Some Asymptotic Properties of Kernel Density Estimation Under Length-Biased and Right-Censored Data

M. Akbari, M. Akbari, and V. Fakoor

Abstract Among the various methods of density estimation, kernel smoothing is particularly appealing for both its simplicity and its interpretability. The main goal of this article is to study the large-sample properties of the kernel density estimator in the setting of length-biased and right-censored data. The almost sure representation of the distribution function estimator will be the key to obtaining the asymptotic representation for the kernel density estimator. This representation enables us to establish the asymptotic normality and uniform consistency of the estimator. A small simulation study is conducted to show how the estimator behaves for finite samples, and an application is also presented using real data.

Keywords Asymptotic normality · Density estimation · Kernel method · Length-biased and right-censored data · Strong representation

1 Introduction

A prevalent cohort sample consists of individuals who have experienced disease incidence but not failure events at the sampling time. Let A_i be the current age of the i-th subject from onset. An individual would be qualified to be included in the sampling population at the recruitment time only if the survival time of i-th

M. Akbari
Department of Statistics, School of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran

Department of Statistics, University of Birjand, Birjand, Iran

M. Akbari
Department of Statistics, University of Mazandaran, Babolsar, Mazandaran, Iran

V. Fakoor (✉)
Department of Statistics, School of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran
e-mail: fakoor@um.ac.ir

H. Doosti (ed.), *Flexible Nonparametric Curve Estimation*,
https://doi.org/10.1007/978-3-031-66501-1_2

subject from onset (T_i) is greater than A_i. Of course, some of these subjects will have censored failure times due to loss of follow-up or survive until the end of the study. This scheme is a left truncation and right censoring (LTRC) model. Note that in situations where the incidence rate is constant, the right-censored survival data collected on a cohort of prevalent studies are termed length-biased. Therefore, the observed data are referred to as length-biased data with right-censoring (LBRC). Let (V, C) be random variables, where V represents the time from recruitment to death or censoring by the termination of study or lost follow-up during the study period, and C denotes the censoring time from enrollment to censoring occurrence. Here, the censoring is never non-informative because the censoring variable $A + C$ and survival time $T = A + V$ share the same A. Due to this, the Kaplan-Meier estimator is not an appropriate estimator of the length-biased survival function (Vardi, 1989).

Several authors have studied nonparametric estimation for the survival function in the presence of LTRC data. Winter and Földes (1988), Tsai et al. (1987) and Wang (1991), have suggested a solution based on a conditional approach for estimating the lifetime distribution from left-truncated and right-censored data. However, when stationarity holds, the estimators based on the conditional approach for the survivor function lose some information (Asgharian et al., 2002; Huang & Qin, 2011). Vardi (1989) considered the problem of finding the nonparametric maximum likelihood estimate (NPMLE) of the length-biased survivor function for LBRC data. Asgharian et al. (2002) derived an estimator for the survival function using Vardi's (1989) estimator. The NPMLE of both the length-biased and unbiased survival functions for LBRC data assumes that the number of censored and uncensored observations in known a priori, which is not valid for a prevalent cohort study. Vardi (1989) noted that the likelihood remains the same if this assumption is not fulfilled. Asgharian and Wolfson (2005) studied the asymptotic properties of NPMLE of the length-biased survival function under the prevalent cohort sample with follow up.

However, the NPMLE of the survival functions for LBRC is inconvenient in practice because it does not provide a closed-form estimator. Luo and Tsai (2009), and then Huang and Qin (2011), proposed an explicit form for the survival estimator that does not lose much efficiency compared to the NPMLE. Additionally, Wang et al. (2017) suggested a new nonparametric estimator of the survival function of the lifetime that is simpler than Huang and Qin's estimator. One of the quantities used in the modeling of lifetime data is the density function, which is related to the survival function. Nonparametric density estimation can be very helpful in exploratory data analysis, and descriptive features of the density estimate, such as the number of modes, the volatility clustering, the skewed property, and the tail behavior.

