5-8-2020

Jackknife empirical likelihood methods for the income inequality lower mean ratio

Li Zhang

JACKKNIFE EMPIRICAL LIKELIHOOD METHODS FOR THE INCOME INEQUALITY LOWER MEAN RATIO

by

LI ZHANG

Under the Direction of Yichuan Zhao, PhD

ABSTRACT

Measuring economic inequality is a significant and meaningful topic in our social system. The Gini index and Pietra ratio are used by many people, but limited to reflecting the sampling distribution. In this thesis, we studied the interval estimates with another measure called the lower-mean ratio u, which was introduced by Elteto and Frigyes (1968). By using jackknife empirical likelihood (JEL), adjusted jackknife empirical likelihood (AJEL), mean jackknife empirical likelihood (MJEL), mean adjusted jackknife empirical likelihood

(MAJEL), and adjusted mean jackknife empirical likelihood (AMJEL) methods, we proposed the interval estimator for u. In the following simulation study, we made a comparison for these methods under different distributions in terms of the coverage probability and the average confidence interval length. The results indicate that MAJEL performs best among these methods for small sample sizes of skewed distribution. For a small sample size of normal distribution, both JEL and MJEL show better performance than the other methods but MJEL is relatively time-consuming. All methods exhibit good performance for a large sample size. The two real data set analyses further illustrate the proposed methods, and the results are consistent with those in the simulation study.

INDEX WORDS: Economic inequality, Gini index, Jackknife empirical likelihood, Adjusted jackknife empirical likelihood, Mean jackknife empirical likelihood, Mean adjusted jackknife empirical likelihood, Adjusted mean jackknife empirical likelihood, Wilk's theorem

JACKKNIFE EMPIRICAL LIKELIHOOD METHODS FOR THE INCOME INEQUALITY LOWER MEAN RATIO

by

LI ZHANG

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

Master

in the College of Arts and Sciences

Georgia State University

2020

JACKKNIFE EMPIRICAL LIKELIHOOD METHODS FOR THE INCOME INEQUALITY LOWER MEAN RATIO

by

LI ZHANG

Committee Chair: Yichuan Zhao

Committee: Jun Kong

Jing Zhang

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2020

DEDICATION

To my family and Georgia State University.

ACKNOWLEDGEMENTS

I want to acknowledge those who have helped me in the completion of this thesis, which would not have been possible without the support of many people.

First and foremost, I would like to thank my advisor professor Yichuan Zhao, who is a nice, patient, and insightful person. With his patient guidance and generous support, I learned all the empirical likelihood methods, which is very helpful to guarantee the successful completion of my thesis research.

I am also grateful to the other committee members for their valuable advice on thesis writing.

I also would like to thank all the other faculty and staff in the Department of Mathematics and Statistics. Without their support, I would not be able to fulfil the requirements of the graduate program. Additionally, I want to thank Dr. Hongyan Jiang for revising the formula format in latex and Faysal Satter for the help on R coding. Moreover, I would like to thank Donald Gregory Harden and Brain Daniel Pidgeon for their proofreading.

Last but not least, I am extremely grateful to my parents, brother, husband, and lovely son for all their support throughout my life. I love you all forever.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

- CI - Confidence Interval
- NA - Normal Approximation
- AJEL - Adjusted Jackknife Empirical Likelihood
- CDF - Cumulative Distribution Function
- PDF - Probability Distribution Function
- JEL - Jackknife Empirical Likelihood
- MEL - Mean Empirical Likelihood
- MJEL - Mean Jackknife Empirical Likelihood
- MAJEL - Mean Adjusted Jackknife Empirical Likelihood
- AMJEL - Adjusted Mean Jackknife Empirical Likelihood

CHAPTER 1

INTRODUCTION

1.1 General introduction

Economic inequality is a social phenomenon showing the unequal distribution of income and opportunity between different groups, which is a common issue for almost every country around the world. Because evaluating economic inequality is always an important topic, many methods, including Gini index, Pietra ratio, Hoover index, Galt score and so on, have been developed to measure income inequality. However, these summary indices cannot reflect the sampling distribution, like the population size and the shared income of the relatively low income group. The measures of Elteto and Frigyes (1968) can address the problems above to some extent.

Elteto and Frigyes (1968) suggested measures depending on the mean income m, the mean income of those earning less than the mean m_1, and the mean income of those earning more than the mean m_2. They proposed the following measures:

$$u = \frac{m}{m_1}, v = \frac{m_2}{m_1}, w = \frac{m_2}{m}.$$

In this thesis, we defined u as the lower-mean ratio. Let $X_1, ..., X_n$ be the *i.i.d.* sample of X.

The estimator of u is

$$\hat{u} = \frac{N\bar{X}}{\sum\limits_{X_i<\bar{X}} X_i},$$

where $\bar{x}$ is the estimator of m, and N is a sum of the number which is smaller than the mean.

The estimator of m_1 is

$$\hat{m}_1 = \frac{\sum\limits_{X_i<\bar{X}} X_i}{N}.$$

For large samples sizes, Gastwirth (1974) established the normal approximation (NA) method for the estimator $\hat{u}$. However, the NA method is not only complicated but also inadequate to be satisfactory for small samples sizes. Here, we adopted some non-parametric methods such as jackknife empirical likelihood methods to improve the performance of measuring income inequality by the lower-mean ratio.

1.2 The review of jackknife empirical likelihood

The first paper to use empirical likelihood-based confidence intervals construction for survival data analysis was reported by Thomas and Grunkemeier (1975). Owen (1988, 1990) used a non-parametric likelihood approach and provided inferences with appealing data-driven and range respecting features. It combines the reliability of non-parametric methods with the effectiveness of the likelihood approach and returns the confidence regions, which respect the boundaries of target support parameters. The regions are invariant under transformations and usually show better performance than the confidence regions obtained from the NA method when the sample size is small (Chen and Keilegom (2009)).

Jing et al. (2009) introduced the jackknife empirical likelihood (JEL) method, which is extremely simple in practice and shows the effectiveness in handling U-statistics. Recently, a lot of applications for the JEL method have been reported. Sang et al. (2019) extended the JEL to K-sample test via the categorical Gini correlation. Bouadoumou et al. (2014) employed the JEL method to obtain the interval estimate for the regression parameter in the accelerated failure time model with censored observations. According to Wang et al. (2013), the JEL test for the equality of two high-dimensional means shows that it has a very robust size across dimensions and has good power. Luo and Qin (2019) employed JEL-based inferences for the Lorenz curve with kernel smoothing. Sreelakshmi et al. (2019) proposed JEL-based inference for single series Gini indices. Alemdjrodo and Zhao (2019) employed JEL to compare two correlated Gini indices in order to reduce the computational cost. Gong et al. (2010) proposed a smoothed JEL for the ROC curve, which enhances the computational efficiency. Cheng and Zhao (2019) successfully employed Bayesian jackknife empirical likelihood (BJEL) for the U- statistics type estimating equations.

