پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان

15th Seminar On Probability and Stochastic Processes
August 30-31, 2025
University of Kurdistan

انجمن آمار ایران

The Law of Iterated Logarithm for the Smooth Estimator of Distribution Function in Length-Biased Data

Masoud Ajami*[1], Mahboubeh Akbari[2]

[1] Department of Statistics, Faculty of Mathematical Sciences, Vali-e-Asr University of Rafsanjan, Rafsanjan, Iran.

[2] Department of Statistics, University of Pretoria, Pretoria, 0002, Gauteng, South Africa.

Abstract

This paper establishes the law of iterated logarithm (LIL) for the smooth estimator of distribution function in length-biased sampling. We prove that under certain regularity conditions, the smooth estimator achieves the same LIL behavior as the non-smooth Cox estimator with rate $2\mu\gamma$. The main result extends previous work on uniform consistency by providing precise asymptotic behavior of the supremum deviation between the smooth estimator and the true distribution function.

Keywords: Length-biased data, smooth estimator, law of iterated logarithm, distribution function, kernel estimation.
Mathematics Subject Classification (2020): 62G05, 62G20.

1 Introduction

Suppose X is a non-negative random variable with probability density function $f(\cdot)$. Length-biased data, which are obtained from length-biased distribution, arise in many practical situations including survival analysis, econometrics, renewal processes, biomedicine and physics. For instance, if the probability of an item being selected for the sample is proportional to its length (X), then the distribution of the observed length is length-biased.

In the case of length-biased data we denote Y_w by Y and $g_w(t,1) = g(t)$ for simplicity. Therefore,

$$g(t) = \frac{tf(t)}{\mu},$$

where $\mu = E(X)$ is assumed finite.

Let f and g have distribution functions (d.f.) F and G, respectively. It is clear that the following relationship between F and its length-biased distribution G holds:

$$G(t) = \frac{1}{\mu}\int_0^t x\, dF(x), \quad t \geq 0.$$

*Corresponding author, m.ajami@vru.ac.ir

Throughout this paper we assume that G is continuous on $\mathbb{R}^+ = [0, \infty)$. From this it can be concluded that F is also continuous. An elementary calculation shows that F is determined uniquely by G, namely

$$F(t) = \mu \int_0^t y^{-1} dG(y), \quad t \geq 0.$$

Let $Y_1, \ldots, Y_n$ be a sample from G. The empirical estimator of F proposed by Cox (1969) is

$$\begin{aligned} F_n^c(t) &= \mu_n \int_0^t y^{-1} dG_n(y) \\ &= \frac{\mu_n}{n} \sum_{i=1}^n \frac{1}{Y_i} I(Y_i \leq t), \end{aligned} \tag{1}$$

where

$$\begin{aligned} \mu_n^{-1} &= \int_0^\infty y^{-1} dG_n(y) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{Y_i}, \end{aligned}$$

and G_n is the empirical d.f. estimator of G given by

$$G_n(t) = \frac{1}{n} \sum_{i=1}^n I(Y_i \leq t),$$

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

Vardi (1982) showed that Cox's estimator is a non-parametric maximum likelihood estimator for the target distribution $F(\cdot)$. Horváth (1985) established strong uniform consistency without rate of $F_n^c(\cdot)$ and showed that under the condition $E(Y^{-1}) < \infty$,

$$\lim_{n \to \infty} \sup_{0 < t < \infty} |F_n^c(t) - F(t)| = 0 \quad a.s. \tag{2}$$

The empirical estimator $F_n^c(\cdot)$ is a step function and discontinuous. In the absence of length-biased data, where F_n is an empirical d.f. of a random sample from F distribution, strong consistency of the smoothed empirical d.f. estimator has been proved by Fernholz (1991). Also, its weak convergence has been proved by Van der Vaart (2000). Since smoothed estimators have better performance compared to non-smoothed estimators, we define the kernel smoothed version of $F_n^s(\cdot)$ through the convolution of two functions $F_n^c(\cdot)$ and CDF $L(\cdot)$.