Some authors have discussed estimation for the density function via various methods. For instance, Devroye and Györfi (1985), Devroye (1987) and Silverman (1986) have explored this topic. Among the various methods of density estimation, kernel smoothing is particularly appealing for both its simplicity and its interpretability. The pioneer of kernel density estimation were Parzen (1962) and Rosenblatt (1956). The kernel density estimation for censored data has been widely studied, and the most important references include Mielniczuk (1986), Marron and Padgett (1987), Lo et al. (1989), and Stute and Wang (1993), among others.

Improvements for truncated data problems have been made by He and Yang (1998). References dealing with strong consistency under left-truncated and censored data can be found in Gijbels and Wang (1993). The bandwidth needed for this estimator can be found in Sánchez-Sellero et al. (1999). Asgharian et al. (2012) established the problem of estimating the density function under the multiplicative censoring model, which arises as a degenerate case in a prevalent cohort setting. Further, Brunel et al. (2009) proposed an estimator for the density function under bias selection and right censoring using the projection method. In this approach, they assume that the censoring variable is independent of the variable suffering from selection bias.

This paper considers an adaptation of kernel density estimation to LBRC data using Wang et al.'s estimator ($\hat{S}_n$). One of the most important properties of the $\hat{S}_n(\cdot)$, employed in our proofs is the strong representation of this estimator as a sum of i.i.d. random variables plus a remainder term. Wang et al. (2017) obtained a remainder term of order o(1), which we will improve. This convergence rate allows us to establish the strong consistency of the kernel density estimator with a rate.

The outline of this paper is as follows. Section 2 establishes the almost sure representation of the $\hat{S}_n$ with a remainder term of order $O\left(n^{-1}\log\log n\right)$. As an application, the kernel estimator of the density function is proposed and some asymptotic results of this estimator are established in Sect. 3. Section 4 deals with some simulation results for the quality of the kernel density estimation and a demonstration of the density function for the Oscar nominees dataset. The proofs of some theorems and preliminary lemmas are relegated to the Appendix.

2 Strong Representation for the PL Estimator

Let $f(\cdot)$ and $F(\cdot)$ denote the density function and distribution function of T and $S(\cdot) = 1 - F(\cdot)$ be its survival function. It is also assumed that the survival distribution function of C is denoted by $G(\cdot)$. In this scheme, it is further assumed that C is independent of (A, V). Given a random sample

$$\{(A_i, Y_i, \Delta_i),\ i=1,\ldots,n\},$$

where $Y_i = A_i + \min(V_i, C_i) = A_i + \tilde{V}_i$ denotes observed lifetime, and $\Delta_i = I(V_i \le C_i)$ indicates whether a lifetime is censored or not, the PL estimator of the survival function F, as defined in Wang et al. (2017), is as follows:

$$\hat{S}_n(t) = \prod_{u\in[0,t]}\left\{1 - \frac{dH_n(u)}{R_n(u)}\right\}, \tag{1}$$

where

$$H_n(t) = n^{-1} \sum_{i=1}^{n} \Delta_i I(Y_i \le t),$$

$$R_n(t) = (2n)^{-1} \sum_{i=1}^{n} \left\{ I(A_i \le t \le Y_i) + \Delta_i I(\tilde{V}_i \le t \le Y_i) \right\}.$$

The goal of this section is to establish an i.i.d. representation for $\hat{S}_n$ in Eq. (1) and obtain the order of the remainder term. Before stating the main results of this section, we introduce some notations. Let us define the functions

$$R(t) = \frac{1}{2} \mathrm{E}[I(A \le t \le Y) + \Delta I(\tilde{V} \le t \le Y)],$$
$$H(t) = \mathrm{E}[\Delta I(Y \le t)],$$
$$w(t) = \int_0^t G(u)du,$$

which R and H can be consistently estimated by R_n and H_n, respectively. Note that $H(\cdot)$ is a sub-distribution function corresponding to $F(\cdot)$, which represents the proportion of uncensored failure events before time t in the presence of length-bias. Therefore, we have

$$R(t) = \mu^{-1} S(t) w(t), \quad dH(t) = \mu^{-1} w(t) f(t) dt. \tag{2}$$

Thus, in view of (2), the cumulative hazard function of T can be derived as

$$\Lambda(t) = \int_0^t \frac{f(u)}{S(u)} du = \int_0^t \frac{\mu^{-1} f(u) w(u)}{\mu^{-1} S(u) w(u)} du = \int_0^t \frac{dH(u)}{R(u)}.$$

