Reduce the computation in jackknife empirical likelihood for comparing two Gini indices

Kangni Alemdjrodo & Yichuan Zhao

Georgia State University

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

1. Objectives

- The Gini index is widely used as a measure of income inequality or wealth inequality in social sciences.
- Methods used to compare two Gini indices are very expensive.
- We propose a new jackknife empirical likelihood (JEL) method to reduce the computation cost.

2. Introduction

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

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

where X and Y are two independent random variables following the same distribution $F(x)$ and $\mu = EX$.

Figure 1: *The Gini index.*

3. Methods

Given i.i.d. data set $\boldsymbol{X} = \{X_1, X_2, ..., X_n\}$, $n \ge 2$, the Gini index defined by (1) can be estimated by the ratio of two U-statistics with kernels $h_1(x,y) = |x-y|$ and $h_2(x) = x$, that is,

$$\widehat{G} = \frac{U_1}{U_2} = \frac{\binom{n}{2}^{-1}\sum_{1\le i<j\le n} h_1(X_i,X_j)}{2n^{-1}\sum_{1\le i\le n} h_2(X_i)}. \tag{2}$$

3.1 Profile Jackknife empirical likelihood

To derive the JEL confidence intervals for the difference of two Gini indices for paired samples, Wang and Zhao (2016), used the following setting:

Let $\{\mathbf{X},\mathbf{Y}\}' = \{(X_1,Y_1)', ..., (X_n,Y_n)'\}$ be i.i.d. bivariate random variables with common distribution function $F(x,y)$. Let $F_1(x) = F(x,\infty)$ and $F_2(y) = F(\infty,y)$ be the marginal distributions for X and Y, and G_1 and G_2 be the corresponding Gini indices associated with $F_1(.)$ and $F_2(.)$, respectively. Let $\Delta = G_1 - G_2$. Define

$$U_n^X(G_1) = \binom{n}{2}^{-1}\sum_{1\le i<j\le n} h(X_i,X_j;G_1), \tag{3}$$

$$U_n^Y(G_2) = \binom{n}{2}^{-1}\sum_{1\le i<j\le n} h(Y_i,Y_j;G_2), \tag{4}$$

where $h(X_i,X_j;G) = (X_i+X_j)G - |X_i - X_j|$

Considering Δ as the parameter of interest and G_2 as a nuisance parameter and using a vector-type jackknife pseudo-values, they developed a profile JEL ratio at the true Δ, in which the estimated nuisance parameter minimizes the JEL ratio when Δ is fixed. However, the computation of the profile JEL could be very costly.

3.2 The proposed Jackknife empirical likelihood

Since equation $U_n^Y(G_2) = 0$ does not depend on Δ, it can be used to find an estimate $\widehat{G}_2$ of G_2. Solving the above equation we obtain a closed form of $\widehat{G}_2$:

$$\widehat{G}_2 = \frac{\binom{n}{2}^{-1}\sum_{1\le i<j\le n}|Y_i - Y_j|}{2\bar{Y}}, \tag{5}$$

where $\bar{Y}$ is the sample mean. Then plugging in $\widehat{G}_2$ in equation $U_n^X(G_1) = 0$, we can apply the JEL method for Δ. Let

$$M_n(\Delta) = \binom{n}{2}^{-1}\sum_{1\le i<j\le n} h(X_i,X_j;\Delta + \widehat{G}_2). \tag{6}$$

To apply JEL as defined by Jing et al. (2009), we define the jackknife pseudo-values as

$$\widehat{V}_i(\Delta) = nM_n(\Delta) - (n-1)M_n^{(-i)}(\Delta), \tag{7}$$

where $M_n^{(-i)}(\Delta)$ is the U-statistic obtained after deleting the i^{th} observation X_i from the sample. The jackknife estimator $M_{n,jack}$ can be viewed as a sample average of approximately independent random variables $\widehat{V}_i$. Then we can apply the standard EL method (see Owen, 2001) to $\widehat{V}_i$, the JEL ratio at Δ can be expressed as

$$R(\Delta) = sup\left\{\prod_{i=1}^n np_i :\ p_i \ge 0,\ \sum_{i=1}^n p_i = 1,\ \sum_{i=1}^n p_i\widehat{V}_i(\Delta) = 0\right\}.$$

