AISM (2006) 58: 675–686
DOI 10.1007/s10463-006-0039-7

Sze-Man Tse

Lorenz curve for truncated and censored data

Received: 6 December 2004 / Revised: 9 September 2005 / Published online: 11 May 2006

Abstract In this paper, we consider a nonparametric estimator of the Lorenz curve and Gini index when the data are subjected to random left truncation and right censorship. Strong Gaussian approximations for the associated Lorenz process are established under appropriate assumptions. A law of the iterated logarithm for the Lorenz curve is also derived. Lastly, we obtain a central limit theorem for the corresponding Gini index.

Keywords Left truncatrion · Right censorship · Product-limit · Quantile process · Strong Gaussian approximations · Lorenz process · Gini index · Law of the iterated logarithm

1 Introduction

Let X be a positive random variable with continuous distribution function F, quantile function Q and finite mean μ. Gastwirth (1971, 1972) defined the Lorenz curve corresponding to X as:

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

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

Work partially supported by NSC Grant 92-2118-M-259-003.

S.M. Tse
Department of Applied Mathematics,
National Dong Hwa University, Hualien, Taiwan
E-mail: tse@server.am.ndhu.edu.tw

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 $(\mathrm{CL})_F = \int_0^1 L_F(u)\,du$ 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 and Gini index, one can use the Lorenz statistics $L_n(y)$ and Gini statistics G_n defined by

$$\begin{aligned} L_n(y) &:= \frac{1}{\bar{X}} \int_0^y Q_n(u)\,\mathrm{d}u, \quad 0 \le y \le 1, \\ G_n &:= 1 - 2 \int_0^1 L_n(u)\,\mathrm{d}u, \end{aligned} \tag{1.1}$$

where $\bar{X}$ is the sample mean and $Q_n(y)$ is the empirical quantile function constructed from i.i.d. sample taken from F. Goldie (1977) proved the uniform consistency of L_n to L_F and derived the weak convergence of the Lorenz process $\sqrt{n}\,[L_n(y) - L(y)]$, $0 \le y \le 1$ to a Gaussian process under suitable conditions. A central limit theorem for the Gini statistics is also obtained. 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.

However, data are often taken under restricted conditions. In this paper, we consider data in the random left truncated and right censored (LTRC) model. In contrast, we shall refer to the model above as the full model. Let X, T and S be independent, positive random variables with continuous distribution functions (df) F, G and L respectively. Moreover, F is differentiable with derivative f. Let $Y = \min(X, S)$ and $\delta = I(X \le S)$. If $Y \ge T$, one observes (Y, T, δ). If $Y < T$, nothing is observed. We think of X as the variable of interest, the observation of which is subjected to right censorship, S, and left truncation, T, mechanisms. δ indicates whether the observed Y is a censored item or not. This is the LTRC model. Denote the df of Y by J. By the independent assumption, we have $1 - J = (1 - F)(1 - L)$. Let (X_i, T_i, S_i), $i = 1, \ldots, N$, be i.i.d. as (X, T, S), where the population size N is fixed but unknown. The empirical data are (Y_i, T_i, δ_i), $i = 1, \ldots, n$, where the number of observed triplets, n, is a

$\text{Bin}(N, \alpha)$ random variable, with $\alpha := P(T \le Y)$. The nonparametric maximum likelihood estimator of F (Tsai et al. 1987) is

$$1 - F_n(t) = \prod_{i:Y_i \le t} \left[1 - \frac{1}{nC_n(Y_i)} \right]^{\delta_i}, \tag{1.2}$$

assuming no ties in the data where $nC_n(z) = \sum_{i=1}^n I\{T_i \le z \le Y_i\}$. We shall refer to F_n as the product-limit (PL) estimator for the LTRC model. The quantile function Q and its empirical counterpart Q_n are defined by

$$Q(p) := \inf\{t : F(t) \ge p\} \quad \text{and} \quad Q_n(p) := \inf\{t : F_n(t) \ge p\} \tag{1.3}$$

for $0 < p < 1$. Gijbels and Wang (1993), established almost sure representation of PL-estimator in terms of sums of normed i.i.d. random processes. Zhou and Yip (1999) initiated and Tse (2003, 2005) established simultaneous strong Gaussian approximations of the PL-process $Z_n(t) = \sqrt{n}\,[\, F_n(t) - F(t)\,]$ and the associated PL-quantile process $\rho_n(p) := \sqrt{n}\, f((Q(p))\,[\, Q(p) - Q_n(p)\,]$ by a two-parameter Gaussian process.