Hence, a natural estimator of Λ, based on n observations $\{(A_i, Y_i, \Delta_i),\ i{=}1, \ldots, n\}$ is given by

$$\hat{\Lambda}_n(t) = \int_0^t \frac{dH_n(u)}{R_n(u)} = \frac{1}{n} \sum_{i:\ Y_i \le t} \frac{\Delta_i}{R_n(Y_i)}. \tag{3}$$

The following theorem provides the i.i.d. representation of $\hat{\Lambda}_n$ to obtain the strong representations for the estimator $\hat{F}_n(t) = 1 - \hat{S}_n(t)$. Let $\tau{=}\inf\{x;\ F(x){=}1\}$ and denote

$$\phi_i(t) = \frac{\Delta_i I(Y_i \le t)}{R(Y_i)} - \frac{1}{2} \int_0^t R^{-2}(u) \{ I(A_i \le u \le Y_i) + \Delta_i I(\tilde{V}_i \le u \le Y_i) \} dH(u). \tag{4}$$

Theorem 1 *If $b < \tau$, then we have uniformly in $0 \le t \le b$,*

$$\hat{\Lambda}_n(t) - \Lambda(t) = n^{-1} \sum_{i=1}^{n} \phi_i(t) + L_n(t),$$

where

$$\sup_{0 \le t \le b} |L_n(t)| = O\left(n^{-3/4} \log n\right) \; a.s.$$

Proof See the Appendix.

Theorem 2 below gives a rate for the strong consistency of the cumulative hazard function estimator.

Theorem 2 *For $b < \tau$, we have*

$$\sup_{0 \le t \le b} |\hat{\Lambda}_n(t) - \Lambda(t)| = O\left(n^{-1/2} (\log \log n)^{1/2}\right), \quad a.s.$$

Proof See the Appendix.

The following theorem is crucial to obtain the convergence rate of the kernel density estimator of f, the density associated with F, in Sect. 3.

Theorem 3 *We have uniformly in $0 \le t \le b < \tau$,*

$$\begin{aligned} \hat{F}_n(t) - F(t) &= (1 - F(t))(\hat{\Lambda}_n(t) - \Lambda(t)) + \mathcal{L}_n(t) \\ &= n^{-1} \sum_{i=1}^{n} (1 - F(t))\phi_i(t) + \mathcal{L}_n(t) \end{aligned}$$

with $\sup_{0 \le t \le b} |\mathcal{L}_n(t)| = O\left(n^{-1} \log \log n\right)$ a.s.

Proof In view of $1 - F(t) = \exp\{-\Lambda(t)\}$ and using Lemma 1, Lemma 2 in the Appendix and Taylor's expansion, we have almost surely

$$\begin{aligned} \hat{F}_n(t) - F(t) =& \bar{F}_n(t) - F(t) + O\left(n^{-1}\right) \\ =& \exp\{-\Lambda(t)\} - \exp\{-\hat{\Lambda}_n(t)\} + O\left(n^{-1}\right) \\ =& \exp\{-\Lambda(t)\}(\hat{\Lambda}_n(t) - \Lambda(t)) \\ & - \frac{\exp\{-\Lambda_n^*(t)\}}{2} (\hat{\Lambda}_n(t) - \Lambda(t))^2 + O\left(n^{-1}\right) \end{aligned}$$

where

$$\min\{\hat{\Lambda}_n(t), \Lambda(t)\} < \Lambda_n^*(t) < \max\{\hat{\Lambda}_n(t), \Lambda(t)\}.$$

The latter inequalities combined with Corollary 2 imply that $\Lambda_n^*(t) \to \Lambda(t)$, a.s., as $n \to \infty$. Since $\exp\{-x\}$ is a continuous function, we get $\exp\{-\Lambda_n^*(t)\} \to \exp\{-\Lambda(t)\}$, $a.s.$ Thus, the proof of the theorem will be completed using Corollary 2 again.

