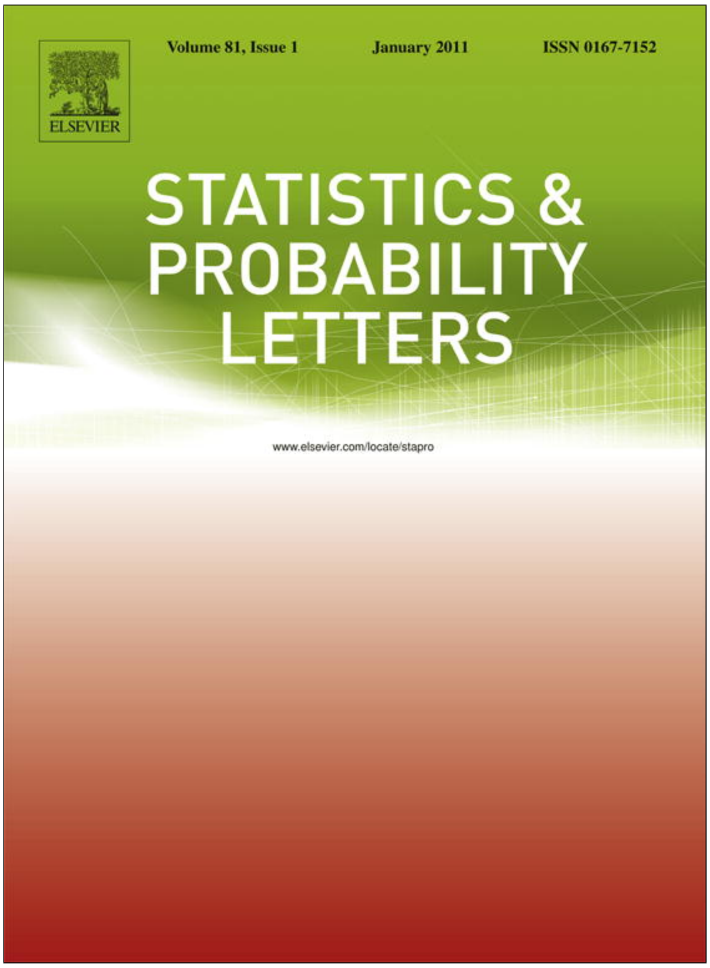
Volume 81, Issue 1
January 2011
ISSN 0167-7152
ELSEVIER
STATISTICS &
PROBABILITY
LETTERS
www.elsevier.com/locate/stapro

Statistics and Probability Letters 81 (2011) 1425–1435

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Asymptotic behaviors of the Lorenz curve and Gini index in sampling from a length-biased distribution

V. Fakoor *, M. Bolbolian Ghalibaf, H.A. Azarnoosh

Department of Statistics, School of Mathematical Sciences, Ferdowsi University of Mashhad, P. O. Box: 1159-91775, Iran

ARTICLE INFO

Article history:
Received 14 December 2010
Received in revised form 15 April 2011
Accepted 18 April 2011
Available online 24 April 2011

Keywords:
Gini index
Length-biased distribution
Lorenz curve
Strong consistency
Strong Gaussian approximation

ABSTRACT

In this work, we consider the nonparametric estimators of the Lorenz curve and Gini index based on a sample from the corresponding length-biased distribution. We show that this estimators are strongly consistent for the associated Lorenz curve and Gini index. Strong Gaussian approximations for the associated Lorenz process are established under appropriate assumptions. We apply the strong Gaussian approximation technique to obtain a functional law for the iterated logarithm for the Lorenz curve. Also, we obtain an asymptotic normality for the corresponding Gini index.

1. Introduction

Pietra (1915) and Gastwirth (1971) independently introduced the *Lorenz curve* corresponding to a non-negative random variable (r.v.) X with a distribution function (d.f.) F, quantile function $Q(p)$ and finite mean $EX = \mu$ as

$$L_F(p) := \mu^{-1} \int_0^p Q(s)\mathrm{d}s, \quad 0 \le p \le 1.$$

In econometrics, with X representing income, $L(p)$ gives the fraction of total income that the holders of the lowest pth fraction of income possess. Most of the measures of income inequality are derived from the Lorenz curve. An important example is the Gini index associated with F defined by

$$G_F := \frac{\int_0^1 [u - L_F(u)]\mathrm{d}u}{\int_0^1 u\mathrm{d}u} = 1 - 2(\mathrm{CL})_F,$$

where $(\mathrm{CL})_F = \int_0^1 L_F(u)\mathrm{d}u$ is the *average Lorenz index* corresponding to F. The Gini index is a ratio of the area between the Lorenz curve and the 45° line to the area under the 45° line. The numerator is usually called the *area of concentration*. Kendall and Stuart (1963) showed that this is equivalent to a ratio of a measure of dispersion to the mean. In general, these notions are useful for measuring concentration and inequality in distributions of resources, and in size distributions. For a list of applications in different areas, we refer the readers to Csörgő and Zitikis (1996a).

To estimate the Lorenz curve and Gini index, one can use the Lorenz statistic $L_n(p)$ and Gini statistic G_n defined by

$$L_n(p) := \mu_n^{-1} \int_0^p Q_n(u)\mathrm{d}u, \quad 0 \le p \le 1,$$

* Corresponding author.
E-mail addresses: fakoor@math.um.ac.ir, vfbafandeh@yahoo.com (V. Fakoor), m_bolbolian@yahoo.com (M.B. Ghalibaf).

0167-7152/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2011.04.013

and

$$G_n := \frac{\int_0^1 [u - L_n(u)]\mathrm{d}u}{\int_0^1 u\mathrm{d}u} = 1 - 2(\mathrm{CL})_n = 1 - 2\int_0^1 L_n(u)\mathrm{d}u,$$

where μ_n is the sample mean and $Q_n(y)$ is the empirical quantile function constructed from an independent and identically distributed (i.i.d.) sample taken from F.