Putting the PL-quantile function in place of the usual quantile function in (1.1), we get an estimator of the Lorenz curve and the Gini index for the LTRC model. Aside from enlarging the scope of application in econometrics, this is also important in other areas, especially in the Medical and Health Sciences where such data are frequently encountered. With such data, Lorenz curves and Gini indices could be applied for measuring inequality between treatment or resources in different geographical regions or income groups. In this paper, we develop these estimators and study their asymptotic properties. In contrast with Goldie (1977) weak convergence approach for the full model, we follow the lead of Csörgő et al. (1986) to build on the strong approximation results of the PL-quantile process in Tse (2005). While requiring stronger condition, the proof for this latter approach is much simpler and the result is more powerful. The left truncation and the right censorship model are included as particular cases. In the absence of both, our result reduces properly to that of Csörgő (1983) for the full model.

2 Notation and preliminary results.

Conditional on the value of n, (Y_i, T_i, δ_i), $i = 1, \dots, n$ are still i.i.d. but with the joint distribution of (Y, T) becomes

$$P\{Y \le y,\ T \le t \mid T \le Y\} = \alpha^{-1} \int_0^y G(t \wedge z)\, \mathrm{d}J(z)$$

for $y, t > 0$. Here and in the following, $\int_a^b = \int_{(a,b]}$ for $0 \le a < b \le \infty$. For $0 \le z < \infty$, let

$$C(z) := \frac{1}{\alpha} P(T \le z \le S)\,[\,1 - F(z-)\,],$$

and $F_1(z) = P(Y \le z,\ \delta = 1 \mid T \le Y)$. For the theorems below, we assume that F, G and L satisfy the condition

$$\int_0^\infty \frac{dF_1(z)}{C^3(z)} < \infty. \tag{2.1}$$

The condition, while not optimal, serves to keep the variances of the limiting processes finite near the lower end points and simplify the proof a great deal.

For any df K, let $a_K = \inf\{z : K(z) > 0\}$ and $b_K = \sup\{z : K(z) < 1\}$ denote the left and right end points of its support. As in the random truncation model (Woodroofe 1985,Gu and Lai 1990), F can be reconstructed only when $a_G \le a_J$ and $b_G \le b_J$. For the sake of simplicity, we assume that $a_G = a_J = 0$ and $b_G \le b_J$ throughout. For $0 < t < b < b_J$, let

$$k(t) := \int_0^t \frac{d\,F_1(u)}{C^2(u)}. \tag{2.2}$$

The next theorem is an adaptation of Theorems 2.1 and 2.2 from Tse (2005) that would suit our purpose latter. The modification makes use of the law of the iterated logarithm for two-parameters Wiener process, a proof of which can be found in Csörgő et al. (1981). The statements are conditional on the observed sample size n.

Theorem 2.1 *Suppose condition (2.1) is satisfied and* $0 < p_0 < p_1 < 1$. *Assume that F is Lipschitz continuous and that F is twice continuously differentiable on* $[\,Q(p_0) - \delta,\ Q(p_1) + \delta\,]$ *for some* $\delta > 0$ *such that f is bounded away from zero there. On a rich enough probability space, one can construct a two-parameter mean zero Gaussian process $B(t, u)$ for $t \ge 0$ and $u \ge 0$ with* $Cov\,[\,B(s, n),\ B(t, m)\,] = \sqrt{n/m}\,k(s)$, *for $n \le m$, $s < t$ such that, almost surely,*

$$\sup_{p_0 \le p \le p_1} |\,\rho_n(p) - (p_1 - p)\,B\big(Q(p), n\big)\,| = O\left(\frac{(\log n)^{3/2}}{n^{1/8}}\right), \tag{2.3}$$

where $0 < Q(p_0) < Q(p_1) < b$. *Moreover,* $B_n(t) = B(t, n)/\sqrt{n}$ *for* $n = 1, 2, \ldots$ *is a sequence of independent and identically distributed mean zero Gaussian processes for* $0 < t < b$, $b < b_J$, *with* $Cov\,[\,B_n(s), B_n(t)\,] = k\big(\min(s, t)\big)$, *for* $0 < s, t < b < b_J$ *such that, almost surely,*

$$\sup_{p_0 \le p \le p_1} |\,\rho_n(p) - (p_1 - p)\,B_n\big(Q(p)\big)\,| = O\left(\frac{\log n}{n^{1/6}}\right), \tag{2.4}$$

where $0 < Q(p_0) < Q(p_1) < b$.