3 Kernel Estimate of Density Function

In this section, we propose kernel density estimation under LBRC data using estimator (1) and establish a strong representation of the estimator under appropriate conditions as an application of Theorem 3.

Let $\{h_n, n \geq 1\}$ be a sequence of positive bandwidths tending to zero, and $K(\cdot)$ be a smooth kernel function. In the sequel, we consider the well-known kernel estimator:

$$\hat{f}_n(t) = h_n^{-1} \int_0^\infty K\left(\frac{t-x}{h_n}\right) d\hat{F}_n(x); \tag{5}$$

we obtain consistency and asymptotic normality of this estimator as an application of the strong representation given in Theorem 3. Assume that the kernel function $K(\cdot)$ is symmetric, of bounded variation on $(-1, 1)$ and satisfies the following conditions:

$$\int_{-1}^1 K(u)du = 1, \quad \int_{-1}^1 uK(u)du = 0, \quad \int_{-1}^1 u^2 K(u)du > 0. \tag{6}$$

According to the second equality in (2), $g(t) = \mu^{-1} f(t)w(t)$ is the density of $H(t)$, and a kernel-type estimate of $g(t)$ is given by $g_n(t) = h_n^{-1} \int K(\frac{t-x}{h_n}) dH_n(x)$. The aim of this section is to give a representation of $\hat{f}_n - \bar{f}_n$ in terms of a sum of random variables whose data are assumed to be LBRC, plus a negligible remainder. For this purpose, we suppose

$$\bar{f}_n(t) = h_n^{-1} \int_0^\infty K\left(\frac{t-x}{h_n}\right) dF(x).$$

Theorem 4 *Suppose the sequence of bandwidths $\{h_n\}$ satisfies*

$$\frac{nh_n}{\log\log n} \to \infty.$$

Under the assumptions of boundary of f on $[a, b]$, we have uniformly in $0 < a \leq t \leq b$,

$$|\hat{f}_n(t) - \bar{f}_n(t)| = \frac{\mu}{w(t)} |g_n(t) - \mathrm{E}[g_n(t)]| + \mathcal{R}_n \quad a.s.$$

where $\mathcal{R}_n = O\left(\frac{\log\log n}{nh_n} \vee \sqrt{\frac{\log\log n}{n}}\right) \quad a.s.$

Proof See the Appendix.

As a result of the Theorem 4, we establish that f_n is uniformly close to f and has asymptotically normal distribution.

Corollary 1 *In addition to the conditions in Theorem 4, assume that $Var(\Delta_i Y_i)$ is bounded. Then by LIL for partial sums, we have*

$$\sup_{a \leq t \leq b} |\hat{f}_n(t) - f(t)| = O\left(h_n^{-1}\sqrt{\frac{\log\log n}{n}} \vee h_n^2\right)$$

Corollary 2 *Under assumption Theorem 4 and assuming*

$$\sigma^2 = \lim_{n\to\infty} nh_n Var[\hat{f}_n(t)] = \frac{\mu}{w(t)} f(t) \int K^2(u)du$$

Slutsky's theorem implies that

$$\sqrt{nh_n}(\hat{f}_n(t) - f(t) \to N(0, \sigma^2)$$

3.1 Bandwidth Selection

An automatic method for determining the optimal window width is cross-validation (CV). We apply the CV approach to obtain an asymptotically optimal bandwidth for the kernel estimator $\hat{f}_n(\cdot)$. A commonly used measure is integrated squared error:

$$\mathrm{ISE}(h_n) := \int_0^\infty (\hat{f}_n(x) - f(x))^2 dx.$$

To get the optimal h_n, it is sufficient to minimize:

$$\int_0^\infty \hat{f}_n^2(x)dx - 2\int_0^\infty \hat{f}_n(x)dF(x) = \int_0^\infty \hat{f}_n^2(x)dx - 2\mathrm{E}(\hat{f}_n). \qquad (7)$$