In addition, the empirical likelihood (EL) may not be properly defined because of the so-called empty set problem (Chen et al. (2008); Tsao and Wu (2013)). Lots of approaches have been developed to improve the accuracy of EL confidence regions and address the issue above as well. The Bootstrap calibration (Owen (1988)) and the Bartlett correction (Chen and Cui (2007)) approaches improved the accuracy of EL confidence regions. The adjusted empirical likelihood (AEL) method (Chen et al. (2008); Liu and Chen (2010); Chen and Liu (2012); Wang et al. (2014)) solved the empty set issue and the low coverage probability problem simultaneously.

1.3 The review of adjusted jackknife empirical likelihood

Based on the EL method, Chen et al. (2008) proposed an adjusted empirical likelihood, in which one or two pseudo-observations are added to the sample to make sure that the convex hull constraint is never violated. Inspired by this method, Zhao, Meng, and Yang (2016) proposed an adjusted JEL for the mean deviation. By doing so, the AJEL method reduces error rates of the proposed jackknife empirical likelihood ratio. AJEL has the same asymptotic distribution with EL. In the thesis, instead of using empirical likelihood, we will employ the jackknife empirical likelihood (JEL) and adjusted jackknife empirical likelihood (AJEL) to get an interval estimate of u.

1.4 The review of mean empirical likelihood

Liang et al. (2019) proposed the mean empirical likelihood (MEL) method, which is relatively more effective in handling small sample sizes when compared with EL. In the MEL method, we will firstly generate a pseudo-data set, then calculate MEL ratios based on the pseudo-data.

1.5 The mean adjusted jackknife empirical likelihood

Based on the advantages of adjusted empirical likelihood and mean jackknife empirical likelihood methods, we propose the mean adjusted jackknife empirical likelihood method (MAJEL) for an interval estimate of u in the thesis.

1.6 The adjusted mean jackknife empirical likelihood

According to the strengths of the two methods i.e., adjusted empirical likelihood and mean jackknife empirical likelihood mentioned above, we also propose the adjusted mean jackknife empirical likelihood method (AMJEL) for an interval estimate of u in this thesis.

1.7 Purpose of the study

In this thesis, we propose jackknife empirical likelihood (JEL), adjusted jackknife empirical likelihood (AJEL), mean jackknife empirical likelihood (MJEL), mean adjusted jackknife empirical likelihood (MAJEL), and adjusted mean jackknife empirical likelihood method (AMJEL) for the inference of u. Then, we construct the confidence interval and calculate the length of the confidence intervals. We will evaluate these methods in terms of the coverage probability and the average length for the confidence interval of u.

The thesis is organized as follows. In Chapter 2, we propose a construction of confidence intervals for u using JEL, AJEL, MJEL, MAJEL and AMJEL. In Chapter 3, we perform an extensive simulation study. In Chapter 4, we apply the proposed methods to two real data sets. In Chapter 5, we make a conclusion for the proposed methodology.

CHAPTER 2

METHODOLOGY

In this chapter, we briefly reviewed a few known methods, including a normal approximation, for estimating u. Then, the jackknife empirical likelihood, adjusted jackknife empirical likelihood, mean jackknife empirical likelihood, mean adjusted jackknife empirical likelihood, and adjusted mean jackknife empirical likelihood are proposed for the interval estimate of u.

2.1 Normal approximation for u

Let $X_1, \ldots, X_n$ be a sequence of i.i.d. random variables from a cumulative distribution function $F(x)$ and density function $f(x)$ with the mean $m = E[X]$, $m_1 = E[X\mathbf{1}(X < m)]$. Then, the estimator of u is,

$$\hat{u} = \frac{N\bar{X}}{\sum\limits_{X_i<\bar{X}} X_i},$$

and

$$\begin{aligned} s_1 &= n^{\frac{1}{2}}\left(\frac{N\bar{X}}{n} - pm\right) \\ &= mn^{\frac{1}{2}}\sum\left(I_i - p\right) + (mf\left(\mathrm{m}\right) + p)n^{\frac{1}{2}}\left(\bar{X} - m\right) + o_p\left(1\right), \end{aligned}$$

$$\begin{aligned} s_2 &= n^{\frac{1}{2}}\left[n^{-1}\sum_{X_i<\bar{X}} X_i - \tau\right] \\ &= n^{\frac{1}{2}}\sum\left(I_iX_i - \tau\right) + mf\left(\mathrm{m}\right)n^{\frac{1}{2}}\left(\bar{X} - m\right) + o_p\left(1\right), \end{aligned}$$

where $p = F(m)$.

Gastwirth (1974) found that the estimator $\hat{u}$ of u was asymptotically normally distributed with mean $\tau^{-1}pm$ and the variance,

$$\frac{v_1}{\tau^2} + v_2\left(\frac{pm}{\tau^2}\right)^2 - \frac{2cpm}{\tau^3},$$

where

$$\begin{cases} \mathrm{I}_i = 1, & X_i < m \\ \mathrm{I}_i = 0, & o.w. \end{cases}$$

$$\tau = \mathrm{E}(\mathrm{I}_i \mathrm{X}_i) = \int_{-\infty}^{m} x dF(x),$$

$$\begin{aligned} v_1 &= \mathrm{var}(s_1) \\ &= m^2 p(1-p) + \sigma^2 (p + mf(m))^2 - 2m\left(mf\left(\mathrm{m}\right) + p\right)\left(pm - \tau\right), \end{aligned}$$

$$\begin{aligned} v_2 &= \mathrm{var}(s_2) \\ &= m^2 f^2\left(m\right)\sigma^2 + \int_{-\infty}^{m} x^2 dF\left(x\right) - \tau^2 + 2mf(m)\int_{-\infty}^{m} x(x-m)dF(x), \end{aligned}$$

$$c = \mathrm{cov}(s_{1,}s_2)$$

$$= m\tau\left(1-p\right) - m^2 f\left(m\right)\left(pm - \tau\right) + mf(m)[mf(m)+p]\sigma^2 + (p + mf(m))\int_{-\infty}^{m} x(x-m)dF(x),$$

and $\sigma^2 = E(X-m)^2$.

Then we constructed a 100(1-α)% normal approximation (NA) based confidence interval for u:

$$R = \{u : \hat{u} \pm Z_{\alpha/2} * SE\},$$

where $Z_{\alpha/2}$ is the upper $\alpha/2$-quantile of standard normal distribution, and

$$SE = \sqrt{\left\{ \frac{\hat{v}_1}{\hat{\tau}^2} + \hat{v}_2 (\frac{\hat{p}\bar{x}}{\hat{\tau}^2})^2 - \frac{2\hat{c}\hat{p}\bar{x}}{\hat{\tau}^3} \right\}},$$

where $\hat{v}_1$ is the estimator of v_1, $\hat{v}_2$ is the estimator of v_2, $\hat{c}$ is the estimator of c, $\hat{p}$ is the estimator of p, and calculated by the following formulas:

$$\hat{f}(x) = \frac{\sum_{i=1}^{n} K(\frac{x-x_i}{h})}{nh},$$

where $h = (4\hat{\sigma}^5/3n)^{1/5} \approx 1.06\hat{\sigma}n^{-1/5}$, see Silverman (1986) for the choice of the bandwidth h. We substitute the estimate value of σ with parameter A with $A = min(\hat{\sigma}, (Q_3 - Q_1)/1.34)$, where Q_1 is the first quantile of data, Q_3 is the third quantile of data, and $K(x) = \exp(-x^2/2)/\sqrt{2\pi}$. Here,