Goldie (1977) proved the uniform consistency of L_n with L_F and derived the weak convergence of the *Lorenz process* $l_n(t) := \sqrt{n}[L_n(t) - L(t)],\ 0 \le t \le 1$, to a Gaussian process under suitable conditions. Csörgő et al. (1986) gave a unified treatment of strong and weak approximations of the Lorenz and other related processes. In particular, they established a strong invariance principle for the Lorenz process, by which Rao and Zhao (1995) derived one of their two versions of the law of the iterated logarithm (LIL) for the Lorenz process. Different versions of the LIL under weaker assumptions are also obtained by Csörgő and Zitikis (1996a, 1997). In Csörgő and Zitikis (1996b), confidence bands for the Lorenz curve that are based on weighted approximations of the Lorenz process are constructed. Csörgő et al. (1987) obtained weak approximations for Lorenz curves under random right censorship. Strong Gaussian approximations for the Lorenz process when data are subject to random right censorship and left truncation are established by Tse (2006), who also derived a functional LIL for the Lorenz process.

Csörgő and Yu (1999) obtained weak approximations for Lorenz curves and their inverses under the assumption of mixing dependence. Glivenko–Cantelli-type asymptotic behavior of the empirical generalized Lorenz curves based on random variables forming a stationary ergodic sequence with deterministic noise were considered by Davydov and Zitikis (2002). Davydov and Zitikis (2003) established a large sample asymptotic theory for the empirical generalized Lorenz curves for when observations are stationary and either short-range or long-range dependent. Strong laws for the generalized absolute Lorenz curves when data are stationary and ergodic sequences were established by Helmers and Zitikis (2005). On the basis of the generalized Lorenz curves, Davydov et al. (2007) proposed a statistical index for measuring the fluctuations of a stochastic process. They developed some of the asymptotic theory of the statistical index for the case where the stochastic process is a Gaussian process with stationary increments and a nicely behaved correlation function. The uniform strong convergence rate of the Lorenz curve estimator under the strong mixing hypothesis is obtained by Fakoor and Nakhaei Rad (in press). They also established a strong Gaussian approximation for the Lorenz process, from which they derived a functional LIL for the Lorenz process, under the assumption of strong mixing. The counterpart of these results for the censored dependent model and truncated dependent model was established by Bolbolian Ghalibaf et al. (2010, 2011).

In this article, we discuss the nonparametric estimator of a Lorenz curve and Gini index from *length-biased* data; to be precise, let $Y_1, Y_2, \dots, Y_n$ be n i.i.d. non-negative random variables from a distribution F^*, defined on $\mathcal{R}^+ = [0, \infty)$. F^* is called a length-biased distribution corresponding to a distribution F (also defined on $\mathcal{R}^+$) if

$$F^*(t) = \mu^{-1}\int_0^t x\mathrm{d}F(x), \quad t \ge 0,$$

where $\mu := \int_0^\infty x\mathrm{d}F(x)$ is assumed to be finite, and hence the density of Y is

$$f^*(t) = \mu^{-1} t f(t), \quad t \ge 0. \tag{1.1}$$

The phenomenon of length bias was first tackled in the context of anatomy by Wicksell (1925) as what he called the corpuscle problem. Length bias was later systematically studied by McFadden (1962) and Blumenthal (1967), then by Cox (1969), in the context of estimation of the distribution of fiber lengths in a fabric.

Length-biased data arise in many practical situations, including econometrics, survival analysis, renewal processes, biomedicine and physics. For instance, if X represents the length of an item and the probability of this item being selected in the sample is proportional to its length, then the distribution of the observed length is length biased. In cross-sectional studies in survival analysis, for example, often the probability of being selected, for a particular subject, is proportional to his/her survival time. Interesting applications of length-biased data can be found in Cox (1969), Patil and Rao (1977, 1978), Colman (1979) and Vardi (1982b). The distribution function, F^*, is, from a slightly different perspective, the distribution of the randomly left truncated r.v.'s Y, in the stationary assumption. If the incidence rate of the event has not changed over time, a stationary form might reasonably describe the incidence of the event; this is equivalent to assuming that the randomly left truncation induced by the sampling is uniform (Wang, 1991).

Throughout this work we assume that F^* is continuous on $\mathcal{R}^+$, from which it follows that F is also continuous. An elementary calculation shows that F is determined uniquely by F^*, namely

$$F(t) = \nu^{-1}\int_0^t y^{-1}\mathrm{d}F^*(y), \quad t \ge 0,$$

where

$$\nu = EY^{-1} = \int_0^\infty y^{-1}\mathrm{d}F^*(y).$$

Cox (1969) and Vardi (1982a) considered the problem of finding a nonparametric maximum likelihood estimate (NPMLE) of F on the basis of a sample $Y_1, Y_2, \ldots, Y_n$ from F^*. The empirical estimator of F can be written in the form

$$F_n(t) = \nu_n^{-1} \int_0^t y^{-1} \mathrm{d}F_n^*(y),$$

where

$$\nu_n = \int_0^\infty y^{-1} \mathrm{d}F_n^*(y),$$

and F_n^*, the empirical estimator of F^*, is given by

$$F_n^*(t) = \frac{1}{n} \sum_{i=1}^n I(Y_i \le t),$$

where $I(A)$ denotes the indicator of the event A.

Suppose that the empirical process is given by

$$\alpha_n(t) = n^{1/2}[F_n(t) - F(t)], \quad t \ge 0.$$

Vardi (1982a) studied the weak convergence of the process α_{m+n} assuming that $\lim_{m,n\to\infty} m(m+n)^{-1} > 0$, where F_{m+n} is based on two independent samples, one a sample of size n from F^* and the other a sample of size m from F. Sen (1984) proved the weak convergence of the process α_n to a Gaussian process assuming that $EY^{-2} < \infty$. Actually, under the more stringent regularity condition that $EY^{-2-\delta} < \infty$, for some $\delta > 0$, he obtained the Bahadur representation of sample quantiles in length-biased sampling. Horváth (1985) established the weak and the strong approximation of the process α_n from a length-biased distribution.