Since Eq. (7) depends on the unknown F through $\mathrm{E}(\hat{f}_n)$, the leave-one-out method can be used as an estimate of $\mathrm{E}(\hat{f}_n)$. Note that $\hat{F}_n(\cdot)$ is a random step function with jumps only at the observed data points Y_i. Let $\{w_i, i = 1, \ldots, n\}$ be the jump at the observed point Y_i that does not have an explicit form, and let $\hat{f}_{-i}$ be the estimator in (5) calculated with all the observed data points except Y_i. Thus, for large n, we have

$$CV(h_n) = \int_0^\infty \hat{f}_n^2(x)dx - 2\hat{\mathrm{E}}(\hat{f}_n)$$
$$= h_n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j K * K\left(\frac{Y_i - Y_j}{h_n}\right) - 2 \sum_{i=1}^{n} w_i \hat{f}_{-i}(Y_i),$$

in which $K * K(t) = \int_0^\infty K(t-x)K(x)dx$. Hence, we can try to determine an optimal bandwidth h_{opt} for the kernel density estimator by

$$h_{opt} = \arg\min_{h_n} CV(h_n). \tag{8}$$

4 Simulations and Data Analysis

4.1 Monte Carlo Simulations

In this section, we carried out some simulations to evaluate the performance of the proposed estimator for the density function under LBRC data. The evaluation of the simulation results is based on the estimated ISE. We consider three different Weibull distributions for the survival variable T: (1) $T \sim W(2, 2)$, (2) $T \sim W(2, 1)$, and (3) $T \sim W(1, 1)$. The Epanechnikov density function

$$K(x) = 3/4(1 - x^2)I(|x| < 1)$$

was used as the kernel function. To generate the LBRC data, following Huang and Qin (2011), we set the recruitment time to be 100, so the onset variable was simulated from a Uniform$(0, 100)$. To investigate the effect of censoring on the estimator, we consider three levels of censoring. The censoring variable (C) generated from an Uniform(0, c), where c is calculated to have approximately 10%, 30% and 51% censoring for each scheme. We use two sample sizes: $n = 50$ and $n = 200$.

We compare the proposed estimator $\hat{f}_n$ with the kernel density estimators based on the NPMLE of Vardi (1989) (denoted by f_n^V) and the product-limit estimator of the survival function proposed by Tsai et al. (1987) (denoted by f_{TJW}). The estimators f_n^V and f_{TJW} can be obtained by replacing the NPMLE of Vardi (1989) and the distribution estimator of Tsai et al. (1987) with $\hat{F}_n(\cdot)$ in (5), respectively. The estimator f_n^V needs to solve integral equations and is the same as the kernel density estimator proposed by Asgharian et al. (2012) under multiplicative censoring because the likelihood obtained under multiplicative censoring has the same form as the likelihood obtained from a cohort study. The estimator f_{TJW} is an estimator that ignores the assumption that the incidence rate is constant.

In all models, we computed the estimators $\hat{f}_n$ and f_n^V from the reference sample of the LBRC setting. The mean ISE (MISE) of the two estimators was estimated based on 1000 replications for different sample sizes and under different levels of censoring. In each replication, the optimal bandwidth (h_{opt}, say) for the density function $\hat{f}_n(\cdot)$ is selected by minimizing the ISE as a function of h, as stated in Sect. 3.1. The comparison of the two estimators is given in Table 1, which indicates the estimated MISE and integrated bias of $\hat{f}_n$ and f_n^V based on 1000 simulation runs for each model. From this table, it is found that the proposed estimator does not lose much efficiency compared to the kernel density estimator based on NPMLE i.e., f_n^V. Increasing the percentage of censoring increases estimated MISE in almost all scenarios. Nevertheless, when the sample size is large and the level of censoring reaches 51%, the MISEs in the simulation are still small and reasonable. The impact of ignoring the information in the length-biased data can be seen in the last column of Table 1.

