $(1 \le i \le n)$

$$\begin{aligned}
\widehat{V}_i &= nU_n - (n-1)U_{n-1}^{(-i)} \qquad (2.13)\\
&= \frac{6}{(n-2)(n-3)}\Bigg[\sum_{i<j<k}\widetilde{h}(\mathbf{Z}_i,\mathbf{Z}_j,\mathbf{Z}_k)\\
&\quad + \sum_{j<i<k}\widetilde{h}(\mathbf{Z}_j,\mathbf{Z}_i,\mathbf{Z}_k) + \sum_{j<k<i}\widetilde{h}(\mathbf{Z}_j,\mathbf{Z}_k,\mathbf{Z}_i)\Bigg] - \frac{2n}{n-3}U_n\\
&= \frac{n(n-1)}{n-3}\frac{1}{n_1}\frac{1}{n_2n_3}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3} I_{\{X_i<Y_j<Z_k\}}I_{\{1\le i\le n_1\}}\\
&\quad + \frac{n(n-1)}{n-3}\frac{1}{n_2}\frac{1}{n_1n_3}\sum_{j=1}^{n_1}\sum_{k=1}^{n_3} I_{\{X_j<Y_{i-n_1}<Z_k\}}I_{\{n_1+1\le i\le n_1+n_2\}}\\
&\quad + \frac{n(n-1)}{n-3}\frac{1}{n_3}\frac{1}{n_1n_2}\sum_{j=1}^{n_1}\sum_{k=1}^{n_2} I_{\{X_j<Y_k<Z_{i-n_1-n_2}\}}I_{\{n_1+n_2+1\le i\le n\}} - \frac{2n}{n-3}U_n\\
&= -\frac{2n}{n-3}U_n + \frac{n(n-1)}{n-3}\Bigg[\frac{V_{i,0,0}}{n_1}I_{\{1\le i\le n_1\}} + \frac{V_{0,i-n_1,0}}{n_2}I_{\{n_1<i\le n_1+n_2\}}\\
&\qquad + \frac{V_{0,0,i-n_1-n_2}}{n_3}I_{\{n_1+n_2<i\le n\}}\Bigg],
\end{aligned}$$

and the jackknife estimate of θ is $\widehat{U}_n(jack) = n^{-1}\sum_{i=1}^n \widehat{V}_i$.

Further, one can easily show that $U_n = \widehat{U}_n(jack)$ and for $1 \le i \le n$,

$$\begin{aligned}
E\widehat{V}_i &= \frac{n\theta}{n-3}\Bigg[\frac{n-2n_1-1}{n_1}I_{\{1\le i\le n_1\}} \qquad (2.14)\\
&\quad + \frac{n-2n_2-1}{n_2}I_{\{n_1<i\le n_1+n_2\}}\\
&\quad + \frac{n-2n_3-1}{n_3}I_{\{n_1+n_2<i\le n\}}\Bigg].
\end{aligned}$$

The following theorem states that Wilks' theorem holds for U_n. Its proof is postponed to Section 2.6.

Theorem 2.2.2 *Assume that*

$$\text{(a)}\quad \sigma^2_{1,0,0} > 0, \sigma^2_{0,1,0} > 0, \textit{ and } \sigma^2_{0,0,1} > 0;$$

$$\text{(b)}\quad 0 < \varliminf_{n\to\infty} (n_1/n_2) \leq \varlimsup_{n\to\infty} (n_1/n_2) < \infty,$$

$$0 < \varliminf_{n\to\infty} (n_2/n_3) \leq \varlimsup_{n\to\infty} (n_2/n_3) < \infty.$$

Then, as $\min(n_1, n_2, n_3) \to \infty$, *at the true value* $\theta = \theta_0$ *we have*

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

From this result, one can construct an approximate $(1 - \alpha)$ level CI for θ_0 as

$$\Theta_c = \{\theta : -2\log R(\theta) \leq c\}, \tag{2.15}$$

where c is chosen to satisfy $P\{\chi^2_1 \leq c\} = 1 - \alpha$.

2.3 Numerical study

In this section, we conduct simulation studies to investigate and compare the performance of our proposed JEL and normal approximations approaches with some other existing methods, normal approximation with Dreiseitl's estimator of variance and bootstrap calibration (See Nakas and Yiannousos, 2004), in the context of constructing of CIs for θ only. We use the following three different criteria to measure the performance of each method.

Table 2.1: $\theta_0 = 0.3407$, $F_1 = N(0,1)$, $F_2 = N(1,1)$ and $F_3 = N(1,2)$

Nominal		0.9	0.95	0.99
level		(cover., alen., clen.)	(cover., alen., clen.)	(cover., alen., clen.)
n_1=15	Normal	(0.8817, 0.3036, 0.3083)	(0.9296, 0.3618, 0.3662)	(0.9766, 0.4754, 0.4784)
n_2=15	JEL	(0.9124, 0.3102, 0.3142)	(0.9588, 0.3709, 0.3739)	(0.9896, 0.4911, 0.4928)
n_3=15	Boot.	(0.8854, 0.2977, 0.3029)	(0.9310, 0.3547, 0.3596)	(0.9808, 0.4661, 0.4689)
	Drei.	(0.8747, 0.2849, 0.2936)	(0.9165, 0.3477, 0.3493)	(0.9677, 0.4513, 0.4534)
n_1=20	Normal	(0.8833, 0.2214, 0.2228)	(0.9356, 0.2638, 0.2651)	(0.9824, 0.3467, 0.3476)
n_2=25	JEL	(0.9160, 0.2263, 0.2276)	(0.9628, 0.2711, 0.2721)	(0.9936, 0.3610, 0.3614)
n_3=30	Boot.	(0.8902, 0.2198, 0.2216)	(0.9404, 0.2621, 0.2634)	(0.9844, 0.3445, 0.3454)
	Drei.	(0.8754, 0.2063, 0.2112)	(0.9215, 0.2514, 0.2537)	(0.9701, 0.3323, 0.3356)
n_1=30	Normal	(0.8912, 0.2097, 0.2112)	(0.9403, 0.2498, 0.2510)	(0.9836, 0.3283, 0.3291)
n_2=30	JEL	(0.9056, 0.2120, 0.2133)	(0.9568, 0.2530, 0.2539)	(0.9928, 0.3339, 0.3343)
n_3=30	Boot.	(0.9057, 0.2098, 0.2110)	(0.9462, 0.2450, 0.2511)	(0.9877, 0.3285, 0.3291)
	Drei.	(0.8826, 0.1917, 0.1929)	(0.9269, 0.2275, 0.2287)	(0.9724, 0.3049, 0.3068)
n_1=35	Normal	(0.8930, 0.1754, 0.1762)	(0.9408, 0.2089, 0.2096)	(0.9858, 0.2746, 0.2749)
n_2=40	JEL	(0.9024, 0.1775, 0.1782)	(0.9542, 0.2120, 0.2125)	(0.9914, 0.2804, 0.2807)
n_3=45	Boot.	(0.8968, 0.1748, 0.1756)	(0.9448, 0.2083, 0.2091)	(0.9870, 0.2737, 0.2742)
	Drei.	(0.8883, 0.1597, 0.1621)	(0.9275, 0.1886, 0.1895)	(0.9808, 0.2635, 0.2672)
n_1=50	Normal	(0.9018, 0.1615, 0.1621)	(0.9433, 0.1924, 0.1929)	(0.9884, 0.2529, 0.2531)
n_2=50	JEL	(0.9122, 0.1626, 0.1632)	(0.9586, 0.1940, 0.1944)	(0.9926, 0.2556, 0.2559)
n_3=50	Boot.	(0.9020, 0.1604, 0.1615)	(0.9522, 0.1911, 0.1922)	(0.9892, 0.2512, 0.2519)
	Drei.	(0.8894, 0.1558, 0.1564)	(0.9388, 0.1856, 0.1861)	(0.9818, 0.2441, 0.2443)

a) Coverage probability: the probability that the true parameter value is contained in the CI. Smaller the difference between the true coverage probability and the nominal one, better the method.

(b) Average length of CIs: CIs with shorter average length are preferred since overly long CIs convey relatively imprecise information about the position of the unknown parameter.

(c) Average length conditional on coverage: average length of all CIs which cover the true parameter value.

We generate L sets of three samples $(j = 1, ..., L)$

$$\{X_1^{(j)}, ..., X_{n_1}^{(j)}\},\ \{Y_1^{(j)}, ..., Y_{n_2}^{(j)}\},\ \{Z_1^{(j)}, ..., Z_{n_3}^{(j)}\},$$

from three different distributions F_1, F_2, F_3. For each set, one can calculate $(1-\alpha)$ level CIs CI_j, $j = 1, ..., L$, using normal approximations (2.7)(Normal), JEL (2.15), Dreiseitl's method (Drei.) and bootstrap (Boot.). Denote the length of CI_j by $|CI_j|$. The Monte Carlo approximation to the coverage probability (cover.), average length (alen.) and average length conditional on coverage (clen.) are given respectively by

$$\text{(i) } L^{-1}\sum_{j=1}^{L} I_{\{\theta \in CI_j\}}, \quad \text{(ii) } L^{-1}\sum_{j=1}^{L} |CI_j|,$$
$$\text{(iii) } L_0^{-1}\sum_{j=1}^{L} |CI_j| I_{\{\theta \in CI_j\}},$$

where $L_0 = \sum_{j=1}^{L} I_{\{\theta \in CI_j\}}$, the total number of CIs covering θ.

Table 2.2: $\theta_0 = 0.6919$, $F_1 = Exp(8)$, $F_2 = Exp(1)$ and $F_3 = Exp(1/4)$

Nominal		0.9	0.95	0.99
level		(cover., alen., clen.)	(cover., alen., clen.)	(cover., alen., clen.)
n_1=15	Normal	(0.8840, 0.2961, 0.3027)	(0.9238, 0.3528, 0.3594)	(0.9688, 0.4636, 0.4687)
n_2=15	JEL	(0.9092, 0.3033, 0.3081)	(0.9524, 0.3638, 0.3677)	(0.9864, 0.4847, 0.4872)
n_3=15	Boot.	(0.8870, 0.2918, 0.2977)	(0.9269, 0.3477, 0.3531)	(0.9711, 0.4569, 0.4619)
	Drei.	(0.8729, 0.2752, 0.2817)	(0.9126, 0.3274, 0.3296)	(0.9508, 0.4332, 0.4368)
n_1=20	Normal	(0.8988, 0.2175, 0.2194)	(0.9304, 0.2593, 0.2608)	(0.9834, 0.3406, 0.3417)
n_2=25	JEL	(0.9134, 0.2221, 0.2232)	(0.9572, 0.2668, 0.2673)	(0.9918, 0.3568, 0.3571)
n_3=30	Boot.	(0.8987, 0.2152, 0.2185)	(0.9435, 0.2409, 0.2430)	(0.9838, 0.3370, 0.3390)
	Drei.	(0.8813, 0.1967, 0.1992)	(0.9259, 0.2347, 0.2363)	(0.9659, 0.3197, 0.3214)
n_1=30	Normal	(0.8966, 0.2050, 0.2067)	(0.9390, 0.2443, 0.2460)	(0.9830, 0.3210, 0.3221)
n_2=30	JEL	(0.9088, 0.2069, 0.2081)	(0.9580, 0.2476, 0.2484)	(0.9926, 0.3284, 0.3287)
n_3=30	Boot.	(0.8968, 0.2022, 0.2046)	(0.9421, 0.2409, 0.2430)	(0.9848, 0.3166, 0.3180)
	Drei.	(0.8825, 0.1877, 0.1906)	(0.9293, 0.2102, 0.2136)	(0.9685, 0.2889, 0.2897)
n_1=35	Normal	(0.8974, 0.1715, 0.1725)	(0.9474, 0.2044, 0.2053)	(0.9846, 0.2686, 0.2691)
n_2=40	JEL	(0.9004, 0.1729, 0.1735)	(0.9594, 0.2069, 0.2074)	(0.9926, 0.2747, 0.2749)
n_3=45	Boot.	(0.9003, 0.1707, 0.1716)	(0.9510, 0.2035, 0.2043)	(0.9862, 0.2674, 0.2679)
	Drei.	(0.8847, 0.1633, 0.1644)	(0.9337, 0.1769, 0.1791)	(0.9713, 0.2557, 0.2564)
n_1=50	Normal	(0.8960, 0.1571, 0.1581)	(0.9446, 0.1872, 0.1882)	(0.9848, 0.2461, 0.2466)
n_2=50	JEL	(0.9034, 0.1575, 0.1581)	(0.9534, 0.1882, 0.1888)	(0.9906, 0.2490, 0.2492)
n_3=50	Boot.	(0.8970, 0.1558, 0.1568)	(0.9463, 0.1856, 0.1867)	(0.9868, 0.2440, 0.2447)
	Drei.	(0.8876, 0.1436, 0.1468)	(0.9392, 0.1602, 0.1633)	(0.9759, 0.2297, 0.2311)

Table 2.3: $\theta_0 = 0.4019$, $F_1 = U(-1, 1)$, $F_2 = Exp(2)$ and $F_3 = Cauchy(1, 2)$

Nominal		0.9	0.95	0.99
level		(cover., alen., clen.)	(cover., alen., clen.)	(cover., alen., clen.)
n_1=15	Normal	(0.8802, 0.3375, 0.3417)	(0.9406, 0.4022, 0.4056)	(0.9802, 0.5285, 0.5308)
n_2=15	JEL	(0.9184, 0.3425, 0.3464)	(0.9598, 0.4084, 0.4113)	(0.9912, 0.5374, 0.5387)
n_3=15	Boot.	(0.8849, 0.3293, 0.3336)	(0.9426, 0.3924, 0.3965)	(0.9818, 0.5158, 0.5185)
	Drei.	(0.8753, 0.3149, 0.3166)	(0.9286, 0.3777, 0.3793)	(0.9696, 0.5005, 0.5019)
n_1=20	Normal	(0.8854, 0.2521, 0.2539)	(0.9412, 0.3003, 0.3020)	(0.9814, 0.3947, 0.3958)
n_2=25	JEL	(0.9100, 0.2553, 0.2572)	(0.9520, 0.3049, 0.3066)	(0.9888, 0.4031, 0.4039)
n_3=30	Boot.	(0.8857, 0.2496, 0.2518)	(0.9469, 0.2974, 0.2994)	(0.9822, 0.3909, 0.3921)
	Drei.	(0.8774, 0.2344, 0.2367)	(0.9316, 0.2814, 0.2835)	(0.9707, 0.3813, 0.3834)
n_1=30	Normal	(0.8856, 0.2352, 0.2368)	(0.9458, 0.2802, 0.2816)	(0.9860, 0.3683, 0.3690)
n_2=30	JEL	(0.8992, 0.2368, 0.2385)	(0.9480, 0.2822, 0.2836)	(0.9908, 0.3710, 0.3716)
n_3=30	Boot.	(0.8872, 0.2327, 0.2342)	(0.9472, 0.2772, 0.2786)	(0.9868, 0.3643, 0.3652)
	Drei.	(0.8777, 0.2178, 0.2189)	(0.9365, 0.2673, 0.2693)	(0.9746, 0.3531, 0.3554)
n_1=35	Normal	(0.8908, 0.1992, 0.2001)	(0.9458, 0.2373, 0.2380)	(0.9858, 0.3119, 0.3123)
n_2=40	JEL	(0.9078, 0.2006, 0.2015)	(0.9548, 0.2392, 0.2400)	(0.9898, 0.3150, 0.3154)
n_3=45	Boot.	(0.8927, 0.1973, 0.1984)	(0.9483, 0.2350, 0.2362)	(0.9872, 0.3089, 0.3095)
	Drei.	(0.8815, 0.1842, 0.1874)	(0.9367, 0.2183, 0.2197)	(0.9753, 0.2982, 0.3003)
n_1=50	Normal	(0.8924, 0.1810, 0.1817)	(0.9464, 0.2157, 0.2162)	(0.9868, 0.2835, 0.2838)
n_2=50	JEL	(0.9070, 0.1818, 0.1825)	(0.9544, 0.2166, 0.2172)	(0.9900, 0.2847, 0.2850)
n_3=50	Boot.	(0.8837, 0.1792, 0.1800)	(0.9482, 0.2136, 0.2143)	(0.9888, 0.2807, 0.2810)
	Drei.	(0.8839, 0.1695, 0.1706	(0.9372, 0.1996, 0.2020)	(0.9777, 0.2655, 0.2672)

In our simulations, various values of the nominal level and sample size were chosen and each experiment was based on $L = 2000$ trials with bootstrap resampling size $B = 400$, generated by routines in R.

Firstly, we consider $F_1 = N(0, 1)$, $F_2 = N(1, 1)$ and $F_3 = N(1, 2)$, which are commonly used in the literature on stress-strength models. In this situation $\theta = 0.3407$. The simulation results for this case are shown in Table 2.1.

Secondly, we select three *Exponential* populations, to see what will happen if the populations are not normal ones, and the results are given in Table 2.2.

Thirdly, we want to check what will happen if the three populations are of different kinds. We choose $F_1 = U(-1, 1)$, $F_2 = Exp(2)$ and $F_3 = Cauchy(1, 2)$. Here, $\theta = 0.4019$ and Table 2.3 contains the simulation results for this case.

Fourthly, we choose $F_1 = Cauchy(1, 2)$, $F_2 = Exp(2)$ and $F_3 = U(-1, 0.5)$. In this case, $\theta = 0.0454$, which is very close to 0 and forces us to choose moderate large sample sizes. This extreme value of θ indicates that the sample contains useful information for discrimination. Table 2.4 is for this special case.

Finally, we consider $F_1 = N(-3, 1)$, $F_2 = Exp(1)$ and $F_3 = Cauchy(6, 1)$, which gives large value of $\theta = 0.9317$, and the results are presented in Tables 2.5.

The following observation can be made from those three tables.