$$\hat{F}_n(x) = \frac{\sum_{i=1}^{n} \mathbf{1}(x_i \leq x)}{n},$$

$$\hat{p} = \hat{F}_n(\bar{x}),$$

$$\hat{v}_1 = \bar{x}^2\hat{p}(1-\hat{p}) + \hat{\sigma}^2\left(\hat{p} + \bar{x}\hat{f}(x)\right)^2 - 2\bar{x}\left(\bar{x}\hat{f}(x) + \hat{p}\right)(\hat{p}\bar{x} - \hat{\tau}),$$

$$\hat{v}_2 = \bar{x}^2\hat{f}^2(\bar{x})\,\hat{\sigma}^2 + \int_{-\infty}^{\bar{x}} \bar{x}^2 d\hat{F}_n(x) - \hat{\tau}^2 + 2\bar{x}\hat{f}(x)\int_{-\infty}^{\bar{x}} x(x-\bar{x})d\hat{F}_n(x),$$

$$\hat{c} = \bar{x}\hat{\tau}(1-\bar{x}) - \bar{x}^2\hat{f}(x)(\hat{p}\bar{x} - \hat{\tau}) + \bar{x}\hat{f}(x)[\bar{x}\hat{f}(x) + \hat{p}]\hat{\sigma}^2 + \left(\hat{p} + \bar{x}\hat{f}(x)\right)\int_{-\infty}^{\bar{x}} x(x-\bar{x})d\hat{F}_n(x).$$

2.2 Jackknife empirical likelihood for u

Let X_1, ..., X_n be $i.i.d$ random vectors in R^p with common distribution function $F(t)$. Ix_i denotes a point mass at x_i. Then, the empirical distribution function is given by:

$$\mathrm{F}_n = \frac{1}{n}\sum_{i=1}^{n} I_{x_i}.$$

F_n is known to be the non-parametric maximum likelihood estimate of F_0 based on X_1, ..., X_n. The likelihood function that F_n maximizes is:

$$L(F) = \prod_{i=1}^{n} F\left\{x_i\right\}.$$

Owen (1988) introduced the empirical likelihood to determine the confidence intervals regardless of the variance. Supposing that we have an $i.i.d.$ sample with $(W_1, \ldots, W_n)$ random variables, then the empirical likelihood ratio function is given by:

$$R(F) = \frac{L(F)}{L(F_n)} = \prod_{i=1}^{n}(np_i).$$

The objective of the empirical likelihood is to construct the confidence intervals for the parameter $u = Eg(W_i)$. Based on Owen's (2001) work, the empirical likelihood for u in this thesis is defined as,

$$L(u) = \max\left\{\prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i g(W_i) = u, p_i \geq 0\right\}.$$

The corresponding empirical likelihood ratio function for u can be defined as:

$$R(u) = \frac{L(u)}{n^{-n}} = \max\left\{\prod_{i=1}^{n} np_i : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i g(W_i) = u, p_i \geq 0\right\}.$$

Even though this empirical likelihood has many advantages in constructing confidence regions, it still has some limitations on more complicated statistics like U-statistics because of computational problems. To address these problems, Jing et al. (2009) proposed the jackknife empirical likelihood for the U-statistics. The jackknife empirical likelihood method combines two non-parametric approaches: jackknife method and empirical likelihood method. The jackknife method was constructed by Quenouille (1956) and further developed by Tukey (1958). The key steps and general context of the JEL method are described as follows. The consistent estimator of the parameter u is given by,

$$T_n(u) = u\hat{m}_1 - \bar{x}.$$

The jackknife pseudo-values are defined as,

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

where $T_{n-1}^{(-i)}(u)$ is computed from the original data set by deleting the i-th observation. Then the jackknife estimator $\hat{T}_{njack}(u)$ is the average of all the pseudo-values,

$$\hat{T}_{njack}(u) = \frac{1}{n}\sum_{i=1}^{n}\hat{V}_i(u).$$

After applying the jackknife pseudo-values to the empirical likelihood, we define,

$$L(u) = \max\left\{\prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i\hat{V}_i(u) = 0, p_i \geq 0\right\}.$$

The corresponding empirical likelihood ratio function for u can be written as,

$$R(u) = \frac{L(u)}{n^{-n}} = \max\left\{\prod_{i=1}^{n} np_i : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i\hat{V}_i(u) = 0, p_i \geq 0\right\}.$$

Using Lagrange multipliers method, we can get,

$$p_i = \frac{1}{n}\frac{1}{1+\lambda\hat{V}_i(u)},$$

where λ satisfies the following equation,

$$f(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{V}_i(u)}{1+\lambda\hat{V}_i(u)} = 0.$$

Then we get the jackknife empirical log-likelihood ratio,

$$-2logR\left(u\right) = 2\sum_{i=1}^{n} log\left\{(1+\lambda\hat{V}_i(u)\right\}.$$

The following theorem explains how Wilk's theorem works for u.

Theorem 1 Assume the regularity conditions that $X_1, ..., X_n$ are $i.i.d.$ with finite mean m and finite σ^2. Let u_0 denote the true value of u. When $n\rightarrow\infty$, -2logR(u_0) converges to χ_1^2 in distribution.

Using Theorem 1, the JEL confidence interval for u is constructed as follows:

$$R_1 = \left\{u : -2logR\left(u\right) \leq \chi_1^2(\alpha)\right\},$$

where $\chi_1^2(\alpha)$ is the upper α-quantile of χ_1^2.

2.3 Adjusted jackknife empirical likelihood for u

The method of adjusted empirical likelihood was developed by Chen et al. (2008). According to Zheng and Yu (2013), the adjusted empirical likelihood is better than the original method because it can reduce the amount of deviation. Also, the adjusted empirical likelihood method can avoid the convex hull restrictions from the jackknife empirical likelihood method. The adjusted jackknife empirical likelihood (AJEL) function for u is defined as,

$$h_i^{ad}(u) = \hat{V}_i(u).$$

Then the AJEL at u is defined as,

$$L^{ad}(u) = \max\left\{\prod_{i=1}^{n+1} p_i : \sum_{i=1}^{n+1} p_i = 1, \sum_{i=1}^{n+1} p_i h_i^{ad}(u) = 0, p_i \geq 0\right\},$$

where $h_{n+1}^{ad}(u) = -a_n \bar{h}^{ad}(u)$ and a_n is a constant number depending on n,

$$\begin{cases} a_n = max(1, log(n)/2) \\ \bar{h}^{ad}(u) = \frac{1}{n}\sum\limits_{i=1}^{n} h_i^{ad}(u). \end{cases}$$

Therefore, we define the adjusted jackknife empirical likelihood ratio for u as,

$$R^{ad}(u) = \max\left\{\prod_{i=1}^{n+1} (n+1)p_i^{ad}, \sum_{i=1}^{n+1} p_i = 1, \sum_{i=1}^{n+1} p_i h_i^{ad}(u) = 0, p_i \geq 0\right\}.$$

Using Lagrange multipliers method, we can get,

$$p_i = \frac{1}{n+1}\frac{1}{1+\lambda h_i^{ad}(u)}, i = 1, \ldots, n+1$$

where λ satisfies the following equation,

$$f(\lambda) = \frac{1}{n+1}\sum_{i=1}^{n+1} \frac{h_i^{ad}(u)}{1+\lambda h_i^{ad}(u)} = 0.$$

Then, we get the adjusted jackknife empirical log-likelihood ratio,

$$-2logR^{ad}(u) = 2\sum_{i=1}^{n+1} log\{1+\lambda h_i^{ad}(u)\}.$$

Combining the approaches by Chen et al. (2008) and Jing et al. (2009), we have the Wilks' theorem holds, as $n \to \infty$.

