Statistics and Probability Letters 145 (2019) 50–56

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Large sample properties of a new measure of income inequality

Xiaoke Zhang *, Joseph L. Gastwirth

Department of Statistics, George Washington University, Washington, DC 20052, USA

ARTICLE INFO

Article history:
Received 14 June 2018
Received in revised form 8 August 2018
Accepted 19 August 2018
Available online 1 September 2018

MSC:
62-G20
62-E20

Keywords:
Functional delta method
Hadamard derivative
Lorenz curve

ABSTRACT

This paper develops the asymptotic properties of a recently proposed measure of inequality emphasizing the status of the poor relative to the rich. A simulation confirms the asymptotic results are applicable in practice.

1. Introduction

The increasing trend in income inequality has become a major issue in public policy. When there is a shift in favor of the upper end, Gastwirth (2014) noted that the increase in the standard measure, the Gini index, does not fully reflect it. He suggested the median replace the mean in its denominator. It is also useful to consider transforms of the Lorenz curve that focus on the status of low-income recipients relative to the high-income ones. In the recent years, new measures for inequality focusing on the poor have been proposed (e.g., Gastwirth, 2016; Prendergast and Staudte, 2017; Davydov and Greselin, 2018; Jasso, 2018). This paper derives the asymptotic properties of a curve and a related index proposed by Gastwirth (2016).

For any value of $q \in [0, 1]$, consider the fraction, $p^*(q)$ of income recipients cumulated from the poorest, needed to have the same amount of income as the top $100q$%. To illustrate the increase in inequality in the United States, Gastwirth (2016) noted that in 1967, the top 5% of income recipients had 17.2% of the total income and the lowest 43.3% had the same amount of income. By 2013, the share of the top 5% reached 22.2% and one needs the lowest 55.5% of households in order that they have the same total income. The analog of the Gini index, twice the area between $p^*(q)$ and the line of equality or AGP, increased from .687 to .788 during the period.

This paper derives the asymptotic properties of the empirical estimators of p^* curve and AGP for independent and identically distributed (i.i.d.) data. The main theoretical result, i.e., Theorem 3.2, shows that the estimated curve of p^* weakly converges to a Gaussian process. A corollary of this result is the asymptotic normality of the estimator of AGP. A simulation study indicates that the asymptotic properties are applicable in samples of 1000, which is smaller than the typical sample sizes (at least 5000) of income surveys.

* Corresponding author.
E-mail address: xkzhang@email.gwu.edu (X. Zhang).

https://doi.org/10.1016/j.spl.2018.08.009
0167-7152/

Due to the fact that the p^* curve is a transform of the Lorenz curve, the functional delta method in terms of the Hadamard derivative of this transform (van der Vaart and Wellner, 1996) will be used to derive the weak convergence of the p^* curve to a Gaussian process. The approach connects the weak convergence for p^* to that for the Lorenz curve, which has been intensively studied (e.g., Goldie, 1977; Gail and Gastwirth, 1978; Beach and Davidson, 1983; Csörgö et al., 1986; Csörgő et al., 1998; Csörgő and Yu, 1999; Zheng, 2002).

The rest of the paper proceeds as follows. The basic definitions and background information are given in Section 2. The theoretical results are presented in Section 3 and Section 4 is devoted to the simulation study.

2. Methodology

Let X be a continuous random variable defined on $[a, b]$ where $0 < a < b$, that represents the income of a population. Denote its cumulative distribution function (cdf) by F and corresponding probability density function (pdf) by f. Assume that $0 < m \leq f(x) \leq M < \infty$ for all $x \in [a, b]$. Therefore, F is continuous and strictly increasing, and the quantile function F^{-1} is also continuous, where $F^{-1}(p) = \inf\{x \in [a, b] : F(x) \geq p\}, p \in [0, 1]$.