(1) As the sample size n increases, all methods improve in terms of all three criteria (i.e., coverage probability, average length and conditional average length),

Table 2.4: $\theta_0 = 0.0454$, $F_1 = Cauchy(1, 2)$, $F_2 = Exp(2)$ and $F_3 = U(-1, 0.5)$

Nominal		0.9	0.95	0.99
level		(cover., alen., clen.)	(cover., alen., clen.)	(cover., alen., clen.)
n_1=20	Normal	(0.8564, 0.0803, 0.0851)	(0.8970, 0.0937, 0.0986)	(0.9432, 0.1200, 0.1239)
n_2=25	JEL	(0.8780, 0.0811, 0.0848)	(0.9274, 0.0948, 0.0982)	(0.9690, 0.1220, 0.1242)
n_3=30	Boot.	(0.8577, 0.0795, 0.0829)	(0.8998, 0.0927, 0.0969)	(0.9459, 0.1181, 0.1227)
	Drei.	(0.8449, 0.0688, 0.0692)	(0.8865, 0.0847, 0.0863)	(0.9338, 0.1074, 0.1088)
n_1=25	Normal	(0.8570, 0.0701, 0.0748)	(0.8984, 0.0835, 0.0883)	(0.9448, 0.1097, 0.1138)
n_2=25	JEL	(0.8840, 0.0709, 0.0745)	(0.9266, 0.0846, 0.0881)	(0.9692, 0.1116, 0.1138)
n_3=25	Boot.	(0.8583, 0.0668, 0.0693)	(0.9005, 0.0826, 0.0856)	(0.9497, 0.1025, 0.1056)
	Drei.	(0.8454, 0.0589, 0.0596)	(0.8872, 0.0752, 0.0774)	(0.9367, 0.0963, 0.0982)
n_1=30	Normal	(0.8634, 0.0638, 0.0672)	(0.9064, 0.0760, 0.0795)	(0.9498, 0.0998, 0.1027)
n_2=30	JEL	(0.8850, 0.0644, 0.0670)	(0.9298, 0.0769, 0.0793)	(0.9756, 0.1013, 0.1028)
n_3=30	Boot.	(0.8672, 0.0615, 0.0664)	(0.9166, 0.0738, 0.0765)	(0.9563, 0.0979, 0.0997)
	Drei.	(0.8516, 0.0478, 0.0477)	(0.8955, 0.0685, 0.0699)	(0.9389, 0.0859, 0.0867)
n_1=35	Normal	(0.8746, 0.0550, 0.0571)	(0.9158, 0.0655, 0.0678)	(0.9606, 0.0861, 0.0879)
n_2=40	JEL	(0.8892, 0.0554, 0.0571)	(0.9356, 0.0660, 0.0678)	(0.9764, 0.0869, 0.0881)
n_3=45	Boot.	(0.8825, 0.0543, 0.0561)	(0.9213, 0.0639, 0.0668)	(0.9708, 0.0853, 0.0872)
	Drei.	(0.8689, 0.0388, 0.0401)	(0.9076, 0.0487, 0.0505)	(0.9451, 0.0714, 0.0734)
n_1=50	Normal	(0.8752, 0.0491, 0.0509)	(0.9190, 0.0585, 0.0603)	(0.9618, 0.0769, 0.0783)
n_2=50	JEL	(0.8904, 0.0495, 0.0509)	(0.9336, 0.0590, 0.0603)	(0.9786, 0.0776, 0.0784)
n_3=50	Boot.	(0.8847, 0.0476, 0.0494)	(0.9287, 0.0571, 0.0597)	(0.9685, 0.0736, 0.0771)
	Drei.	(0.8695, 0.0287, 0.0299)	(0.9092, 0.0392, 0.0402)	(0.9547, 0.0687, 0.0698)

Table 2.5: $\theta_0 = 0.9317$, $F_1 = N(-3, 1)$, $F_2 = Exp(1)$ and $F_3 = Cauchy(6, 1)$

Nominal level		0.9 (cover., alen., clen.)	0.95 (cover., alen., clen.)	0.99 (cover., alen., clen.)
n_1=15	Normal	(0.6518, 0.1680, 0.2409)	(0.6805, 0.2001, 0.2844)	(0.6895, 0.2630, 0.3702)
n_2=15	JEL	(0.7987, 0.2013, 0.2317)	(0.8382, 0.2461, 0.2810)	(0.8843, 0.3409, 0.3795)
n_3=15	Boot.	(0.6627, 0.1628, 0.2276)	(0.6911, 0.1941, 0.2707)	(0.7028, 0.2551, 0.3497)
	Drei.	(0.6434, 0.1562, 0.2164)	(0.6765, 0.1833, 0.2598)	(0.6812, 0.2497, 0.3362)
n_1=20	Normal	(0.8154, 0.1325, 0.1504)	(0.8493, 0.1579, 0.1790)	(0.8655, 0.2075, 0.2337)
n_2=25	JEL	(0.8358, 0.1431, 0.1569)	(0.8844, 0.1750, 0.1909)	(0.9134, 0.2435, 0.2630)
n_3=30	Boot.	(0.8215, 0.1261, 0.1451)	(0.8528, 0.1502, 0.1735)	(0.8708, 0.1975, 0.2270)
	Drei.	(0.8078, 0.1192, 0.1221)	(0.8336, 0.1474, 0.1489)	(0.8576, 0.1943, 0.1968)
n_1=30	Normal	(0.8256, 0.1323, 0.1499)	(0.8516, 0.1577, 0.1785)	(0.8677, 0.2073, 0.2338)
n_2=30	JEL	(0.8492, 0.1394, 0.1537)	(0.8727, 0.1691, 0.1858)	(0.9169, 0.2318, 0.2510)
n_3=30	Boot.	(0.8376, 0.1265, 0.1452)	(0.8596, 0.1507, 0.1734)	(0.8712, 0.1981, 0.2262)
	Drei.	(0.8169, 0.1200, 0.1236)	(0.8395, 0.1426, 0.1443)	(0.8585, 0.1919, 0.1953)
n_1=35	Normal	(0.8366, 0.1106, 0.1194)	(0.9033, 0.1318, 0.1400)	(0.9394, 0.1732, 0.1819)
n_2=40	JEL	(0.8963, 0.1150, 0.1186)	(0.9262, 0.1396, 0.1449)	(0.9535, 0.1908, 0.1975)
n_3=45	Boot.	(0.8470, 0.1104, 0.1196)	(0.9096, 0.1315, 0.1406)	(0.9408, 0.1729, 0.1812)
	Drei.	(0.8271, 0.1002, 0.1043)	(0.8861, 0.1231, 0.1257)	(0.9317, 0.1659, 0.1675)
n_1=50	Normal	(0.8587, 0.1056, 0.1130)	(0.9095, 0.1258, 0.1325)	(0.9537, 0.1654, 0.1716)
n_2=50	JEL	(0.8989, 0.1083, 0.1106)	(0.9447, 0.1309, 0.1348)	(0.9619, 0.1775, 0.1830)
n_3=50	Boot.	(0.8647, 0.1035, 0.1102)	(0.9136, 0.1234, 0.1309)	(0.9599, 0.1621, 0.1678)
	Drei.	(0.8495, 0.0869, 0.0895)	(0.8902, 0.1179, 0.1193)	(0.9477, 0.1489, 0.1497)

and the differences between these four gradually diminish.

(2) In terms of coverage probabilities, except Dreiseitl's method, the other three approaches are very competitive when the populations are of the same kind, but JEL is the best when θ is close to 0 or 1.

(3) In Tables 2.1-2.3, except the JEL method, the others are often more anti-conservative than JEL since their lengths are shorter. In Table 2.4 with small value 0.0454 of θ and Table 2.5 with $\theta = 0.9317$, all methods are clearly anti-conservative but JEL has best coverage probabilities. As the sample size becomes large, all methods improve but still remain anti-conservative in Tables 2.4-2.5.

(4) In Tables 2.1-2.5, Dreiseitl's method always produces shortest length of CIs, follow by bootstrap method, then normal approximation with jackknife estimator of variance and JEL.

In summary, in terms of coverage probability, JEL is always the best among these methods but normal approximation with jackknife estimator of variance is easy to implement.

2.4 Applications to real data

In this section, we apply our proposed statistical methods to some real examples in human health research. The first data set we will refer to below is contained in

Andrew and Herzberg (1985). The second data set was collected from a research on Alzheimer's disease (AD), and one can refer to Zhou and Castellucio (2004) and Koepsell et al. (2008) for more details about this data set.

2.4.1 Chemical and overt diabetes data

Diabetes is a disease which causes the body not to produce or properly use insulin which is an essential hormone converting sugar, starches and other food into energy. Diabetes is destructive. By destroying circulation to the heart, brain and kidneys, it increases the risk of heart attack, stroke and kidney failure. Therefore, it is important to correctly diagnose diabetes.

Basically, diabetes mellitus could be diagnosed using fasting plasma glucose level, or plasma glucose after a $75g$ oral glucose load as in a glucose tolerance test, or insulin resistance ability.

The set of data considered here was once used by Reaven and Miller (1979) to examine the relationship between chemical subclinical and overt non-ketotic diabetes in 145 non-obese adults. The subjects were clinically classified into three populations, with 76 being normal, 36 diagnosed as chemical diabetic and 33 overt diabetes. Five measurements for each individual were included in the data. They are relative weight, glucose intolerance, insulin response to oral glucose, insulin resistance (IR) and fasting plasma glucose (PLG), of which the IR was measured

Table 2.6: PLG, $\hat{\theta} = 0.7299$

Level	0.9	0.95	0.99
JEL	(0.6376, 0.8127)	(0.6180, 0.8282)	(0.5775, 0.8585)
Norm.	(0.6441, 0.8157)	(0.6276, 0.8322)	(0.5955, 0.8643)
Dre.	(0.6475, 0.8123)	(0.6317, 0.8281)	(0.6009, 0.8589)
Boot.	(0.6447, 0.8217)	(0.6277, 0.8387)	(0.5946, 0.8718)

Table 2.7: IR, $\hat{\theta} = 0.7161$

Level	0.9	0.95	0.99
JEL	(0.6198, 0.7942)	(0.5991, 0.8074)	(0.5665, 0.8320)
Norm.	(0.6296, 0.8027)	(0.6130, 0.8192)	(0.5806, 0.8516)
Dre.	(0.6335, 0.7987)	(0.6177, 0.8145)	(0.5868, 0.8455)
Boot.	(0.6231, 0.8008)	(0.6061, 0.8178)	(0.5728, 0.8511)

by the steady state plasma glucose (SSPG) determined after chemical suppression of endogenous insulin secretion.

We will apply our proposed methods, together with bootstrap calibration and normality based on Dreiseitl's variance estimator to construct CIs for the parameter we are interested in and check if the subjects were correctly classified. Here, as an illustration, we only consider the data sets of two symptoms: the PLG and the IR.

First, let X, Y and Z be the PLG measured in the normal, chemical diabetic and overt diabetic groups, respectively. Usually, $X < Y < Z$. so it is interesting to estmate $P(X < Y < Z)$, the probability that the level of glucose in the chemical diabetic group is higher than that in the normal group but lower than that in the

Table 2.8: MMSE, $\hat{\theta} = 0.3644$

Level	0.9	0.95	0.99
JEL	(0.3448, 0.3838)	(0.3410, 0.3875)	(0.3336, 0.3947)
Norm.	(0.3449,0.3840)	(0.3412, 0.3877)	(0.3338, 0.3950)
Boot.	(0.3450, 0.3824)	(0.3414, 0.3860)	(0.3343, 0.3930)

overt diabetic group. An estimator of $\theta = P(X < Y < Z)$ is given by $\hat{\theta} = 0.7299$. Employing the four methods to the data, the 90%, 95% and 99% CIs for θ are presented in Table 2.6.

Next, if X, Y and Z are respectively the IR measured in the normal, chemical diabetic and overt diabetic groups, then the corresponding estimator of $P(X < Y < Z)$ is $\hat{\theta} = 0.7161$. The 90%, 95% and 99% CIs for θ obtained via the four methods are respectively given in Table 2.7.

2.4.2 Alzheimer's disease

Alzheimer's disease (AD) is the most common form of dementia, and generally it is diagnosed in people over 65 years of age. In this example, all subjects were 65 years or above and had taken the Mini-Mental State Examination (MMSE) within 2 years before death. The examination was based on the extent of neuritic plaques and neurofibrillary tangles, the hallmarks of AD, at brain autopsy. Based on the frequency of both plaques and tangles in the neocortex, the patients were classified into one of the three different disease classes. Class I include subjects with a high likelihood of dementia being due to AD, Class II include subjects with intermediate likelihood of dementia, and Class III include subjects with low likelihood. Each patient underwent the test and a continuous test result was recorded thereafter. Totally, 3728 results were included in this analysis where the sample size for the three individual classes are 2283, 850 and 595, respectively.

We are interested in how accurately the test results are able to classify patients into the three categories. Let let X, Y and Z be the MESS results reported in Class I, II and III, respectively. The estimated value of $\theta = P(X < Y < Z)$, the probability that the frequency of both plaques and tangles in the neocortex in class II is higher than that in class I but lower than that in class III, is $\hat{\theta} = 0.3644$. This value indicates that the probability that the test correctly classifies three random subjects from the population, each from one of the three stages of AD, is about 36 percent. Such an overall accuracy measure can then be compared to other tests with similar diagnostic aims. The 90%, 95% and 99% CIs for θ obtained via the four methods except Dreiseitl's are provided in Table 2.8, since the sample sizes here are quite large and Dreiseitl's is too computationally expensive and less efficient in doing inference.

2.4.3 Summary

In summary, from above analysis, we can conclude that, compared to the naive test with $\theta_0 = 1/6$, the test based on all four methods here are efficient in terms of the value of θ as well as confidence interval, that is, the tests are statistically significant.

2.5 Conclusion

In this chapter, we proposed two statistical methods, normal approximations and the JEL method, to make statistical inference for the volume under the three-class ROC surface. We used three-sample U-statistics as unbiased estimators of these volumes and calculated their corresponding jackknifed variances. The normal approximation was based on the studentized three-sample U-statistics with jackknife estimator of variance. The computation involved here is not much complex, therefore the U-statistics methodology is applicable. Performance of these two proposed methods is compared with some existing techniques such as bootstrap calibration and normality based on Dreiseitl's variance estimator. The simulation studies suggest that our methods produce very nice statistical results. Both the bootstrap and Dreiseitl's methods are quite computationally intensive, and they are even worse as the sample sizes become large. However, our proposed methods largely relieve computation burden and run significantly faster than the other two as observed in simulation studies as well as real data analysis. Although we can not theoretically show the inequality of the two variance estimators based on jackknife and Dreiseitl's methods, interestingly, our auxiliary numerical simulations reveal that Dreiseitl's variance estimate always tends to be smaller than the jackknife one. Further, we are only interested in statistical properties of the global index $\theta = P(X < Y < Z)$, and for this reason we do not touch much of the detailed three-class classification problems.

2.6 Proof of Theorem 2.2.2

In this section, we provide the technique details to prove Theorem 2.2.2. Before proceeding to the proof of Theorem 2.2.2, we list some results that will be used. Referring to the proofs below, without loss of generality, we may assume that $n_1 \leq n_2 \leq n_3$ thereafter.

As a direct consequence of Theorem 2.2.1, we have

$$U_n - \theta_0 = O_p(n_1^{-1/2}). \tag{2.16}$$

The following Lemma guarantees the existence and uniqueness of the solution to equation (2.11).

Lemma 2.6.1 *Suppose that* $\sigma^2_{1,0,0} > 0$, $\sigma^2_{0,1,0} > 0$, $\sigma^2_{0,0,1} > 0$, $\liminf_{n\to\infty}(n_1/n_2) > 0$, *and* $\liminf_{n\to\infty}(n_2/n_3) > 0$. *Then as* $n_1 \to \infty$, *we have*

$$P\left\{\min_{1\leq i\leq n}(\widehat{V}_i - E\widehat{V}_i) < 0 < \max_{1\leq i\leq n}(\widehat{V}_i - E\widehat{V}_i)\right\} \longrightarrow 1.$$

Proof. It suffices to prove that

$$P\left\{\min_{1\leq i\leq n}(\widehat{V}_i - E\widehat{V}_i) \geq 0\right\} \to 0 \quad \text{and} \quad P\left\{\max_{1\leq i\leq n}(\widehat{V}_i - E\widehat{V}_i) \leq 0\right\} \to 0.$$

We only prove the second one since the first one can be done similarly.

Let $\xi_{ni} = \psi(\widehat{V}_i - E\widehat{V}_i)$, where $\psi(x)$ is nondecreasing, twice differentiable with bounded first and second derivatives such that

$$\psi(x) = \begin{cases} 0, & \text{if } x \le 0, \\ a(x), & \text{if } 0 < x < \delta, \\ 1, & \text{if } x \ge \delta. \end{cases}$$

with $0 < a(x) < 1$ for $0 < x < \delta$. Then similar to the proof in Jing et al. (2009), we can show that

$$P\left\{\max_{1\le i\le n}(\widehat{V}_i - E\widehat{V}_i) \le 0\right\} \le \frac{\sum_{i=1}^n \mathrm{Var}(\xi_{ni}) + \sum_{i\ne j}\mathrm{Cov}(\xi_{ni}, \xi_{nj})}{(\sum_{i=1}^n E\xi_{ni})^2},$$

and it suffices to show that, for any $i, j \in \{1, ..., n\}$ and $i < j$,

$$\text{(i) } \mathrm{Var}(\xi_{ni}) \le 1; \quad \text{(ii) } \lim_{n\to\infty} E\xi_{ni} \ge c > 0; \quad \text{(iii) } \mathrm{Cov}(\xi_{ni}, \xi_{nj}) \overset{n\to\infty}{\longrightarrow} 0.$$

Proof of (i). This is obvious since $\mathrm{Var}(\xi_{ni}) \le E\xi_{ni}^2 \le 1$;

Proof of (ii). From (2.3), simple calculations show that

$$V_{i,0,0} - \theta = h_1(X_i) - \theta + \frac{1}{n_2 n_3}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}[I_{\{X_i<Y_j<Z_k\}} - h_1(X_i)];$$
$$V_{0,j,0} - \theta = h_2(Y_j) - \theta + \frac{1}{n_1 n_3}\sum_{i=1}^{n_1}\sum_{k=1}^{n_3}[I_{\{X_i<Y_j<Z_k\}} - h_2(Y_j)];$$
$$V_{0,0,k} - \theta = h_3(Z_k) - \theta + \frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}[I_{\{X_i<Y_j<Z_k\}} - h_3(Z_k)],$$

where

$$h_1(x) = P(x < Y < Z),$$

$$h_2(y) = P(X < y < Z), \; h_3(z) = P(X < Y < z).$$

From (2.13) and (2.14), for $i : 1 \leq i \leq n_1$,

$$\begin{aligned}
\widehat{V}_i - E\widehat{V}_i &= \frac{n}{n-3}\left[\frac{n-1}{n_1}(V_{i,0,0} - \theta) - 2(U_n - \theta)\right] \\
&= \frac{n(n-1)}{n_1(n-3)}[h_1(X_i) - \theta] - \frac{2n}{n-3}(U_n - \theta) \\
&\quad + \frac{n(n-1)}{(n-3)n_1 n_2 n_3}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}[I_{\{X_i<Y_j<Z_k\}} - h_1(X_i)] \\
&\stackrel{d}{=} g_1(X_i) + R_{ni}^{(1)},
\end{aligned}$$

where

$$g_1(x) = \frac{n(n-1)}{n_1(n-3)}[h_1(x) - \theta].$$

By Taylor expansion, we have

$$\begin{aligned}
\xi_{ni} &= \psi(\widehat{V}_i - E\widehat{V}_i) = \psi[g_1(X_i) + R_{ni}^{(1)}] \\
&= \psi[g_1(X_i)] + \psi'[g_1(X_i)]R_{ni}^{(1)} + \eta_i(R_{ni}^{(1)})^2,
\end{aligned}$$

where $|\eta_i| < C$ for some constant C. In the sequel, C is always used to denote positive constant, which may vary on different occasions. Then we have, as $n_1 \to \infty$,

$$\begin{aligned}
E\xi_{ni} &= E\{\psi[g_1(X_i)]\} + E\{\psi'[g_1(X_i)]R_{ni}^{(1)}\} + E[\eta_i(R_{ni}^{(1)})^2] \qquad (2.17) \\
&= E\{\psi[g_1(X_i)]\} + E[\eta_i(R_{ni}^{(1)})^2] \\
&\to E\{\psi[g_1(X_i)]\}
\end{aligned}$$

since

$$
\begin{aligned}
E(R_{ni}^{(1)})^2 &\leq 2\left(\frac{n(n-1)}{n_1(n-3)}\right)^2 \cdot \frac{1}{n_2 n_3} E[I_{\{X_i<Y_j<Z_k\}} - h_1(X_i)]^2 \qquad (2.18) \\
&\quad +8\left(\frac{n}{n-3}\right)^2 (U_n - \theta)^2 \\
&\leq C(n_2 n_3)^{-1}(\theta + \sigma_{1,0,0}^2) + o(n_1^{-2}) \\
&\to 0
\end{aligned}
$$

But $Eg_1(X_i) = 0$ and $\sigma_{1,0,0}^2 > 0$, we get that $P\{g_1(X_i) > 0\} > 0$, which in turn implies that $E\psi[g_1(X_i)] > 0$.