Theorem 2 Assume the regularity conditions that $X_1, ..., X_n$ are *i.i.d.* with finite mean m

and finite σ^2. Let u_0 denote the true parameter value. When $n \rightarrow \infty$, -2logR^{ad} (u_0) converges to the χ_1^2 in distribution.

Using Theorem 2, the AJEL confidence interval for u is constructed as follows:

$$R_2 = \left\{ u : -2logR^{ad}\left(u\right) \leq \chi_1^2(\alpha) \right\},$$

where $\chi_1^2(\alpha)$ is the upper α-quantile of χ_1^2.

2.4 Mean jackknife empirical likelihood for u

In order to improve the accuracy of the empirical likelihood confidence interval for small sample sizes, Liang et al. (2019) introduced the mean empirical likelihood (MEL). We firstly generated a pseudo data set using the means of the observed values and then applied it for the empirical likelihood analysis. In this thesis, we define the mean jackknife empirical likelihood (MJEL) pseudo value as follows,

$$U_i(u) = \frac{\hat{V}_j(u) + \hat{V}_k(u)}{2}, i = 1, \ldots, N, 1 \leq j \leq k \leq n,$$

where $N = n(n+1)/2$.

Therefore, the mean jackknife empirical likelihood ratio for u is,

$$R^m = \max \left\{ \prod_{i=1}^{N} Np_i : \sum_{i=1}^{N} p_i = 1, \sum_{i=1}^{N} p_i U_i(u) = 0, p_i \geq 0 \right\}.$$

Then we get the mean jackknife empirical log-likelihood ratio:

$$\frac{-2logR^m\left(u\right)}{n+1} = \frac{2}{n+1} \sum_{i=1}^{N} log\{1 + \lambda U_i(u)\},$$

where λ satisfies the following equation,

$$f(\lambda) = \sum_{i=1}^{N} \frac{U_i(u)}{1 + \lambda U_i(u)} = 0.$$

Let u_0 denote the true value of u. Based on the method by Liang et al. (2019), we have the following theorem.

Theorem 3 Let u_0 denote the true parameter value. Under the same conditions in Theorem 1, $-2logR^m\,(u_0)/(n+1)$ converges to the χ_1^2 in distribution, as $n\to\infty$.

Using Theorem 3, the MJEL confidence interval for u is constructed as follows:

$$R_3 = \left\{ u : \frac{-2logR^m(u)}{n+1} \leq \chi_1^2(\alpha) \right\},$$

where $\chi_1^2(\alpha)$ is the upper α-quantile of χ_1^2.

2.5 Mean adjusted jackknife empirical likelihood for u

Based on adjusted empirical likelihood and mean jackknife empirical likelihood methods, we propose the mean adjusted jackknife empirical likelihood method for u. The mean adjusted jackknife empirical likelihood (MAJEL) pseudo value for u is defined as,

$$W_i(u) = \frac{\hat{h}_j^{ad}(u) + \hat{h}_k^{ad}(u)}{2}, i = 1, \ldots, M, 1 \leq j \leq k \leq n+1,$$

where $M = (n+2)(n+1)/2$.

Therefore, the mean adjusted jackknife empirical likelihood ratio for u is,

$$R^{maj} = \max \left\{ \prod_{i=1}^{M} Mp_i : \sum_{i=1}^{M} p_i = 1, \sum_{i=1}^{M} p_i W_i(u) = 0, p_i \geq 0 \right\}.$$

Then we get the mean adjusted jackknife empirical log-likelihood ratio:

$$\frac{-2logR^{maj}\,(u)}{n+2} = \frac{2}{n+2} \sum_{i=1}^{M} log\{1 + \lambda W_i(u)\},$$

where λ satisfies the following equation,

$$f(\lambda) = \sum_{i=1}^{M} \frac{W_i(u)}{1 + \lambda W_i(u)} = 0.$$

Theorem 4 Let u_0 denote the true parameter value. Under the same conditions in Theorem 1, $-2logR^{maj}(u_0)/(n+2)$ converges to the χ_1^2 in distribution, as $n \to \infty$.

Using Theorem 4, the MAJEL confidence interval for u is constructed as follows:

$$R_4 = \left\{ u : \frac{-2logR^{maj}(u)}{n+2} \leq \chi_1^2(\alpha) \right\},$$

where $\chi_1^2(\alpha)$ is the upper α-quantile of χ_1^2.

2.6 Adjusted mean jackknife empirical likelihood for u

For the mean jackknife empirical likelihood (MJEL), we add one point U_{N+1} which is defined as,

$$U_{N+1}(u) = -a_{N+1}\bar{U}(u),$$

where a_{N+1} is a constant number depending on n,

$$\begin{cases} a_{N+1} = max(1, log(N+1)/2) \\ \bar{U}(u) = \frac{1}{N}\sum\limits_{i=1}^{N} U_i(u). \end{cases}$$

Then the adjusted mean jackknife empirical likelihood (AMJEL) ratio for u is defined as,

$$R^{amj}(u) = \max\left\{ \prod_{i=1}^{N+1}(N+1)p_i, \sum_{i=1}^{N+1} p_i = 1, \sum_{i=1}^{N+1} p_i U_i(u) = 0, p_i \geq 0 \right\}.$$

Using Lagrange multipliers method, we can get,

$$p_i = \frac{1}{N+1}\frac{1}{1+\lambda U_i(u)}, i = 1, \ldots, N+1$$

where λ satisfies the following equation,

$$f(\lambda) = \frac{1}{N+1} \sum_{i=1}^{N+1} \frac{U_i(u)}{1 + \lambda U_i(u)} = 0.$$

Then we get the adjusted mean jackknife empirical log-likelihood ratio,

$$\frac{-2logR^{amj}\left(u\right)}{n+1} = \frac{2}{n+1} \sum_{i=1}^{N+1} log\{1 + \lambda U_i(u)\}.$$

When $n \to \infty$, we have the Wilks' theorem holds.

Theorem 5 Let u_0 denote the true parameter value. Under the same conditions in Theorem 1, $-2logR^{amj}\left(u_0\right)/(n+1)$ converges to the χ_1^2 in distribution, as $n \to \infty$.

Using Theorem 5, the AMJEL confidence interval for u is constructed as follows:

$$R_5 = \left\{ u : \frac{-2logR^{amj}(u)}{n+1} \leq \chi_1^2(\alpha) \right\},$$

where $\chi_1^2(\alpha)$ is the upper α-quantile of χ_1^2.

