Pak. J. Statist.
2011 Vol. 27(3), 231-239

ASYMPTOTIC BEHAVIORS OF THE LORENZ CURVE UNDER STRONG MIXING

Vahid Fakoor[1] and **Najmeh Nakhaei Rad**[2]
[1] Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran. Email: fakoor@math.um.ac.ir
[2] Department of Statistics, Islamic Azad University, Mashhad, Iran. Email: najme.rad@gmail.com

ABSTRACT

In this paper, we consider a nonparametric estimator of the Lorenz curve when data are showing some kind of dependence. The uniform strong convergence rate of the estimator under strong mixing hypothesis is obtained. Strong Gaussian approximation for the associated Lorenz process are established under appropriate assumptions. A law of the iterated logarithm for the Lorenz process is also derived.

KEYWORDS

Law of the iterated logarithm, Lorenz process, Quantile process, Strong Gaussian approximation, Strong mixing, Strong uniform consistency

1. INTRODUCTION AND PRELIMINARIES

Pietra (1915) and Gastwirth (1971) independently introduced the *Lorenz curve* corresponding to a non-negative random variable X with a distribution function F, quantile function $Q(p) := \inf\{t: F(t) \geq p\}$ and finite mean $EX = \mu$ as:

$$L_F(y) := \frac{1}{\mu}\int_0^y Q(u)du, 0 \leq y \leq 1.$$

In econometrics, with X representing income, $L(y)$ gives the fraction of total income that the holders of the lowest yth fraction of income possesses. 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

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

where $(CL)_F = \int_0^1 L_F(u)du$ is the *cumulative Lorenz curve* corresponding to F. This 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, one can use the *Lorenz statistic* $L_n(y)$ defined by

$$L_n(y) := \frac{1}{\mu_n} \int_0^y Q_n(u)du, 0 \leq y \leq 1,$$

where μ_n is the sample mean and $Q_n(p) := \inf\{t: F_n(t) \geq p\}$ is the empirical quantile function constructed from i.i.d. sample taken from F (F_n is the empirical distribution function).

Goldie (1977) proved the uniform consistency of L_n to L_F and derived the weak convergence of the *Lorenz process* $l_n(y) := \sqrt{n}[L_n(y) - L(y)], 0 \leq y \leq 1$ 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 curve. 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. Strong Gaussian approximations for the Lorenz process when data are subject to random right censorship and left truncation are established by Tse (2006), he is also derived a functional LIL for the Lorenz process.

However, in most economic situations, the basic sequence of observations may not be independent. It is more realistic to assume some form of dependence among the data are observed. Csörgő and Yu (1999), obtained weak approximations for Lorenz statistic and its inverse 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 large sample asymptotic theory for the empirical generalized Lorenz curves when observations are stationary and either short-range or long-range dependent. Strong laws for generalized absolute Lorenz curves when data are stationary and ergodic sequences established by Helmers and Zitikis (2005). Based on 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 in the case where the stochastic process is a Gaussian process with stationary increments and a nicely behaved correlation function.

It is the purpose of this pa per to derive strong uniform consistency of the Lorenz statistic and strong Gaussian approximation for Lorenz process, for the case in which data are assumed to be strong mixing. As a result of our strong Gaussian approximation, we obtain a functional LIL for the Lorenz process.

In this paper we consider the strong mixing dependence, which amounts to a form of asymptotic independence between the past and the future as shown by its definition.

Definition 1:

Let $\{X_i, i \geq 1\}$ *denote a sequence of random variables. Given a positive integer* m, *set*

$$\alpha(m) = \sup_{k\geq 1}\{|P(A \cap B) - P(A)P(B)|;\ A \in \mathcal{F}_1^k, B \in \mathcal{F}_{k+m}^{\infty}\}, \quad (1.1)$$

where $\mathcal{F}_i^k$ denote the σ-field of events generated by $\{X_j;\ i \leq j \leq k\}$. The sequence is said to be strong mixing (α-mixing) if the mixing coefficient $\alpha(m) \to 0$ as $m \to \infty$.

