Statistics and Probability Letters xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Jackknife empirical likelihood confidence interval for the Gini index

Q1 Dongliang Wang [a], Yichuan Zhao [b,*], Dirk W. Gilmore [c]

[a] Department of Public Health and Preventive Medicine, SUNY Upstate Medical University, 750 East Adams Street, Syracuse, NY 13210, United States
[b] Department of Mathematics and Statistics, Georgia State University, 30 Pryor Street, Atlanta, GA 30303, United States
[c] Lifeline Capital Management, LLC, 1979 Lakeside Parkway, Atlanta, GA 30084, United States

ARTICLE INFO

Article history:
Received 2 June 2015
Received in revised form 29 September 2015
Accepted 30 September 2015
Available online xxxx

Keywords:
Confidence interval
Lorenz curve
Total time on test
Wilks' theorem

ABSTRACT

Jackknife empirical likelihood for the Gini index is derived. Adjusted jackknife empirical likelihood and bootstrap calibration are further investigated. The resulting interval estimators are comparable to existing empirical likelihood methods in terms of coverage accuracy, but yield much shorter intervals.

1. Introduction

The Gini index, originated from the Gini mean difference, has been widely recognized as one of the most commonly used measures for assessing distributional inequality. The Gini index can also be expressed as the area between the 45-degree line and the Lorenz curve. The Lorenz curve, proposed by Lorenz (1905), is another commonly used measure of distributional inequality and the 45-degree line represents perfect equality.

A variety of statistical methodologies have been developed for interval estimation of the Gini index with particular focus on obtaining reliable standard error estimator with no distributional assumptions. Most recently, Davidson (2009) proposed an asymptotic approximation of the plug-in estimator of the Gini index as a sum of i.i.d. variables and derived a reliable variance estimator via delta method. The feasibility of estimating the standard error via ordinary least square regression was discussed by Ogwang (2000), Giles (2004), Modarres and Gastwirth (2006) and Giles (2006). A range of commonly used nonparametric methods have also been exploited for estimating the standard error of the Gini index estimator, such Q2 as the asymptotic theory of U-statistics by Hoeffding (1948), the jackknife method by Yitzhaki (1991) and Karagiannis and Kovacevic (2000) and the bootstrap method by Mills and Zandvakili (1997) and Biewen (2002). In addition, parametric methods for estimating the sampling distribution of the Gini index estimator can be found in Moothathu (1985, 1990), Q3 Chotikapanich and Griffiths (2002), Abdul-Sathar et al. (2005) and Wang et al. (2015). Nonparametric variance estimation of the Gini index under more complicated sampling schemes has been discussed in Sandström et al. (1988) and Bhattacharya Q4 (2007).

* Corresponding author.
E-mail address: yichuan@gsu.edu (Y. Zhao).

http://dx.doi.org/10.1016/j.spl.2015.09.026
0167-7152/

Please cite this article in press as: Wang, D., et al., Jackknife empirical likelihood confidence interval for the Gini index. Statistics and Probability Letters (2015), http://dx.doi.org/10.1016/j.spl.2015.09.026

Recently empirical likelihood methods, formally proposed by Owen (1988, 2001), were developed for interval estimation of the Gini index. The empirical likelihood ratio statistic, proposed by Qin et al. (2010), followed an asymptotic scaled asymptotic chi-square distribution, where the scale needed to be estimated in practice. Peng (2011) derived the empirical likelihood over transformed data. The transformation essentially cuts the sample size in half and the resulting interval estimator seemed to be unnecessarily wide.

The objective of this paper is to develop the jackknife empirical likelihood (JEL) methods for the Gini index. The rest of the paper is organized as follows. In Section 2 new JEL confidence intervals for the Gini index are proposed. In Section 3 a simulation study is carried out to compare the performance of our JEL methods with other existing methods. The proposed JEL methods are illustrated with a real data set in Section 4, followed by conclusions and discussions in Section 5.