Using (1) we can define the smoothed estimator F_n^s through the convolution of two functions F_n^c and d.f. L, i.e.

$$\begin{aligned} F_n^s(t) &= \int_0^\infty L\left(\frac{t-y}{h_n}\right) dF_n^c(y) \\ &= \frac{\mu_n}{n} \sum_{i=1}^n \frac{1}{Y_i} L\left(\frac{t - Y_i}{h_n}\right), \end{aligned} \tag{3}$$

where h_n is an arbitrary sequence of smoothing parameters (or bandwidths) that satisfies $\lim_{n \to \infty} h_n = 0$ and L is the cumulative form of a kernel density function K, i.e.

$$L(t) = \int_0^t K(x)\, dx.$$

Smoothed estimators have better performance compared to non-smoothed estimators. Jahanshahi *et al.* (2017) investigated uniform consistency and asymptotic normality of (3) and constructed a one-sample Kolmogorov-type goodness of fit test with this estimator for length-biased data. Also Zamini *et al.* (2024) established a Berry-Esseen type bound for the smoothed estimator in this setting. Recently Ajami *et al.* (2024) have proved the extended Glivenko-Cantelli theorem for $F_n^c(\cdot)$ and $F_n^s(\cdot)$.

The purpose of this paper is to establish the law of iterated logarithm (LIL) for the smooth estimator $F_n^s(\cdot)$ in length-biased sampling, generalizing the strong consistency result in (2). The proof relies on key assumptions about the kernel function, bandwidth, and distribution properties, which are outlined below:

Assumptions.

A1. The kernel function K is symmetric, of bounded variation on $(-1,1)$ with total variation $V_K < \infty$. In addition $K(t) = 0$ if $t \notin (-1,1)$ and satisfies the following conditions:
$\int_{-1}^{1} K(t)dt = 1$,
$\int_{-1}^{1} tK(t)dt = 0$,
$\int_{-1}^{1} t^2K(t)dt = m < \infty$,

A2. $\int_0^\infty u^{-2}G^{1/r}(u)du < \infty$, for some $r > 2$.

A3. $\sup_{0<t<\infty} |f^{'}(t)| \leq c$.

A4. $E(Y^{-2}) = \int_0^\infty y^{-2}\mathrm{d}G(y) < \infty$.

A5. $\int_0^\infty u^{-2-\delta}dG(u) < \infty$, for some $\delta > 0$.

Discussion on the Assumptions. Assumptions **A1** and **A2** are widely used in kernel estimation. Assumption **A3** is not a strong assumption for h_n but it is an essential assumption for validation of the results. If we choose $h_n = n^{-\alpha}$ for $3^{-1} < \alpha < 2^{-1}$, **A1** is satisfied.

2 Main Results

In this section we discuss the LIL property for the smooth estimator F_n^s. Before that, in the following Lemma, we derive the property of the LIL for the non-smooth Cox's estimator.

Lemma 2.1. *Suppose that Assumption* (**A**4) *is satisfied. If we let*

$$b_n = (2\log\log n)^{-1/2},$$

and

$$\alpha_n(t) = \sqrt{n}(F_n^c(t) - F(t))$$

then

$$\limsup_{n\to\infty} \sup_{0<t<\infty} b_n|\alpha_n(t)| = 2\mu\gamma \quad a.s. \tag{4}$$

where $\gamma = \int_0^\infty y^{-2}\mathrm{d}G(y) - \mu^{-2}$.

Proof. We define $H(t) = \int_0^t y^{-1}\mathrm{d}G(y)$ and $H_n(t) = \int_0^t y^{-1}\mathrm{d}G_n(y)$, so we can write

$$H_n(t) - H(t) = \frac{1}{n}\sum_{i=1}^{n}\left[Y_i^{-1}I\left(Y_i \leq t\right) - E\left(Y^{-1}I(Y \leq t)\right)\right], \tag{5}$$

and define

$$\xi_{(t)}^2 = \mathrm{Var}\left(Y_i^{-1}I\left(Y_i \leq t\right)\right) = \int_0^t y^{-2}dG(y) - \left(\int_0^t y^{-1}dG(y)\right)^2. \tag{6}$$