Similarly, we can show that $E\psi[g_2(Y_j)] > 0$ and $E\psi[g_3(Z_k)] > 0$ for $j = n_1 + 1, \ldots, n_1 + n_2$ and $k = n_1 + n_2 + 1, \ldots, n$, respectively. This proves (ii).

Proof of (iii). By Taylor expansion,

$$
\psi(\widehat{V}_i - E\widehat{V}_i) = \begin{cases} \psi[g_1(X_i)] + \lambda_{1i} R_{ni}^{(1)}, & \text{if } 1 \leq i \leq n_1, \\ \psi[g_2(Y_{i-n_1})] + \lambda_{2i} R_{ni}^{(2)}, & \text{if } n_1 + 1 \leq i \leq n_1 + n_2, \\ \psi[g_3(Z_{i-n_1-n_2})] + \lambda_{3i} R_{ni}^{(3)}, & \text{if } n_1 + n_2 + 1 \leq i \leq n. \end{cases}
$$

where $|\lambda_{li}| \leq C$ for $l = 1, 2, 3$. Therefore, if $1 \leq i, j \leq n_1$, as $n_1 \to 0$, we have

$$
\begin{aligned}
E[\xi_{ni}\xi_{nj}] &= E\{(\psi[g_1(X_i)] + \lambda_{1i}R_{ni}^{(1)})(\psi[g_1(X_j)] + \lambda_{1j}R_{nj}^{(1)})\} \qquad (2.19) \\
&= E\{\psi[g_1(X_i)]\psi[g_1(X_j)]\} + E\{\lambda_{1i}R_{ni}^{(1)}\psi[g_1(X_j)]\} \\
&\quad +E\{\lambda_{1j}R_{nj}^{(1)}\psi[g_1(X_i)]\} + E\{\lambda_{1i}\lambda_{1j}R_{ni}^{(1)}R_{nj}^{(1)}\} \\
&\to E\{\psi[g_1(X_i)]\psi[g_1(X_j)]\} \\
&= E\{\psi[g_1(X_i)]\} \cdot E\{\psi[g_1(X_j)]\},
\end{aligned}
$$

since in virtue of (2.18), the definition of $\psi(x)$ and Cauchy inequality, we have

$$|E\{\lambda_{1i}R_{ni}^{(1)}\psi[g_1(X_j)]\}| \leq CE|R_{ni}^{(1)}| \leq C\sqrt{E(R_{ni}^{(1)})^2} \to 0,$$

similarly, $|E\{\lambda_{1j}R_{nj}^{(1)}\psi[g_1(X_i)]\}| \to 0$, and $|E[R_{ni}^{(1)}R_{nj}^{(1)}]| \leq \sqrt{E(R_{ni}^{(1)})^2E(R_{nj}^{(1)})^2} \to 0$. Now the fact $\text{Cov}(\xi_{ni}, \xi_{nj}) = E(\xi_{ni}\xi_{nj}) - E(\xi_{ni})E(\xi_{nj})$, together with (2.17) and (2.19) lead to $\text{Cov}(\xi_{ni}, \xi_{nj}) \to 0$.

The other cases can be proven similarly without difficulty. This concludes the proof of (iii). □

Lemma 2.6.2 *Let* $S_n = n^{-1}\sum_{i=1}^{n}(\widehat{V}_i - E\widehat{V}_i)^2$. *Under the conditions of* Lemma 2.6.1, *as* $n_1 \to \infty$, $S_n = nS^2_{n_1,n_2,n_3} + o(1)$ *a.s.*.

Proof. From (2.13) and (2.14), for $1 \leq i \leq n_1$, we have

$$\begin{aligned} \widehat{V}_i - E\widehat{V}_i &= \widehat{V}_i - \frac{n(n-1)}{n_1(n-3)}U_n + \frac{n(n-2n_1-1)}{n_1(n-3)}(U_n - \theta) \\ &= \frac{n(n-1)}{n_1(n-3)}(V_{i,0,0} - U_n) + \frac{n(n-2n_1-1)}{n_1(n-3)}(U_n - \theta), \end{aligned}$$

then, together with (2.3), we have

$$\begin{aligned} \sum_{i=1}^{n_1}(\widehat{V}_i - E\widehat{V}_i)^2 &= \left[\frac{n(n-1)}{n_1(n-3)}\right]^2 \sum_{i=1}^{n_1}(V_{i,0,0} - U_n)^2 \\ &\quad + \left[\frac{n(n-2n_1-1)}{n_1(n-3)}\right]^2 n_1(U_n - \theta)^2. \end{aligned}$$

Similarly, we have

$$\begin{aligned} \sum_{i=n_1+1}^{n_1+n_2}(\widehat{V}_i - E\widehat{V}_i)^2 &= \left[\frac{n(n-1)}{n_2(n-3)}\right]^2 \sum_{j=1}^{n_2}(V_{0,j,0} - U_n)^2 \\ &\quad + \left[\frac{n(n-2n_2-1)}{n_2(n-3)}\right]^2 n_2(U_n - \theta)^2, \end{aligned}$$

and

$$\sum_{i=n_1+n_2+1}^{n}(\widehat{V}_i - E\widehat{V}_i)^2 = \left[\frac{n(n-1)}{n_3(n-3)}\right]^2\sum_{k=1}^{n_3}(V_{0,0,k} - U_n)^2 + \left[\frac{n(n-2n_3-1)}{n_3(n-3)}\right]^2 n_3(U_n-\theta)^2.$$

Combining the previous three equalities and using (2.16), we get

$$\begin{aligned} S_n &= \frac{1}{n}\sum_{i=1}^{n}(\widehat{V}_i - E\widehat{V}_i)^2 \\ &= n\left(\frac{n-1}{n-3}\right)^2\left[\frac{1}{n_1^2}\sum_{i=1}^{n_1}(V_{i,0,0} - \overline{V}_{\cdot,0,0})^2 \right. \\ &\quad \left. +\frac{1}{n_2^2}\sum_{j=1}^{n_2}(V_{0,j,0} - \overline{V}_{0,\cdot,0})^2 + \frac{1}{n_3^2}\sum_{k=1}^{n_3}(V_{0,0,k} - \overline{V}_{0,0,\cdot})^2\right] \\ &\quad +\frac{n}{(n-3)^2}\left[\frac{n-2n_1-1}{n_1} + \frac{n-2n_2-1}{n_2} + \frac{n-2n_3-1}{n_3}\right](U_n-\theta)^2 \\ &= n\widehat{\mathrm{Var}}(jack) + o(1) \quad a.s. \\ &= nS^2_{n_1,n_2,n_3} + o(1) \qquad a.s. \end{aligned} \tag{2.20}$$

from which the lemma follows. □

Lemma 2.6.3 *Let*

$$\widetilde{H}_n = \max_{1\le i\le n_1<j\le n_1+n_2<k\le n}|h(X_i;Y_j;Z_k)|$$

and assume that $Eh^2(X_1;Y_1;Z_1) < \infty$. *Then under the conditions of* Lemma 2.6.1, $\widetilde{H}_n = o(n^{1/2})$ *a.s.*.

Proof. By (3.5), we can rewrite

$$\widetilde{H}_n = \max_{1\le i\le n_1<j\le n_1+n_2<k\le n}|h(X_i;Y_j;Z_k)| = \max_{1\le i<j<k\le n}|\widetilde{K}(\mathbf{Z}_i,\mathbf{Z}_j,\mathbf{Z}_k)|,$$

where

$$\widetilde{K}(\mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_k) = h(X_i; Y_{j-n_1}; Z_{k-n_1-n_2}) I_{\{1 \le i \le n_1 < j \le n_1+n_2 < k \le n\}}.$$

Then it is equivalent to prove that

$$n^{-1/2} \max_{1 \le i < j < k \le n} |\widetilde{K}(\mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_k)| = o(n^{1/2}) \quad a.s.$$

Note that

$$\begin{aligned} & n^{-1/2} \max_{1 \le i < j < k \le n} |\widetilde{K}(\mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_k)| \\ = \; & n^{-1/2} \max_{1 < k \le n} \{\max_{j<k} \{\max_{i<j} |\widetilde{K}(\mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_k)|\}\} \\ = \; & n^{-1/2} \max_{1 < k \le n} \sqrt{k} \{k^{-1/2} \max_{j<k} \{\max_{i<j} |\widetilde{K}(\mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_k)|\}\}, \end{aligned}$$

so it suffices to prove that

$$n^{-1/2} \max_{1 < j < n} \{\max_{i<j} |\widetilde{K}(\mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_n)|\} \to 0 \quad a.s.,$$

but

$$\begin{aligned} & n^{-1/2} \max_{1 < j < n} \{\max_{i<j} |\widetilde{K}(\mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_n)|\} \\ = \; & n^{-1/2} \max_{1 < j < n} \sqrt{j} \{j^{-1/2} \max_{i<j} |\widetilde{K}(\mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_n)|\} \\ \le \; & j^{-1/2} \max_{i<j} |\widetilde{K}(\mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_n)|, \end{aligned}$$

hence, we only need to show that

$$(n-1)^{-1/2} \max_{1 \le i \le n-1} |\widetilde{K}(\mathbf{Z}_i, \mathbf{Z}_{n-1}, \mathbf{Z}_n)| \to 0 \quad a.s.$$

Now by a chaining argument, it suffices to show that

$$2^{-n/2} \max_{1 \le i \le 2^n} |\widetilde{K}(\mathbf{Z}_i, \mathbf{Z}_{2^n}, \mathbf{Z}_{2^n+1})| = o(n^{1/2}) \quad a.s.$$

However, using Chebyshev inequality, for each $\varepsilon > 0$, we have

$$
\begin{aligned}
& \sum_{n=1}^{\infty} P\left\{\max_{1\le i\le 2^n} |\widetilde{K}(\mathbf{Z}_i, \mathbf{Z}_{2^n}, \mathbf{Z}_{2^n+1})| \ge \varepsilon 2^{2^n}\right\} \\
\le & \sum_{n=1}^{\infty} 2^n P\left\{|\widetilde{K}(\mathbf{Z}_1, \mathbf{Z}_{n_1+1}, \mathbf{Z}_{n_1+n_2+1})| \ge \varepsilon 2^{n/2}\right\} \\
= & \sum_{n=1}^{\infty}\sum_{m=n}^{\infty} 2^n P\{2^{(m+1)/2} > \varepsilon^{-1}|\widetilde{K}(\mathbf{Z}_1, \mathbf{Z}_{n_1+1}, \mathbf{Z}_{n_1+n_2+1})| \ge 2^{m/2}\} \\
= & \sum_{m=1}^{\infty}\sum_{n=1}^{m} 2^n P\{2^{(m+1)/2} > \varepsilon^{-1}|\widetilde{K}(\mathbf{Z}_1, \mathbf{Z}_{n_1+1}, \mathbf{Z}_{n_1+n_2+1})| \ge 2^{m/2}\} \\
\le & \sum_{m=1}^{\infty} 2^{m+1} P\{2^{(m+1)/2} > \varepsilon^{-1}|\widetilde{K}(\mathbf{Z}_1, \mathbf{Z}_{n_1+1}, \mathbf{Z}_{n_1+n_2+1})| \ge 2^{m/2}\} \\
\le & \ 2\varepsilon^{-2} E\widetilde{K}^2(\mathbf{Z}_1, \mathbf{Z}_{n_1+1}, \mathbf{Z}_{n_1+n_2+1}) < \infty.
\end{aligned}
$$

which, by the Borel-Cantelli Lemma, in turn implies that

$$2^{-n/2} \max_{1\le i\le 2^n} |\widetilde{K}(\mathbf{Z}_i, \mathbf{Z}_{2^n}, \mathbf{Z}_{2^n+1})| = o(n^{1/2}) \ \ a.s.,$$

therefore the proof is completed. □

The following corollary follows directly from Lemma 2.6.3.

Corollary 2.6.1 *If* $H_n = \max_{1\le i\le n_1<j\le n_1+n_2<k\le n} I_{\{X_i<Y_j<Z_k\}}$, *then Under the conditions of* Lemma 2.6.1, $H_n = o(n^{1/2})$ *a.s.*

Remark 2.6.1 *In fact, since*

$$\max_{1\le i\le n_1<j\le n_1+n_2<k\le n} I_{\{X_i<Y_j<Z_k\}} \le 1,$$

the conclusion $H_n = o(n^{1/2})$ *is a direct consequence of* $n^{-1/2}H_n \le n^{-1/2} \to 0$ *as* $n \to \infty$.

Lemma 2.6.4 *Let* $Q_n = \max_{1\le i\le n}|\widehat{V}_i - \theta|$. *Under the conditions of* Lemma 2.6.1, $Q_n = o(n^{1/2})$ *a.s. and* $n^{-1}\sum_{i=1}^{n}|\widehat{V}_i - E\widehat{V}_i|^3 = o(n^{1/2})$ *a.s.*.

Proof. Noting that $\underline{\lim}_{n\to\infty}(n_1/n_2) > 0$ and $\underline{\lim}_{n\to\infty}(n_2/n_3) > 0$ imply $n_2 \le Cn_1$ and $n_3 \le C'n_2$, respectively, for some positive constants C and C'. For any $i : 1 \le i \le n_1$, by (2.3), (2.13) and (2.14), we have

$$\begin{aligned}
|\widehat{V}_i - E\widehat{V}_i| &= \left|\frac{n(n-1)}{n_1(n-3)}V_{i,0,0} - \frac{n(n-2n_1-1)}{n_1(n-3)}\theta\right| \\
&= \left|\frac{n}{n-3}\frac{n-1}{n_1}\frac{n_2}{n_3}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3} I_{\{X_i<Y_j<Z_k\}}\right. \\
&\quad \left. -\frac{2n}{n-3}\frac{1}{n_1n_2n_3}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3} I_{\{X_i<Y_j<Z_k\}} - \frac{n(n-2n_1-1)}{n_1(n-3)}\theta\right| \\
&\le \frac{n}{n-3}\frac{n-1}{n_1}H_n + \frac{2n}{n-3}H_n + \frac{n}{n-3}\frac{n-2n_1-1}{n_1}|\theta| \\
&\le 4(CC'+1)H_n + 4H_n + 4C|\theta|.
\end{aligned}$$

Similarly, for any $n_1 + 1 \le i \le n_1 + n_2$ and $n_1 + n_2 + 1 \le i \le n$, we also have $|\widehat{V}_i - E\widehat{V}_i| \le 4(CC'+1)H_n + 4H_n + 4C|\theta|$. Combining the three parts together, we get that

$$|\widehat{V}_i - E\widehat{V}_i| \le 4(CC'+2)H_n + 4C|\theta|$$

holds for any $1 \le i \le n$, and hence

$$Q_n = o(n^{1/2}) \qquad a.s. \tag{2.21}$$

follows from $H_n = o(n^{1/2})$.

For the second assertion, From Theorem 2.2.1 and Lemma 2.6.2, we have

$$\begin{aligned} S_n &= nS^2_{n_1,n_2,n_3} + o(1) = \frac{n}{n_1}\sigma^2_{1,0,0} + \frac{n}{n_2}\sigma^2_{0,1,0} + \frac{n}{n_3}\sigma^2_{0,0,1} \\ &\leq (1 + C + CC')\sigma^2_{1,0,0} + (2 + C')\sigma^2_{0,1,0} + 3\sigma^2_{0,0,1} + o(1) \qquad a.s. \end{aligned}$$

Now, together with the first assertion, with probability 1

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n}|\widehat{V}_i - E\widehat{V}_i|^3 &\leq \frac{1}{n}\sum_{i=1}^{n}|\widehat{V}_i - E\widehat{V}_i|^2 \times Q_n \\ &\leq [(1 + C + CC')\sigma^2_{1,0,0} + (2 + C')\sigma^2_{0,1,0} + 3\sigma^2_{0,0,1} + o(1)] \\ &\quad \times o(n^{1/2}) \\ &= o(n^{1/2}). \end{aligned}$$

which completes the proof. □

Proof of Theorem 2.2.2. By Lemma 2.6.1, the solution to equation (2.11) exists and is unique. We next show that this solution γ satisfies $|\gamma| = O_p(n^{-1/2})$. Noting that, (2.11) together with the fundamental inequality $|x \pm y| \geq |x| - |y|$ leads to

$$\begin{aligned} 0 = |f(\gamma)| &= \frac{1}{n}\left|\sum_{i=1}^{n}(\widehat{V}_i - E\widehat{V}_i) - \gamma\sum_{i=1}^{n}\frac{(\widehat{V}_i - E\widehat{V}_i)^2}{1 + \gamma(\widehat{V}_i - E\widehat{V}_i)}\right| \\ &\geq \frac{|\gamma|S_n}{1 + |\gamma|Q_n} - \frac{1}{n}\left|\sum_{i=1}^{n}\widehat{V}_i - \theta_0\right|. \end{aligned}$$

By (2.16), the second term is $O_p(n_1^{-1/2})$. By (2.20), $S_n = nS^2_{n_1,n_2,n_3} + o(1)$ $a.s.$, it follows that

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

hence by (2.20) again, $|\gamma| = O_p(n^{-1/2})$. Further, if let $\beta_i = \gamma(\widehat{V}_i - E\widehat{V}_i)$, then

$$\begin{aligned}\max_{1\le i\le n}|\beta_i| &= |\gamma|\max_{1\le i\le n}|\widehat{V}_i - E\widehat{V}_i| \\ &= O_p(n^{-1/2})o(n^{1/2}) = o_p(1). \qquad (2.22)\end{aligned}$$

On the one hand, expanding (2.11), we get

$$\begin{aligned}0 &= f(\gamma) \\ &= \frac{1}{n}\sum_{i=1}^{n}\widehat{V}_i - \theta_0 - \gamma S_n + \frac{1}{n}\sum_{i=1}^{n}\frac{(\widehat{V}_i - E\widehat{V}_i)\beta_i^2}{1+\beta_i},\end{aligned}$$

where the last term is bounded by

$$\frac{1}{n}\sum_{i=1}^{n}\frac{|\widehat{V}_i - E\widehat{V}_i|^3}{|1+\beta_i|}\gamma^2 = o(n^{1/2})O_p(n^{-1})O_p(1) = o_p(n^{-1/2}).$$

Therefore, we may write

$$\gamma = \left(\frac{1}{n}\sum_{i=1}^{n}\widehat{V}_i - \theta_0\right)S_n^{-1} + \tau = (U-\theta_0)S_n^{-1} + \tau \qquad (2.23)$$

where $|\tau| = o_p(n^{-1/2})$.

On the other hand, in virtue of (2.22) and by a Taylor expansion, we have $\log(1+\beta_i) = \beta_i - \beta_i^2/2 + \alpha_i$, where for some finite $A > 0$,

$$P\{|\alpha_i| \le A|\beta_i|^3, 1 \le i \le n\} \to 1$$

as $n \to \infty$. Then plugging (2.23) and (2.10) into (2.9), we get

$$\begin{aligned}-2\log R(\theta_0) &= -2\sum_{i=1}^{n}\log(np_i) = 2\sum_{i=1}^{n}\log(1+\beta_i) \\ &= 2n\gamma(U-\theta_0) - nS_n\gamma^2 + 2\sum_{i=1}^{n}\alpha_i \\ &= \frac{n(U-\theta_0)^2}{S_n} - nS_n\tau^2 + 2\sum_{i=1}^{n}\alpha_i,\end{aligned}$$

where

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

$$|\sum_{i=1}^{n}\alpha_i| \le A|\gamma|^3\sum_{i=1}^{n}|\widehat{V}_i - \theta_0|^3 = O_p(n^{-3/2})o(n^{3/2}) = o_p(1),$$

and by Theorem 2.2.1 and Lemma 2.6.2, as $n \to \infty$,

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

Hence, from Slutsky's theorem, we have $-2\log R(\theta_0) \to_d \chi^2_1$, which concludes the proof of Theorem 2.2.2.