We assume that the underlying d.f. F admits a unique pth quantile $Q(p)$ defined by

$$Q(p) = \inf\{t \in \mathcal{R}; F(t) \ge p\}, \quad 0 < p < 1.$$

Let $Y_{n:1} \le \cdots \le Y_{n:n}$ be the order statistics corresponding to $Y_1, \ldots, Y_n$. Then, the sample estimator corresponding to $Q(p)$ is defined by $Q_n(p)$ $(=Y_{n:k})$ where k $(=k_n)$ is a suitably chosen (random) integer, depending on all the order statistics and defined by

$$k_n = \max\left\{k; \sum_{i=1}^k Y_{n:i}^{-1} \le p\left(\sum_{i=1}^n Y_{n:i}^{-1}\right)\right\}, \quad 0 < p < 1.$$

The main aim of this work is to derive a strong Gaussian approximation of the Lorenz process for a sample from the corresponding length-biased distribution. As a result, we obtain the law of the iterated logarithm for the Lorenz curve. We show that estimators of the Lorenz curve and Gini index are strongly consistent. Also, we obtain asymptotic normality for the corresponding Gini index.

Now we introduce some assumptions that are used to state our results gathered below for easy reference.

Assumptions:

(1) The d.f. F has a continuous probability density function (p.d.f.) f in some neighbourhood of $Q(p)$ and $f(Q(p))$ is strictly positive and finite for all $0 < p < 1$.
(2) Suppose that $\vartheta_{2+\delta} = E(Y^{-2-\delta}) = y^{-2-\delta}\mathrm{d}F^*(y) = \frac{1}{\mu}\int_0^\infty y^{-1-\delta}\mathrm{d}F(y) < \infty$, for some $\delta > 0$, and $\sup\{|f'(x)|; x \in \mathcal{R}^+\} < \infty$.
(3) $\vartheta(r) = \int_0^\infty (F^*(y))^{1/r} y^{-2} \mathrm{d}y < \infty$, for some $r > 2$.

Remark 1. According to (1.1), and noting that $0 < Q(p) < \infty$, we have

$$0 < \mu f^*(Q(p)) = Q(p) f(Q(p)) < \infty, \quad 0 < p < 1. \tag{1.2}$$

Remark 2. Suppose that $r = 2 + \delta$; then Assumption (3) is slightly stronger than Assumption (2), i.e., $\vartheta_{2+\delta} < \infty$, for some $\delta > 0$.

The layout of the work is as follows. In Section 2, we obtain the strong uniform consistency of $Q_n(.)$ and the strong Gaussian approximation for the normed quantile process $\rho_n(p) := \sqrt{n} f(Q(p))[Q(p) - Q_n(p)]$. This preliminary discussion is necessary for achieving the establishment of the main results. Section 3 contains asymptotic behaviors of the estimator of the Lorenz curve. Strong uniform consistency and the strong Gaussian approximation of the estimator of the Lorenz curve provided and a law of the iterated logarithm for the Lorenz curve are derived. Asymptotic behaviors of the estimator of the Gini index are presented in Section 4. In this section we establish strong consistency and asymptotic normality of the estimator of the Gini index. Some proofs of the main results are deferred to the Appendix.

2. Preliminaries

In the following we need the uniform strong consistency of the estimator Q_n; in the next lemma we show the uniform strong consistency with the rate of the estimator Q_n.

Lemma 1. *Under Assumptions* (1) *and* (2), *we have*

$$\sup_{0<p<1} |Q_n(p) - Q(p)| = O\left(\sqrt{\frac{\log\log n}{n}}\right) \quad a.s.$$

Proof. It follows from the discussion after (4.12) in Sen (1984). □

Strong approximations for α_n will be derived from the well-known approximations of the empirical process

$$\beta_n(t) = n^{1/2}[F_n^*(t) - F^*(t)], \quad t \geq 0.$$

Without loss of generality we can assume that our probability space $(\Omega, \mathcal{A}, P)$ is so rich that the approximations of Komlós et al. (1975) hold, and we have

$$\sup_{t\geq 0} |\beta_n(t) - K(t,n)| = O\left(n^{-\frac{1}{2}}(\log n)^2\right) \quad \text{a.s.},$$

and

$$\sup_{t\geq 0} |\beta_n(t) - B_n(t)| = O\left(n^{-\frac{1}{2}}\log n\right) \quad \text{a.s.},$$

where $K(t, n)$ is a two-parameter Gaussian process with zero mean and covariance

$$E[K(x,n)K(y,m)] = (mn)^{-1/2}(m \wedge n)[F^*(x\wedge y) - F^*(x)F^*(y)], \quad (a \wedge b = \min\{a,b\}),$$

and $\{B_n(t), t \geq 0\}$ for $n = 1, 2, \ldots$ is a sequence of mean zero Gaussian process with covariance

$$E[B_n(x)B_n(y)] = F^*(x\wedge y) - F^*(x)F^*(y).$$

Using $K(t, n)$, Horváth (1985) defined the process $\Gamma(t, n)$ for approximation α_n such that

$$\Gamma(t,n) = \nu^{-1}\int_0^t y^{-1}\mathrm{d}K(y,n) - \nu^{-1}F(t)\int_0^\infty y^{-1}\mathrm{d}K(y,n).$$

It is easy to check that $\{\Gamma(t, n), t \geq 0\}$ is a Gaussian process with zero mean and covariance

$$E[\Gamma(x,n)\Gamma(y,m)] = (mn)^{-1/2}(m\wedge n)[\sigma(x\wedge y) - F(x)\sigma(y) - F(y)\sigma(x) + F(x)F(y)\sigma], \tag{2.1}$$

where

$$\sigma(t) = \nu^{-2}\int_0^t y^{-2}\mathrm{d}F^*(y),$$

and

$$\sigma = \lim_{t\to\infty}\sigma(t) = \nu^{-2}\int_0^\infty y^{-2}\mathrm{d}F^*(y).$$