We shall apply Theorem 2.1 to establish strong approximation results for the Lorenz process from which a LIL for the Lorenz curve will be derived.

3 Main theorems

Theorem 2.1 gives strong Gaussian approximations of the PL-quantile process over restricted intervals $[p_0, p_1]$ for fixed $0 < p_0 < p_1 < 1$. The restriction reflects the presence of left truncation and right censorship, in the absence of which we can take $p_0 = 0$ and $p_1 = 1$ under appropriate assumptions of F. For the LTRC model, we present modified definitions of relevant notions over $[p_0, p_1]$ in such a way that results for the full model are recovered when $p_0 = 0$ and $p_1 = 1$. In addition, we assume throughout that F has a finite expectation μ.

In the full model, the total time on test transform curve corresponding to a continuous distribution F on $[0, \infty)$, $H_F^{-1}(u)$, is defined for $u \in [0, 1]$ as (see e.g. Langberg et al., (1980)

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

Obviously, $H_F^{-1}(u) \le H_F^{-1}(1) := \lim_{u\uparrow 1} H_F^{-1}(u) = \mu$. For the LTRC model, we modify the definition of $H_F^{-1}(u)$ as

$$H_F^{-1}(u) = (p_1 - u)\,Q(u) + \int_{p_0}^u Q(y)\,\mathrm{d}y, \quad u \in [p_0, p_1]. \tag{3.1}$$

As $p_0 \downarrow 0$ and $p_1 \uparrow 1$, $H_F^{-1}(p_1) \to \int_0^1 Q(y)\,dy = \mu$. We can regard $H_F^{-1}(p_1)$ as a surrogate for the finite mean μ. A natural estimator for $H_F^{-1}(u)$ is

$$H_n^{-1}(u) = (p_1 - u)\,Q_n(u) + \int_{p_0}^u Q_n(y)\,\mathrm{d}y, \quad u \in [p_0, p_1]. \tag{3.2}$$

where $Q_n(u)$ is the right continuous version of the PL-quantile function. Lemma 3.1 says that this estimator is uniformly consistent over $[p_0, p_1]$. The following lemmas hold in the probability space of Theorem 2.1.

Lemma 3.1 *Suppose the conditions of Theorem 2.1 are satisfied and $E\,X = \mu$ is finite. We have, almost surely,*

$$\lim_{n\to\infty} \sup_{p_0 \le u \le p_1} |\,H_n^{-1}(u) - H_F^{-1}(u)\,| = 0.$$

Proof Here and in the following, we shall write sup for $\sup_{p_0 \le u \le p_1}$ unless otherwise stated. By Lemma 3.3 of Tse (2005), we have,

$$\begin{aligned}\sup \mid H_n^{-1}(u) - H_F^{-1}(u) \mid \; &\le \; \sup \int_{p_0}^{u} \mid Q_n(y) - Q(y) \mid \mathrm{d}y \\ &\quad + \sup \Big[\, (p_1 - u) \, |Q_n(u) - Q(u)| \, \Big] \\ &= O\left(\sqrt{\frac{\log \log n}{n}} \right)\end{aligned}$$

almost surely. □

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

$$t_n(u) := \sqrt{n} \, [\, H_n^{-1}(u) - H_F^{-1}(u) \,], \quad u \in [p_0, p_1]. \tag{3.3}$$

Lemma 3.2 and its corollary characterize the asymptotic limit of $t_n(u)$.

Lemma 3.2 *Suppose the conditions of Theorem 2.1 are satisfied and $E\,X = \mu$ is finite. On a rich enough probability space, there exist a sequence of independent and identically distributed mean zero Gaussian processes $\{B_n(t), 0 < t < b\}$ for $b < b_J$, with $Cov\,[\, B_n(s), B_n(t) \,] = k\big(\min(s,t)\big)$, for $0 < s, t < b < b_J$ such that, almost surely,*

$$\begin{aligned}\sup_{p_0 \le u \le p_1} &\left| t_n(u) - \left(\int_{p_0}^{u} \frac{(p_1 - y)\, B_n(Q(y))}{f\big(Q(y)\big)} \mathrm{d}y + \frac{(p_1 - u)^2 \, B_n\big(Q(u)\big)}{f\big(Q(u)\big)} \right) \right| \\ &= O\left(\frac{\log n}{n^{1/6}} \right).\end{aligned}$$