Among various mixing conditions used in the literature, α-mixing is reasonably weak and has many practical applications (see, e.g. Doukhan (1994) or Cai (1998, 2001, for more details). In particular, Masry and Tjostheim (1995) proved that, both ARCH processes and nonlinear additive AR models with exogenous variables, which are particularly popular in finance and econometrics, are stationary and α-mixing.

For convenient reference, the basic conditions on F, and the assumptions used in this paper from which the various results are obtained, are gathered together here.

Assumptions:

(1) Suppose that $\{X_n, n \geq 1\}$ is a non-negative sequence of stationary α-mixing random variables with continuous distribution function F and finite mean $EX = \mu$.

(2) $\alpha(n) = O(n^{-5-\epsilon})$ for some $0 < \epsilon \leq \frac{1}{4}$.

(3) Assume that F satisfies the Csörgő-Révész conditions, i.e.,

 (i) $F(x)$ is twice differentiable on (a, b), where

 $$a = \sup\{x: F(x) = 0\}, b = \inf\{x: F(x) = 1\}, 0 \leq a < b \leq \infty;$$

 (ii) $F' = f \neq 0$ on (a, b);

 (iii) for some $\gamma > 0$ we have

 $$\sup_{a<x<b} F(x)\big(1 - F(x)\big)\frac{|f'(x)|}{f^2(x)} = \sup_{0<s<1} s(1 - s)\frac{|f'(Q(s))|}{f^2(Q(s))} \leq \gamma.$$

(4) $\liminf s \downarrow 0\, f(Q(s)) > 0$, and $\liminf s \uparrow 1\, f(Q(s)) > 0$.

(5) $\limsup s \downarrow 0\, |f'(Q(s))| < \infty$, and $\limsup s \uparrow 1\, |f'(Q(s))| < \infty$.\$

In the next section, we present our main results.

2. ASYMPTOTIC BEHAVIORS OF LORENZ CURVE

We first introduce the following Gaussian process, which plays an important role to present the main results of the study. Let

$$g_n(s) = I(U_n \leq s) - s, n \geq 1,$$

where $\{U_n; n \geq 1\}$ is a uniform on $[0,1]$ strictly stationary strong mixing sequence of random variables. Define for $0 \leq s, s' \leq 1$,

$$\Gamma(s, s') = Cov\big(g_1(s), g_1(s')\big) + \sum_{n=2}^{\infty}\big[Cov\big(g_1(s), g_n(s')\big) + Cov\big(g_1(s'), g_n(s)\big)\big].$$

A separable Gaussian process $\{K(s,t): 0 \leq s \leq 1, t \geq 0\}$ is called a Kiefer process if it satisfies $K(s, 0) = K(1, t) = K(0, t) = 0, E\big(K(s,t)\big) = 0$, and has covariance function

$\Gamma^*(t, t', s, s') = \min(t, t')\,\Gamma(s, s')$, for $t, t' \geq 0$ and $0 \leq s, s' \leq 1$.

We now restate below a strong approximation by Yu (1996) for the normed quantile process $\rho_n(p) := \sqrt{n} f(Q(p))[Q(p) - Q_n(p)]$.

Theorem 1:

(Yu, 1996) Let $\{X_n, n \geq 1\}$ be a stationary α-mixing sequence of random variables with common continuous distribution function F. Assume that F satisfies the Csörgő-Révész conditions and Assumption 2 holds. Then there exists a Kiefer process $K(s, n)$ defined on the same probability space as $\rho_n(s)$ with covariance function $\Gamma^*(n, n', s, s') = \min(n, n')\,\Gamma(s, s')$ and a constant $\lambda > 0$ depending only on ε such that

$$\sup_{\delta_n \leq s \leq 1-\delta_n} \left| \rho_n(s) - \frac{K(s,n)}{\sqrt{n}} \right| = O\left((\log n)^{-\lambda} \right) \quad a.s., \tag{2.1}$$

where $\delta_n = n^{-\frac{1}{2}} (\log n)^{\lambda} (\log \log n)$.