Applying the Lagrange multiplier technique, we obtain the log-likelihood

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

Let $g(x;\Delta,G_2) = Eh(x,X_2;\Delta + G_2)$ and $\sigma_g^2(\Delta,G_2) = Var(g(X_1;\Delta,G_2))$. We have the Wilks theorem for the JEL as follows :

Theorem 1: If $EX_1^2 < \infty$ and $\sigma_g^2(\Delta,G_2) > 0$, then

$$-2\log R(\Delta) \xrightarrow{D} \chi_1^2, \quad as\ n \to \infty, \tag{8}$$

where χ_1^2 is a standard chi-squared random variable with one degree of freedom.

Following this theorem, an asymptotic $100\,(1-\alpha)\,\%$ confidence interval for Δ is given by

$$\{\tilde{\Delta} : -2\log R(\tilde{\Delta}) \le \chi_{1-\alpha}^2(1)\},$$

where $\chi_{1-\alpha}^2(1)$ is the $100\,(1-\alpha) -$ percentile of the chi-square distribution with one degree of freedom.

4. Results

4.1 Simulation Study

We compare the performance of the proposed JEL (JEL-AZ) method with the JEL method in Wang and Zhao (2016) (JEL-WZ) through simulations. We also investigate the adjusted jackknife empirical likelihood (AJEL) and the bootstrap-calibration (JEL-Boot) to improve coverage accuracy for small samples.

Table 1: Comparison of coverage probabilities and average lengths of the confidence intervals for different jackknife empirical likelihood methods; $X,Y \sim U(0,1)$.

	$n_1=n_2=50$		$n_1=n_2=100$		$n_1=n_2=150$	
	$1-\alpha=0.90$	$1-\alpha=0.95$	$1-\alpha=0.90$	$1-\alpha=0.95$	$1-\alpha=0.90$	$1-\alpha=0.95$
JEL-WZ	0.884 (0.084)	0.940 (0.108)	0.889 (0.054)	0.942 (0.070)	0.897 (0.044)	0.948 (0.057)
JEL-AZ	0.903 (0.114)	0.952 (0.136)	0.895 (0.072)	0.939 (0.093)	0.896 (0.057)	0.948 (0.073)
AJEL-WZ	0.896 (0.089)	0.950 (0.116)	0.898 (0.061)	0.946 (0.080)	0.902 (0.049)	0.951 (0.061)
AJEL-AZ	0.928 (0.134)	0.966 (0.162)	0.929 (0.084)	0.966 (0.106)	0.922 (0.064)	0.960 (0.081)
JEL-WZ-Boot	0.914 (0.086)	0.957 (0.109)	0.893 (0.055)	0.951 (0.071)	0.895 (0.043)	0.949 (0.059)
JEL-AZ-Boot	0.905 (0.115)	0.953 (0.137)	0.899 (0.073)	0.943 (0.094)	0.898 (0.060)	0.950 (0.075)

Table 2: Comparison of computing times (in seconds) for different jackknife empirical likelihood methods; $X,Y \sim U(0,1)$.