CHAPTER 3

SIMULATION STUDY

In this chapter, we reported the performance of different methods, including JEL, AJEL, MJEL, MAJEL, and AMJEL, on measuring u of finite samples. Meanwhile, we compared the results with that from the NA method under the normal distribution, Weibull distribution, log-normal distribution, and exponential distribution, respectively. We also estimated u with different simple sizes: 5, 20, 50, 80, and 100. The number of repetitions we performed on each sample size was 5,000. In addition, we estimated the coverage probability of confidence intervals by checking whether $-2logR(u_0)$ is less than $\chi_1^2(\alpha)$ or not. We used three different significance levels ($\alpha = 0.1$, 0.05, 0.01) in this study, and calculated the average length of confidence intervals for each u at all situations.

3.1 Simulation under the normal distribution N(3,1)

The result of the coverage probability is shown in Table 3.1. Generally, from the small sample size to the big sample size, AJEL, MAJEL, and AMJEL always show over coverage at all three confidence levels, while the others (NA, JEL, and MJEL) give good performance and their coverage probabilities increase with the sample size at different confidence levels. Especially, when the sample size n equals to 5, NA performs worst with a coverage probability of 73.0%.

The result of the average interval length is shown in Table 3.2. For a specific sample size, the average interval length of all six methods shortens with a decreased confidence level. For all three confidence levels, each method has an increasing average interval length when the sample size becomes smaller. In general, for a fixed sample size, MAJEL has the longest average interval length while NA has the shortest average length at all three confidence levels among the six methods.

3.2 Simulation under the exponential distribution exp(1)

In general, each method has an increasing coverage probability when the sample size is increased at a specific confidence level (Table 3.3). Among these methods, for a fixed sample size, MAJEL shows the best coverage probability even though it has a little bit over coverage for a large sample size ($n = 100$), while NA performs worst at the same confidence level.

We get the similar observations for the average interval length as shown in Table 3.4. At a specific confidence level, MAJEL gives the longest average interval length and NA exhibits the shortest average length for a fixed sample size among all the six methods.

3.3 Simulation under the Weibull distribution weibull(2,1)

The results of the coverage probability in Table 3.5 and the average interval length in Table 3.6 illustrate the same observations as it shows in Section 3.2 except for that no methods have over coverage in this section.

3.4 Simulation under the log-normal distribution log-norm(0,1)

As shown in Table 3.7 and Table 3.8, we can make the same conclusions with Section 3.3. Briefly, among all the six methods, MAJEL gives the best coverage probability and longest average interval length, while NA performs worst and has the shortest average interval length at all three confidence levels for a specific sample size.

Table (3.1) The coverage probability under the normal distribution

	CL=90%					CL =95%					CL=99%				
n	5	20	50	80	100	5	20	50	80	100	5	20	50	80	100
NA	73.0%	85.9%	88.6%	88.9%	89.3%	80.2%	91.1%	94.2%	94.4%	94.6%	87.4%	96.3%	97.5%	98.3%	98.5%
JEL	76.5%	86.2%	88.6%	89.1%	89.6%	82.7%	91.6%	94.5%	94.6%	94.7%	87.7%	96.6%	97.8%	98.5%	98.7%
$AJEL$	91.8%	91.1%	91.7%	91.8%	92.3%	100.0%	95.6%	96.0%	96.2%	96.6%	100.0%	99.2%	99.3%	99.4%	99.6%
$MJEL$	77.9%	86.6%	89.0%	89.3%	89.8%	83.5%	91.8%	94.7%	94.9%	94.9%	88.1%	96.8%	98.0%	98.7%	98.9%
$MAJEL$	91.2%	91.6%	92.0%	92.4%	93.1%	95.8%	96.0%	96.5%	96.7%	96.8%	99.3%	99.4%	99.5%	99.6%	99.7%
$AMJEL$	91.0%	91.2%	92.0%	92.1%	92.8%	95.4%	95.8%	96.1%	96.4%	96.7%	99.2%	99.3%	99.4%	99.5%	99.7%

Table (3.2) The average interval length under the normal distribution

	CL=90%					CL =95%					CL=99%				
n	5	20	50	80	100	5	20	50	80	100	5	20	50	80	100
NA	1.157	0.689	0.440	0.411	0.401	1.349	0.818	0.540	0.517	0.456	1.682	1.139	0.754	0.607	0.525
JEL	1.166	0.703	0.456	0.413	0.408	1.355	0.838	0.545	0.521	0.458	1.714	1.165	0.793	0.614	0.531
$AJEL$	1.728	0.921	0.702	0.515	0.425	2.268	0.958	0.825	0.612	0.535	2.225	1.199	1.120	0.812	0.708
$MJEL$	1.172	0.712	0.492	0.441	0.411	1.376	0.843	0.575	0.538	0.478	1.734	1.186	0.875	0.621	0.608
$MAJEL$	1.712	1.020	0.832	0.549	0.519	2.174	1.112	0.879	0.655	0.581	2.218	1.590	1.271	0.911	0.795
$AMJEL$	1.709	1.015	0.807	0.543	0.517	2.170	1.107	0.876	0.650	0.578	2.131	1.487	1.159	0.903	0.787

Table (3.3) The coverage probability under the exponential distribution

	CL=90%					CL =95%					CL=99%				
n	5	20	50	80	100	5	20	50	80	100	5	20	50	80	100
NA	43.3%	64.6%	78.2%	85.5%	86.6%	51.5%	71.8%	86.1%	90.1%	91.6%	60.7%	82.4%	91.5%	92.3%	94.8%
JEL	49.9%	80.6%	84.2%	85.5%	86.8%	55.6%	83.6%	90.5%	91.8%	92.1%	62.6%	93.8%	94.1%	94.6%	95.1%
$AJEL$	74.5%	87.4%	87.8%	89.3%	89.5%	79.7%	88.7%	92.5%	94.7%	94.9%	85.4%	95.5%	97.7%	98.5%	98.8%
$MJEL$	54.4%	81.3%	85.2%	87.1%	87.8%	64.3%	88.3%	90.9%	93.3%	94.2%	70.1%	94.6%	95.1%	95.2%	97.2%
$MAJEL$	80.3%	88.1%	88.8%	89.6%	90.1%	88.3%	94.4%	94.7%	94.8%	95.0%	94.2%	98.3%	98.7%	98.9%	99.1%
$AMJEL$	75.5%	87.7%	88.2%	89.4%	89.6%	79.9%	88.9%	92.8%	94.8%	94.9%	87.4%	96.4%	98.2%	98.5%	98.9%