If, in addition to the Csörgő - Révész conditions, we also assume that F satisfies the Assumptions (4) and (5), then we have

$$\sup_{0 \leq s \leq 1} \left| \rho_n(s) - \frac{K(s,n)}{\sqrt{n}} \right| = O\left((\log n)^{-\lambda} \right) \ a.s. \tag{2.2}$$

An implication of Theorem 1 is the law of the iterated logarithm for empirical quantile process, i.e.,

$$\sup_{0 \leq t \leq 1} |Q_n(t) - Q(t)| = O\left(\frac{\log \log n}{n} \right) \ a.s. \tag{2.3}$$

2.1 Strong Uniform Consistency

Theorem 2 below proves the uniform strong consistency with rate of the estimator L_n.

Theorem 2:

Suppose that Assumptions (1)-(5) are satisfied. Then

$$\sup_{0 \leq t \leq 1} |L_n(t) - L_F(t)| = O\left(\frac{\log \log n}{n} \right) \ a.s. \tag{2.4}$$

An elementary computation shows that,

$$L_n(t) - L_F(t) = \frac{1}{\mu_n} \int_0^t [Q_n(s) - Q(s)]\,ds - \frac{(\mu_n - \mu)}{\mu_n} L_F(t). \tag{2.5}$$

It is easy to see that,

$$\mu_n - \mu = \int_0^1 [Q_n(s) - Q(s)]\,ds. \tag{2.6}$$

Now, using (2.3), (2.5) and (2.6), we obtain the result.

2.2 Strong Gaussian Approximation

In Theorem 3 below, we construct a two parameter mean zero Gaussian process that strongly uniformly approximate the empirical process $l_n(t)$.

Theorem 3:

Suppose that Assumptions (1)-(5) are satisfied. Then there exists a Kiefer process, $K(s,t)$, defined on the same probability space as the sequence $\{X_i, i \geq 1\}$ with covariance function $\Gamma^{*}(n,n',s,s') = \min(n,n')\,\Gamma(s,s')$, for $0 < n,n' < \infty$, and $0 \leq s,s' \leq 1$, such that,

$$\sup_{0<p<1}\left| l_n(p) - \frac{1}{\mu}\left(\int_0^p \frac{K(y,n)/\sqrt{n}}{f(Q(y))}dy - L_F(p)\int_0^1 \frac{K(y,n)/\sqrt{n}}{f(Q(y))}dy \right)\right| = O\left((\log n)^{-\lambda}\right) \quad a.s., \tag{2.7}$$

for some $\lambda > 0$.

Proof:

See the Appendix.

2.3 Functional LIL

An immediate result of an almost sure invariance principle of the form (2.7) is a functional law of the iterated logarithm for Lorenz curve. Let $b_n = (2\log\log n)^{\frac{1}{2}}$, $D[a,b]$ be the space of functions on $[a,b]$ that are right continuous and have left limits.

Theorem 4:

Suppose that Assumptions (1)-(5) are satisfied. On a rich enough probability space, $\frac{l_n(\cdot)}{b_n}$ is almost surly relatively compact in $D[0,1]$ with respect to the supremum norm and its set of limit points is

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

where B is the unit ball in the reproduce kernel Hilbert space $H\left(\Gamma^*\right)$.

Proof:

Theorem 4 follows at once from (2.7) and Theorem A in Berkes and Philipp (1977).