Table 1 Simulation results of MISE and integrated bias (in the parenthesis) for nonparametric estimators based on 1000 replications

Unbiased distribution	n	$C\%$	f_n^V (NPMLE)	$\hat{f}_n$ (Proposed)	f_{TJW}
$\exp(-t^2/4)$	50	10	0.054(0.051)	0.058(0.053)	0.068(0.064)
		30	0.058(0.053)	0.062(0.053)	0.073(0.061)
		51	0.086(0.047)	0.077(0.048)	0.099(0.069)
	200	10	0.017(0.049)	0.022(0.050)	0.018(0.050)
		30	0.018(0.049)	0.018(0.050)	0.020(0.056)
		51	0.021(0.052)	0.023(0.052)	0.025(0.053)
$\exp(-t^2)$	50	10	0.086(0.168)	0.093(0.169)	0.105(0.170)
		30	0.101(0.168)	0.103(0.169)	0.115(0.169)
		51	0.107(0.173)	0.120(0.175)	0.163(0.179)
	200	10	0.030(0.164)	0.032(0.166)	0.034(0.165)
		30	0.033(0.163)	0.032(0.165)	0.037(0.166)
		51	0.037(0.164)	0.041(0.165)	0.043(0.166)
$\exp(-t)$	50	10	0.248(0.058)	0.306(0.059)	0.434(0.062)
		30	0.259(0.057)	0.331(0.057)	0.437(0.058)
		51	0.253(0.057)	0.276(0.057)	0.442(0.058)
	200	10	0.136(0.060)	0.136(0.060)	0.163(0.063)
		30	0.176(0.058)	0.225(0.060)	0.260(0.064)
		51	0.149(0.057)	0.185(0.058)	0.180(0.058)

4.2 Real Data

We applied the methods discussed in the previous sections to the Oscar nominees' database, which is available in the supplementary materials by Han et al. (2011). The database collects performers' information from the First Academy Award (May 16, 1929) to July 25, 2007. One of the interesting aims of this section is to analyze the survival of Oscar nominees after their last nomination and estimate the kernel density function for their survival time.

The dataset contains nine variables, but we focused on the most important variables for our purpose: date of birth, date of death, and nomination date. If the nomination date was after the date of death, we dropped the record. We computed age at each nomination, and there are 825 performers in total. We used the age at the last nomination as the truncation time A. It is obvious that the survival time is left-truncated by the age at the last nomination. According to the formal test proposed by Addona and Wolfson (2006) ($W_n = 0.279$, p-value=0.454), it is reasonable to consider this dataset as length-biased data. The censoring date for this database was July 25, 2007, and the censoring rate was about 46%.

Figure 1 shows the density estimates of the lifetime of Oscar nominees using different survival functions and the CV function of the proposed estimator using several different bandwidths. The optimal smoothing parameter in all estimators was chosen to minimize the cross-validation score. It can be seen that the proposed density estimation $\hat{f}_n$ is closer to the density estimation with the NPMLE survival function than f_{TJW}. The resulting CV function for $\hat{f}_n$ shows that the optimal bandwidth in the figure is close to 2.