From Gastwirth (1971), the Lorenz curve, denoted by L, is defined by

$$L(p) = \frac{1}{\mu} \int_0^p F^{-1}(t)\, dt, \quad p \in [0, 1], \tag{1}$$

where $\mu = EX$. For each $p \in [0, 1]$, $L(p)$ represents the share of all income received by the poorest $100p\%$. From the assumptions on f and F, L is continuous and strictly increasing on $[0, 1]$, and thus L^{-1} is continuous, where $L^{-1}(p) = \inf\{t \in [0, 1] : L(t) \geq p\}, p \in [0, 1]$.

To quantify economic inequality focusing on low-income groups, Gastwirth (2016) proposed the following measure:

$$p^*(q) = L^{-1}\left(1 - L\left(1 - q\right)\right), \quad q \in [0, 1]. \tag{2}$$

For each $q \in [0, 1]$, $p^*(q)$ is the fraction of income recipients cumulated from the poorest whose total income equals the total income of the top $100q\%$.

The Gini index, the ratio of the mean difference to twice the mean, also equals twice the area between the Lorenz curve and the line of equality, i.e., the identity function $D(q) = q, q \in [0, 1]$. Analogously we define a new index, denoted by AGP, which is twice the area between the p^* curve and the line of equality:

$$\text{AGP} = 2\int_0^1 \left\{p^*(q) - q\right\}\, dq = 2\int_0^1 p^*(q)\, dq - 1. \tag{3}$$

Based on an i.i.d. sample $\{X_i : i = 1, \ldots, n\}$, the natural estimator for p^*, denoted by p_n^*, is as follows:

$$p_n^*(q) = L_n^{-1}\left(1 - L_n\left(1 - q\right)\right), \quad q \in [0, 1].$$

Here L_n is an estimator of L defined by

$$L_n(p) = ([np] + 1 - np)\,\frac{\sum_{i=0}^{[np]} X_{(i)}}{\sum_{j=0}^{n} X_{(j)}} + (np - [np])\,\frac{\sum_{i=0}^{[np]+1} X_{(i)}}{\sum_{j=0}^{n} X_{(j)}}, \quad p \in [0, 1],$$

where $[np]$ is the integer part of np, $X_{(1)} \leq \cdots \leq X_{(n)}$ are the order statistics of $\{X_i : i = 1, \ldots, n\}$, and $X_{(0)} = X_{(n+1)} = 0$. Geometrically L_n is the line obtained by linear interpolation between the points $\{\sum_{i=0}^{j} X_{(i)}/\sum_{i=0}^{n} X_{(i)} : j = 0, 1, \ldots, n\}$. Define $L_n^{-1}(p) = \inf\{t \in [0, 1] : L_n(t) \geq p\}, p \in [0, 1]$.

We adopt the same L_n as in Gastwirth (1972) and Goldie (1977) for convenience so that L_n and L_n^{-1} are both continuous and strictly increasing, although discontinuous estimators for L and L^{-1} are also available (e.g., Taguchi, 1968; Gail and Gastwirth, 1978; Csörgö et al., 1986; Csörgő and Yu, 1999).

Similar to (3), the natural estimator for AGP is defined by

$$\text{AGP}_n = 2\int_0^1 p_n^*(q)\, dq - 1.$$

3. Theoretical results

For scalars a and b, denote $a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$. For any function g defined on the interval $[0, 1]$, the norm $\|g\|_\infty = \sup_{t\in[0,1]}|g(t)|$. The notation "$\rightsquigarrow$" represents weak convergence.

Denote the set of all real numbers by $\mathbb{R}$. Let $\mathbb{C}[0, 1]$ be the space of all continuous functions on $[0, 1]$ with the norm $\|\cdot\|_\infty$ and define $\mathbb{G}[0, 1] = \{g \in \mathbb{C}[0, 1] : g(0) = 0, g(1) = 1, and g is strictly increasing\}$. For any $g \in \mathbb{G}[0, 1]$, define $g^{-1}(p) = \inf\{x \in [0, 1] : g(x) \geq p\}, p \in [0, 1]$.

We first develop the strong consistency of both p_n^* and AGP_n.