Proof By (3.1), (3.2), (2.4) and the definition of the PL-quantile process,

$$\begin{aligned}t_n(u) &= \sqrt{n} \int_{p_0}^{u} \big[Q_n(y) - Q(y) \big] \mathrm{d}y + \sqrt{n}\,(p_1 - u) \,[\, Q_n(u) - Q(u) \,] \\ &= \int_{p_0}^{u} \frac{(p_1 - y)\, B_n\big(Q(y)\big)}{f\big(Q(y)\big)} \mathrm{d}y + \frac{(p_1 - u)^2 \, B_n\big(Q(u)\big)}{f\big(Q(y)\big)} + O\left(\frac{\log n}{n^{1/6}} \right)\end{aligned}$$

almost surely. The lemma is proved. □

The next corollary is a direct consequence of Lemma 3.2.

Corollary 3.3 *Suppose the conditions of Theorem 2.1 are satisfied and* $E\,X = \mu$ *is finite. We have*

$$\sup_{p_0 \le u \le p_1} \sqrt{n}\,\left| H_n^{-1}(u) - H_F^{-1}(u) \right|$$
$$\xrightarrow{\mathcal{D}} \sup_{p_0 \le u \le p_1} \left| \int_{p_0}^{u} \frac{(p_1 - y)\, B(Q(y))}{f\big(Q(y)\big)}\, \mathrm{dy} + \frac{(\mathrm{p}_1 - \mathrm{u})^2\, \mathrm{B}\big(\mathrm{Q(u)}\big)}{\mathrm{f}\big(\mathrm{Q(u)}\big)} \right|,$$

where $B(t)$ *is a Gaussian process distributed as* $B_n(t)$ *in Theorem 2.1.*

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

$$W_F(u) := \frac{H_F^{-1}(u)}{H_F^{-1}(p_1)}, \qquad W_n(u) := \frac{H_n^{-1}(u)}{H_n^{-1}(p_1)},$$
$$w_n(u) := \sqrt{n}\left[\, W_n(u) - W_F(u)\, \right] \tag{3.4}$$

for $u \in [p_0, p_1]$. Also, the cumulative total time on test transform and its empirical counterpart are defined by

$$V_F := \int_{\mathrm{p}_0}^{\mathrm{p}_1} W_F(y)\, \mathrm{d}y = \frac{1}{H_F^{-1}(p_1)} \int_{p_0}^{p_1} H_F^{-1}(y)\, \mathrm{d}y,$$
$$V_n := \int_{p_0}^{p_1} W_n(y)\, \mathrm{d}y. \tag{3.5}$$

Lemmas 3.4 and 3.5 give the uniform consistency of $W_n(u)$ and strong approximation of the scaled total time on test empirical process by limiting process respectively. Lemma 3.6 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 3.5.

Lemma 3.4 *Suppose the conditions of Theorem 2.1 are satisfied and* $E\,X = \mu$ *is finite. We have almost surely,*

$$\lim_{n \to \infty} \sup_{p_0 \le u \le p_1} \left| W_n(u) - W_F(u) \right| = 0.$$

Proof By triangular inequality, the left hand side is bounded by

$$\sup \left| \frac{H_n^{-1}(u)}{H_n^{-1}(p_1)} - \frac{H_n^{-1}(u)}{H_F^{-1}(p_1)} \right| + \sup \left| \frac{H_n^{-1}(u)}{H_F^{-1}(p_1)} - \frac{H_F^{-1}(u)}{H_F^{-1}(p_1)} \right|$$
$$\le \sup \left| H_n^{-1}(u) \frac{H_F^{-1}(p_1) - H_n^{-1}(p_1)}{H_n^{-1}(p_1)\, H_F^{-1}(p_1)} \right| + \sup \left| \frac{1}{H_F^{-1}(p_1)} \big[H_F^{-1}(u) - H_n^{-1}(u) \big] \right| \mapsto 0$$

almost surely by Lemma 3.1. □

Lemma 3.5 *Suppose the conditions of Theorem 2.1 are satisfied and $E\,X = \mu$ is finite. On a rich enough probability space, there exist a sequence of independent and identically distributed mean zero Gaussian processes $\{B_n(t), 0 < t < b\}$ for $b < b_J$, with $Cov\,[\,B_n(s), B_n(t)\,] = k\big(\min(s,t)\big)$, for $0 < s, t < b < b_J$ such that, almost surely,*