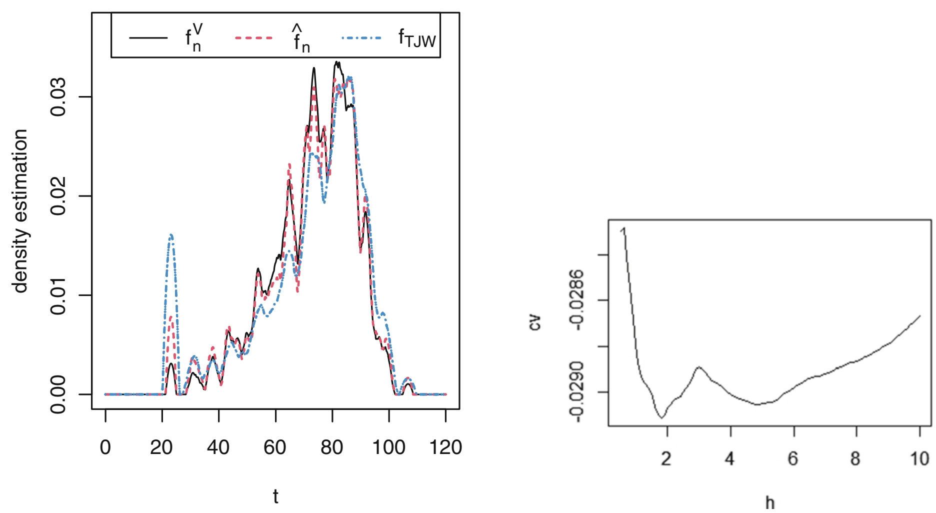

Fig. 1 Kernel density estimators and CV for the Oscar nominees data. Left: density estimator using NPMLE survival function (solid line), proposed density estimator (dashed line), and density estimator using TJW survival estimator (dash-dotted line). Right: cross-validation as a function of bandwidth for proposed density estimator

Appendix

This section first presents some preliminary lemmas that are used in the proofs of the main results, and then gives the proof of the theorems. Regarding Theorem 3, we need to make a slight modification to the product-limit estimator $\hat{F}_n$ to safeguard against $\log 0$ when taking logarithms of $1 - \hat{F}_n(t)$. In the following lemma, we show that the estimator $\bar{F}_n$ behaves in the same way as $\hat{F}_n$, where

$$\bar{F}_n(t) = 1 - \prod_{i:Y_i \le t} \left\{ 1 - \frac{\Delta_i}{nR_n(Y_i) + 1} \right\}. \tag{9}$$

Lemma 1 *If $\int_0^b dH(u)/R^2(u) < \infty$, then uniformly in $0 \le t \le b < \tau$, we have*

$$\bar{F}_n(t) - \hat{F}_n(t) = O\left(n^{-1}\right), \quad a.s.$$

Proof According to (1), one has

$$\bar{F}_n(t) - \hat{F}_n(t) = \prod_{i:Y_i \le t} \left\{ 1 - \frac{\Delta_i}{nR_n(Y_i)} \right\} - \prod_{i:Y_i \le t} \left\{ 1 - \frac{\Delta_i}{nR_n(Y_i) + 1} \right\}.$$

Then applying $|\prod_{i=1}^n c_i - \prod_{i=1}^n d_i| \le \sum_{i=1}^n |c_i - d_i|,\ |c_i|, |d_i| \le 1$, we have

$$\begin{aligned} \bar{F}_n(t) - \hat{F}_n(t) &\le \sum_{i:Y_i \le t} n^{-2} \frac{\Delta_i}{R_n^2(Y_i)} \\ &\le n^{-1} \int_0^b \frac{1}{R_n^2(u)} dH_n(u) \\ &\le n^{-1} \sup_{0<u<b} \left| \frac{R^2(u)}{R_n^2(u)} \right| \int_0^b \frac{1}{R^2(u)} dH_n(u). \end{aligned}$$

Note that since $\sum_{j=1}^n I(A_j \le Y_i \le Y_j) \ge 1$ for any $i = 1, \ldots, n$, conditions $|1 - \frac{\Delta_i}{nR_n(Y_i)}| \le 1$ and $|1 - \frac{\Delta_i}{nR_n(Y_i)+1}| \le 1$ hold. From the strong law of large numbers (SLLN), as $n \to \infty$,

$$\int_0^b \frac{1}{R^2(u)} dH_n(u) \to \int_0^b \frac{1}{R^2(u)} dH(u) < \infty.$$

Further, by LIL for empirical distribution functions, we have

$$\sup_{0<t<b} |R_n(t) - R(t)| = O\left(n^{-1/2}\sqrt{\log\log n}\right) \quad a.s. \tag{10}$$

Consequently, it follows that

$$\sup_{0<u<b}\left|\frac{R^2(u)}{R_n^2(u)}\right| \le 1+\sup_{0<u<b}\frac{|R_n^2(u)-R^2(u)|}{R_n^2(u)}$$
$$\le 1+2\sup_{0<u<b}\frac{|R_n(u)-R(u)|}{R_n^2(u)}$$
$$= 1+O\left(n^{-1/2}\sqrt{\log\log n}\right) \quad a.s. \tag{11}$$

Thus, the desired conclusion follows.