Theorem 1 (*Theorem 4.2. from Horváth, 1985*)**.** *Suppose that Assumption* (3) *is satisfied. On a suitably enlarged probability space, there exist two-parameter mean zero Gaussian processes* $\{\Gamma(t, n), t \geq 0\}$ *with covariance* (2.1) *such that*

$$\sup_{t\geq 0} |\alpha_n(t) - \Gamma(t,n)| = O(n^{-\lambda}) \quad a.s.,$$

for any $0 < \lambda < 1/2 - 1/r$, *for some* $r > 2$. □

Also, by using $\{B_n(t), t \geq 0\}$ we can define the process $\Gamma_n(t)$ for approximation α_n such that

$$\Gamma_n(t) = \nu^{-1}\int_0^t y^{-1}\mathrm{d}B_n(y) - \nu^{-1}F(t)\int_0^\infty y^{-1}\mathrm{d}B_n(y).$$

It is easy to check that $\{\Gamma_n(t), t \geq 0\}$ is a sequence of Gaussian processes with zero mean and covariance

$$E[\Gamma_n(x)\Gamma_n(y)] = \sigma(x\wedge y) - F(x)\sigma(y) - F(y)\sigma(x) + F(x)F(y)\sigma. \tag{2.2}$$

The next theorem is an adaptation of Theorem 1 that would suit our purpose better.

Theorem 2. *Suppose that Assumption* (3) *is satisfied. On a suitably enlarged probability space, there exist a sequence of mean zero Gaussian processes* $\{\Gamma_n(t), t \geq 0\}$ *with covariance* (2.2) *such that*

$$\sup_{t \geq 0} |\alpha_n(t) - \Gamma_n(t)| = O(n^{-\lambda}) \quad a.s.,$$

for any $0 < \lambda < 1/2 - 1/r$*, for some* $r > 2$. □

We now construct two mean zero Gaussian processes that strongly uniformly approximate the normed quantile process $\rho_n(p)$.

Theorem 3. *Suppose that Assumptions* (1)–(3) *are satisfied. On a suitably enlarged probability space, there exist two-parameter mean zero Gaussian processes* $\{\Gamma(t, n), t \geq 0\}$ *with covariance* (2.1) *such that*

$$\sup_{0 \leq p \leq 1} |\rho_n(p) - \Gamma(Q(p), n)| = O\left(\left(n^{-\frac{1}{4}} \log n\right) \vee (n^{-\lambda})\right) \quad a.s., \tag{2.3}$$

for any $0 < \lambda < 1/2 - 1/r$*, for some* $r > 2$*, where* $a \vee b = \max\{a, b\}$*. Moreover, there exist a sequence of mean zero Gaussian processes* $\{\Gamma_n(t), t \geq 0\}$ *with covariance* (2.2) *such that*

$$\sup_{0 \leq p \leq 1} |\rho_n(p) - \Gamma_n(Q(p))| = O\left(\left(n^{-\frac{1}{4}} \log n\right) \vee (n^{-\lambda})\right) \quad a.s.,$$

for any $0 < \lambda < 1/2 - 1/r$*, for some* $r > 2$.

Proof. See the Appendix. □

3. Asymptotic behaviors of the Lorenz statistic

In this section, we obtain strong uniform consistency and the strong Gaussian approximation for $L_n(p)$, and making use of the strong approximation, we shall derive the functional LIL of $L_n(p)$.

3.1. Strong uniform consistency

Theorem 4 proves the strong uniform consistency of L_n. It tells us how fast L_n converges to L_F.

Theorem 4. *Suppose that the conditions of Lemma 1 are satisfied. Then, we have*

$$\sup_{0<p<1} |L_n(p) - L_F(p)| = O\left(\sqrt{\frac{\log \log n}{n}}\right) \quad a.s. \tag{3.1}$$

Proof. An elementary computation shows that

$$L_n(p) - L_F(p) = \frac{1}{\mu_n} \int_0^p [Q_n(y) - Q(y)] \mathrm{d}y - \frac{\mu_n - \mu}{\mu_n} L_F(p), \quad p \in (0, 1).$$

Further, it follows from law of the iterated logarithm that

$$\mu_n - \mu = O\left(\sqrt{\frac{\log \log n}{n}}\right) \quad \text{a.s.} \tag{3.2}$$

Now, by using Lemma 1 and (3.2), we obtain the result. □

3.2. Strong Gaussian approximation

Theorem 5 gives two strong Gaussian approximations for the *Lorenz process* over interval [0, 1].

Theorem 5. *Suppose that the conditions of Theorem 3 are satisfied. On a suitably enlarged probability space, there exist two-parameter mean zero Gaussian processes* $\{\Gamma(t, n),\ t \geq 0\}$ *with covariance* (2.1) *such that*

$$\sup_{0 \leq p \leq 1} \left| l_n(p) - \frac{1}{\mu}\left(\int_0^p \frac{\Gamma(Q(y), n)}{f(Q(y))} \mathrm{d}y - L_F(p) \int_0^1 \frac{\Gamma(Q(y), n)}{f(Q(y))} \mathrm{d}y\right)\right| = O\left(\left(n^{-\frac{1}{4}} \log n\right) \vee (n^{-\lambda})\right) \quad a.s., \tag{3.3}$$

for any $0 < \lambda < 1/2 - 1/r$, *for some* $r > 2$. *Moreover, there exist a sequence of mean zero Gaussian processes* $\{\Gamma_n(t), t \geq 0\}$ *with covariance* (2.2) *such that*

$$\sup_{0\leq p\leq 1}\left|l_n(p) - \frac{1}{\mu}\left(\int_0^p \frac{\Gamma_n(Q(y))}{f(Q(y))}\mathrm{d}y - L_F(p)\int_0^1 \frac{\Gamma_n(Q(y))}{f(Q(y))}\mathrm{d}y\right)\right| = O\left(\left(n^{-\frac{1}{4}}\log n\right) \vee (n^{-\lambda})\right) \quad a.s.,$$

for any $0 < \lambda < 1/2 - 1/r$, *for some* $r > 2$.