3. APPENDIX

In establishing Theorem 3, we were aided by some ideas found in Tse (2006). The *total time on test transform* curve corresponding to a continuous distribution F on $[0,\infty)$, $H_F^{-1}(p)$, is defined for $p \in [0,1]$ as (see e.g. Langberg et al., 1980),

$$H_F^{-1}(p) := \int_0^p (1-y)\,d\,Q(y) = (1-p)Q(p) + \int_0^p Q(y)dy \tag{3.1}$$

$$Q(0) = 0\,.$$

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

$$H_n^{-1}(p) = (1-p)Q_n(p) + \int_0^p Q_n(y)\,dy, \qquad p \in [0,1]$$

Lemma 1 proves that this estimator is uniform strong consistent for H_F^{-1}.

Lemma 1:

Suppose that Assumptions (1)-(5) are satisfied. Then, we have

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

Proof:

By (2.3), we have

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

We define the normed total time on test empirical process $t_n(p)$ by

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

Lemma 2:

Suppose that Assumptions (1)-(5) are satisfied. Then there exists a Kiefer process, $K(s,t)$, defined on the same probability space as the sequence $\{X_i, i \ge 1\}$ with covariance function $\Gamma'(n,n',s,s') = \min(n,n')\Gamma(s,s')$, for $0 < n,n' < \infty$, and $0 \le s,s' \le \infty$, such that

$$\sup_{0\le p\le 1} \left| t_n(p) - \left(\int_0^p \frac{K(y,n)/\sqrt{n}}{f(Q(y))} dy + (1-p) \frac{K(p,n)/\sqrt{n}}{f(Q(p))} \right) \right| = O((\log n)^{-\lambda}) \quad a.s.,$$

for some $\lambda > 0$.

Proof:

By (3.2), (3.4) and the definition of $p_n(p)$, we have

$$t_n(p) = \sqrt{n}(1-p)[Q_n(p) - Q(p)] + \sqrt{n}\int_0^p [Q_n(y) - Q(y)]dy$$
$$= (1-p)\frac{K(p,n)/\sqrt{n}}{f(Q(p))} + \int_0^p \frac{K(y,n)/\sqrt{n}}{f(Q(y))}dy + O((\log n)^{-\lambda}) \quad a.s.$$

So, we obtain the result.

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

$$W_F(p) := \frac{H_F^{-1}(p)}{\mu}, \quad W_n(p) := \frac{H_n^{-1}(p)}{\mu_n}$$

and

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

for $p \in [0,1]$. The following lemmas give the uniform consistency of $W_n(p)$ and strong approximation of the scaled total time on test empirical process respectively. Their proofs can be done along the lines of lemma 3.4 and lemma 3.5 of Tse (2006). We therefore omit the proofs.

Lemma 3:

Suppose that Assumptions (1)-(5) are satisfied. Then, we have

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

Lemma 4:

Suppose that Assumptions (1)-(5) are satisfied. Then there exists a Kiefer process $K(s,t)$, defined on the same probability space as the sequence $\{X_i, i \ge 1\}$ with covariance function $\Gamma^*(n,n',s,s') = \min(n,n')\Gamma(s,s')$, for $0 < n,n' < \infty$, and $0 \le s,s' \le 1$, such that,

$$\sup_{0 \le p \le 1}\left|w_n(p) - \frac{1}{\mu}\left(\int_0^p \frac{K(y,n)/\sqrt{n}}{f(Q(y))}dy + (1-p)\frac{K(p,n)/\sqrt{n}}{f(Q(p))}\right) - \frac{H_F^{-1}(p)}{\mu}\int_0^1 \frac{K(y,n)/\sqrt{n}}{f(Q(y))}dy\right|$$
$$= O((\log n)^{-\lambda}) \quad a.s.,$$

for some $\lambda > 0$.

Proof of Theorem 3:

By definition of the Lorenz curve and by using (3.1) and (3.2) we have

$$W_F(p) = \frac{(1-p)Q(p)}{\int_0^1 Q(u)du} + L_F(p) \tag{3.3}$$

We have also

$$W_n(p) = \frac{(1-p)Q_n(p)}{\int_0^1 Q_n(u)\,du} + L_n(p), \qquad p \in [0,1]. \tag{3.4}$$

Substituting (3.3) and (3.4) in Lemma 4 we obtain the result.