2. Results

Let $F(\cdot) = P(X \le x)$ be the cumulative distribution function of a non-negative random variable X. The Gini index G is formally defined as

$$G = \frac{1}{\mu}\int_0^\infty (2F(x) - 1)xdF(x) = \frac{\mathrm{E}|X - Y|}{2\mathrm{E}X},$$

where X and Y are two independent random variables following the same distribution $F(x)$ and $\mu = \mathrm{E}X = \int_0^\infty xdF(x)$.

Let $\boldsymbol{X} = [X_1, X_2, \ldots, X_n]$, $n \ge 2$ be i.i.d. random variables with common distribution function $F(x)$, and $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$ be the corresponding order statistics. A commonly used nonparametric estimator for G is derived as

$$G_n = \frac{1}{\bar{X}}\int_0^\infty xd\hat{F}^2(x) - 1 = \frac{2}{n^2\bar{X}}\sum_{i=1}^n X_{i:n}\left(i - \frac{1}{2}\right) - 1, \tag{2.1}$$

by plugging in the empirical distribution function $\hat{F}(x) = 1/n\sum_{i=1}^n I(X_i \le x)$, where $\bar{X}$ is the sample mean. Alternatively, the Gini index can be estimated by

$$G_U = (2\bar{X})^{-1}\binom{n}{2}^{-1}\sum_{1\le i_1<i_2\le n}|X_i - X_j|, \tag{2.2}$$

which is a ratio of two U-statistics with the kernels $h(x, y) = |x-y|$ and $h(x) = x$, respectively. To our best knowledge, none of the existing jackknife empirical likelihood methods, including the JEL method proposed by Jing et al. (2009), can handle two U-statistics with kernels of different degrees.

2.1. Jackknife empirical likelihood

We start this section by defining a new estimating equation for G, namely,

$$U_n(\tilde{G}) = \binom{n}{2}^{-1}\sum_{1\le i_1<i_2\le n} h(X_{i_1}, X_{i_2}; \tilde{G}) = 0,$$

where $h(X_{i_1}, X_{i_2}; \tilde{G}) = (X_{i_1} + X_{i_2})\tilde{G} - |X_{i_1} - X_{i_2}|$. At the true value G of the Gini index, $U_n(G)$ is in the form of a U-statistic with a symmetric kernel h of degree $m = 2$ and it is clear that $\mathrm{E}h(X_{i_1}, X_{i_2}; G) = 0$. The definition of $U_n(G)$ allows us to readily establish the Wilks' theorem of the jackknife empirical likelihood under the framework developed by Jing et al. (2009). Towards this end, the jackknife pseudo-values can be defined as

$$\hat{V}_i(G) = nU_n(G) - (n-1)U_{n-1}^{(-i)}(G), \tag{2.3}$$

where $U_{n-1}^{(-i)}(G)$ is the U-statistic with the ith observation X_i deleted. As indicated by Arvesen (1969), the U-statistic $U_n(G)$ can thus be rewritten as an average of the jackknife pseudo-values

$$U_n(G) = \frac{1}{n}\sum_{i=1}^n \hat{V}_i(G).$$

We can then follow the jackknife empirical likelihood method in Jing et al. (2009). Let $\boldsymbol{\pi}' = (\pi_1, \ldots, \pi_n)$ be the probability vector over the jackknife pseudo-values and $H(u) = 1/n\sum_{i=1}^n \pi_i I\{\hat{V}_i(G) \le u\}$ be the corresponding empirical distribution function, where $I(\cdot)$ is the indicator function. The jackknife empirical likelihood ratio at G can be expressed as

$$R(G) = \max\left\{\prod_{i=1}^n n\pi_i : \pi_i \ge 0, \sum_{i=1}^n \pi_i = 1, \sum_{i=1}^n \pi_i\hat{V}_i(G) = 0\right\}.$$