Proof. See the Appendix. □

3.3. The functional LIL

The next theorem gives a functional LIL for the Lorenz process. We work on the probability space of Theorem 5. Let $C(0, 1)$ be the space of continuous functions on [0, 1] and $\mathcal{B}$ be Starassen's set of absolutely continuous functions:

$$\mathcal{B} = \left\{g|g : [0, 1] \to \mathcal{R}, g(0) = 0, \int_0^1 (g'(x))^2\mathrm{d}x \leq 1\right\}.$$

Theorem 6. *Suppose that conditions of* Theorem 5 *are satisfied. On a rich enough probability space,* $l_n(.)/\sqrt{2\log\log n}$ *is almost surely relatively compact in* $C(0, 1)$ *with respect to the supremum norm and its set of limit points is*

$$\mathcal{G} = \left\{g_h : g_h(p) = \frac{1}{\mu}\left(\int_0^p \frac{h(y)}{f(Q(y))}\mathrm{d}y - L_F(p)\int_0^1 \frac{h(y)}{f(Q(y))}\mathrm{d}y\right),\ 0 \leq p \leq 1,\ h \in \mathcal{H}\right\},$$

where

$$\mathcal{H} = \left\{h : [0, 1] \to \mathcal{R},\ h(u) = g(\sigma(Q(u))) - F(Q(u))\sqrt{\sigma} : g \in \mathcal{B}\right\}.$$

Proof. Observe that process $A(.) = \Gamma(Q(.), u)$ over [0, 1] for $u \geq 0$ is equal in distribution to the process

$$\left\{u^{-\frac{1}{2}}[W(\sigma(Q(.)), u) - F(Q(.))W(\sigma, u)],\ u \geq 0\right\},$$

over [0, 1], where $W(t, u)$ is a standard two-parameter Wiener process. Hence, for $u = n$ where n are natural numbers, $A(y)/\sqrt{2\log\log n}$ is relatively compact in $C(0, 1)$ and set of limit points is $\mathcal{H}$ from the standard functional LIL for a two-parameter Wiener process (Theorem 1.14.1 in Csörgő and Révész, 1981). (3.3) then gives the desired result. □

4. Asymptotic behaviors of the Gini statistic

In this section, we shall give the strong consistency for the Gini statistic G_n. Also, we obtain asymptotic normality of G_n.

4.1. Strong consistency

In the following theorem, we prove the strong consistency of the Gini statistic G_n. It tells us how fast G_n converges to the Gini index G_F.

Theorem 7. *Suppose the conditions of* Lemma 1 *are satisfied. Then, we have*

$$|G_n - G_F| = O\left(\sqrt{\frac{\log\log n}{n}}\right) \quad a.s. \tag{4.1}$$

Proof. We can write

$$|G_n - G_F| = 2\int_0^1 |L_F(u) - L_n(u)|\mathrm{d}u,$$

and by using (3.1), we obtain the result. □

4.2. Asymptotic normality

In the next theorem we obtain asymptotic normality for the normed Gini sequence $\sqrt{n}[G_n - G_F]$.

Theorem 8. *Suppose that the conditions of Theorem 3 are satisfied. Then, we have*

$$\sqrt{n}[G_n - G_F] \xrightarrow{\mathcal{D}} \frac{1}{\mu}\left(2(\mathrm{CL})_F \int_0^1 \frac{\Gamma(Q(y))}{f(Q(y))}\mathrm{d}y - \int_0^1\int_0^p \frac{\Gamma(Q(y))}{f(Q(y))}\mathrm{d}y\mathrm{d}p - \int_0^1 \frac{(1-p)\Gamma(Q(p))}{f(Q(p))}\mathrm{d}p\right),$$

where $\Gamma(t)$ is a Gaussian process distributed as $\Gamma_n(t)$ in Theorem 2.

Proof. See the Appendix. □

Remark 3. As a consequence of the preceding theorem we can compute the confidence interval for G_F:

$$\left[G_n - z_{\alpha/2}\frac{\widehat{\sigma}}{\sqrt{n}} G_n + z_{\alpha/2}\frac{\widehat{\sigma}}{\sqrt{n}}\right],$$

where $z_{\alpha/2}$ is the $(1-\alpha/2)$-quantile of a standard normal distribution and $\widehat{\sigma}^2$ is a convergent estimator of variance

$$\frac{1}{\mu}\left(2(\mathrm{CL})_F \int_0^1 \frac{\Gamma(Q(y))}{f(Q(y))}\mathrm{d}y - \int_0^1\int_0^p \frac{\Gamma(Q(y))}{f(Q(y))}\mathrm{d}y\mathrm{d}p - \int_0^1 \frac{(1-p)\Gamma(Q(p))}{f(Q(p))}\mathrm{d}p\right).$$

Acknowledgments

The authors are very grateful to the editor of Statistics and Probability Letters, Professor Hira Koul, and the anonymous referees for valuable comments and suggestions.

Appendix

In establishing Theorem 3, we need the following lemma.

Lemma 2. *Suppose that Assumptions (2), (3) and (1.2) are satisfied. Then,*

$$\sup_{t,s\in J_n} |\alpha_n(t) - \alpha_n(s)| = O\left(n^{-\frac{1}{4}}\log n\right) \quad a.s.,$$

where $J_n = \left\{t : Q_n(p) - Cn^{-\frac{1}{2}} \le t \le Q_n(p) + Cn^{-\frac{1}{2}}\right\}$, and C is a positive number.