And also we define

$$\gamma^2 = \xi_{(\infty)}^2 = \int_0^\infty y^{-2}dG(y) - \mu^{-2}. \tag{7}$$

Now according to the classical LIL

$$\limsup_{n\to\infty} \sup_{0<t<\infty} \frac{\sqrt{n}[H_n(t) - H(t)]}{\sqrt{2\xi_{(t)}^2\log\log n}} = 1 \quad a.s, \tag{8}$$

and

$$\limsup_{n\to\infty} \frac{\sqrt{n}[\mu - \mu_n]}{\sqrt{2\gamma\mu\log\log n}} = 1 \quad a.s, \tag{9}$$

so under Assumption (**A**4) we have $E(Y^{-2}) < \infty$ then we can write

$$(F_n^c(t) - F(t)) \quad = \quad (1) + (2) + (\mu_n - \mu)[H_n(t) - H(t)] \tag{10}$$

in which

$$(1) = \mu[H_n(t) - H(t)], \tag{11}$$

and

$$(2) = (\mu_n - \mu)H(t). \tag{12}$$

On the other hand, based on (8) and (9) one can see that

$$\limsup_{n\to\infty} \sup_{0<t<\infty} [\mu - \mu_n][H_n(t) - H(t)] = O\Big(\frac{\log\log n}{n}\Big) \quad a.s. \tag{13}$$

Now based on the LIL property for (1) and (2), it can be shown that

$$\begin{aligned}\Big\{\frac{n}{\log\log n}\Big\}^{1/2} |F_n^c(t) - F(t)| &= \mu\Big\{\frac{n}{\log\log n}\Big\}^{1/2}[H_n(t) - H(t)] \\ &+ \Big\{\frac{n}{\log\log n}\Big\}^{1/2}(\mu_n - \mu) + O\Big(\frac{\sqrt{\log\log n}}{n}\Big) \\ &= \sqrt{2}\mu\xi^2_{(t)} + \sqrt{2}\mu\gamma + O\Big(\frac{\sqrt{\log\log n}}{n}\Big) \quad a.s. \end{aligned} \tag{14}$$

So, based on (14) the proof is completed. □

Lemma 2.2. *Suppose that Assumptions* (**A**2), (**A**3) *and* (**A**5) *are satisfied. Then*

$$\sup_{t,s\in J_n} |\alpha_n(t) - \alpha_n(s)| = O(n^{-\frac{1}{4}}\log n), \quad a.s., \tag{15}$$

where $J_n = \{t, s : |t - s| \le Cn^{-\frac{1}{2}}\}$, *and* C *is a positive number.*

Proof. cf. Lemma 2 Fakoor *et al.* (2011). □

Theorem 2.3. *Under Assumptions* (**A**1), (**A**3), $\gamma < \infty$ *and*

$$\lim_{n\to\infty} \frac{\sqrt{n}h_n^2}{\sqrt{\log\log n}} = 0, \tag{16}$$

then

$$\limsup_{n\to\infty} \sup_{0<t<\infty} \Big\{\frac{n}{2\log\log n}\Big\}^{1/2} |F_n^s(t) - F(t)| = 2\mu\gamma \quad a.s.$$

Proof. Using integration by parts, for any $t > 0$ and large n, one can see that

$$\begin{aligned}\sqrt{n}(F_n^s(t) - F(t)) &= \sqrt{n}\int_{-1}^{1} \{F_n^c(t - h_nu) - F(t)\}K(u)du \\ &= \int_{-1}^{1} \{\alpha_n(t - h_nu) - \alpha_n(t)\}K(u)du \\ &+ \int_{-1}^{1} \alpha_n(t)K(u)du + \sqrt{n}\int_{-1}^{1} \{F(t - h_nu) - F(t)\}K(u)du \\ &= \int_{-1}^{1} \{V_1(t) + V_2(t) + V_3(t)\}K(u)du. \end{aligned} \tag{17}$$

First, based on Lemma 2.1, we obtain that

$$\limsup_{n\to\infty}\sup_{t>0} b_n\Big|\int_{-1}^{1} V_2(t)K(u)du\Big| = \limsup_{n\to\infty} \sup_{0<t<\infty} b_n\alpha_n(t) = 2\mu\gamma \quad a.s., \tag{18}$$

and secondly according to (16) and using Taylor expansion, we have

$$\limsup_{n\to\infty}\sup_{t>0} b_n\Big|\int_{-1}^{1} V_3(t)K(u)du\Big| = \sup_{t>0} f''(t)\frac{\sqrt{n}h_n^2}{b_n}\Big|\int_{-1}^{1} u^2K(u)du\Big| = 0. \tag{19}$$