Table (3.4) The average interval length under the exponential distribution

	CL=90%					CL =95%					CL=99%				
n	5	20	50	80	100	5	20	50	80	100	5	20	50	80	100
NA	0.919	0.620	0.467	0.374	0.323	1.112	0.750	0.507	0.451	0.390	1.243	0.987	0.638	0.527	0.491
JEL	0.921	0.622	0.471	0.378	0.326	1.156	0.752	0.510	0.453	0.391	1.292	0992	0.644	0.530	0.495
$AJEL$	1.448	0.910	0.551	0.418	0.365	1.713	1.114	0.675	0.501	0.446	2.190	1.541	0.943	0.709	0.618
$MJEL$	0.956	0.681	0.482	0.389	0.337	1.154	0.781	0.512	0.474	0.402	1.371	1.114	0.718	0.601	0.581
$MAJEL$	1.451	0.926	0.616	0.420	0.416	1.862	1.305	0.716	0.521	0.516	2.358	1.629	0.982	0.725	0.710
$AMJEL$	1.449	0.915	0.561	0.419	0.373	1.751	1.189	0.682	0.510	0.474	2.259	1.569	0.968	0.713	0.684

Table (3.5) The coverage probability under the Weibull distribution

	CL=90%					CL =95%					CL=99%				
n	5	20	50	80	100	5	20	50	80	100	5	20	50	80	100
NA	40.8%	53.0%	64.9%	83.8%	87.3%	48.8%	62.3%	72.6%	88.4%	90.7%	54.2%	68.1%	79.6%	91.2%	92.2%
JEL	41.9%	58.4%	70.3%	84.9%	87.3%	48.9%	63.7%	79.4%	88.9%	90.8%	57.6%	69.0%	85.5%	92.2%	92.8%
$AJEL$	78.5%	80.4%	82.9%	88.4%	88.6%	81.7%	84.2%	85.4%	93.2%	93.8%	86.4%	88.5%	90.3%	95.1%	95.3%
$MJEL$	48.2%	70.1%	77.2%	84.9%	88.3%	54.2%	75.9%	81.8%	90.3%	91.5%	65.1%	71.4%	88.5%	92.7%	93.3%
$MAJEL$	81.9%	87.2%	88.0%	89.2%	89.5%	85.0%	89.2%	92.4%	93.8%	94.2%	90.2%	93.1%	94.5%	98.2%	98.5%
$AMJEL$	81.1%	84.9%	86.5%	88.7%	89.1%	83.8%	85.9%	89.8%	93.5%	94.0%	88.2%	91.2%	93.3%	96.8%	97.7%

Table (3.6) The average interval length under the Weibull distribution

	CL=90%					CL =95%					CL=99%				
n	5	20	50	80	100	5	20	50	80	100	5	20	50	80	100
NA	1.968	1.521	1.427	0.936	0.898	2.011	2.005	1.502	0.975	0.916	2.179	1.826	1.631	0.992	0.956
JEL	1.973	1.771	1.459	0.941	0.902	2.039	2.012	1.601	0.980	0.919	2.190	1.865	1.649	0.998	0.960
$AJEL$	2.126	1.907	1.543	1.009	0.986	2.706	2.107	1.659	1.012	0.997	2.879	2.028	1.928	1.019	1.002
$MJEL$	1.938	1.776	1.478	0.982	0.978	2.147	2.006	1.608	0.996	0.961	2.368	1.897	1.702	1.003	0.972
$MAJEL$	2.371	2.016	1.921	1.026	1.008	2.920	2.264	1.796	1.260	1.012	2.961	2.168	1.981	1.268	1.037
$AMJEL$	2.213	1.989	1.697	1.012	0.998	2.819	2.198	1.702	1.101	1.003	2.917	2.098	1.961	1.113	1.016

Table (3.7) The coverage probability under the log-normal distribution

	CL=90%					CL =95%					CL=99%				
n	5	20	50	80	100	5	20	50	80	100	5	20	50	80	100
NA	44.5%	53.9%	59.7%	71.2%	82.5%	53.6%	62.7%	65.2%	79.8%	90.5%	65.0%	70.4%	81.2%	85.2%	96.1%
JEL	45.4%	59.4%	60.5%	77.3%	82.7%	56.2%	63.5%	69.4%	84.0%	91.2%	67.6%	78.7%	83.5%	89.4%	96.6%
$AJEL$	60.5%	76.4%	79.9%	84.1%	86.5%	71.8%	81.2%	84.3%	90.7%	93.7%	82.4%	84.5%	89.7%	93.5%	97.8%
$MJEL$	46.1%	62.0%	65.2%	80.9%	83.2%	57.2%	67.9%	71.9%	85.3%	92.0%	69.1%	81.4%	85.5%	90.9%	97.2%
$MAJEL$	72.6%	82.1%	86.5%	87.0%	88.4%	81.4%	89.1%	93.0%	93.5%	94.5%	89.7%	92.1%	94.5%	96.8%	98.3%
$AMJEL$	68.6%	79.8%	82.6%	85.4%	87.1%	77.6%	86.5%	89.9%	91.5%	93.9%	87.7%	89.0%	92.6%	94.9%	98.1%

Table (3.8) The average interval length under the log-normal distribution

	CL=90%					CL =95%					CL=99%				
n	5	20	50	80	100	5	20	50	80	100	5	20	50	80	100
NA	1.613	0.968	0.597	0.421	0.294	1.939	1.154	0.712	0.502	0.350	2.271	1.523	0.940	0.661	0.461
JEL	1.764	1.063	0.601	0.430	0.302	1.991	1.171	0.714	0.518	0.363	2.304	1.541	0.954	0.670	0.464
$AJEL$	1.926	1.067	0.643	0.476	0.381	2.306	1.207	0.798	0.612	0.431	2.779	1.624	0.992	0.712	0.608
$MJEL$	1.738	0.976	0.607	0.434	0.325	2.047	1.177	0.724	0.564	0.367	2.318	1.581	0.958	0.694	0.518
$MAJEL$	2.032	1.184	0.754	0.538	0.412	2.508	1.328	0.839	0.680	0.487	2.901	1.894	1.084	0.803	0.680
$AMJEL$	1.995	1.102	0.719	0.492	0.408	2.386	1.298	0.812	0.638	0.458	2.825	1.715	1.016	0.769	0.638

CHAPTER 4

REAL DATA ANALYSIS

In the real data analysis, we utilized data sets with small sample sizes to illustrate the proposed methods.

The NA, JEL, AJEL, MJEL, MAJEL, and AMJEL methods were separately applied to analyze the data sets. Then, we calculated the confidence interval length and the confidence interval bounds of the point estimate of u at three confidence levels (90%, 95%, and 99%).

4.1 Median household income analysis

The data set is about median household income, which gives the average annual household income by race in New Jersey, Puerto Rico, New Hampshire, District of Columbia, West Virginia, Mississippi, and Maryland from 2013 to 2017. Here, we just analyzed the data from the year 2017 with the simple size $n = 7$. As shown in the histogram (Figure 5 in the appendices), we can see that the distribution of household income is skewed left. The p-value of the goodness-of-fit test for the Weibull distribution is 0.3451, which is well above 0.05. The skewness of the data is -0.3288. All these results support that the data set follows a Weibull distribution. The point estimate of u is 1.892. The confidence interval (CI) for the u was calculated using NA, JEL, AJEL, MJEL, MAJEL, and AMJEL, respectively. The lengths and the bounds of the CI were displayed and compared with previous simulation results in Table 3.6.

The results in Table 4.1 indicate that MAJEL has the longest CI lengths at all three confidence levels, which is consistent with the observations in the simulation study (Table 3.6).