Proof. Assuming that $s \le t$, an elementary computation shows that

$$\begin{aligned} F_n(t) - F(t) - F_n(s) + F(s) &= \nu_n^{-1}\int_s^t y^{-1}\mathrm{d}[F_n^*(y) - F^*(y)] + (\nu_n^{-1} - \nu^{-1})\int_s^t y^{-1}\mathrm{d}F^*(y) \\ &= \nu_n^{-1}\left\{t^{-1}[F_n^*(t) - F^*(t)] - s^{-1}[F_n^*(s) - F^*(s)]\right\} \\ &\quad + \nu_n^{-1}\int_s^t y^{-2}[F_n^*(y) - F^*(y)]\mathrm{d}y + (\nu_n^{-1} - \nu^{-1})\int_s^t y^{-1}\mathrm{d}F^*(y) \end{aligned}$$

where, by (3.2), (1.2) and the definition of J_n,

$$\sup\left\{\left|(\nu_n^{-1} - \nu^{-1})\int_s^t y^{-1}\mathrm{d}F^*(y)\right|;\ x, y \in J_n\right\} = O(n^{-1}),$$

along with (1.2) and the Bahadur (1966) representation, we have

$$\sup_{t,s\in J_n} |F_n^*(t) - F^*(t) - F_n^*(s) + F^*(s)| = O\left(n^{-\frac{3}{4}}\log n\right) \quad \text{a.s.}$$

Therefore, we conclude that

$$\sup_{t,s\in J_n} |\alpha_n(t) - \alpha_n(s)| = O\left(n^{-\frac{1}{4}}\log n\right) \quad \text{a.s.} \quad \square$$

Proof of Theorem 3. Suppose that $s = Q_n(p)$ and $t = Q(p)$; Lemma 1 yields $s, t \in J_n$. Applying Lemma 2 gives

$$F_n(Q_n(p)) - F_n(Q(p)) = F(Q_n(p)) - F(Q(p)) + O\left(n^{-\frac{3}{4}}\log n\right) \quad \text{a.s.} \tag{A.1}$$

It is easy to see that $F_n(Q_n(p))$ can be replaced by p up to $O(n^{-1})$. For the right hand side, a Taylor expansion of the first term about Q(p) up to the second-order term gives

$$f(Q(p))[Q_n(p) - Q(p)] + O([Q_n(p) - Q(p)]^2) + O\left(n^{-\frac{3}{4}}\log n\right) \quad \text{a.s.}$$

Invoking Lemma 1 and rearranging terms in (A.1), we have

$$\sqrt{n}f(Q(p))[Q_n(p) - Q(p)] = \sqrt{n}[p - F_n(Q(p))] + O\left(n^{-\frac{1}{4}}\log n\right) \quad \text{a.s.}$$

Since F is continuous, $F(Q(p)) = p$. Recalling the definitions of the process α_n and quantile process ρ_n, we have

$$\rho_n(p) - \alpha_n(Q(p)) = O\left(n^{-\frac{1}{4}}\log n\right) \quad \text{a.s.} \tag{A.2}$$

Finally, by using Theorem 2 and (A.2) we obtain (2.3). Similarly, on replacing $\Gamma_n(Q(p))$ with $\Gamma(Q(p), n)$, this completes the proof of the theorem. □

Langberg et al. (1980) define the *total time on test transform* curve corresponding to a continuous distribution F on $\mathcal{R}^+$, $H_F^{-1}(p)$ for $p \in [0, 1]$ as

$$H_F^{-1}(p) = \int_0^p (1-y)\mathrm{d}Q(y) = (1-p)Q(p) + \int_0^p Q(y)\mathrm{d}y, \quad Q(0) = 0. \tag{A.3}$$

Obviously, $H_F^{-1}(p) \le H_F^{-1}(1) := \lim_{p\uparrow 1} H_F^{-1}(p) = \mu$. A natural estimator for $H_F^{-1}(p)$ is

$$H_n^{-1}(p) = (1-p)Q_n(p) + \int_0^p Q_n(y)\mathrm{d}y, \quad p \in [0, 1]. \tag{A.4}$$

Lemma 3 proves that this estimator is strongly uniformly consistent for H_F^{-1}.

Lemma 3. *Suppose that the conditions of Lemma 1 are satisfied. Then, we have*

$$\sup_{0\le p\le 1} |H_n^{-1}(p) - H_F^{-1}(p)| = O\left(\sqrt{\frac{\log\log n}{n}}\right) \quad \textit{a.s.}$$

Proof. By Lemma 1, we have

$$\begin{aligned}\sup_{0\le p\le 1} |H_n^{-1}(p) - H_F^{-1}(p)| &\le \sup_{0\le p\le 1}[(1-p)|Q_n(p) - Q(p)|] + \sup_{0\le p\le 1}\int_0^p |Q_n(y) - Q(y)|\mathrm{d}y \\ &= O\left(\sqrt{\frac{\log\log n}{n}}\right) \quad \text{a.s.} \quad \square\end{aligned}$$

Next, define the *normed total time on test* empirical process $t_n(p)$ by

$$t_n(p) := \sqrt{n}[H_n^{-1}(p) - H_F^{-1}(p)], \quad p \in [0, 1].$$

Lemma 4 characterizes the asymptotic limit of $t_n(p)$.

Lemma 4. *Suppose that the conditions of Theorem 3 are satisfied. Then, we have*

$$\sup_{0\le p\le 1}\left| t_n(p) - \left(\int_0^p \frac{\Gamma_n(Q(y))}{f(Q(y))}\mathrm{d}y + \frac{(1-p)\Gamma_n(Q(p))}{f(Q(p))}\right)\right| = O\left(\left(n^{-\frac{1}{4}}\log n\right) \vee (n^{-\lambda})\right) \quad \textit{a.s.},$$

for any $0 < \lambda < 1/2 - 1/r$, *for some* $r > 2$.