Chapter 3

Interval Estimation of the Hypervolume under ROC Manifold

3.1 Introduction

The Receiver Operating Characteristic (ROC) curves and the Area Under the ROC Curve (AUC) (Zhou et al. (2002)) are the standard methods to evaluate the accuracy of numerical diagnostic tests for two-category classification (e.g. diseased and non-diseased). Many real applications involve more than two categories. As will be shown in the tissue biomarker examples in Section 3.4, it is sometimes more relevant to differentiate multiple stages or subtypes of a disease rather than to merely distinguish between a disease or non-disease state. Thus, an extended ROC analysis capable of multi-category classification is in demand.

Scurfield (1996) brought out the mathematical definition of proper ROC measures for more than two categories. The ROC curves are extended to ROC surfaces for three-category classification and ROC manifolds for multi-category classification. The corresponding extensions of AUC are Volume under the ROC Surface (VUS) and Hypervolume under the ROC Manifold (HUM), respectively. Mossman (1999) introduced the concept of three-way ROC analysis into medical studies. Nakas and Yiannoutsos (2004) considered the estimation of VUS for ordered three-category classification by using the U-statistic theory. Li and Fine (2008) further proposed the estimation of HUM for unordered classification by following the probabilistic interpretation and applied the HUM as a model selection criterion in microarray study.

The empirical likelihood method was first introduced to construct confidence intervals for population means (Owen (1988), Owen (1990)), which enjoys many advantages over other nonparametric methods, such as automatic determination of the confidence region by the sample and transformation respecting, easy incorporation of side information and Bartlete correctability.

In this chapter, we focus on inferences of HUM for a k-category classification. As is shown in Li and Fine (2008), HUM may be interpreted as the probability

$$P(X_1 < X_2 < ... < X_k),$$

where X_i is a random variable standing for the test value for a subject randomly selected from the ith category. We assume that X_j tends to take a higher value than X_i when $j > i$ in the above expression and the probability is the largest among all $P(X_{t_1} < X_{t_2} < ... < X_{t_k})$ where $(t_1, t_2, \cdots, t_k)$ is any permutation of $(1, 2, \cdots, k)$. HUM reduces to AUC when $k = 2$ and to VUS when $k = 3$. A test with a larger HUM value would be preferred since it could correctly sort out the order of k test values each from one of the k categories with greater probability. The estimation of HUM can be carried out straight forwardly by constructing appropriate k-sample U-statistics. Asymptotic results for U-statistic could be applied in this case to provide confidence interval based inferences for HUM. However, it is noticed from our extensive numerical simulations that such an asymptotic confidence interval may not achieve the nominal coverage probability for a sample of small or moderate size. One alternative is to use a bootstrap resampling technique which usually requires intensive computation. One may also try to apply the usual empirical likelihood method to the k-sample U-statistics under consideration, but the computation burden will be very heavy as we need to solve several simultaneous nonlinear equations. One can also refer to Section 1.2.1 for explanations.

Therefore, we propose a Jackknife empirical likelihood (JEL) approach to overcome the above-mentioned difficulty. JEL introduced by Jing et al. (2009) is a fantastic marriage of two popular nonparametric approaches, jackknife and empirical likelihood method. The key idea of JEL is to turn the statistic of interest into a sample based on jackknife pseudo-values and apply Owen's EL method for the mean of those jackknife pseudo-values. As we will show, JEL maintains satisfactory small-sample accuracy, and also largely relieves the computation burden since we simply need to solve only one single nonlinear equation instead of many. To illustrate the procedure of JEL, we describe it for general one-sample U-statistics as follows.

Let $\mathbf{Z}_1, ..., \mathbf{Z}_n$ be independent (not necessarily identically distributed) r.v's and

$$T_n = T(\mathbf{Z}_1, ..., \mathbf{Z}_n) = \binom{n}{m}^{-1} \sum_{1 \le i_1 < ... < i_m \le n} h(X_{i_1}, ..., X_{i_m})$$

be a one-sample U-statistic of degree m as an unbiased estimator of the parameter θ. Define the jackknife pseudo-values by

$$\widehat{V}_i = nT_n - (n-1)T_{n-1}^{(-i)},$$

where

$$T_{n-1}^{(-i)} = \binom{n-1}{m}^{-1} \sum_{(n-1,m)}^{(-i)} h(\mathbf{Z}_{j_1}, ..., \mathbf{Z}_{j_m}),$$

here and after, $\sum_{(n-1,m)}^{(-i)}$ denotes the summation over all possible indices $(j_1, ..., j_m)$ chosen from $(1, ..., i-1, i+1, ..., n)$, subject to the restriction $1 \le j_1 < ... < j_m \le n$.

The jackknife estimator of θ is defined to be the average of the pseudo-values:

$$\widehat{T}_n(jack) \cong n^{-1} \sum_{i=1}^{n} \widehat{V}_i.$$

Let $\mathbf{p} = (p_1, ..., p_n)$ be a probability vector, i.e., $\sum_{i=1}^{n} p_i = 1$ and $p_i \ge 0$ for $1 \le i \le n$. Let $G_{\mathbf{p}}(x) = \sum_{i=1}^{n} p_i I_{\{\widehat{V}_i \le x\}}$ be the d.f. which assigns probability p_i to the ith pseudo-value $\widehat{V}_i$ and consider the mean functional $\vartheta(G_{\mathbf{p}}) = \sum_{i=1}^{n} p_i \widehat{V}_i$. The JEL, evaluated at θ, is

$$L(\theta) = \max\left\{\prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1,\ p_i \geq 0,\ \vartheta(G_{\mathbf{p}}) = \theta_{\mathbf{p}}\right\}$$

with $\theta_{\mathbf{p}} = \sum_{i=1}^{n} p_i E\widehat{V}_i$.

Since $\prod_{i=1}^{n} p_i$, restricted to the constraint $\sum_{i=1}^{n} p_i = 1$, attains its maximum n^{-n} at $p_i = n^{-1}$, we define the JEL ratio at θ by

$$R(\theta) = \max\left\{\prod_{i=1}^{n}(np_i) :\ \sum_{i=1}^{n} p_i(\widehat{V}_i - E\widehat{V}_i) = 0\right\}. \tag{3.1}$$

Using Lagrange multiplier methods, when

$$\min_{1\leq i\leq n}(\widehat{V}_i - E\widehat{V}_i) < 0 < \max_{1\leq i\leq n}(\widehat{V}_i - E\widehat{V}_i),$$

the above maximum is attained at

$$p_i = \frac{1}{n} \cdot \frac{1}{1+\gamma(\widehat{V}_i - E\widehat{V}_i)}, \tag{3.2}$$

where γ satisfies

$$f(\gamma) \equiv \frac{1}{n}\sum_{i=1}^{n} \frac{\widehat{V}_i - E\widehat{V}_i}{1+\gamma(\widehat{V}_i - E\widehat{V}_i)} = 0. \tag{3.3}$$

After substituting the p_i's into (3.1) by those obtained in (3.2) and taking the logarithm of $R(\theta)$, we get the nonparametric jackknife empirical log-likelihood ratio

$$\log R(\theta) = -\sum_{i=1}^{n} \log\{1+\gamma(\widehat{V}_i - E\widehat{V}_i)\}.$$

If one can find the asymptotic distribution of the jackknife empirical log-likelihood ratio, a $(1-\alpha)$-level confidence interval for θ can be then constructed. The superiority of JEL over the usual empirical likelihood is apparent, since the optimization problem becomes under linear constraints only.

In Section 3.2, we provide our methodology for making statistical inferences for HUM. Necessary implementation procedures and key technical results are included. In Section 3.3, we conduct extensive numerical studies to assess the performance of our proposed methods. In Section 3.4, a real example is analyzed to illustrate our methods. We offer some concluding remarks in Section 3.6.

3.2 Methodology and results

3.2.1 Asymptotic Normal approximations

Let $(X_{1,1},...,X_{1,n_1})$, $(X_{2,1},...,X_{2,n_2})$,...,$(X_{k,1},...,X_{k,n_k})$ be samples from k different populations for X_1, $X_2, \cdots X_k$, each with d.f.'s $F_1,...,F_k$, respectively. In practice, these observations could be the diagnostic test results for subjects from the k categories. Denote $n = \sum_{i=1}^{k} n_i$ to be the total sample size. We usually assume that these k samples are independent.

To estimate the parameter of interest, we may consider a U-statistic of degree

$(1, ..., 1)$ with a kernel $h(x_1; ...; x_k)$,

$$U_n = \frac{1}{\prod_{j=1}^{k} n_j} \sum_{i_1=1}^{n_1} ... \sum_{i_k=1}^{n_k} h(X_{1,i_1}; ...; X_{k,i_k}),$$

which is consistent and unbiased for $\theta = Eh(X_1; ...; X_k)$. In our problem of estimating HUM, we choose $h(x_1; ...; x_k)$ to be the indicator function $I_{\{x_1<...<x_k\}}$ for estimating the parameter $\theta = P(X_{1,1} < ... < X_{k,1})$.

Denote $\sigma^2 = E(U_n - \theta)^2$. We have a central limit theorem (CLT) for U_n, i.e., $(U_n - \theta)/\sigma \to_d N(0, 1)$ as $\min(n_1, ..., n_k) \to \infty$, where "$\to_d$" means convergence in distribution.

Since σ^2 is usually unknown, we need to replace σ^2 by its estimator. A consistent estimator $\hat{\sigma}^2$ of σ^2 can be constructed as follows. Denote $U_{n_t-1}^{(t,-i)}$ as the U-statistic after deleting $X_{t,i}$ (the i-th datum point in the t-th sample) for $t = 1, 2, ..., k$ and $i = 1, \ldots, n_t$, given by

$$\left((n_t - 1) \prod_{j=1, j\neq t}^{k} n_j\right)^{-1} \sum_{i_1=1}^{n_1} \cdots \sum_{i_t=1, i_t\neq i}^{n_t} \cdots \sum_{i_k=1}^{n_3} I_{\{X_{1,i_1}<...<X_{t,i_t}<...<X_{k,i_k}\}},$$

and for the t-th sample

$$V_t^{(-i)} = n_t U_n - (n_t - 1) U_{n_t-1}^{(t,-i)}.$$

Some simple calculations show that, for $t = 1, ..., k$, $\overline{V}_t^{(-i)} = U_n$, where $\overline{V}_t^{(-i)}$ is the average of $V_t^{(-i)}$ in the t-th sample. We may then propose a consistent estimator of $\text{Var}(U_n)$ as

$$\hat{\sigma}^2 = \sum_{t=1}^{k} \frac{1}{n_t(n_t - 1)} \sum_{i=1}^{n_t} (V_t^{(-i)} - \overline{V}_t^{(-i)})^2.$$

To state the main results for asymptotic approximation, we define

$$\sigma_t^2 = \mathrm{Var}(g_t(X_{t1}))$$

$$g_t(x) = P(X_{1,1} < \cdots < X_{t-1,1} < x < X_{t+1,1} < \cdots < X_{k,1}) - \theta.$$

Theorem 3.2.1 *Assume that for $t = 1, \ldots, k$, $\sigma_t^2 > 0$, and let $S_{n_k}^2 = \sum_{t=1}^{k} n_t^{-1}\sigma_t^2$. Then, we have*

$$(U_n - \theta)/S_{n_k} \xrightarrow{d} N(0,1) \ \textit{as} \ \min(n_1, \ldots, n_k) \to \infty,$$

and

$$\hat{\sigma}^2 - S_{n_k}^2 = o_p((\min(n_1, \ldots, n_k))^{-1}).$$

One may refer to p151-153 of Koroljuk and Borovshich (1994) for a proof of this theorem. By Theorem 3.2.1, we have a CLT for the Studentized U_n, i.e.,

$$(U_n - \theta)/\hat{\sigma} \to_d N(0,1)$$

as $\min(n_1, \ldots, n_k) \to \infty$. This enables us to construct a $100(1-\alpha)\%$ level asymptotic confidence interval for θ as

$$(U_n - z_{\alpha/2}\hat{\sigma},\ U_n + z_{\alpha/2}\hat{\sigma}). \tag{3.4}$$

3.2.2 JEL for the k-sample U-statistic U_n

To apply the JEL method to the k-sample U-statistic U_n in our problem, let

$$(\mathbf{Z}_1, \cdots, \mathbf{Z}_n) = (X_{1,1}, \cdots, X_{1,n_1}, X_{2,1}, \cdots, X_{2,n_2}, ..., X_{k,1}, \cdots, X_{k,n_k}),$$

and re-formulate U_n as

$$U_n = \binom{n}{k}^{-1} \sum_{1 \le i_1 < i_2 < ... < i_k \le n} \tilde{h}(\mathbf{Z}_{i_1}, \mathbf{Z}_{i_2}, ..., \mathbf{Z}_{i_k}),$$

where

$$\tilde{h}(\mathbf{Z}_{i_1}, \mathbf{Z}_{i_2}, ..., \mathbf{Z}_{i_k}) = \frac{\binom{n}{k}}{\prod_{j=1}^{k} n_j} I_{\{X_{1,i_1} < X_{2,i_2-n_1} < ... < X_{k,i_k - \sum_{j=1}^{k-1} n_j}\}}$$

for $1 \le i_1 \le n_1 < i_2 \le n_1 + n_2 < ... \le \sum_{j=1}^{k-1} n_j < i_k \le n$, and 0 otherwise.

Similar to the three-sample U-statistics case in Section 2.2.2, for $1 \le i \le n$, we define

$$\begin{aligned} U_{n-1}^{(-i)} &= U(\mathbf{Z}_1, ..., \mathbf{Z}_{i-1}, \mathbf{Z}_{i+1}, ..., \mathbf{Z}_n) \\ &= \binom{n-1}{k}^{-1} \sum_{(n-1,k)}^{(-i)} \tilde{h}(\mathbf{Z}_{i_1}, \mathbf{Z}_{i_2}, ..., \mathbf{Z}_{i_k}), \end{aligned}$$

where $\sum_{(n-1,k)}^{(-i)}$ denotes the summation over all possible indices $(i_1, ..., i_k)$ chosen from $(1, ..., i-1, i+1, ..., n)$, subject to the restriction $1 \le i_1 < ... < i_k \le n$.

It follows that the jackknife pseudo-values are $(1 \le i \le n)$

$$\begin{aligned} \widehat{V}_i &= nU_n - (n-1)U_{n-1}^{(-i)} \\ &= -\frac{(k-1)n}{n-k} U_n + \frac{n(n-1)}{(n-k)} \sum_{t=1}^{k} \frac{V_t^{(-i)}}{n_t} I_{\{\sum_{s=0}^{t-1} n_s < i \le \sum_{s=1}^{t} n_s\}}, \end{aligned}$$

whereafter $n_0 = 0$. The jackknife estimate of θ is then $\widehat{U}_n(jack) = n^{-1} \sum_{i=1}^{n} \widehat{V}_i$. Further, one can easily show that $U_n = \widehat{U}_n(jack)$ and

$$E\widehat{V}_i = \frac{n\theta}{n-k} \sum_{t=1}^{k} \frac{n-(k-1)n_t - 1}{n_t} I_{\{\sum_{s=0}^{t-1} n_s < i \le \sum_{s=1}^{t} n_s\}}.$$

Let $p = (p_1, ..., p_n)$ be a probability vector, G_p be the distribution function which assigns probability p_i to the ith pseudo-value $\widehat{V}_i$ and consider the mean functional $\vartheta(G_p) = \sum_{i=1}^n p_i \widehat{V}_i$. The JEL, evaluated at θ, is

$$L(\theta) = \max\left\{\prod_{i=1}^n p_i : \sum_{i=1}^n p_i = 1,\ p_i \geq 0,\ \vartheta(G_p) = \theta_p\right\}$$

with $\theta_p = \sum_{i=1}^n p_i E\widehat{V}_i$, and the JEL ratio at θ is defined by

$$R(\theta) = \max\left\{\prod_{i=1}^n (np_i) : \ \sum_{i=1}^n p_i(\widehat{V}_i - E\widehat{V}_i) = 0\right\}. \tag{3.5}$$

Using the Lagrange multiplier, when

$$\min_{1\leq i\leq n}(\widehat{V}_i - E\widehat{V}_i) < 0 < \max_{1\leq i\leq n}(\widehat{V}_i - E\widehat{V}_i),$$

the maximum in (3.5) is attained at

$$\hat{p}_i = \frac{1}{n} \cdot \frac{1}{1 + \gamma(\widehat{V}_i - E\widehat{V}_i)} \tag{3.6}$$

where γ satisfies

$$f(\gamma) \equiv \frac{1}{n}\sum_{i=1}^n \frac{\widehat{V}_i - E\widehat{V}_i}{1 + \gamma(\widehat{V}_i - E\widehat{V}_i)} = 0. \tag{3.7}$$

After substituting the $\hat{p}_i$'s into (3.5) and taking the logarithm of $R(\theta)$, we arrive at the following nonparametric jackknife empirical log-likelihood ratio

$$\log R(\theta) = -\sum_{i=1}^n \log\{1 + \gamma(\widehat{V}_i - E\widehat{V}_i)\}.$$

The next theorem states that Wilks' theorem holds for U_n. Its proof is deferred to Section 3.6.

Theorem 3.2.2 *Assume that for $t = 1, ..., k$, $\sigma_t^2 > 0$ and for $t = 2, ..., k$,*

$$0 < \underline{\lim}_{n\to\infty} (n_{t-1}/n_t) \le \overline{\lim}_{n\to\infty} (n_{t-1}/n_t) < \infty.$$

Then, as $\min(n_1, ..., n_k) \to \infty$, at the true value θ_0, we have

$$-2\log R(\theta_0) \xrightarrow{d} \chi_1^2.$$

From Theorem 3.2.2, one can construct an approximate $100(1-\alpha)\%$ level confidence interval for θ_0 as

$$\Theta_c = \{\theta : -2\log R(\theta) \le c\}, \tag{3.8}$$

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

3.3 Simulation study

In this section, we conduct simulation studies to investigate and compare the performance of JEL, normal approximation (Norm.) and bootstrap calibration of normal approximation (Boot.) in the construction of confidence intervals for θ. We use the three criteria (coverage probability, average length of confidence intervals and average length conditional on coverage) proposed in Section 2.3 to assess the performance of each method.