ACKNOWLEDGEMENTS

The authors would like to thank two referees for their valuable suggestions and commons. The first author was supported by a grant from Ferdowsi University of Mashhad; (No. MS88097FAK).

REFERENCES

1. Berkes, I. and Philipp, W. (1977). An Almost Sure Invariance Principle for the Empirical Distribution Function of Mixing Random Variables. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 41, 115-137.
2. Cai, Z. (1998). Asymptotic properties of Kaplan-Meier estimator for censored dependent data. *Statist. and Probab.Lett.,* 37, 381-389.
3. Cai, Z. (2001). Estimating a distribution function for censored time series data. *Journal of Multivariate Analysis,* 78, 299-318.
4. Csörgő, M., Csörgő, S. and Horváth, L. (1986). *An asymptotic theory for empirical reliability and concentration processes*. Lecture Notes in Statistics, vol. 33, Springer Berlin Heidelberg New York.
5. Csörgő, M. and Yu, H. (1999). Weak approximations for empirical Lorenz curves and their Goldie inverses of stationary observations. *Advances in Applied Probability.* 31, 698-719.
6. Csörgő, M. and Zitikis, R. (1996a). Strassen's LIL for the Lorenz curve. *Journal of Multivariate Analysis*, 59, 1-12.
7. Csörgő, M. and Zitikis, R. (1996b). Confidence bands for the Lorenz curve and Goldie curves. In *A volume in honor of Samuel Kotz*. New York: Wiley.
8. Csörgő, M. and Zitikis, R. (1997). On the rate of strong consistency of Lorenz curves. *Statist. and Probab. Lett.,* 34, 113-121.
9. Davydov. Y, Khoshnevisan, D., Shi, Z. and Zitikis, R. (2007). Convex rearrangements, generalized Lorenz curves, and correlated Gaussian data. *J. Statist. Plann. and Infer.,* 137, 915-934.
10. Davydov, Y. and Zitikis, R. (2002). Convergence of generalized Lorenz curves based on stationary ergodic random sequences with deterministic noise. *Statist. and Probab. Lett.,* 59, 329-340.
11. Davydov, Y. and Zitikis, R. (2003). Generalized Lorenz curves and convexifications of stochastic processes. *J. Appl. Probab.* 40(4), 906-925.
12. Doukhan, P. (1994). Mixing: Properties and Examples. *Lect. Notes in Statist*. 61, Springer Verlag.
13. Gastwirth, J.L. (1971). A general definition of the Lorenz curve. *Econometrica,* 39, 1037-1039.

14. Goldie, C.M. (1977). Convergence theorems for empirical Lorenz curve and their inverses. *Advances in Applied Probability*, 9, 765-791.
15. Helmers, R. and Zitikis, R. (2005). Strong laws for generalized absolute Lorenz curves when data are stationary and ergodic sequences. *Proc. Amer. Math. Soc*. 133, 3703-3712.
16. Kendall, M.G. and Stuart, A. (1963). *The advanced theory of statistics I*. (2nd. ed.) Charles Griffen and Company, London.
17. Langberg, N.A., Leon, R.V. and Proschan, F. (1980). Characterization of nonparametric classes of life distributions. *Annals of Probability*, 8, 1163-1170.
18. Masry, E., Tjostheim, D. (1995). Nonparametric estimation and identification of nonlinear ARCH time series: Strong convergence and asymptotic normality. *Econ. Theor*., 11, 258-289.
19. 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.
20. 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.
21. Tse, S.M. (2006). Lorenz curve for truncated and censored data. *Ann. Inst. of Statist. Math.,* 58, 675-686.
22. Yu, Hao. (1996). A note on strong approximation for quantile processes of strong mixing sequences. *Statist. and Probab. Lett.,* 30, 1-7.