Proof. By (A.3), (A.4), Theorem 3 and the definition of the PL-quantile process,

$$\begin{aligned} t_n(p) &= \sqrt{n}\int_0^p |Q_n(y) - Q(y)|\mathrm{d}y + \sqrt{n}(1-p)[Q_n(p) - Q(p)] \\ &= \int_0^p \frac{\Gamma_n(Q(y))}{f(Q(y))}\mathrm{d}y + \frac{(1-p)\Gamma_n(Q(p))}{f(Q(p))} + O\left(\left(n^{-\frac{1}{4}}\log n\right) \vee (n^{-\lambda})\right) \quad \text{a.s.}\end{aligned}$$

The lemma is proved. □

Next, we define the *scaled total time on test transform*, its statistic and the associated empirical process corresponding to F. We have

$$W_F(p) := \frac{H_F^{-1}(p)}{\mu}, \qquad W_n(p) := \frac{H_n^{-1}(p)}{\mu_n} \tag{A.5}$$

and

$$w_n(p) := \sqrt{n}[W_n(p) - W_F(p)],$$

for $p \in [0, 1]$. Also, the cumulative total time on test transform and its empirical counterpart are defined by

$$V_F := \int_0^1 W_F(y)\mathrm{d}y = \frac{1}{\mu_n}\int_0^1 H_F^{-1}(y)\mathrm{d}y, \tag{A.6}$$

$$V_n := \int_0^1 W_n(y)\mathrm{d}y. \tag{A.7}$$

The following two lemmas give the uniform consistency of $W_n(p)$ and the strong approximation of the scaled total time on test empirical process respectively. Lemma 7 is a central limit theorem for the normed cumulative total time on test sequence $v_n := \sqrt{n}[V_n - V_F]$ that follows directly from Lemma 6.

Lemma 5. *Suppose that the conditions of Lemma 1 are satisfied. Then, we have*

$$\sup_{0\le p\le 1} |W_n(p) - W_F(p)| = O\left(\sqrt{\frac{\log\log n}{n}}\right) \quad a.s.$$

Proof. By (A.5) and the triangular inequality, the left hand side is bounded by

$$\sup_{0\le p\le 1}\left|\frac{H_n^{-1}(p)}{\mu_n} - \frac{H_n^{-1}(p)}{\mu}\right| + \sup_{0\le p\le 1}\left|\frac{H_n^{-1}(p)}{\mu} - \frac{H_F^{-1}(p)}{\mu}\right| = \sup_{0\le p\le 1}\left|H_n^{-1}(p)\frac{\mu-\mu_n}{\mu_n\mu}\right| + \sup_{0\le p\le 1}\left|\frac{1}{\mu}\left[H_F^{-1}(p) - H_n^{-1}(p)\right]\right|$$
$$= O\left(\sqrt{\frac{\log\log n}{n}}\right),$$

almost surely by Lemma 3 and (3.2). $\square$

The proof of the following lemma can be given along the lines of Lemma 3.5 of Tse (2006).

Lemma 6. *Suppose that the conditions of Theorem 3 are satisfied. Then, almost surely,*

$$\sup_{0\le p\le 1}\left|w_n(p) - \frac{1}{\mu}\left(\int_0^p \frac{\Gamma_n(Q(y))}{f(Q(y))}\mathrm{d}y + \frac{(1-p)\Gamma_n(Q(p))}{f(Q(p))}\right) + \frac{H_F^{-1}(p)}{\mu^2}\int_0^1 \frac{\Gamma_n(Q(y))}{f(Q(y))}\mathrm{d}y\right|$$
$$= O\left(\left(n^{-\frac{1}{4}}\log n\right) \vee (n^{-\lambda})\right),$$

for any $0 < \lambda < 1/2 - 1/r$, *for some* $r > 2$. $\square$

Lemma 7. *Suppose that the conditions of Theorem 3 are satisfied. Then, we have*

$$v_n \xrightarrow{\mathcal{D}} \frac{1}{\mu}\left(\int_0^1\int_0^p \frac{\Gamma(Q(y))}{f(Q(y))}\mathrm{d}y\mathrm{d}p + \int_0^1 \frac{(1-p)\Gamma(Q(p))}{f(Q(p))}\mathrm{d}p - V_F\int_0^1 \frac{\Gamma(Q(y))}{f(Q(y))}\mathrm{d}y\right),$$

where $\Gamma(t)$ *is a Gaussian process distributed as* $\Gamma_n(t)$ *in Theorem 2.*

Corollary 1. *It is easy to show that* $V_F = 2(\mathrm{CL})_F$, *i.e., the cumulative total time on test transform* V_F *is twice the cumulative Lorenz curve* $(\mathrm{CL})_F$. *Hence, Lemma 7 is a central limit theorem that can also be interpreted as*

$$\frac{1}{2}v_n = \sqrt{n}\left[\frac{1}{2}V_n - (\mathrm{CL})_F\right],$$

and $\frac{1}{2}V_n$ *is a consistent estimator for* $(\mathrm{CL})_F$.

Proof of Theorem 5. By the definition of the Lorenz curve corresponding to F in our model and by using (A.3) and (A.4) we have

$$W_F(p) = \frac{(1-p)Q(p)}{\int_0^1 Q(u)\mathrm{d}u} + L_F(p). \tag{A.8}$$

We have also

$$W_n(p) = \frac{(1-p)Q_n(p)}{\int_0^1 Q_n(u)\mathrm{d}u} + L_n(p), \quad p \in [0, 1]. \tag{A.9}$$

Substituting (A.8) and (A.9) in Lemma 6 we obtain the result. □

Corollary 2. *Suppose the conditions of Theorem 3 are satisfied. Then, almost surely,*

$$\sup_{0\le p\le 1}\left|\sqrt{n}[\mu_n L_n(p) - \mu L_F(p)] - \int_0^p \frac{\Gamma_n(Q(y))}{f(Q(y))}\mathrm{d}y\right| = O\left(\left(n^{-\frac{1}{4}}\log n\right) \vee (n^{-\lambda})\right),$$

for any $0 < \lambda < 1/2 - 1/r$, *for some* $r > 2$.