4.2 Income and education analysis

In the data set “income and education”, there are 20 observations. From the histogram (Figure 6 in the appendices), we can see that the distribution of household income is skewed. The p-value of the goodness-of-fit test for exponential distribution is 0.1946, which is also well above 0.05. The skewness of the data is 1.267. Therefore, the data set fits an exponential distribution. The point estimate of u is 2.989. The CI for u was calculated using NA, JEL, AJEL, MJEL, MAJEL, and AMJEL, respectively. The lengths and the bounds of the CI were also calculated and compared with the simulation results in Table 3.4.

As shown in Table 4.2, MAJEL and AMJEL have very close results, and both of them have a longer CI length than MJEL, AJEL, JEL and NA at all three confidence levels. MAJEL still has the longest CI lengths, which is consistent with the results in the simulation study (Table 3.4).

Table (4.1) Median household income analysis

	CL=90%			CL =95%			CL=99%		
n=7	length	LB	UB	length	LB	UB	Length	LB	UB
NA	1.450	1.167	2.617	1.668	1.058	2.726	1.794	0.995	2.789
JEL	1.504	1.089	2.593	1.712	0.908	2.620	1.825	0.801	2.626
AJEL	1.782	0.989	2.771	1.898	0.945	2.843	2.125	0.790	2.915
MJEL	1.525	1.092	2.617	1.754	0.982	2.736	1.984	0.802	2.786
MAJEL	1.891	0.927	2.818	2.018	0.942	2.960	2.287	0.784	3.071
AMJEL	1.815	0.938	2.753	1.910	0.940	2.850	2.185	0.790	2.975

Table (4.2) Income and education analysis

	CL=90%			CL =95%			CL=99%		
n=20	length	LB	UB	length	LB	UB	Length	LB	UB
NA	1.466	2.256	3.722	2.078	1.950	4.028	2.402	1.788	4.190
JEL	1.487	2.249	3.736	2.092	1.948	4.040	2.475	1.741	4.216
AJEL	1.887	2.212	4.099	2.197	1.915	4.112	2.636	1.719	4.355
MJEL	1.752	2.235	3.987	2.101	1.926	4.027	2.612	1.728	4.341
MAJEL	2.017	2.128	4.145	2.312	1.902	4.214	2.870	1.684	4.554
AMJEL	2.011	2.125	4.136	2.306	1.904	4.210	2.861	1.689	4.550

CHAPTER 5

CONCLUSIONS

In this thesis, we defined the lower mean ratio income inequality u, which was introduced by Elteto and Frigyes (1968) and used for measuring the economic inequality. For large samples, Gastwirth (1974) established the normal approximation (NA) method for the estimator $\hat{u}$. In order to obtain a more accurate confidence interval for u, we employed the NA method, and compared it with JEL, AJEL, MJEL, MAJEL, and AMJEL methods.

A simulation study was performed to assess the performance of the methods above in terms of the coverage probability and the average interval length. The results indicate that the performance of all methods can be improved by increasing the sample size. Among the six methods, MAJEL has the longest average interval length. When the sample size is small, for a skewed distribution, MAJEL shows the best performance and NA gives the worst performance. For a normal distribution, both JEL and MJEL have good performance, but MJEL is more time-consuming.

In addition, for a skewed distribution, the performance of NA is even worse than that in the symmetric normal distribution. Also, AJEL, AMJEL, and MAJEL show over coverage under the normal distribution N(3,1). Finally, the real data analyses of two data sets illustrate that results are consistent with those in the simulation study.

Further more, the proposed JEL methods and NA method that is developed by Gastwirth (1974) can be applied to the other measures v and w, which are proposed by Elteto and Frigyes (1968).

REFERENCES

Asadi, M., Berred, A. (2012). Properties and estimation of the mean past lifetime. Statistics, 46(3):405-417.

Alemdjrodo, K., Zhao, Y. (2019). Reduce the computation in jackknife empirical likeli- hood for comparing two correlated Gini indices. Journal of Nonparametric Statistics, 31(4):849-866.

Bouadoumou, M., Zhao, Y., and Lu, Y. (2014). Jackknife empirical likelihood for the accelerated failure time model with censored data. Communications in Statistics - Simulation and Computation, 104:1811-1832.

Chen, J. and Liu, Y. (2012). Adjusted empirical likelihood with high-order one-sided coverage precision. Statistics and Its Interface, 5:281-292.

Chen, J., Variyath, A., and Abraham, B. (2008). Adjusted empirical likelihood and its properties. J Comput Graph Stat, 17:426-443.

Chen, S. X. and Cui, H. (2007). On the second-order properties of empirical likelihood with moment restrictions. Journal of Econometrics, 141:492-516.

Chen, S. X. and Keilegom, I. V. (2009). A review on empirical likelihood methods for regression. Test, 18(33):415-447.

Cheng, Y. and Zhao, Y. (2019). Bayesian jackknife empirical likelihood. Biometrika, 106:981–988.

Elteto, O. and Frigyes, E. (1968). New income inequality measures as efficient tools for casual analysis and planning. Econometrica, 36:383-396.

Gastwirth, Joseph L (1974). Large sample theory of some measures of income inequality. Econometrica, 42:191-196.

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

Jayasinghe, CL, and Zeephongsekul, P. (2013). Nonparametric smooth estimation of the expected inactivity time function. Journal of Statistical Planning and Inference, 143:911–928.

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

Liang, W., Dai, H., and He, S. (2019). Mean empirical likelihood. Computational Statistics and Data Analysis, 138:155-169.

Luo, S., and Qin G. (2019). Jackknife empirical likelihood-based inferences for Lorenz curve with kernel smoothing. Communications in Statistics-Theory and Methods, 48:559-582.

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

Owen, A. (1990). Empirical likelihood and confidence regions. Ann. Statist, 18:90–120.

Owen, A. (1991). Empirical likelihood for linear models. Ann. Statist. 19:1725–1747.

Owen, A. (2001). Empirical likelihood. London: Chapman Hall/CRC.

Quenouille, M. (1956). Notes on bias in estimation. Biometrika, 10:353-360.

Sang, Y., Dang, X., and Zhao, Y. (2019). Jackknife empirical likelihood approach for K-sample tests, preprint.

Satter, F., and Zhao, Y. (2018). Nonparametric interval estimation for the mean of a zero-inflated population. Communications in Statistics - Simulation and Computation, inpress, 1-9.

Silverman, B. (1986). Density Estimation for Statistics and Data Analysis, Monographs on Statistics and Applied Probability, London, CRC.

Sreelakshmi, N., Kattumannil, K., and Rituparna, Sen. (2019). Jackknife empirical likelihood-based inference for S-Gini indices, Communications in Statistics-Simulation and Computation, 1-17, in press.

Tukey, J.W. (1958). Bias and confidence in not quite large samples (Abstract) Ann. Math. Statist. 29: 614.

Vali Z. (2017). On confidence intervals for the mean past lifetime function under random censorship. Journal of Statistical Computation and Simulation, 87:2153-2160.

Wang, D., Wu, T., and Zhao, Y. (2019). Penalized empirical likelihood for the sparse Cox regression model. Journal of Statistical Planning and Inference, 201:71-85.