In each experiment, we generate L sets of four samples ($j = 1, \ldots, L$)

$$\{X_1^{(j)}, \ldots, X_{n_1}^{(j)}\}, \ \{X_1^{(j)}, \ldots, X_{n_2}^{(j)}\}, \ \{X_1^{(j)}, \ldots, X_{n_3}^{(j)}\} \text{ and } \{X_1^{(j)}, \ldots, X_{n_4}^{(j)}\},$$

Table 3.1: $F_1 = N(0,1)$, $F_2 = N(6,1)$, $F_3 = N(9,1)$, $F_4 = N(12,1)$ and $\theta_0 = 0.9662$

Nominal Level		0.9	0.95	0.99
Size	Methods	(cover., alen., clen.)	(cover., alen., clen.)	(cover., alen., clen.)
n_1=15	JEL	(0.809, 0.0902, 0.1015)	(0.884, 0.1098, 0.1191)	(0.917, 0.1507, 0.1613)
n_2=15	Norm.	(0.759, 0.0821, 0.0998)	(0.794, 0.0978, 0.1169)	(0.853, 0.1286, 0.1473)
n_3=15	Boot.	(0.766, 0.0741, 0.0954)	(0.804, 0.0882, 0.0882)	(0.868, 0.1158, 0.1476)
n_4=15				
n_1=35	JEL	(0.860, 0.0534, 0.0568)	(0.914, 0.0648, 0.0682)	(0.958, 0.0886, 0.0913)
n_2=35	Norm.	(0.837, 0.0518, 0.0564)	(0.878, 0.0617, 0.0666)	(0.909, 0.0812, 0.0861)
n_3=35	Boot.	(0.859, 0.0514, 0.0568)	(0.892, 0.0612, 0.0612)	(0.924, 0.0804, 0.0866)
n_4=35				
n_1=45	JEL	(0.866, 0.0456, 0.0473)	(0.914, 0.0554, 0.0569)	(0.976, 0.0757, 0.0769)
n_2=45	Norm.	(0.846, 0.0447, 0.0474)	(0.886, 0.0533, 0.0559)	(0.932, 0.0701, 0.0722)
n_3=50	Boot.	(0.852, 0.0444, 0.0476)	(0.903, 0.0529, 0.0529)	(0.939, 0.0696, 0.0726)
n_4=50				
n_1=55	JEL	(0.885, 0.0432, 0.0447)	(0.934, 0.0524, 0.0538)	(0.976, 0.0713, 0.0724)
n_2=55	Norm.	(0.860, 0.0426, 0.0452)	(0.907, 0.0507, 0.0535)	(0.945, 0.0666, 0.0692)
n_3=55	Boot.	(0.860, 0.0421, 0.0450)	(0.912, 0.0501, 0.0501)	(0.957, 0.0658, 0.0691)
n_4=55				

from the populations F_1, F_2, F_3 and F_4 respectively. For each set, we calculate $(1-\alpha)$ level confidence intervals CI_j $(j = 1,\ldots,L)$, using JEL (3.8), normal approximation (3.4) and bootstrap calibration (Li and Fine (2008)). Denote the length of CI_j by $|CI_j|$. The Monte Carlo approximation to the coverage probability (cov.), average length (alen.) and average length conditional on coverage (clen.) are respectively given by

$$\text{(i) } L^{-1}\sum_{j=1}^{L} I_{\{\theta\in CI_j\}},\ \text{(ii) } L^{-1}\sum_{j=1}^{L}|CI_j|,\ \text{(iii) } L_0^{-1}\sum_{j=1}^{L}|CI_j|I_{\{\theta\in CI_j\}},$$

where $L_0 = \sum_{j=1}^{L} I_{\{\theta\in CI_j\}}$, the total number of confidence intervals covering θ.

In our simulations, various values of the nominal level and sample size were chosen and each experiment was based on $L = 2000$ trials with bootstrap resampling size $B = 400$, generated by routines in R. We only present two cases in this paper. In the first case, we consider $F_1 = N(0,1)$, $F_2 = N(6,1)$, $F_3 = N(9,1)$, and $F_4 = N(12,1)$. The true HUM is $\theta_0 = 0.9662$ in this case and the simulation results are shown in Table 3.1. In the second case, we choose exponential distributions $F_1 = Exp(8)$, $F_2 = Exp(1)$, $F_3 = Exp(1/4)$, and $F_4 = Exp(1/16)$. The true HUM is $\theta_0 = 0.5239$ and Table 3.2 contains the simulation results for this case.

The following observation can be made from the two tables.

(1) As the sample size n increases, all methods improve in terms of all three criteria (i.e., coverage probability, average length and conditional average length), and the differences among those three methods gradually disappear.

Table 3.2: F_1=$Exp(8)$, F_2=$Exp(1)$, F_3=$Exp(1/4)$, F_4=$Exp(1/16)$, θ_0=0.5239

Nominal Level		0.9	0.95	0.99
Size	Methods	(cover., alen., clen.)	(cover., alen., clen.)	(cover., alen., clen.)
n_1=15	JEL	(0.913, 0.3092, 0.3118)	(0.961, 0.3701, 0.3719)	(0.991, 0.4911, 0.4921)
n_2=15	Norm.	(0.882, 0.3017, 0.3047)	(0.937, 0.3595, 0.3620)	(0.988, 0.4725, 0.4736)
n_3=15	Boot.	(0.912, 0.3091, 0.3117)	(0.959, 0.3683, 0.3731)	(0.992, 0.4840, 0.4853)
n_4=15				
n_1=35	JEL	(0.906, 0.1917, 0.1924)	(0.957, 0.2290, 0.2294)	(0.992, 0.3024, 0.3026)
n_2=35	Norm.	(0.891, 0.1905, 0.1912)	(0.942, 0.2270, 0.2275)	(0.993, 0.2984, 0.2986)
n_3=35	Boot.	(0.904, 0.1928, 0.1936)	(0.960, 0.2297, 0.2303)	(0.989, 0.3019, 0.3020)
n_4=35				
n_1=45	JEL	(0.903, 0.1603, 0.1608)	(0.954, 0.1915, 0.1919)	(0.991, 0.2530, 0.2530)
n_2=45	Norm.	(0.894, 0.1598, 0.1603)	(0.944, 0.1904, 0.1908)	(0.992, 0.2502, 0.2503)
n_3=50	Boot.	(0.904, 0.1616, 0.1618)	(0.956, 0.1925, 0.1927)	(0.991, 0.2530, 0.2531)
n_4=50				
n_1=55	JEL	(0.900, 0.1497, 0.1502)	(0.951, 0.1788, 0.1791)	(0.991, 0.2359, 0.2360)
n_2=55	Norm.	(0.905, 0.1497, 0.1501)	(0.954, 0.1784, 0.1787)	(0.988, 0.2344, 0.2346)
n_3=55	Boot.	(0.902, 0.1514, 0.1519)	(0.955, 0.1804, 0.1825)	(0.991, 0.2371, 0.2373)
n_4=55				

(2) In terms of coverage probability, all three methods are very competitive but JEL is the best.

(3) In terms of average length, bootstrap calibration seems to be the best, followed by the normal approximation and then the JEL.

(4) In Table 3.1, the JEL, normal approximations and bootstrap calibration methods are always anti-conservative. The normal approximation method is always more anti-conservative than the JEL and bootstrap calibration methods since it has shorter length of confidence intervals. As n increases, all three methods improve, but are still anti-conservative. In Table 3.2, the normal approximation method is always anti-conservative while the other methods are often conservative.

From above observation, we conclude that in terms of finite sample coverage probabilities, the JEL method is always the best among these three approaches.

3.4 Application to tissue biomarkers of synovitis

We now apply our proposed method to a real example about tissue biomarkers for synovitis which is known as the medical condition for inflammation of the synovial membrane. Pessler et al. (2008a and b) examined the differential ability in the expression of synovial tissue markers. The authors identified eight immunohistochemical synovial biomarkers which may be used to differentiate among several

Table 3.3: Sample sizes for synovitis data.

Category	Sample size
Non-inflamed healthy control (X_1)	23
Non-inflamed orthopedic arthropathies (X_2)	26
Osteoarthritis (X_3)	18
Early undifferentiated arthritis (X_4)	10
Rheumatoid arthritis with active disease (X_5)	11
Chronic septic arthritis (proven by positive bacterial culture, X_6)	24

inflammatory and non-inflammatory arthropathies and normal synovium. We revisit their data to illustrate our methods. In this paper the eight biomarkers are denoted by TM (Total mononuclear cells), Ki67 (proliferating cells), CD15 (neutrophilic granulocytes), VWF (vascular endothelium, polyclonal rabbit IgG), CD38 (plasma cells), CD68 (macrophages), CD3 (T cells, antibody clone PS1) and CD20 (B cells, L-26), respectively. Each marker has a continuous score and may be regarded as a diagnostic test.

The outcome for each patient was verified by review of operative and arthroscopy reports, pathology reports and the patients' hospital records. In this data set, the patient outcome involves six different categories. The sample sizes for these categories are summarized in Table 3.3. The scientific question is to quantify how often the tissue marker can differentiate the six categories and find out those desirable tissue markers with the best diagnostic accuracy.

To compare across the multiple categories, Pessler et al. (2008a and b) chose to conduct pairwise ROC analysis for all two-category pairs and reported the pairwise AUC values accordingly. This approach is rather limited since it produced multiple AUC values for each of the eight markers. It is difficult to compare the overall diagnostic accuracy of the markers from such AUC values.

We thus applied the methods introduced in this paper to compute the HUM values θ. According to its probabilistic interpretation, HUM indicates how often the marker sorts the six categories correctly and therefore is an appropriate measure in this case. We note that exact definitions of θ may vary for different markers. In practice, we should always select the θ which is the largest among all $6! = 720$ permutations of the category index $(1, 2, 3, 4, 5, 6)$. The estimated HUM's and associated 95% confidence intervals are presented in Table 3.4.

HUM values sort the eight markers for their differentiability among the six categories. In this case, TM, Ki67, and CD15 are the top three markers with the highest HUM values. We notice that a naive marker which randomly sorts six categories only has a HUM value ≈ 0.001. The three markers are 100 times more accurate than a naive marker. These markers have also been recognized in Pessler et al. (2008a and b) for having good individual pairwise AUC values. However, previously one could not conclude the overall superiority of these markers. Our calculation results confirm their overall quality.

The interval estimation also provides further insight on the diagnostic accuracy

Table 3.4: 95% confidence intervals by JEL and Norm.

Rank	Marker	Definition of HUM	θ	CI (JEL)	CI (Norm.)
1	TM	$P_{\{X_4<X_1<X_3<X_2<X_6<X_5\}}$	0.144	(0.046, 0.243)	(0.052, 0.236)
2	Ki67	$P_{\{X_4<X_6<X_2<X_1<X_3<X_5\}}$	0.106	(0.007, 0.206)	(0.015, 0.197)
3	CD15	$P_{\{X_4<X_5<X_2<X_1<X_3<X_6\}}$	0.101	(0., 0.199)	(0.019, 0.182)
4	VWF	$P_{\{X_5<X_1<X_2<X_3<X_4<X_6\}}$	0.069	(0.004, 0.141)	(0.004, 0.133)
5	CD38	$P_{\{X_5<X_1<X_3<X_2<X_6<X_4\}}$	0.051	(0.008, 0.094)	(0.012, 0.090)
6	CD68	$P_{\{X_5<X_1<X_3<X_2<X_6<X_4\}}$	0.047	(0., 0.115)	(0., 0.104)
7	CD3	$P_{\{X_4<X_3<X_2<X_1<X_6<X_5\}}$	0.033	(0.002, 0.067)	(0.003, 0.063)
8	CD20	$P_{\{X_5<X_3<X_1<X_2<X_6<X_4\}}$	0.011	(0., 0.027)	(0., 0.024)

of the tissue markers. The normal confidence intervals are quite different from the JEL confidence intervals and may not be appropriate due to the small sample sizes. In fact all the intervals obtained from normal approximation tend to be unnecessarily narrower than those obtained from JEL. The confidence intervals for CD68, CD3 and CD20 all include the value 0.001 and therefore are not statistically significantly different from a useless test. These markers may not be helpful for clinical practice and should not deserve the same amount of research attention as those markers with higher HUM values.

3.5 Discussion

We considered a new inference technique for diagnostic medicine when the number of categories of disease status are more than two. The methodology introduced here may have broader applications for various classification tasks involved in finance, economics and engineering etc.

There is one potential technical difficulty with the estimation of HUM. When there are k categories, usually we have to determine the most sensible HUM values by choosing the one with the largest numeric value among all $k!$ possible orders of categories. This selection process could be potentially time-consuming. However, in most medical problems that we come across with, the number of categories are usually less than ten. Sometimes it may be also advisable to combine certain similar categories when the samples are not large enough to provide valid statistical inferences.

3.6 Proof of Theorem 3.2.2

In this part, we provide the technique details to prove Theorem 3.2.2. Referring to the proofs below, without loss of generality, we may assume that $n_1 \leq \cdots \leq n_k$ thereafter.

As a direct consequence of Theorem 3.2.1, we have

$$U_n - \theta_0 = O_p(n_1^{-1/2}). \tag{3.9}$$

The following Lemma guarantees the existence and uniqueness of the solution to equation (3.7).

Lemma 3.6.1 *Assume that for* $t = 1, \ldots, k$, $\sigma_t^2 > 0$, *and* $\liminf_{n\to\infty}(n_{t-1}/n_t) > 0$ *for* $t = 2, \ldots, k$. *Then as* $n_1 \to \infty$, *we have*

$$P\left\{\min_{1\le i\le n}(\widehat{V}_i - E\widehat{V}_i) < 0 < \max_{1\le i\le n}(\widehat{V}_i - E\widehat{V}_i)\right\} \longrightarrow 1.$$

Proof. The proof is quite similar to the three-sample U-statistics case and can be directly extended to the k-sample problem. The reader is referred to Lemma 2.6.1 for details.

Let $S_n = n^{-1}\sum_{i=1}^n(\widehat{V}_i - E\widehat{V}_i)^2$ and $Q_n = \max\limits_{1\le i\le n}|\widehat{V}_i - E\widehat{V}_i|$. Under the conditions of Lemma 3.6.1, some calculations show that, similar to Lemma 2.6.2 and Lemma 2.6.4, as $n_1 \to \infty$, with probability 1

$$S_n = nS_{n_k}^2 + o(1), \quad Q_n = o(n^{1/2}) \tag{3.10}$$

and

$$n^{-1}\sum_{i=1}^n |\widehat{V}_i - E\widehat{V}_i|^3 \le S_n \times Q_n \le o(n^{1/2}). \tag{3.11}$$

Proof of Theorem 3.2.2. By Lemma 3.6.1, the solution to equation (3.7) exists and is unique. We next show that this solution γ satisfies $|\gamma| = O_p(n^{-1/2})$.

Noting that, (3.7) together with the fundamental inequality $|x \pm y| \geq |x| - |y|$ leads to

$$\begin{aligned} 0 = |f(\gamma)| &= \frac{1}{n}\left|\sum_{i=1}^{n}(\widehat{V}_i - E\widehat{V}_i) - \gamma\sum_{i=1}^{n}\frac{(\widehat{V}_i - E\widehat{V}_i)^2}{1+\gamma(\widehat{V}_i - E\widehat{V}_i)}\right| \\ &\geq \frac{|\gamma| S_n}{1+|\gamma| Q_n} - \frac{1}{n}\left|\sum_{i=1}^{n}\widehat{V}_i - \theta_0\right|. \end{aligned}$$

By (3.9), the second term is $O_p(n_1^{-1/2})$. By (3.10), $S_n = nS_{n_k}^2 + o(1)$ $a.s.$, it follows that

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

hence by (3.10) again, $|\gamma| = O_p(n^{-1/2})$. Further, if let $\beta_i = \gamma(\widehat{V}_i - E\widehat{V}_i)$, then

$$\begin{aligned} \max_{1\leq i\leq n}|\beta_i| &= |\gamma|\max_{1\leq i\leq n}|\widehat{V}_i - E\widehat{V}_i| \\ &= O_p(n^{-1/2})o(n^{1/2}) = o_p(1). \end{aligned} \tag{3.12}$$

On the one hand, expanding (3.7), we get

$$\begin{aligned} 0 &= f(\gamma) \\ &= \frac{1}{n}\sum_{i=1}^{n}\widehat{V}_i - \theta_0 - \gamma S_n + \frac{1}{n}\sum_{i=1}^{n}\frac{(\widehat{V}_i - E\widehat{V}_i)\beta_i^2}{1+\beta_i}, \end{aligned}$$

where the last term is bounded by

$$\frac{1}{n}\sum_{i=1}^{n}\frac{|\widehat{V}_i - E\widehat{V}_i|^3}{|1+\beta_i|}\gamma^2 = o(n^{1/2})O_p(n^{-1})O_p(1) = o_p(n^{-1/2}).$$

Therefore, we may write

$$\gamma = \left(\frac{1}{n}\sum_{i=1}^{n}\widehat{V}_i - \theta_0\right)S_n^{-1} + \tau = (U_n - \theta_0)S_n^{-1} + \tau \tag{3.13}$$

where $|\tau| = o_p(n^{-1/2})$.

On the other hand, using (3.12) and by a Taylor expansion, we have

$$\log(1+\beta_i) = \beta_i - \beta_i^2/2 + \alpha_i,$$

where for some finite constant $A > 0$, $P\{|\alpha_i| \le A|\beta_i|^3, 1 \le i \le n\} \to 1$ as $n \to \infty$. Then plugging (3.13) and (3.6) into (3.5), we get

$$\begin{aligned} -2\log R(\theta_0) &= -2\sum_{i=1}^{n}\log(np_i) = 2\sum_{i=1}^{n}\log(1+\beta_i) \\ &= 2n\gamma(U_n - \theta_0) - nS_n\gamma^2 + 2\sum_{i=1}^{n}\alpha_i \\ &= \frac{n(U_n-\theta_0)^2}{S_n} - nS_n\tau^2 + 2\sum_{i=1}^{n}\alpha_i, \end{aligned}$$

where

$$|-nS_n\tau^2| = n(nS_{n_k}^2 + o(1))o_p(n^{-1}) = o_p(1),$$

$$|\sum_{i=1}^{n}\alpha_i| \le A|\gamma|^3\sum_{i=1}^{n}|\widehat{V}_i - \theta_0|^3 = O_p(n^{-3/2})o(n^{3/2}) = o_p(1)$$

and by Theorem 3.6.1 and (3.10), as $n \to \infty$,

$$\frac{n(U_n-\theta_0)^2}{S_n} \xrightarrow{d} \chi_1^2.$$

Now from Slutsky's theorem, we have $-2\log R(\theta_0) \to_d \chi_1^2$, which completes the proof.

Chapter 4

Empirical Likelihood for Compound Poisson Sum

4.1 Introduction

Let $\{X_j\}_{j=1}^{\infty}$ be a sequence of i.i.d. r.v.'s with common d.f. F. Define a renewal counting process $\{N(t), t > 0\}$ by $N(t) = \max\{k : T_k \leq t\}$, where T_k is the occurrence time of X_k. Then $N(t)$ can be interpreted as the number of occurrences X_k in $(0, t]$. Further, suppose that $\{N(t), t > 0\}$ is independent of the sequence $\{X_j\}_{j=1}^{\infty}$ and write

$$S_{N(t)} = \sum_{j=1}^{N(t)} X_j,$$

then the stochastic process $\{S_{N(t)}, t > 0\}$ is called a *renewal reward process* (for definiteness, we assume that $S_{N(t)} = 0$ if $N(t) = 0$). When $\{N(t), t > 0\}$ is a Poisson process, the renewal reward process $S_{N(t)}$ is termed as a *compound Poisson process* (CPP), which is frequently used to describe phenomena in the field of applied probability when a single Poisson process fails to do so.

One example of CPPs is in spatial study in physics. Consider the energy received by some region of the surface of the Earth from cosmic particles up to time t. Let $N(t)$ be the total number of particles arrived up to time t and X_j the energy of the jth particle, then $S_{N(t)}$ is the total energy received by the region up to time t.

Another typical example is in mining industry. Denote by $N(t)$ the number of disasters up to time t, and X_j the number of death in the jth disaster. Usually, $\{N(t), t > 0\}$ can be assumed to be a Poisson process. Then, $S_{N(t)}$ is the total number of death in all the disasters up to time t.