Proof of Theorem 8. By definition, we can write

$$G_F := \frac{\int_0^1 [u - L_F(u)]\mathrm{d}u}{\int_0^1 u\mathrm{d}u} = 1 - 2(\mathrm{CL})_F = 1 - V_F,$$

and hence we have

$$V_n - V_F = V_n - 2(\mathrm{CL})_F = V_n - (1 - G_F),$$

and the central limit theorem for υ_n holds also for $\sqrt{n}[V_n - 2(\mathrm{CL})_F]$ and $\sqrt{n}[V_n - (1 - G_F)]$. In particular, the cumulative total time on test statistic V_n is a consistent estimator of $1 - G_F$. Thus, an estimator for the Gini index is the Gini statistic defined by

$$G_n := 1 - V_n.$$

Therefore the theorem is proved. □

References

Bahadur, R.R., 1966. A note on quantiles in large samples. Annals of Mathematical Statistics 37, 577–580.
Blumenthal, S., 1967. Proportional sampling in life length studies. Technometrics 9, 205–218.
Bolbolian Ghalibaf, M., Fakoor, V., Azarnoosh, H. A., 2010. Asymptotic behaviors of the Lorenz curve for left truncated and dependent data. Submitted for publication.
Bolbolian Ghalibaf, M., Fakoor, V., Azarnoosh, H.A., 2011. Asymptotic behaviors of the Lorenz curve for censored data under strong mixing. Communications in Statistics—Theory and Methods 40, 323–332.
Colman, R., 1979. An introduction to mathematical sterology. Memoirs No. 3, Dept. of Theoretical Statistics, Univ. of Aarhus, Denmark.
Cox, D.R., 1969. In: Johnson, N.L., Smith, H. (Eds.), Some Sampling Problems in Technology. New Developments in Survey Sampling. Wiley, New York.
Csörgő, M., Révész, P., 1981. Strong Approximation in Probability and Statistics. Academic Press, New York.
Csörgő, M., Csörgő, S., Horváth, L., 1986. An asymptotic theory for empirical reliability and concentration processes. In: Lecture Notes in Statistics, vol. 33. Springer, Berlin, Heidelberg, New York.
Csörgő, M., Csörgő, S., Horváth, L., 1987. Estimation of total time on test transforms and Lorenz curves under random censorship. Statistics 18, 77–97.
Csörgő, M., Yu, H., 1999. Weak approximations for empirical Lorenz curves and their Goldie inverses of stationary observations. Advances in Applied Probability 31, 698–719.
Csörgő, M., Zitikis, R., 1996a. Strassen's LIL for the Lorenz curve. Journal of Multivariate Analysis 59, 1–12.
Csörgő, M., Zitikis, R., 1996b. Confidence bands for the Lorenz curve and Goldie curves. In: A volume in honor of Samuel Kotz. Wiley, New York.
Csörgő, M., Zitikis, R., 1997. On the rate of strong consistency of Lorenz curves. Statistics and Probability Letters 34, 113–121.
Davydov, Y., Zitikis, R., 2002. Convergence of generalized Lorenz curves based on stationary ergodic random sequences with deterministic noise. Statistics and Probability Letters 59, 329–340.
Davydov, Y., Zitikis, R., 2003. Generalized Lorenz curves and convexifications of stochastic processes. Journal of Applied Probability 40 (4), 906–925.
Davydov, Y., Khoshnevisan, D., Shic, Z., Zitikis, R., 2007. Convex rearrangements, generalized Lorenz curves, and correlated Gaussian data. Journal of Statistical Planning and Inference 137, 915–934.
Fakoor, V., Nakhaei Rad, N., 2011. Strong Gaussian approximation of Lorenz curve under strong mixing. Pakistan Journal of Statistics (in press).
Gastwirth, J.L., 1971. A general definition of the Lorenz curve. Econometrica 39, 1037–1039.
Goldie, C.M., 1977. Convergence theorems for empirical Lorenz curve and their inverses. Advances in Applied Probability 9, 765–791.
Helmers, R., Zitikis, R., 2005. Strong laws for generalized absolute Lorenz curves when data are stationary and ergodic sequences. Proceedings of the American Mathematical Society 133, 3703–3712.
Horváth, L., 1985. Estimation from a length-biased distribution. Statistics & Decisions 3, 91–113.
Kendall, M.G., Stuart, A., 1963. The Advanced Theory of Statistics I, 2nd ed. Charles Griffen and Company, London.
Komlós, J., Major, P., Tusnády, G., 1975. An approximation of partial sums of independent r.v.'s and the sample d.f.U.. Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete 32, 111–131.
Langberg, N.A., Leon, R.V., Proschan, F., 1980. Characterization of nonparametric classes of life distributions. Annals of Probability 8, 1163–1170.
McFadden, J.A., 1962. On the lengths of intervals in a stationary point process. Journal of the Royal Statistical Society: Series B 24, 364–382.

Patil, G.P., Rao, C.R., 1977. The weighted distributions: a survey their applications. In: Krishnaiah, P.R. (Ed.), Applications of Statistics. North-Holland, Amesterdam, pp. 383–405.
Patil, G.P., Rao, C.R., 1978. Weighted distributions and size-biased of sampling with applications to wildlife populations and human families. Biometrics 34, 179–189.
Pietra, G., 1915. Delle relazioni fra indici di variabilitá, note I e II. Atti del Reale Istituto Veneto di Scienze, Lettere ed. Arti. 74, 775–804.
Rao, C.R., Zhao, L.C., 1995. Strassen's law of the iterated logarithm for the Lorenz curves. Journal of Multivariate Analysis 54, 239–252.
Sen, P.K., 1984. On asymptotic representations for reduced quantiles in sampling from a length-biased distribution. Calcutta Statistical Association Bulletin 33, 59–67.
Tse, S.M., 2006. Lorenz curve for truncated and censored data. Annals of the Institute of Statistical Mathematics 58, 675–686.
Vardi, Y., 1982a. Nonparametric estimation in the presence of length-bias. The Annals of Statistics 10, 616–620.
Vardi, Y., 1982b. Nonparametric estimation in renewal processes. The Annals of Statistics 10, 772–785.
Wang, M.C., 1991. Nonparametric estimation from cross-sectional survival data. Journal of the American Statistical Association 86, 130–143.
Wicksell, S.D., 1925. The corpuscle problem. A mathematical study of a biometric problem. Biometrika 17 (1/2), 84–99.