$$\sup_{p_0 \le u \le p_1} \left| w_n(u) - \frac{1}{H_F^{-1}(p_1)} \left(\int_{p_0}^{u} \frac{(p_1 - y)\, B_n(Q(y))}{f\big(Q(y)\big)}\, \mathrm{d}y + \frac{(p_1 - u)^2\, B_n\big(Q(u)\big)}{f\big(Q(u)\big)} \right) \right.$$
$$\left. + \frac{H_F^{-1}(u)}{\big(H_F^{-1}(p_1)\big)^2} \int_{p_0}^{p_1} \frac{(p_1 - y)\, B_n(Q(y))}{f\big(Q(y)\big)}\, \mathrm{d}y \right| = O\left(\frac{\log n}{n^{1/6}} \right).$$

Proof By triangular inequality, the left hand side is bounded by

$$\sup \left| \frac{H_n^{-1}(u)\sqrt{n}\big[H_F^{-1}(p_1) - H_n^{-1}(p_1)\big]}{H_n^{-1}(p_1)\, H_F^{-1}(p_1)} + \frac{H_F^{-1}(u)}{\big(H_F^{-1}(p_1)\big)^2} \int_{p_0}^{p_1} \frac{(p_1 - y)\, B_n(Q(y))}{f\big(Q(y)\big)}\, \mathrm{d}y \right|$$
$$+ \sup \left| \frac{\sqrt{n}\big[H_n^{-1}(u) - H_F^{-1}(u)\big]}{H_F^{-1}(p_1)} - \frac{1}{H_F^{-1}(p_1)} \left(\int_{p_0}^{u} \frac{(p_1 - y)\, B_n(Q(y))}{f\big(Q(y)\big)}\, \mathrm{d}y \right. \right.$$
$$\left. \left. + \frac{(p_1 - u)^2\, B_n\big(Q(u)\big)}{f\big(Q(u)\big)} \right) \right| = O\left(\frac{\log n}{n^{1/6}} \right)$$

almost surely by Lemmas 3.1 and 3.2. □

Obviously, with (2.3) in place of (2.4) in the proofs of Lemmas 3.2 and 3.5, we get companion statements of the results in terms of the two-parameter Gaussian process in Theorem 2.1. The latter will be used in Theorem 3.7 below.

Lemma 3.6 *Suppose the conditions of Theorem 2.1 are satisfied and $E\,X = \mu$ is finite. We have*

$$v_n \xrightarrow{\mathcal{D}} \frac{1}{H_F^{-1}(p_1)} \left(\int_{p_0}^{p_1} \int_{p_0}^{u} \frac{(p_1 - y)\, B(Q(y))}{f\big(Q(y)\big)}\, \mathrm{d}y\, \mathrm{d}u + \int_{p_0}^{p_1} \frac{(p_1 - u)^2\, B\big(Q(u)\big)}{f\big(Q(u)\big)} \mathrm{d}u \right.$$
$$\left. - V_F \int_{p_0}^{p_1} \frac{(p_1 - y)\, B(Q(y))}{f\big(Q(y)\big)}\, \mathrm{d}y \right)$$

where $B(\cdot)$ is a Gaussian process distributed as $B_n(\cdot)$ of Theorem 2.1.

Next, we define the Lorenz curve corresponding to F in the LTRC model by

$$L_F := \frac{\int_{p_0}^{y} Q(u)\, \mathrm{d}u}{\int_{p_0}^{p_1} Q(u)\, \mathrm{d}u}, \quad y \in [p_0, p_1]. \tag{3.6}$$

Note that from (3.1), (3.4) and (3.6), we have

$$W_F(y) = \frac{(p_1 - y)\, Q(y)}{\int_{p_0}^{p_1} Q(u)\,\mathrm{d}u} + L_F(y). \tag{3.7}$$