The other popular example appears in actuarial applications. Let $N(t)$ be the number of claims up to time t and X_j the amount of the jth claim, then $S_{N(t)}$ is accumulation of money claims up to time t.

One may also expect more applications of CPPs in applied fields such as finance, risk theory; e.g., see Helmers et al (2003) for some developments on compound Poisson sums and their relevance in finance. Excellent interpretations and more examples of CPPs may be found in Karlin and Taylor (1981, p426), and Parzen

(1967, p129-130); see also Gnedenko and Korolev (1996) for the general theory of random sums.

It is well known that for a renewal reward process $\{S_{N(t)}, t > 0\}$, if $N(t)/t$ converges in probability to a constant or, more generally, to a positive r.v. (Rényi (1957), Blum et al (1963)), then $S_{N(t)}$ is asymptotically normally distributed, i.e.,

$$\frac{S_{N(t)} - EN(t)EX_1}{\sqrt{EN(t)EX_1^2}} \to_d N(0,1), \quad \text{as } t \to \infty$$

where "$\to_d$" means convergence in distribution and $N(0,1)$ denotes a standard normal r.v. Especially, when $\{N(t), t > 0\}$ is a Poisson process with rate $\lambda > 0$, independent of the i.i.d. r.v.'s $X_1, X_2, ...$ with mean $\mu = EX_1$ and variance $\sigma^2 = \text{Var}(X_1) > 0$, we can use this asymptotic normality to construct confidence intervals (CIs) for $\lambda\mu$. But the main problem is that, as pointed out by Helmers (2003), the usual normal approximation for compound Poisson sums usually performs very badly because, typically in insurance applications, the distribution of the X_i is highly skewed to the right. This urges for better methods, e.g. the bootstrap or Edgeworth/saddlepoint approximations (see, Babu et al. (2003) for results on Edgeworth expansion and Jing et al. (2009) on saddlepoint approximation), to construct more accurate confidence intervals for $\lambda\mu$. One can also consider a Studentized CPP, which is motivated by the fact that a natural consistent estimator of the variance of $(S_{N(t)} - \lambda\mu t)$ is given by $\Delta_{N(t)} = \sum_{j=1}^{N(t)} X_j^2$. Therefore, one can construct approximate $(1-\alpha)$ level CIs for $\lambda\mu$ as

$$\left(t^{-1}S_{N(t)} - z_{\alpha/2}t^{-1}\Delta_{N(t)}^{1/2},\ t^{-1}S_{N(t)} + z_{\alpha/2}t^{-1}\Delta_{N(t)}^{1/2}\right), \tag{4.1}$$

from the fact that

$$\frac{S_{N(t)} - \lambda\mu t}{\sqrt{\Delta_{N(t)}}} \to_d N(0,1), \text{ as } t \to \infty, \tag{4.2}$$

where $\Phi(z_{\alpha/2}) = 1 - \alpha/2$.

Babu et al. (2003) establishes Edgeworth expansions for the Studentized compound Poisson processes, when the distribution of X_1 is absolutely continuous and $EX_1^6 < \infty$. From this we can also construct CIs for $\lambda\mu$ as

$$\left(\frac{S_{N(t)}}{t} - \frac{(z_{\alpha/2} - \hat{p}_{\alpha/2})}{t}\sqrt{\Delta_{N(t)}},\ \frac{S_{N(t)}}{t} + \frac{(z_{\alpha/2} + \hat{p}_{\alpha/2})}{t}\sqrt{\Delta_{N(t)}}\right), \tag{4.3}$$

where

$$\begin{aligned} \hat{p}_{\alpha/2} &= [(2\hat{\mu}_3 - \bar{X}^3_{N(t)} + 3\bar{X}_{N(t)}(\hat{v}^2 + \hat{\sigma}^2)z^2_{\alpha/2}) \\ &\quad + \hat{\mu}_3 + (\bar{X}_{N(t)})^3 + 3\bar{X}_{N(t)}\hat{\sigma}^2]/(6\hat{v}^3\sqrt{N(t)}), \end{aligned}$$

with

$$\bar{X}_{N(t)} = \frac{S_{N(t)}}{N(t)}, \quad \hat{\sigma}^2 = \frac{1}{N(t)}\sum_{j=1}^{N(t)}(X_j - \bar{X}_{N(t)})^2,$$

$$\hat{v}^2 = \frac{\Delta_{N(t)}}{N(t)}, \quad \hat{\mu}_3 = \frac{1}{N(t)}\sum_{j=1}^{N(t)}(X_j - \bar{X}_{N(t)})^3.$$

Kegler (2007) uses

$$\left(\exp\left\{\log\left(\frac{S_{N(t)}}{t}\right) - z_{\alpha/2}\sqrt{\frac{\Delta_{N(t)}}{S^2_{N(t)}}}\right\}, \exp\left\{\log\left(\frac{S_{N(t)}}{t}\right) + z_{\alpha/2}\sqrt{\frac{\Delta_{N(t)}}{S^2_{N(t)}}}\right\}\right) \tag{4.4}$$

as CI for $\lambda\mu$. However, this method is applicable only when $S_{N(t)} > 0$.

In this chapter, we propose to use Owen's EL method to construct CIs. In Section 4.2, we present our main result. Simulation studies are presented in Section

4.3 to illustrate and compare the performance of our method with other methods. A small real data set is analyzed in Section 4.4 and the proofs are provided in Section 4.5.

4.2 Methodology and results

For the reader's convenience, we briefly describe the EL procedure for CPPs as follows.

Let $\mathbf{p} = (p_1, ..., p_n)$ be a probability vector and $G_{\mathbf{p}}$ be the d.f. which assigns probability p_i to the ith atom X_i. Since $\lambda\mu t = E\left(\sum_{j=1}^{N(t)} X_j\right)$, we can argue that

$$E\left(\sum_{j=1}^{N(t)} X_j \middle| N(t) = n\right) \approx \lambda\mu t.$$

This leads us to consider the following EL

$$L(\theta|N(t) = n) = \max\left\{\prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1, p_i \geq 0, \vartheta(G_{\mathbf{p}}) = \theta t/n\right\},$$

where $\vartheta(G_{\mathbf{p}}) = \sum_{i=1}^{n} p_i X_i$ and $\theta = \lambda\mu$.

The corresponding EL ratio is

$$\Re(\theta|N(t) = n) = \max\left\{\prod_{i=1}^{n}(np_i) : \sum_{i=1}^{n} p_i = 1, p_i \geq 0, \vartheta(G_{\mathbf{p}}) = \theta t/n\right\}. \qquad (4.5)$$

Applying Lagrange multiplier method, when $\min_i X_i < \theta t/n < \max_i X_i$, we have

$$p_i = \frac{1}{n} \cdot \frac{1}{1 + \gamma(X_i - \theta t/n)},$$

where γ satisfies

$$f(\gamma) \equiv \frac{1}{n}\sum_{i=1}^{n} \frac{X_i - \theta t/n}{1+\gamma(X_i - \theta t/n)} = 0. \tag{4.6}$$

After plugging the p_i's back into (4.5) and taking the logarithm of $\Re(\theta)$, we get the nonparametric empirical log-likelihood ratio conditional on $N(t) = n$,

$$\log \Re(\theta | N(t) = n) = -\sum_{i=1}^{n} \log[1 + \gamma(X_i - \theta t/n)].$$

Let

$$\omega_{N(t)} = \frac{\sum_{i=1}^{N(t)} \left(X_i - \bar{X}_{N(t)}\right)^2}{\sum_{i=1}^{N(t)} X_i^2}$$

and $\theta_0 = \mu_0\lambda_0$ be the true value of θ. After removing the condition $N(t) = n$, we can get Wilks' theorem for the *adjusted* empirical log-likelihood ratio.

Theorem 4.2.1 *Assume that $EX_1^2 < \infty$ and $\sigma^2 > 0$, then at the true value $\theta = \theta_0$, as $t \to \infty$,*

$$-2\omega_{N(t)} \log \Re(\theta_0) \to_d \chi_1^2.$$

From Theorem 4.2.1, one can construct an approximate $(1-\alpha)$ level CI for $\lambda\mu$ as

$$\Theta_c = \{\theta : -2\omega_{N(t)} \log \Re(\theta_0) \le c\}, \tag{4.7}$$

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

Table 4.1: $F = Exp(1/2)$ and $\lambda_0 = 0.5$

Level		0.9	0.95
Size		(cov., avecov., concov., alen., clen.)	(cov., avecov., concov., alen., clen.)
t=35	EL	(0.861, 0.803, 0.770, 1.072, 1.075)	(0.913, 0.715, 0.714, 1.277, 1.281)
	Norm.	(0.856, 0.796, 0.759, 1.078, 1.130)	(0.904, 0.704, 0.675, 1.284, 1.341)
	E.E.	(0.858, 0.796, 0.790, 1.078, 1.083)	(0.910, 0.709, 0.707, 1.284, 1.287)
	Keg.	(0.876, 0.773, 0.767, 1.134, 1.143)	(0.934, 0.677, 0.673, 1.380, 1.387)
t=45	EL	(0.869, 0.934, 0.932, 0.931, 0.932)	(0.921, 0.824, 0.789, 1.118, 1.120)
	Norm.	(0.857, 0.910, 0.873, 0.942, 0.982)	(0.908, 0.809, 0.782, 1.122, 1.160)
	E.E.	(0.866, 0.922, 0.916, 0.942, 0.948)	(0.917, 0.817, 0.815, 1.122, 1.125)
	Keg.	(0.887, 0.905, 0.896, 0.979, 0.989)	(0.932, 0.785, 0.781, 1.186, 1.193)
t=55	EL	(0.880, 1.027, 1.012, 0.857, 0.871)	(0.923, 0.896, 0.853, 1.030, 1.031)
	Norm.	(0.880, 1.014, 0.982, 0.867, 0.896)	(0.913, 0.883, 0.856, 1.034, 1.065)
	E.E.	(0.865, 0.997, 0.991, 0.867, 0.972)	(0.918, 0.888, 0.884, 1.034, 1.038)
	Keg.	(0.877, 0.979, 0.970, 0.896, 0.904)	(0.935, 0.864, 0.857, 1.082, 1.090)
t=60	EL	(0.879, 1.073, 1.071, 0.819, 0.820)	(0.929, 0.951, 0.907, 0.977, 0.978)
	Norm.	(0.873, 1.058, 1.031, 0.825, 0.846)	(0.923, 0.939, 0.915, 0.983, 1.008)
	E.E.	(0.877, 1.064, 1.061, 0.825, 0.827)	(0.929, 0.945, 0.945, 0.983, 0.983)
	Keg.	(0.893, 1.051, 1.044, 0.849, 0.855)	(0.942, 0.919, 0.917, 1.025, 1.026)
t=65	EL	(0.897, 1.137, 1.132, 0.789, 0.792)	(0.937, 0.992, 0.952, 0.945, 0.946)
	Norm.	(0.882, 1.110, 1.084, 0.794, 0.814)	(0.925, 0.977, 0.957, 0.946, 0.967)
	E.E.	(0.895, 1.116, 1.112, 0.794, 0.797)	(0.931, 0.983, 0.981, 0.946, 0.949)
	Keg.	(0.722, 1.098, 1.090, 0.816, 0.822)	(0.942, 0.958, 0.955, 0.984, 0.986)

4.3 Simulation study

In this section, we conduct simulation studies to investigate and compare the performance of EL, normal approximation (Norm.), Edgeworth expansion (E.E.) approximation and Kegler's method (Keg.) in the construction of CIs for $\lambda\mu$. The CIs based on EL, normal approximation, Edgeworth expansion and Kegler's methods are given by (4.7), (4.1), (4.3) and (4.4) respectively.

We use the five criteria (coverage probability, average length of CIs and average length conditional on coverage) proposed in Section 2.3 and coverage probability per each average/conditional length (larger the value, better the interval and the method), to assess the performance of each method.

We generate L sets of Poisson number $N^{(j)}(t)$ $(j = 1, \ldots, L\ )$ from a Poisson process with parameter λ, accompanied by a sample $\{X_1^{(j)}, \ldots, X_{N(t)}^{(j)}\}$ from distribution F. For each set, one calculate $(1-\alpha)$ level CIs CI_j, $j = 1, \ldots, L$, using normal approximation, EL method, Kegler's formula and Edgeworth expansion approximation.

Denote the length of CI_j by $|CI_j|$. The Monte Carlo approximation to the coverage probability (cov.), average length (alen.), average length conditional on coverage (clen.), coverage probability per each average length (avecov.) and cover-

Table 4.2: $F = N(1,1)$ and $\lambda_0 = 10$

Level		0.9	0.95
Size		(cov., avecov., concov., alen., clen.)	(cov., avecov., concov., alen., clen.)
t=15	EL	(0.903, 0.342, 0.341, 2.634, 2.648)	(0.950, 0.304, 0.301, 3.125, 3.156)
	Norm.	(0.907, 0.337, 0.336, 2.679, 2.690)	(0.952, 0.298, 0.297, 3.192, 3.202)
	E.E.	(0.904, 0.337, 0.338, 2.679, 2.672)	(0.949, 0.297, 0.297, 3.192, 3.194)
	Keg.	(0.902, 0.237, 0.236, 3.812, 3.816)	(0.957, 0.240, 0.210, 4.554, 4.558)
t=25	EL	(0.895, 0.435, 0.427, 2.056, 2.062)	(0.951, 0.389, 0.388, 2.445, 2.451)
	Norm.	(0.891, 0.430, 0.428, 2.077, 2.085)	(0.947, 0.383, 0.382, 2.476, 2.483)
	E.E.	(0.892, 0.429, 0.429, 2.077, 2.078)	(0.949, 0.383, 0.383, 2.476, 2.477)
	Keg.	(0.902, 0.305, 0.305, 2.949, 2.951)	(0.946, 0.269, 0.269, 3.520, 3.519
t=35	EL	(0.890, 0.505, 0.504, 1.762, 1.766)	(0.942, 0.452, 0.451, 2.084, 2.089)
	Norm.	(0.889, 0.506, 0.505, 1.758, 1.762)	(0.939, 0.448, 0.447, 2.095, 2.101)
	E.E.	(0.890, 0.506, 0.506, 1.758, 1.759)	(0.942, 0.449, 0.449, 2.095, 2.096)
	Keg.	(0.896, 0.360, 0.360, 2.487, 2.489)	(0.946, 0.319, 0.319, 2.966, 2.966)
t=45	EL	(0.904, 0.585, 0.584, 1.545, 1.548)	(0.951, 0.517, 0.516, 1.839, 1.843)
	Norm.	(0.905, 0.585, 0.583, 1.546, 1.549)	(0.953, 0.516, 0.515, 1.842, 1.847)
	E.E.	(0.904, 0.585, 0.585, 1.546, 1.546)	(0.951, 0.516, 0.516, 1.842, 1.844)
	Keg.	(0.902, 0.411, 0.411, 2.193, 2.194)	(0.955, 0.365, 0.365, 2.616, 2.616)
t=55	EL	(0.900, 0.646, 0.645, 1.396, 1.397)	(0.951, 0.575, 0.574, 1.656, 1.659)
	Norm.	(0.905, 0.643, 0.642, 1.400, 1.401)	(0.953, 0.571, 0.571, 1.668, 1.669)
	E.E.	(0.902, 0.646, 0.646, 1.400, 1.400)	(0.952, 0.571, 0.570, 1.668, 1.667)
	Keg.	(0.895, 0.450, 0.450, 1.987, 1.987)	(0.949, 0.401, 0.401, 2.369, 2.369)

Table 4.3: $F = U(0,1)$ and $\lambda_0 = 15$

Level		0.9	0.95
Size		(cov., avecov., concov., alen., clen.)	(cov., avecov., concov., alen., clen.)
t=35	EL	(0.899, 0.477, 0.475, 1.885, 1.888)	(0.950, 0.423, 0.422, 2.246, 2.250)
	Norm.	(0.902, 0.475, 0.473, 1.900, 1.903)	(0.952, 0.420, 0.420, 2.264, 2.267)
	E.E.	(0.901, 0.474, 0.474, 1.900, 1.900)	(0.951, 0.420, 0.420, 2.264, 2.264)
	Keg.	(0.904, 0.474, 0.474, 1.905, 1.905)	(0.954, 0.420, 0.420, 2.272, 2.272)
t=45	EL	(0.899, 0.617, 0.616, 1.457, 1.460)	(0.950, 0.546, 0.545, 1.737, 1.740)
	Norm.	(0.903, 0.613, 0.612, 1.469, 1.472)	(0.949, 0.542, 0.541, 1.751, 1.754)
	E.E.	(0.902, 0.615, 0.614, 1.469, 1.471)	(0.950, 0.542, 0.542, 1.751, 1.753)
	Keg.	(0.904, 0.614, 0.613, 1.472, 1.474)	(0.951, 0.542, 0.541, 1.755, 1.756)
t=55	EL	(0.900, 0.731, 0.731, 1.232, 1.232)	(0.952, 0.649, 0.649, 1.469, 1.470)
	Norm.	(0.895, 0.723, 0.722, 1.243, 1.244)	(0.956, 0.645, 0.645, 1.481, 1.482)
	E.E.	(0.898, 0.722, 0.723, 1.243, 1.243)	(0.953, 0.644, 0.642, 1.481, 1.481)
	Keg.	(0.901, 0.723, 0.723, 1.245, 1.244)	(0.952, 0.642, 0.642, 1.484, 1.483)
t=60	EL	(0.899, 0.826, 0.826, 1.085, 1.086)	(0.955, 0.738, 0.737, 1.294, 1.295)
	Norm.	(0.903, 0.822, 0.822, 1.096, 1.097)	(0.958, 0.734, 0.733, 1.306, 1.306)
	E.E.	(0.901, 0.823, 0.823, 1.096, 1.096)	(0.955, 0.731, 0.730, 1.306, 1.306)
	Keg.	(0.902, 0.822, 0.822, 1.097, 1.097)	(0.954, 0.730, 0.730, 1.308, 1.308)
t=65	EL	(0.901, 0.911, 0.911, 0.981, 0.982)	(0.949, 0.808, 0.808, 1.170, 1.171)
	Norm.	(0.894, 0.905, 0.904, 0.991, 0.992)	(0.947, 0.802, 0.801, 1.181, 1.182)
	E.E.	(0.897, 0.907, 0.907, 0.991, 0.992)	(0.947, 0.801, 0.801, 1.181, 1.182)
	Keg.	(0.901, 0.907, 0.906, 0.992, 0.993)	(0.948, 0.801, 0.801, 1.183, 1.183)

age probability per each conditional length (concov.) are respectively given by

$$\text{(i) } L^{-1}\sum_{j=1}^{L} I_{\{\theta_0\in CI_j\}}, \quad \text{(ii) } L^{-1}\sum_{j=1}^{L} |CI_j|, \quad \text{(iii) } L_0^{-1}\sum_{j=1}^{L} |CI_j| I_{\{\theta_0\in CI_j\}},$$

$$\text{(IV) } \frac{\sum_{j=1}^{L} I_{\{\theta_0\in CI_j\}}}{\sum_{j=1}^{L} |CI_j|} = \frac{\text{(i)}}{\text{(ii)}}, \quad \text{(V) } \frac{L^{-1}\sum_{j=1}^{L} I_{\{\theta_0\in CI_j\}}}{L_0^{-1}\sum_{j=1}^{L} |CI_j| I_{\{\theta_0\in CI_j\}}} = \frac{\text{(i)}}{\text{(iii)}}.$$

where $L_0 = \sum_{j=1}^{L} I_{\{\theta_0\in CI_j\}}$, the total number of CIs covering θ_0.