Wang, H. and Zhao, Y. (2009). A comparison of some confidence intervals for the mean quality-adjusted lifetime with censored data. Computational Statistics and Data Analysis, 53:2733-2739.

Wang, L., Chen, J., and Pu, X. (2014). Resampling calibrated adjusted empirical likelihood, Canadian Journal of Statistics, 43: 42-59.

Wang, R., Peng, L., and Qi, Y. (2013). Jackknife empirical likelihood test for equality of two high dimensional means. Statistica Sinica, 23 (2):667-690.

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

Zhao, Meng, and Yang (2016). Jackknife empirical likelihood inference for the mean absolute deviation. Computational Statistics and Data Analysis, 91:92-101.

Zhao, Y., and Qin, G. (2006). Inference for the mean residual life function via empirical likelihood. Communications in Statistics—Theory and Methods, 35:1025-1036.

Zheng, M., and Yu, W. (2013). Empirical likelihood method for multivariate Cox regression. Computational Statistics. 28:1241-1267.

APPENDICES

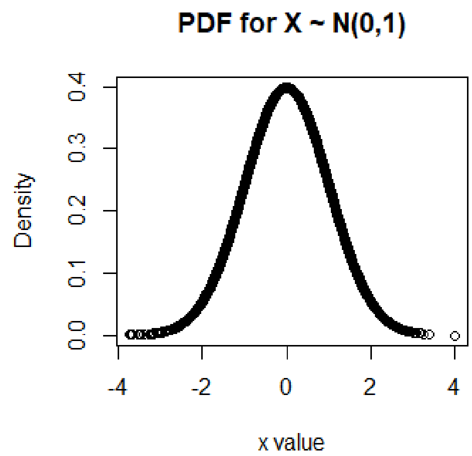

Figure (1) Normal distribution PDF

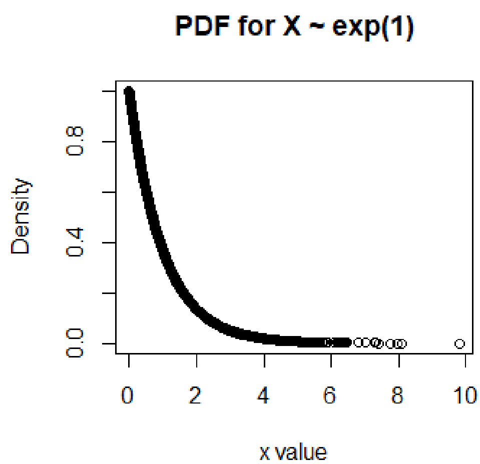

Figure (2) Exponential distribution PDF

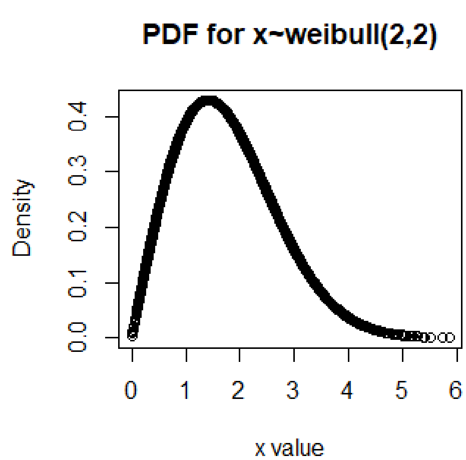

Figure (3) Weibull distribution PDF

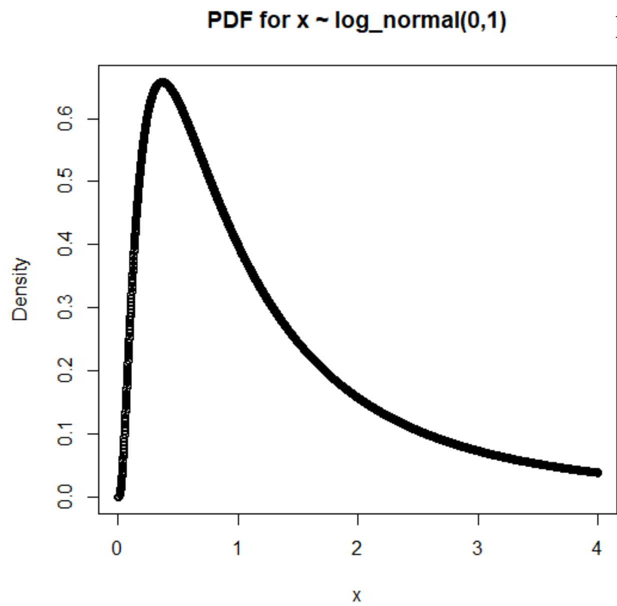

Figure (4) Log-normal distribution PDF

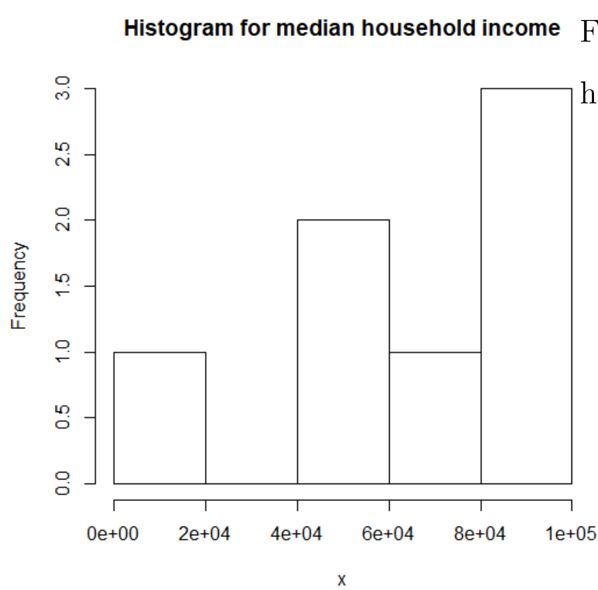

Figure (5) Histogram for the median household income data

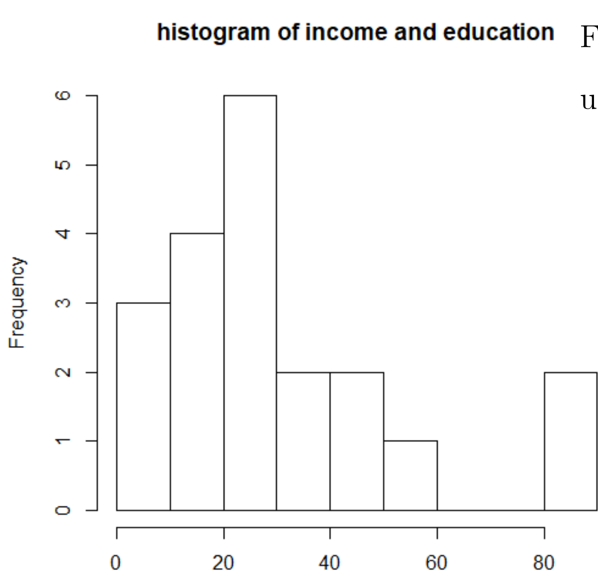

Figure (6) Histogram for the income and education data