Hence, defining the cumulative Lorenz curve corresponding to F as

$$(CL)_F := \int_{p_0}^{p_1} L_F(y)\,\mathrm{d}y,$$

we see that

$$\begin{aligned}\int_{p_0}^{p_1} L_F(y)\,\mathrm{d}y &= \frac{1}{\int_{p_0}^{p_1} Q(u)\,\mathrm{d}u} \int_{p_0}^{p_1}\int_{\mathrm{p}_0}^{\mathrm{y}} Q(u)\,\mathrm{d}u\,\mathrm{d}y \\ &= \frac{1}{\int_{p_0}^{p_1} Q(u)\,\mathrm{d}u} \int_{p_0}^{p_1}\int_{u}^{p_1} \mathrm{d}y\, Q(u)\,\mathrm{d}u = \frac{1}{\int_{p_0}^{p_1} Q(u)\,\mathrm{d}u} \int_{p_0}^{p_1} (p_1 - u)\, Q(u)\,\mathrm{d}u.\end{aligned}$$

Therefore, we have $V_F = 2\,(CL)_F$, i.e., the cumulative total time on test transform V_F is twice the cumulative Lorenz curve $(\mathrm{CL})_F$. Hence, Lemma 3.6 is a central limit theorem that can also be interpreted for

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

and $\frac{1}{2} V_n$ is a consistent estimator for $(\mathrm{CL})_F$. An estimator for the Lorenz curve is the Lorenz statistics $L_n(u)$ defined by

$$L_n(y) := \frac{\int_{p_0}^{y} Q_n(u)\,\mathrm{d}u}{\int_{p_0}^{p_1} Q_n(u)\,\mathrm{d}u}, \quad y \in [p_0, p_1],$$

and we have also the relationship

$$W_n(y) = \frac{(p_1 - y)\, Q_n(y)}{\int_{p_0}^{p_1} Q_n(u)\,\mathrm{d}u} + L_n(y), \quad y \in [p_0, p_1].$$

The empirical Lorenz process l_n is defined by

$$l_n(y) := \sqrt{n}\left[L_n(y) - L_F(y) \right], \quad y \in [p_0, p_1].$$

Similarly to Lemma 3.5 and recalling the remark after Lemma 3.5, we have

Theorem 3.7 *Suppose the conditions of Theorem 2.1 are satisfied and $E\,X = \mu$ is finite. On a rich enough probability space, one can construct a two-parameter mean* *zero Gaussian process $B(t,u)$ for $t \geq 0$ and $u \geq 0$ with Cov* $[\, B(s,n), B(t,m)\,] = \sqrt{n/m}\, k(s)$, *for $n \leq m$, $s < t$ such that, almost surely,*

$$\sup_{p_0 \leq u \leq p_1} \left| l_n(u) - \frac{1}{H_F^{-1}(p_1)} \left(\int_{p_0}^{u} \frac{(p_1 - y)\, B(Q(y), n)}{f\big(Q(y)\big)}\, \mathrm{d}y \right. \right.$$
$$\left.\left. - L_F(u) \int_{p_0}^{p_1} \frac{(p_1 - y)\, B(Q(y), n)}{f\big(Q(y)\big)}\, \mathrm{d}y \right) \right| = O\left(\frac{(\log n)^{3/2}}{n^{1/8}} \right), \quad (3.8)$$

where $0 < Q(p_0) < Q(p_1) < b$. Moreover, $B_n(t) = B(t,n)/\sqrt{n}$ for $n = 1, 2, \ldots$ is a sequence of independent and identically distributed mean zero Gaussian processes for $0 < t < b$, $b < b_J$, with Cov $[\, B_n(s), B_n(t)\,] = k\big(\min(s,t)\big)$, for $0 < s$, $t < b < b_J$ such that, almost surely,

$$\sup_{p_0 \leq u \leq p_1} \left| l_n(u) - \frac{1}{H_F^{-1}(p_1)} \left(\int_{p_0}^{u} \frac{(p_1 - y)\, B_n(Q(y))}{f\big(Q(y)\big)}\, \mathrm{d}y \right. \right.$$
$$\left.\left. - L_F(u) \int_{p_0}^{p_1} \frac{(p_1 - y)\, B_n(Q(y))}{f\big(Q(y)\big)}\, \mathrm{d}y \right) \right| = O\left(\frac{\log n}{n^{1/6}} \right). \quad (3.9)$$

The weak convergence implication of this result generalizes that of Goldie (1977) to the LTRC model, albeit under more stringent condition.