In our simulations, various values of λ, the nominal level α and time variable t were chosen and each experiment was based on $L = 5000$ trials, generated by routines in R. We only present four different cases here.

Firstly, we consider an exponential distribution $F = Exp(1/2)$, and the simulation results are shown in Table 4.1.

Secondly, we choose $F = N(1, 1)$, to see what will happen if the population is a normal one, and the results are given in Table 4.2.

Thirdly, we want to check the performance of the methods if the population is uniform $F = U(0, 1)$, and Table 4.3 contains the simulation results for this case.

Finally, since there are no continuity conditions imposed on the random variables, we choose a discrete population $F = Binomial(20, 0.05)$ and the results are presented in Table 4.4.

The following observation can be made from those tables.

(1) As the time t increases, all the methods, improve in terms of all five criteria (i.e., coverage probability, average length, conditional average

Table 4.4: $F = Binomial(20, 0.05)$ and $\lambda_0 = 20$

Level		0.9	0.95
Size		(cov., avecov., concov., alen., clen.)	(cov., avecov., concov., alen., clen.)
t=35	EL	(0.898, 0.172, 0.172, 5.215, 5.223)	(0.950, 0.152, 0.152, 6.243, 6.252)
	Norm.	(0.903, 0.170, 0.170, 5.290, 5.307)	(0.945, 0.150, 0.149, 6.303, 6.322)
	Keg.	(0.904, 0.170, 0.170, 5.305, 5.308)	(0.951, 0.150, 0.150, 6.329, 6.335)
t=45	EL	(0.899, 0.222, 0.222, 4.011, 4.016)	(0.949, 0.196, 0.196, 4.806, 4.813)
	Norm.	(0.898, 0.219, 0.219, 4.101, 4.104)	(0.945, 0.193, 0.193, 4.887, 4.897)
	Keg.	(0.899, 0.219, 0.219, 4.108, 4.111)	(0.949, 0.194, 0.193, 4.899, 4.903)
t=55	EL	(0.899, 0.267, 0.266, 3.374, 3.377)	(0.949, 0.232, 0.232, 4.044, 4.047)
	Norm.	(0.894, 0.258, 0.257, 3.467, 3.468)	(0.943, 0.228, 0.228, 4.131, 4.136)
	Keg.	(0.899, 0.529, 0.259, 3.472, 3.472)	(0.943, 0.228, 0.228, 4.139, 4.141)
t=60	EL	(0.900, 0.300, 0.300, 2.963, 2.963)	(0.947, 0.266, 0.266, 3.555, 3.556)
	Norm.	(0.901, 0.295, 0.294, 3.059, 3.058)	(0.945, 0.259, 0.259, 3.645, 3.647)
	Keg.	(0.902, 0.294, 0.294, 3.062, 3.061)	(0.946, 0.259, 0.259, 3.650, 3.650)
t=65	EL	(0.902, 0.338, 0.337, 2.674, 2.676)	(0.952, 0.297, 0.297, 3.207, 3.208)
	Norm.	(0.914, 0.330, 0.330, 2.770, 2.773)	(0.959, 0.290, 0.290, 3.301, 3.304)
	Keg.	(0.914, 0.330, 0.330, 2.773, 2.773)	(0.960, 0.290, 0.290, 3.305, 3.306)

length and coverage probability per each average/conditional length), and the difference among them gradually disappear.

(2) In terms of coverage probability, all the methods are very competitive, but the EL method always gives the best coverage probabilities.

(3) In terms of conditional average length on coverage, the EL method seems to be the best, followed by the normal approximation and Edgeworth expansion, and then by Kegler's method.

(4) In terms of coverage probabilities per each average length, the EL method always produces the best results,, followed by Edgeworth expansion and normal approximation. Kegler's method seems to be the worst especially when considering a normal population.

(5) In Table 4.2, all the methods are often anti-conservative, and Kegler's method is conservative in Table 4.3. The EL method is always more anti-conservative than the other methods since it has shorter length of CIs.

In summary, in terms of coverage probabilities, length of CIs and coverage probability per each average/conditional length, the EL method is always the best among these four approaches.

4.4 Application to coal-mining disasters data

In this section, we apply our proposed method to a real example about coal-mining disasters. The data set we will refer to is contained in Andrew and Herzberg (1985).

Coal-mining disaster is an accident that occurs in the process of mining coal. Thousands of miners die from mining accidents each year, most occur in developing countries and rural parts of developed countries.

Mining accidents might be due to various causes, including leaks of poisonous or explosive natural gases, dust explosions, collapsing of mine stopes, flooding and so on.

The data set we use here is about coal-mining disasters in Britain involving 10 or more men killed, caused by explosions of fire-damp or coal-dust, from 15 March 1851 to 22 March 1962 inclusive. We are interested in the average number of death in each year. Based on a year unit, if there are any disasters in some year, then as a whole, we assume that the number of disasters in that year is 1, and otherwise 0. We count the number of death as the summation of all death in each disaster during that year. Let t be the number of years starting from 15 March 1851, $X_j(t)$ (j=1,...,) be the number of death in the jth ($j = 1, \ldots, N(t)$) disaster with mean μ and $N(t)$ be the number of disasters up to the tth year. Apparently, $\{N(t), t > 0\}$ can naturally be assumed to be a Poisson process with parameter λ. Under the above assumptions, there are in all 79 disasters in record in the 112 years.

Table 4.5: CIs by EL, Normality, Edgeworth expansion and Kegler's method

Confidence level	0.9	0.95	0.99
EL	(68, 115)	(65, 121)	(59, 134)
Norm.	(65, 111)	(61, 115)	(52, 123)
E.E	(68, 114)	(65, 120)	(60, 131)
Keg.	(67, 114)	(64, 120)	(58, 132)

Therefore, on a yearly basis, the 90%, 95% and 99% CIs for $\lambda\mu$, average death in each year, via the EL method, normal approximations, Edgeworth expansion and Kegler's method, can be respectively obtained, and they are listed in Table 4.5.

That is, for instance, using the EL method, we are 90% confidential that the yearly number of death is between 68 and 115, from 15 March 1851 to 22 March 1962 inclusive.

4.5 Proof of Theorem 4.2.1

In this section, we will prove Theorem 4.2.1. Before proceeding to the proof, we provide some technical lemmas.

Our first lemma guarantees the existence and uniqueness of γ in equation (4.6).

Lemma 4.5.1 *Suppose that $EX_1^2 < \infty$ and $\sigma^2 > 0$. Then as $t \to \infty$, we have*

$$P\left(\min_{1\le i\le N(t)} X_i < \theta t/N(t) < \max_{1\le i\le N(t)} X_i\right) \to 1.$$

Proof. It suffices to show that $P(\min_{1\le i\le N(t)}(X_i - \theta t/N(t)) \ge 0) \to 0$ and $P(\max_{1\le i\le N(t)}(X_i - \theta t/N(t)) \le 0) \to 0$. We only prove the second one since the first one can be done similarly. When $\mu = 0$, the proof is trivial. Therefore, we only need to consider the case $\mu \ne 0$.

Noting that for small $\delta > 0$, we have

$$\begin{aligned}
& P\left(\max_{1\le i\le N(t)} (X_i - \theta t/N(t)) \le 0\right) \\
= \ & P\left(\max_{1\le i\le N(t)} (X_i - \mu) \le \theta t/N(t) - \mu \le 0\right) \\
\le \ & P\left(\max_{1\le i\le N(t)} (X_i - \mu) \le \delta/2\right) + P(\theta t/N(t) - \mu > \delta/2).
\end{aligned}$$

As $\lim_{t\to\infty}(t/N(t)) = 1/\lambda$ $a.s.$ leads to

$$P(|t/N(t) - 1/\lambda| > \epsilon) \to 0 \ \text{ as } t \to \infty$$

for any $\epsilon > 0$, it follows that

$$\begin{aligned}
0 \le P(\theta t/N(t) - \mu > \delta/2) &\le P(|\theta t/N(t) - \mu| > \delta/2) \\
&= P(|t/N(t) - 1/\lambda| > \delta/(2\lambda|\mu|)) \to 0
\end{aligned}$$

by taking $\epsilon = \delta/(2\lambda|\mu|)$. Therefore, we only need to prove that

$$P\left(\max_{1\le i\le N(t)} (X_i - \mu) \le \delta/2\right) \to 0.$$

To this end, letting $\xi_i = \psi(X_i - \mu)$, where $\psi(x)$ is a nondecreasing such that

$$\psi(x) = \begin{cases} 0, & \text{if } \ x \le \delta/2, \\ a(x), & \text{if } \ \delta/2 < x < \delta, \\ 1, & \text{if } \ x \ge \delta. \end{cases}$$

with $0 < a(x) < 1$ for $\delta/2 < x < \delta$. Then we have

$$\begin{aligned} P\left(\max_{1\le i\le N(t)} (X_i - \mu) \le \delta/2\right) &= P(X_1 - \mu \le \delta/2, ..., X_N(t) - \mu \le \delta/2) \\ &= P(\xi_1 = 0, ..., \xi_{N(t)} = 0) \\ &= P\left(\sum_{i=1}^{N(t)} (\xi_i - E\xi_i) = -N(t)E\xi_1\right) \\ &\le \frac{\mathrm{Var}(\xi_1)}{(E\xi_1)^2} \cdot \sum_{n=1}^{\infty} \frac{(\lambda t)^n e^{-\lambda t}}{n \cdot n!}. \end{aligned}$$

Noting that, on one hand for any positive r.v. X, the inequality

$$E\left(\frac{1}{X}\right) \ge \frac{1}{EX}$$

follows from Schwarz inequality. On the other hand, for any $x \ge 1$, the inequality

$$\frac{1}{x} \le \frac{1}{x+1} + \frac{3}{(x+1)(x+2)}$$

holds, and it follows that

$$E\left(\frac{1}{X}\right) \le E\left(\frac{1}{X+1}\right) + E\left(\frac{3}{(X+1)(X+2)}\right).$$

Therefore, we have

$$\sum_{n=1}^{\infty} \frac{(\lambda t)^n e^{-\lambda t}}{n \cdot n!} = O(t^{-1}), \quad \text{as} \ \ t \to \infty$$

since

$$\begin{aligned} O(t^{-1}) = \frac{1}{\lambda t} &\leq \sum_{n=1}^{\infty} \frac{(\lambda t)^n e^{-\lambda t}}{n \cdot n!} \\ &\leq \frac{1}{\lambda} + \frac{3}{(\lambda t)^2}(1 - e^{-\lambda t} - \lambda t e^{-\lambda t}) = O(t^{-1}). \end{aligned}$$

Now, it suffices to prove that

$$(1)\ \mathrm{Var}(\xi_1) \leq 1 \ \text{ and } \ (2)\ \lim_{n\to\infty} E\xi_1 \geq c > 0.$$

The first assertion is trivial since $\mathrm{Var}(\xi_i) \leq E\xi_1^2 \leq 1$.

For (2), since $EX_1 = \mu$ and $\sigma^2 > 0$, we easily get

$$\lim_{\delta'\downarrow 0} P(X_1 - \mu > \delta') = P(X_1 - \mu > 0) > 0$$

for some $\delta' \downarrow 0$. Therefore, there exists some δ'' such that for any $\varepsilon : 0 < \varepsilon < \delta''$,

$$P(X_1 - \mu > \varepsilon) > 0.$$

Let $\varepsilon = \delta/2$, it follows that

$$\begin{aligned} E\xi_1 &= Ea(X_1 - \mu)I_{\{X_1 - \mu > \delta/2\}} + (1 - Ea(X_1 - \mu))P(X_1 - \mu \geq \delta) \\ &\geq Ea(X_1 - \mu)I_{\{X_1 - \mu > \varepsilon\}} > 0. \end{aligned}$$

□

Remark 4.5.1 *From* Roy and Tiku (1962), *we have the first-order negative moment of a Poisson r.v. as*

$$\sum_{n=1}^{\infty} \frac{(\lambda t)^n e^{-\lambda t}}{n \cdot n!} \approx \frac{1}{(\lambda t - 1)(1 - e^{-\lambda t})} = O(t^{-1}) \qquad \textit{as} \quad t \to \infty$$

Now, noting that $EX_1^2 < \infty$ is equivalent to

$$\sum_{n=1}^{\infty} P(X_n^2 > \epsilon n) < \infty,$$

for any $\epsilon > 0$. It then follows from the Borel-Cantelli lemma that

$$P(|X_n| \leq \epsilon n^{1/2}) = 1,$$

which in turn implies that

$$\max_{1 \leq i \leq n} |X_i - \mu| = o(n^{1/2}) \qquad a.s. \tag{4.8}$$

Also note that

$$\lim_{t \to \infty} t/N(t) = 1/\lambda, \tag{4.9}$$

and

$$W_{N(t)} = \max_{1 \leq i \leq N(t)} |X_i - \theta t/N(t)| \leq \max_{1 \leq i \leq N(t)} |X_i - \mu| + |\mu - \theta t/N(t)|. \tag{4.10}$$

Combining (4.8), (4.9) and (4.10), we have

$$W_{N(t)} = o(t^{1/2}) \qquad a.s.$$

Lemma 4.5.2 *Let* $S^2 = \sum_{i=1}^{N(t)} (X_i - \theta t/N(t))^2/N(t)$. *If* $EX_1^2 < \infty$, *then* $S^2 = \sigma^2 + o(1)$ *almost surely.*

Proof. Note that

$$\begin{aligned} S^2 &= \frac{1}{N(t)} \sum_{i=1}^{N(t)} (X_i - \theta t/N(t))^2 \\ &= \frac{1}{N(t)} \sum_{i=1}^{N(t)} (X_i - \bar{X}_{N(t)})^2 + (\bar{X}_{N(t)} - \theta t/N(t))^2, \end{aligned}$$

now the result follows easily from the strong law of large numbers and the fact that

$$
\begin{aligned}
&(\bar{X}_{N(t)} - \theta t/N(t))^2 \\
&= (\bar{X}_{N(t)} - \mu)^2 + 2(\mu - \theta t/N(t))(\bar{X}_{N(t)} - \mu) + (\mu - \theta t/N(t))^2 \\
&= o(1),
\end{aligned}
$$

since $\lim_{t\to\infty} t/N(t) = 1/\lambda$. □

From the above estimates, with probability 1

$$
\begin{aligned}
\frac{1}{N(t)} \sum_{i=1}^{N(t)} |X_i - \theta t/N(t)|^3 &\leq W_{N(t)} \times S^2 \\
&= o(t^{1/2}) \times [\sigma^2 + o(1)] \\
&= o(t^{1/2}).
\end{aligned}
$$

Proof of Theorem 4.2.1. From Lemma 4.5.1, it follows that with probability tending to 1, the true value $\theta_0 t/N(t)$ satisfies

$$\min_{1\leq i\leq N(t)} X_i \leq \theta_0 t/N(t) \leq \max_{1\leq i\leq N(t)} X_i.$$

When this is true, the solution to equation (4.6) exists and is unique. We now show that the root of (4.6) is $\gamma = O_p(t^{-1/2})$. To this end, noting that

$$\begin{aligned}
0 = |f(\gamma)| &= \frac{1}{N(t)}\left|\sum_{i=1}^{N(t)} \frac{X_i - \theta_0 t/N(t)}{1+\gamma(X_i - \theta_0 t/N(t))}\right| \\
&= \frac{1}{N(t)}\left|\sum_{i=1}^{N(t)} \left[(X_i - \theta_0 t/N(t)) - \frac{\gamma(X_i - \theta_0 t/N(t))^2}{1+\gamma(X_i - \theta_0 t/N(t))}\right]\right| \\
&= \frac{1}{N(t)}\left|\sum_{i=1}^{N(t)}(X_i - \theta_0 t/N(t)) - \gamma\sum_{i=1}^{N(t)} \frac{(X_i - \theta_0 t/N(t))^2}{1+\gamma(X_i - \theta_0 t/N(t))}\right| \\
&\geq \frac{|\gamma|}{N(t)}\sum_{i=1}^{N(t)} \frac{(X_i - \theta_0 t/N(t))^2}{1+\gamma(X_i - \theta_0 t/N(t))} - \frac{1}{N(t)}\left|\sum_{i=1}^{N(t)}(X_i - \theta_0 t/N(t))\right| \\
&\geq \frac{|\gamma| S^2}{1+|\gamma| W_{N(t)}} - \frac{1}{N(t)}\left|\sum_{i=1}^{N(t)}(X_i - \theta_0 t/N(t))\right|.
\end{aligned}$$

Following from Corollary 2.8 in von Chossy and Rappl (1983), the second term is $O_p(t^{-1/2})$. Recalling Lemma 4.5.2, $S^2 = \sigma^2 + o(1)$ $a.s.$, it follows that

$$\frac{|\gamma|}{1+|\gamma| W_{N(t)}} = O_p(t^{-1/2}),$$

hence we have

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

For simplicity, write $\beta_i = \gamma(X_i - \theta_0 t/N(t))$ where γ is the root of equation (4.6). Then,

$$\max_{1\leq i\leq N(t)} |\beta_i| = O_p(t^{-1/2})o(t^{1/2}) = o_p(1).$$

Expanding (4.6), we have

$$\begin{aligned} 0 = f(\gamma) &= \frac{1}{N(t)} \sum_{i=1}^{N(t)} (X_i - \theta_0 t/N(t))(1 - \beta_i + \beta_i^2/(1+\beta_i)) \\ &= \frac{1}{N(t)} \sum_{i=1}^{N(t)} X_i - \theta_0 t - S^2\gamma + \frac{1}{N(t)} \sum_{i=1}^{N(t)} \frac{\beta_i^2 (X_i - \theta_0 t/N(t))}{1+\beta_i}, \end{aligned}$$

where the final term is bounded by

$$\frac{1}{N(t)} \sum_{i=1}^{N(t)} \frac{|X_i - \theta_0 t/N(t)|^3 \gamma^2}{|1+\beta_i|} = o(t^{1/2}) O_p(t^{-1}) O_p(1) = o_p(n^{-1/2}).$$

Therefore, we may write

$$\gamma = \frac{\bar{X}_{N(t)} - \theta_0 t/N(t)}{S^2} + \tau, \quad \text{where } |\tau| = o_p(t^{-1/2}).$$

Further, we have the expansion

$$\log(1+\beta_i) = \beta_i - \beta_i^2/2 + \zeta_i$$

where for some finite $C > 0$,

$$P(|\zeta_i| \le C|\beta_i|^3, 1 \le i \le N(t)) \to 1$$

as $t \to \infty$.