Finally based on Lemma 2.2

$$\begin{aligned}\limsup_{n\to\infty}\sup_{t>0} b_n|\int_{-1}^{1} V_1(t)K(u)du\Big| &\le \limsup_{n\to\infty} b_n \sup_{t>0}\Big|\alpha_n(t - uh_n) - \alpha_n(t)\Big|\int_{-1}^{1} K(u)du \\ &\le \limsup_{n\to\infty} \frac{\log n}{n^{\frac{1}{4}}\sqrt{2\log\log n}}, \end{aligned}$$

so

$$\limsup_{n\to\infty}\sup_{t>0} b_n\Big|\int_{-1}^{1} V_1(t)K(u)du\Big| = 0 \quad a.s. \tag{20}$$

Now, from (17) and using the triangle inequality, it can be stated that

$$\begin{aligned}\limsup_{n\to\infty}\sup_{t>0} b_n n^{1/2}\Big|F_n^s(t)-F(t)\Big| \;&\leq\; \limsup_{n\to\infty}\sup_{t>0} b_n\Big|\int_{-1}^{1} V_2(t)K(u)du\Big| \\ &-\; \limsup_{n\to\infty}\sup_{t>0} b_n\Big|\int_{-1}^{1}\{V_1(t)+V_3(t)\}K(u)du\Big|. \end{aligned}\tag{21}$$

So from (18)-(20) it is obvious that

$$\limsup_{n\to\infty}\sup_{0<t<\infty}\Big\{\frac{2n}{\log\log n}\Big\}^{1/2}|F_n^s(t)-F(t)| \leq 2\mu\gamma \quad a.s. \tag{22}$$

On the other hand, to prove the lower bound, using the reversed triangle inequality we derive from (17),

$$\begin{aligned}\limsup_{n\to\infty}\sup_{t>0} b_n n^{1/2}\Big|F_n^s(t)-F(t)\Big| \;&\geq\; \limsup_{n\to\infty}\sup_{t>0} b_n\Big|\int_{-1}^{1} V_2K(u)du\Big| \\ &-\; \limsup_{n\to\infty}\sup_{t>0} b_n\Big|\int_{-1}^{1}\{V_1+V_3\}K(u)du\Big|. \end{aligned}\tag{23}$$

By reapplying formulas (18)-(20), we have

$$\limsup_{n\to\infty}\sup_{0<t<\infty}\Big\{\frac{2n}{\log\log n}\Big\}^{1/2}|F_n^s(t)-F(t)| \geq 2\mu\gamma \quad a.s. \tag{24}$$

So based on (22) and (24), the proof is completed.

□

References

Ajami M., Zamini R., and Jahanshahi S. M. A. (2024), The extended Glivenko-Cantelli property for Kernel-Smoothed estimator of the cumulative distribution function in the length-biased sampling, *Journal of Mahani Mathematical Research*, **13**(2), 535–545.

Cox D. R. (1969), Some sampling problems in technology. In: Johnson N. L. and Smith H. Jr., *New Developments in Survey Sampling*, Wiley, New York, pp. 506–527.

Fernholz L. T. (1991), Almost sure convergence of smoothed empirical distribution functions, *Scandinavian Journal of Statistics*, **22**, 255–262.

Horváth L. (1985), Estimation from a length-biased distribution, *Statistics & Risk Modeling*, **3**(1–2), 91–114.

Jahanshahi S. M. A., Rad A. and Fakoor V. (2017), Goodness-of-fit test under length-biased sampling, *Communications in Statistics–Theory and Methods*, **46**(15), 7580–7592.

Fakoor, V., Bolbolian,M. and Azarnoosh, H. (2011), Asymptotic behaviors of the Lorenz curve and Gini index in sampling from a length-biased distribution, *Statistics and Probability Letters*, **9**(81), 1425–1435.

Van der Vaart A. W. (2000), *Asymptotic Statistics*, Cambridge University Press.

Vardi Y. (1982), Nonparametric estimation in renewal processes, *The Annals of Statistics*, **10**(3), 772–785.

Zamini R., Ajami M. and Fakoor V. (2024), Berry-Esseen bound for smooth estimator of distribution function under length-biased data, *Communications in Statistics–Theory and Methods*, **53**(5), 1800–1809.