Recalling from (3.1) and (3.2) that $H_F^{-1}(p_1) = \int_{p_0}^{p_1} Q(y)\, dy$ and $H_n^{-1}(p_1) = \int_{p_0}^{p_1} Q_n(y)\, \mathrm{d}y$, we also have

Corollary 3.8 *Suppose the conditions of Theorem 2.1 are satisfied and $E\, X = \mu$ is finite. On a rich enough probability space, there exist a sequence of independent and identically distributed mean zero Gaussian processes $\{B_n(t), 0 < t < b\}$ for $b < b_J$, with* Cov $[\, B_n(s), B_n(t)\,] = k\big(\min(s,t)\big)$, *for $0 < s$, $t < b < b_J$ such that, almost surely,*

$$\sup_{p_0 \leq u \leq p_1} \left| \sqrt{n}\big[\, H_n^{-1}(p_1)\, L_n(u) - H_F^{-1}(p_1)\, L_F(u)\,\big] - \int_{p_0}^{u} \frac{(p_1 - y)\, B_n(Q(y))}{f\big(Q(y)\big)}\, \mathrm{d}y \right|$$
$$= O\left(\frac{\log n}{n^{1/6}} \right).$$

The next theorem gives a LIL for the Lorenz curve. We work on the probability space of Theorem 2.1. Let $b_n = (2n \log \log n)^{1/2}$, $D[a,b]$ be the space of functions on $[a,b]$ that are right continuous and have left limits and S be Strassen's set of absolutely continuous functions

$$S = \left\{ g \mid g : [0,1] \to \mathbf{R},\ g(0) = 0,\ \int_0^1 \big(g'(x)\big)^2\, \mathrm{d}x \leq 1 \right\}.$$

Theorem 3.9 *Suppose the conditions of Theorem 2.1 are satisfied and E X is finite. On a rich enough probability space, $l_n(\cdot)/b_n$ is almost surely relatively compact in $D[p_0, p_1]$ with respect to the supremum norm and its set of limit points is*

$$\mathcal{G} = \Bigg\{ g_h : g_h(u) = \frac{1}{H_F^{-1}(p_1)} \Bigg(\int\limits_{p_0}^{u} \frac{h(y)}{f(Q(y))}\, \mathrm{d}y - L_F(u) \int\limits_{p_0}^{p_1} \frac{h(y)}{f(Q(y))}\, \mathrm{d}y \Bigg), \quad p_0 \le u \le p_1, \quad h \in \mathcal{H} \Bigg\},$$

where

$$\mathcal{H} = \Bigg\{ h : [p_0, p_1] \to \mathbf{R},\ h(u) = l\big(Q(p_1)\big)^{1/2} (p_1 - u)\ g\bigg(\frac{l(Q(u))}{l(Q(p_1))} \bigg) : g \in S \Bigg\}.$$

Proof Observe that the process $A(\cdot) = (p_1 - \cdot)\, B(Q(\cdot), u)$ over $[p_0, p_1]$ for $u \ge 0$ is equal in distribution to the process

$$\Bigg\{ l\big(Q(p_1)\big)^{1/2} (p_1 - \cdot) u^{1/2}\, W\bigg(\frac{l(Q(\cdot))}{l(Q(p_1))}, u \bigg), \quad u \ge 0 \Bigg\}$$

over $[p_0, p_1]$, where $W(t, u)$ is a standard two-parameter Wiener process. Hence, for $u = n$ where n are natural numbers, $A(y)/b_n$ is relatively compact in $D[p_0, p_1]$ and its set of limit points is $\mathcal{H}$ from the standard functional LIL for a two-parameter Wiener process (Theorem 1.14.1 in Csörgő et al. 1981). (3.8) then gives the desired result. □

Alternatively using Goldie (1977) arguments, the LIL can also be retrived from Gu and Lai (1990). In the absence of both truncation and censorship, Csörgő and Zitikis (1997) showed that at least finite $E\,X^2$ is needed when $p_0 = 0$ and $p_1 = 1$. Lastly, we define the Gini index for the LTRC model as

$$G_F := \frac{\int_{p_0}^{p_1} [\, u - L_F(u)\,]\, \mathrm{d}u}{\int_{p_0}^{p_1} u\, \mathrm{d}u} = \frac{(p_1 - p_0)^2/2 - (\mathrm{CL})_F}{(p_1 - p_0)^2/2}.$$