Now substituting γ, we have

$$
\begin{aligned}
&-2\omega_{N(t)} \log \Re(\theta_0) \\
&= -2\omega_{N(t)} \sum_{i=1}^{N(t)} \log(N(t)p_i) = 2\omega_{N(t)} \sum_{i=1}^{N(t)} \log(1+\beta_i) \\
&= \omega_{N(t)} \left(2\sum_{i=1}^{N(t)} \beta_i - \sum_{i=1}^{N(t)} \beta_i^2 + 2\sum_{i=1}^{N(t)} \zeta_i \right) \\
&= \omega_{N(t)} \left(\frac{N(t)(\bar{X}_{N(t)} - \theta_0 t/N(t))^2}{S^2} - N(t)S^2\tau^2 + 2\sum_{i=1}^{N(t)} \zeta_i \right) \\
&= \frac{\sum_{i=1}^{N(t)}(X_i - \bar{X}_{N(t)})^2}{\sum_{i=1}^{N(t)}(X_i - \theta_0 t/N(t))^2} \cdot \left(\frac{\sum_{i=1}^{N(t)} X_i - \theta_0 t}{\sqrt{\sum_{i=1}^{N(t)} X_i^2}} \right)^2 \\
&\quad - N(t)\omega_{N(t)} S^2\tau^2 + 2\omega_{N(t)} \sum_{i=1}^{N(t)} \zeta_i.
\end{aligned}
$$

For the first term, on one hand

$$
\frac{\sum_{i=1}^{N(t)}(X_i - \bar{X}_{N(t)})^2}{\sum_{i=1}^{N(t)}(X_i - \theta_0 t/N(t))^2} \to 1,
$$

since both $\sum_{i=1}^{N(t)}(X_i - \bar{X}_{N(t)})^2$ and $\sum_{i=1}^{N(t)}(X_i - \theta_0 t/N(t))^2$ tend to σ^2. On the other hand, we have

$$
\left(\frac{\sum_{i=1}^{N(t)} X_i - \theta_0 t}{\sqrt{\sum_{i=1}^{N(t)} X_i^2}} \right)^2 \to_d \chi_1^2
$$

by the central limit theorem for Studentized CPP.

The second term is bounded by

$$
|-N(t)\omega_{N(t)} S^2\tau^2| = O_p(t) \left(\sigma^2/EX_1^2 + o(1)\right) (\sigma^2 + o(1)) o_p(t^{-1}) = o_p(1).
$$

For the final term

$$\begin{aligned}\left|\omega_{N(t)}\sum_{i=1}^{N(t)}\zeta_i\right| &\le O_p(t)\left(\sigma^2/EX_1^2+o(1)\right)\sum_{i=1}^{N(t)}|\zeta_i| \\ &\le C|\gamma|^3\sum_{i=1}^{N(t)}|X_i-\theta_0 t/N(t)|^3 = O_p(t^{-3/2})o(t^{3/2}) = o_p(1).\end{aligned} \tag{4.11}$$

Combining the above estimations, from Slutsky's theorem, we have

$$-2\omega_{N(t)}\log\Re(\theta_0)\rightarrow_d \chi_1^2,$$

which completes the proof. □

Chapter 5

Conclusions and Further Research

5.1 Conclusions

The purpose of this thesis is to make statistical inference for VUS $P(X < Y < Z)$, HUM $P(X_1 < X_2 \cdots < X_k)$ and unit mean $\lambda\mu$ of compound Poisson sum by constructing confidence intervals. We used three-sample and multi-sample U-statistics as unbiased estimators of VUS and HUM, respectively, and calculate their corresponding jackknifed variances. The normal approximation was based on the studentized three-sample and multi-sample U-statistics with the jackknife estimator of variance. There might be one potential technical difficulty with the estimation of HUM. When there are k categories, usually we have to determine the most sensible HUM values by choosing the one with the largest numeric value among all $k!$ possible orders of categories. This selection process could be potentially time-consuming. However, in most medical problems that we come across with, the number of categories are usually less than ten. Sometimes it may be also advisable to combine certain similar categories when the samples are not large enough to provide valid statistical inferences.

By employing the JEL method to three-sample and multi-sample U-statistics, we first explain, and thereafter theoretically prove that the jackknife empirical log-likelihood ratio converges to χ_1^2. JEL is again proved to be very efficient in dealing with nonlinear statistics, eg. U-statistics in this thesis, as it largely relieves computation burden that one will surely encounter in the usual empirical likelihood procedure. For compound Poisson sum, it seems hard to connect with Owen's empirical likelihood at first sight. However, by making use of properties of conditional expectations, we can do an approximation. That is, we assume that $E(\sum_{j=1}^{N(t)} X_j | N(t) = n) \approx \lambda\mu t$ and consider the EL

$$L(\theta|N(t) = n) = \max\left\{\prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1, p_i \geq 0, \sum_{i=1}^{n} p_i X_i = \lambda\mu t/n\right\}.$$

By utilizing Owen's EL to the mean functional $\sum_{i=1}^{n} p_i X_i$, we derive asymptotic distribution for the *adjusted* empirical log-likelihood ratio, which is also χ_1^2, and construct construct confidence intervals for $\lambda\mu$. Although the validity of this assumption is arguable, at least it enables us to apply Owen's empirical likelihood, which is easy to implement and provides more precise statistical results than some other methods, to compound Poisson sum and obtain some beautiful results. The simulation outcomes confirm that the performance of our proposed method is much better than some existing methods in terms of some statistical criteria we are interested in, and thus it can be relied on to make statistical inference.

5.2 Further Research

With respect to the rapid development and fruitful results of empirical likelihood recently, there is still much work to do based on our current work.

1. In this thesis, although we provide easy and effective tools to make statistical inference for VUS and HUM, we do not touch much of the detailed three and multi-class classification problems, which might be more useful in applications.

2. To obtain those pseudo-values, we remove the i-th data X_i from a large sample Z containing all sample points. It will be interesting to consider deleting k data values, each $X_{j,i}$ from the j-th sample $\mathbf{X_j}$ at a time, where $j = 1, \ldots, k$ and $i = 1, \ldots, n_j$ for a k-sample U-statistic as defined in Chapter 3.

3. In this study, all the research work is done for one dimensional population. In view of the paper of Chen et al. (2009), the empirical likelihood can work for large dimensional data when $p = o(n^{1/2})$ where n is the sample size and p is dimension of the data. We conjecture

that the empirical likelihood method for high dimensional data is still applicable when p is just proportional to n.

4. For parameters defined by estimating equations, Qin and Lawless (1994) derived asymptotic results under smooth conditions. When the estimating functions are replaced by U-type statistics with smooth kernels, the JEL method can be used to extend the work of Qin and Lawless (1994) with only slight modifications. However, if the kernels of the U-statistics are not necessarily smooth with respect to the parameters of interest or auxiliary parameters, will JEL still work?

Bibliography

[1] Anderson, E. (1935). The Irises of the Gaspé Peninsula. *Bulletin of the American Iris Society* **59,** 2-5.

[2] Andrew D.F. and Herzberg A.M. (1985). *Data: A Collection of Problems from Many Fields for the student and Research Worker*. Springer, New York.

[3] Arvesen, J. N. (1969). Jackknifing U-statistics. *Annals of Mathematical Statistics* **40,** 2076-2100.

[4] Babu, G.J., Singh, K. and Yang, Y.N. (2003). Edgeworth expansions for compound Poisson processes and the bootstrap. *Ann. Inst. Statist. Math*, **55**, 83-94.

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

[6] Blum, J. R., Hanson, D. L. and Rosenblatt, J. I. (1963). On the central limit theorem for the sum of a random number of independent random variables. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, **1**, 389-393.

[7] Cao, R. and Van Keilegom, I. (2006). Empirical likelihood for two-sample problems via nonparametric density eatimation. *The Canadian Journal of Statistics* **34**, 61-77.

[8] Chan, N. H. and Ling, S. (2006). Empirical likelihood for GRACH models. *Econometric Theory* **22**, 403-428.

[9] Chandra, S. and Owen, D. B. (1975). On estimating the reliability of a component subject to several different stresses (strengths). *Naval Research Logistics Quarterly* **22,** 31-39.

[10] Chen, J. and Qin, J. (1993). Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika* **80**, 107-116.

[11] Chen, J.H. and Sitter, R.R. (1999). A pseudo empirical likelihood approach to the effective use of auxiliary information in complex surveys. *Statistica Sinica* **9**, 385-406.

[12] Chen, J., Peng, L. and Zhao, Y. (2009). Empirical likelihood based confidence intervals for copulas. *Journal of Multivariata Analysis* **100**, 137-151.

[13] Chen, S. X. and Hall, P. (1993). Smoothed empirical likelihood confidence intervals for quantiles. *Annals of Statistics* **21**, 1166-1181.

[14] Chen, S. and Gao, J. (2007). An adaptive empirical likelihood test for parametric time series regression models. *Journal of Econometrics* **141**, 950-972.

[15] Chen, S., Peng, L. and Qin, Y. (2009). Empirical likelihood methods for high dimension. *Biometrika* **96**, 711-722.

[16] Diciccio, T. S., Hall, P. and Romono, J. (1991). Empirical likelihood is Bartelett correctable. *Annals of Statistics* **19**, 1053-1061.

[17] Dreiseitl, S., Ohno-Machado, L. and Binder, M. (2000). Comparing three-class diagnostic tests by three-way ROC analysis. *Medical Decision Making* **20,** 323-331.

[18] Dutta, K. and Srivastava, G. L. (1987). An n-standby system with $P(X < Y < Z)$. *IAPQR Transactions* **12,** 95-97.

[19] Gnedenko, B.V. and Korolev, V.Y. (1996). *Random Summation.* CRC Press.

[20] Guggenberger, P. and Smith, R. J. (2008). Generalized empirical likelihood tests in time series models with potential identification failure. *Journal of Econometrics* **142**, 134-161.

[21] Hall, P. and LaScala, B. (1990). Methodology and Algorithms of empirical likelihood. *International Statistical Review* **58,** 109-127.

[22] Hanley, J. A. (1989). Receiver operating characteristic (ROC) methodology: the state of the art. Critical review. *Diagnostic imaging* **29,** 307-335.

[23] Helmers, R. and Tarigan, B. (2004). Compound sums and their applications in finance. *Centrum voor Wiskunde en Informatica.* Available at: *http://db.cwi.nl/rapporten/abstract.php?abstractnr=1336*

[24] Hlawka, P. (1975). Estimation of the parameter $p = P(X < Y < Z)$. *Prace Nauk. Inst. Mat. Politechn. Wroclaw. No. 11, Ser. Stu.* **i** *Materialy No. 10 Problemy rachunku prawdopodobienstwa.,* 55-65 (in Polish).

[25] Halmos, P. R. (1946). The theory of unbiased estimation. Ann. Math. Stat., **17**, 34-43.

[26] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics* **19,** 293-325.

[27] Ivshin, V. V. (1998). On the estimation of the probabilities of a double linear inequality in the case of uniform and two-parameter exponential distributions. *Journal of Mathematical Sciences* **88,** 819-827.

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

[29] Jing, B. Y., Wang, Q. Y., Wang, X. P. and Zhou, W. (2009). Saddlepoint approximation for studentized compound Poisson sum. (submitted)

[30] Johnson, R. A. (1988). Stress-strength models for reliability. In *Handbooks of Statistics*, vol. 7, Krishnaiah PR, Rao CR (eds). Elsevier/North-Holland: New York/Amsterdan, 27-54.

[31] Karlin, S. and Taylor, H. (1981). *A Second Course in Stochastic Processes.* Academic Press, California.

[32] Kegler, S.R. (2007). Applying the compound Poisson process model to the reporting of injury-related mortality rates. Available at: *http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1828152*

[33] Keziou, A. and Leoni-Aubin, S. (2008). On empirical likelihood for semiparametric two-sample density ratio models. *Journal of Statistical Planning and Inference* **104**, 1224-1232.

[34] Koepsell T.D., Chi Y.Y., Zhou X.H., Lee W.W., Ramos E.M., and Kull W.A. (2007). An alternative method for estimating efficacy of the AN1792 Vaccine for Alzheimer's Disease. Neurology. **68,** 1868-1872.

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

[36] Kotz, S., Lumelskii, Y. and Pensky, M. (2003). *The Stress-strength Model and its Generalizations: Theory and Applications*. World Scientific, Singapore.

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

[38] Li, J. and Fine, J. P. (2008). ROC analysis for multiple classes and multiple categories and its application in microarray study. *Biostatistics* **9**, 566-576.

[39] Lu, X. and Qi, Y. (2004). Empirical likelihood for the additive risk model. *Probability and Mathematical Statistics* **24**, 419-431.

[40] Miwa, T., Hayter, A. J. and Liu, W. (2000). Calculations of level probabilities for normal random variables with unequal variances with applications to Bartholomew's test in unbalanced one-way models. *Computational Statistics and Data Analysis* **34,** 17-32.

[41] Mossman,D.(1999). Three-way ROCs. *Medical Decision Making* **19**, 78-89.

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

[43] Nakas, C. T. and Alonzo, T. A. (2007). ROC graphs for assessing the ability of a diagnostic marker to detect three disease classes with an umbrella ordering. *Biometrics* **63,** 603-609.

[44] Newcombe R. G. (2006). Confidence intervals for an effect size measure based on the Mann-Whitney statistic. Part 1: generals issues and tail-area-based methods. *Statistics in medicine,* **25** 543-557.

[45] Newcombe R. G. (2006). Confidence intervals for an effect size measure based on the Mann-Whitney statistic. Part 2: asymptotic methods and evaluation. *Statistics in medicine,* **25** 559-573.

[46] Nordman, D. J. and Lahiri, S. N. (2006) A frequency domain empirical likelihood for short- and long-range dependence. *Annals of Statistics* **34**, 3019-3050.

[47] Nordman, D. J., Sibbertsen, P. and Lahiri, S. N. (2007). Empirical likelihood confidence intervals for the mean of a long-range dependent process. *Journal of Time Series analysis* **28**, 576-599.

[48] Obuchowshi, N. A., Applegate, K. E., Goske, M. J., Arheart, K. L., Myers, M. T. and Morrison, S. (2001).. The differential diagnosis for multiple diseases: comparison with the binary-truth state experiment in two empirical studies. *Academic Radiology* **8,** 947-954.

[49] Otsu, T. (2006). Generalized empirical likelihood inference for nonlinear and time series models under weak identification. *Econometric Theory* **22**, 513-527.

[50] Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237-249.

[51] Owen, A. (1990). Empirical likelihood ratio confidence regions. *Annals of Statistics* **18**, 90-120.

[52] Owen, A. B. (1991). Empirical likelihood for linear models. *Annals of Statistics* **19**, 1725-1747.

[53] Owen, A. B. (2001). *Empirical Likelihood.* Chapman and Hall, London.

[54] Parzen, E. (1967). *Stochastic Processes*. 3rd edition: Holden-Day, San Francisco, USA.

[55] Pepe, M. S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction.* Oxford University Press, New York.

[56] Pessler F, Ogdie A, Diaz-Torne C, Dai L, Yu X, Einhorn E, et al. (2008a). Subintimal Ki-67 as a synovial tissue biomarker for inflammatory arthropathies. *Ann Rheum Dis* **67**(2), 162-7.

[57] Pessler F, Dai L, Diaz-Torne C, Ogdie A, Gomez-Vaquero C, Paessler ME, et al. (2008b). Increased angiogenesis and cellular proliferation as hallmarks of the synovium in chronic septic arthritis. *Arthritis Rheum* **59**(8), 1137-46.

[58] Qin, J. (1993). Empirical likelihood in biased sample problems. *Annals of Statistics* **21**, 1182-1196.

[59] Qin, J. and Lawless, J. F. (1994). Empirical likelihood and general estimating equations. *Annals of Statistics* **22**, 300-325.

[60] Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika* **43**, 353-360.

[61] Reaven, G. M. and Miller, R. G. (1979). An attempt to define the nature of chemical diabetes using a multimensional analysis. *Diabetologia* **16**, 17-24.

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

[63] Rényi, A. (1957). On the asymptotic distribution of the sum of a random number of independent random variables. *Acta Math. Acad. Sci. Hung.* **8**, 193-199.

[64] Roy, J. and Tiku, M. L. (1962). A Laguerre Series approximation to the sampling distribution of variance. *Sankhyā* **24**, 181-184.

[65] Sampat M.P., Patel A.C., Wang Y., Gupta S., Kan C.W., Bovik A.C., and Markey M.K. (2009). Indexes for three-class classification performance assessment-an empirical comparison. *IEEE transcation on information technology in biomedicine* **13,** 300-312.

[66] Scurfield, D.K (1996). Multiple-event forced-choice tasks in the theory of signal detectability. *Journal of Mathematical Psychology* **40,** 253-269.

[67] Sen, P.K. (1960). On some convergence properties of U-statistics. *Calcutte Stat. Assoc. Bull.* **10**, 1-18.

[68] Shao, J. and Tu, D. (1995). The Jackknife and Bootstrap. Springer-Verlag.

[69] Shi, X. (1984). The approximate independence of jackknife pseudo-values and the bootstrap methods. *Journal of the Wuhan Institute of Hydraulic and Electric Engineering* **2,** 83-90.

[70] Singh, N. (1980). On the estimation of $Pr(X_1 < Y < X_2)$. *Communications in Statistics. A. Theory and Methods* **9,** 1551-1561.

[71] Swets, J.A. and Pichett, R. M. (1982). *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory.* Academic Press, New York.

[72] Thomas, D. R. and Grunkemeirt, G. L. (1975). Confidence interval estimation of survival probabilities for censored data. Journal of the American Statistical Association **70**, 865-871.

[73] Turkey, J. W. (1958). Bias and confidence in not-quite large samples. *Annals of Mathematical Statistics* **29,** 614.

[74] von Chossy, R. and Rappl, G. (1983). Some approximation methods for the distribution of random sums. *Insurance: Mathematics and Economics*, **2,** 251-270.

[75] Wang, Q. H. and Jing, B. Y. (1999). Empirical likelihood for partial linear model with fixed designs. *Statistics & Probability Letters* **41**, 425-433.

[76] Wang, Q. H. and Jing, B. Y. (2003). Empirical likelihood for partial linear model. *Annals of the Institute of Statistical Mathematics* **55**, 585-595.

[77] Xue, L. and Zhu, L. (2007). Empirical likelihood for a varying coefficient model with longitudinal data. *Journal of American Statistical Association* **102,** 642-654.

[78] You, J., Chen, G. and Zhou, Y. (2006). Block empirical likelihood for longitudinal partially linear regression models. *The Canadian Journal of Statistics* **34,** 79-96.

[79] Zhang, B. (1997a). Quantile processes in the presence of auxiliary information. *Annals of the Institute of Statistical Mathematics* **49**, 35-55.

[80] Zhang, B. (1997b). Empirical likelihood confidence intervals for M-functionals in the presence of auxiliary information. *Statistics & Probability Letters* **32**, 87-97.

[81] Zhao, Y. and Jian, W. (2007). Analysis of longitudinal data in the case-control studies via empirical likelihood. *Communications in Statistics-Simulation and Computation* **36,** 565-578.

[82] Zhao, Y. and Chen, F. (2008). Empirical likelihood inference for censored median regression model via nonparametric kernal estimation. *Journal of Multivariate Analysis* **99,** 215-231.

[83] Zhao, Y. and Yang. S. (2008). Empirical likelihood inference for censored median regression with weighted empirical hazaed functions. *Annals of the Institute of Statistical Mathematics* **60,** 441-457.

[84] Zhong, B. and Rao, J. N. K. (2000). Empirical likelihood inference under stratified random sampling using auxiliary population information. *Biometrika* **87**, 929-938.

[85] Zhou, X. H., Obuchowski, N. A., McClish, D. K. (2002). *Statistical Methods in Diagnostic Medicine.* Wiley: New York.

[86] Zhou X.H., Castellucio P. (2004). Adjusting for non-ignorable verification bias in clinical studies for Alzheimer's disease. *Statistics in Medicine* **23,** 221-230.

[87] Zhou, Y. and Liang, H. (2005). Empirical-likelihood-based semiparametric inference for the treatment in the two-sample problem with censoring. *Biometrika* **92,** 271-282.

[88] Zhou, W. and Jing, B. Y. (2003). Adjusted empirical likelihood method for the quantiles. *Annals of the Institute of Statistical Mathematics* **55**, 689-703.

[89] Zhou, W. (2008). Statistical inference for $P(X < Y)$. *Statistics in Medicine* **27**, 257-279.