	$n_1=n_2=50$		$n_1=n_2=100$		$n_1=n_2=150$	
	$1-\alpha=0.90$	$1-\alpha=0.95$	$1-\alpha=0.90$	$1-\alpha=0.95$	$1-\alpha=0.90$	$1-\alpha=0.95$
JEL-WZ	5294	6239	5998	6219	5513	6236
JEL-AZ	344	428	463	502	437	532
AJEL-WZ	3835	4991	4835	5603	4675	6114
AJEL-AZ	357	453	511	547	456	612
JEL-WZ-Boot	31460	31751	73171	67799	254544	220765
JEL-AZ-Boot	6131	7033	20735	19471	46992	44879

Based on the table 1 and 2 make the following conclusions:

- All the coverage probabilities tend to their nominal levels as the sample size increases.
- JEL-AZ methods have better coverage than JEL-WZ ones
- The running times of JEL-WZ simulations are 5 to 20 times these of JEL-AZ's.

4.2 Real Application

We apply the proposed methods to estimate the real GDP (Gross Domestic Product) per capita in constant dollars expressed in international prices, base year 1985 for the years 1970 and 1990. These data sets are extracted from the Penn World Tables data (Summers & Heston (1995)). In this case, the Gini index measures the dispersion of real GDP across the 108 countries for which data are available.

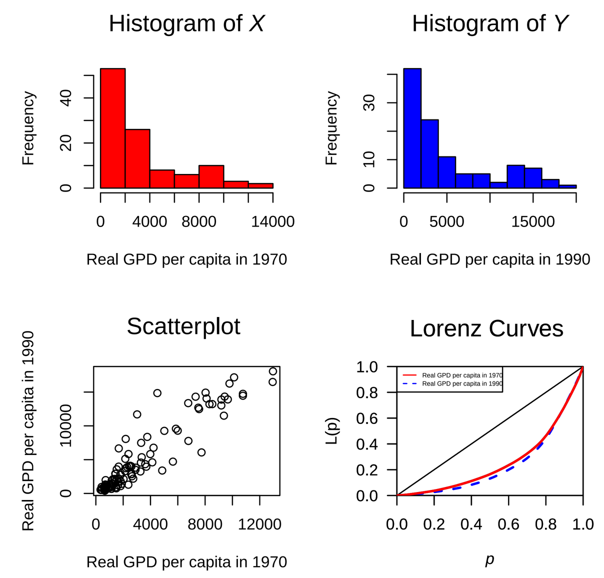

Figure 2: *Exploratory plots for the data set.*

Figure 1 shows that the data at years 1970 and 1990:

- Seem to be exponentially distributed: their estimated Gini indexes (0.488 and 0.525) are very close to 0.5, the Gini index from exponential distribution.
- Are highly correlated: the correlation coefficient is 0.931.

Table 3: Point and interval estimates for the parameters of interest from the data set.

	Point estimate				Interval estimate	
	G_1 at 1970	G_2 at 1990	$G_1 - G_2$		$1-\alpha=0.90$	$1-\alpha=0.95$
Plug-in	0.484	0.520	-0.037	JEL-WZ	$(-0.060,-0.011)$	$(-0.056,-0.015)$
U-statistic	0.488	0.525	-0.037	JEL-AZ	$(-0.071,-0.001)$	$(-0.066,-0.007)$
				AJEL-WZ	$(-0.059,-0.010)$	$(-0.055,-0.015)$
				AJEL-AZ	$(-0.110,-0.010)$	$(-0.098,-0.017)$

Table 3 confirms that JEL-AZ methods confidence intervals for the difference between the Gini indices at years 1970 and 1990 are wider than these of JEL-WZ methods but in both cases the confidence intervals are small enough to conclude that there is a significant difference between the two indices ($G_1 < G_2$).

5. Conclusions

- JEL-AZ methods require less computation than JEL-WZ ones.
- They have better coverage probability than JEL-WZ methods.
- Can be applied to a wide range of paired data sets.

References

1. Wang, D., Zhao, Y. (2016). *Jackknife empirical likelihood for comparing two Gini indices.* The Canadian Journal of Statistics, 44(1):102-119.

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

3. Owen, A. (2001). *Empirical Likelihood.* Chapman and Hall, London.

Contact: `yichuan@gsu.edu; kalemdjrodo1@student.gsu.edu`