In the absence of truncation and censorship, $p_0 = 0$ and $p_1 = 1$, the original definition of Gini index is recovered. Since

$$G_F = \frac{(p_1 - p_0)^2 - 2(\mathrm{CL})_F}{(p_1 - p_0)^2} = \frac{(p_1 - p_0)^2 - V_F}{(p_1 - p_0)^2},$$

we have

$$V_n - V_F = V_n - (p_1 - p_0)^2\, [1 - G_F\,] = V_n - (p_1 - p_0)^2\, 2(\mathrm{CL})_F,$$

and the central limit theorem for v_n holds also for $\sqrt{n}[\, V_n - (p_1 - p_0)^2\, (1 - G_F)\,]$ and $\sqrt{n}[\, V_n - (p_1 - p_0)^2\, 2(\mathrm{CL})_F\,]$. In particular, the cumulative total time on test statistics $V_n/(p_1 - p_0)^2$ is a consistent estimator of $1 - G_F$. Thus, an estimator for the Gini index is the Gini statistics defined by

$$G_n := 1 - \frac{V_n}{(p_1 - p_0)^2}.$$

Theorem 3.10 *Suppose the conditions of Theorem 2.1 are satisfied and $E X = \mu$ is finite.*

We have

$$\sqrt{n}[G_n - G_F] \xrightarrow{\mathcal{D}} \frac{(p_1 - p_0)^2}{H_F^{-1}(p_1)} \Bigg(2(\mathrm{CL})_F \int_{p_0}^{p_1} \frac{(p_1 - y)\, B(Q(y))}{f\big(Q(y)\big)}\, \mathrm{d}y - \int_{p_0}^{p_1} \int_{p_0}^{u} \frac{(p_1 - y)\, B(Q(y))}{f\big(Q(y)\big)}\, \mathrm{d}y\, \mathrm{d}u - \int_{p_0}^{p_1} \frac{(p_1 - u)^2\, B\big(Q(u)\big)}{f\big(Q(u)\big)}\, \mathrm{d}u \Bigg).$$

Acknowledgements The author would like to thank the editor and two referees for their valuable suggestions and comments.

References

Csörgő, M. (1983). *Quantile processes with statistical applications.* Philadelphia: SIAM

Csörgő, M., Csörgő, S., Horváth, L. (1986). *An asymptotic theory for empirical reliability and concentration processes.* Lecture Notes in Statistics, vol. 33, Springer Berlin Heidelberg New York:

Csörgő, M., Révész, P. (1981). *Strong approximations in probability and statistics.* New York: Academic

Csörgő, M., Zitikis, R. (1995). Mean residual life processes. *Annals of Statistics 24,* 1717–1739.

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.* New York: Wiley

Csörgő, M., Zitikis, R. (1997). On the rate of strong consistency of Lorenz curves. *Statistics & Probability Letters 34,* 113-121.

Gastwirth, J. L. (1971). A general definition of the Lorenz curve. *Econometrica 39,* 1037–1039.

Gastwirth, J. L. (1972). The estimation of the Lorenz curve and Gini index. *Review of Econometric Statistics 54,* 306–316.

Gijbels, I., Wang, T. J. (1993). Strong representations of the survival function estimator for truncated and censored data with applications. *Journal of Multivariate Analysis 47,* 210–229.

Goldie, C. M. (1977). Convergence theorems for empirical Lorenz curve and their inverses. *Advances in Applied Probability 9,* 765–791.

Gu, M.G., Lai, T.L. (1990). Functional laws of the iterated logarithm for the product-limit estimator of a distribution function under random censorship or truncation. *Annals of Probability 18,* 160–189.

Kendall, M. G., Stuart, A. (1963). *The advanced theory of statistics I.* (2nd. ed.) Charles Griffen and Company, London.

Langberg, N.A., Leon, R.V., Proschan, F. (1980). Characterization of nonparametric classes of life distributions. *Annals of Probability 8,* 1163–1170.

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.

Tsai, W. Y., Jewel, N.P., Wang, M.C. (1987). A note on the product-limit estimator under right censoring and left truncation. *Biometrika 74,* 883–886.

Tse, S. M. (2003). Strong Gaussian approximations in the left truncated and right censored model. *Statistica Sinica. 13,* 275–282.

Tse, S. M. (2005). Quantile processes for left truncated and right censored data. *Annals of Institute of Statistical Mathematics 57,* 61–69.

Woodroofe, M. (1985). Estimating a distribution function with truncated data. *Annals of Statistics 13,* 163–177.

Zhou, Y., Yip, S.F. (1999). A strong representation of the product-limit estimator for left truncated and right censored data. *Journal of Multivariate Analysis 69,* 261–280.