Lemma 2 *Under the assumption of Lemma 1, we have uniformly in* $0 \le t \le b < \tau$,

$$1-\bar{F}_n(t) = e^{-\hat{\Lambda}_n(t)} + O\left(n^{-1}\right) \quad a.s.$$

Proof Using $|e^{-x}-e^{-y}| \le |x-y|,\ x, y \ge 0$ and expanding log expression, we have

$$|1-\bar{F}_n(t) - \exp\{-\hat{\Lambda}_n(t)\}| \le |\log(1-\bar{F}_n(t)) + \hat{\Lambda}_n(t)|$$
$$= \left|\sum_{i:Y_i\le t}\log\left(1-\frac{\Delta_i}{nR_n(Y_i)+1}\right)+\sum_{i:\ Y_i\le t}\frac{\Delta_i}{nR_n(Y_i)}\right|$$
$$= \left|\sum_{i:Y_i\le t}\frac{\Delta_i}{nR_n(Y_i)(nR_n(Y_i)+1)}-\sum_{i:Y_i\le t}\sum_{m=2}^{\infty}\frac{\Delta_i}{m(nR_n(Y_i)+1)^m}\right|, \tag{12}$$

Therefore

$$(12) \le \sum_{i:Y_i\le t}\frac{\Delta_i}{n^2R_n^2(Y_i)}.$$

Thus by SLLN and (11), we get uniformly in $0 \le t \le b < \tau$

$$|1-\bar{F}_n(t) - e^{-\hat{\Lambda}_n(t)}| \le n^{-1}\sup_{0<u<b}\left|\frac{R^2(u)}{R_n^2(u)}\right|\int_0^b\frac{dH_n(u)}{R^2(u)} = O\left(n^{-1}\right) \quad a.s.$$

This completes the proof of the lemma.

Lemma 3 *For any* $b < \tau$, *we have*

$$\sup_{0<t\le b}\left|\int_0^t\left(\frac{1}{R_n(u)}-\frac{1}{R(u)}\right)d[H_n(u)-H(u)]\right| = O\left(n^{-3/4}\log n\right) \quad a.s.$$

Proof Let us divide the interval $[0, b]$ into subintervals $[x_i, x_{i+1}]$; $i = 1, \ldots, \nu_n$ such that $0{=}x_1{<}x_2{<}\cdots{<}x_{\nu_n+1}{=}b$ and $\nu_n = O(n^{1/2}/\sqrt{\log\log n})$. Observe that

$$\begin{aligned}
&\sup_{0\le t\le b}\Big|\int_0^t (\frac{1}{R_n(u)} - \frac{1}{R(u)})d[H_n(u) - H(u)]\Big| \\
&\le \max_{1\le i\le \nu_n} \sum_{j=1}^{i-1} \int_{x_j}^{x_{j+1}} \Big|(\frac{1}{R_n(u)} - \frac{1}{R(u)})d[H_n(u) - H(u)]\Big| \\
&\quad \cdot \max_{1\le i\le \nu_n} \sup_{x_i\le t\le x_{i+1}} \int_{x_i}^{t} \Big|(\frac{1}{R_n(u)} - \frac{1}{R(u)})d[H_n(u) - H(u)]\Big| \\
&=: I + II.
\end{aligned}$$

Equation (10) and the modulus of continuity of the empirical process imply that

$$\begin{aligned}
I \le \sup_{0\le t\le b}\Big|\frac{1}{R_n(t)} - \frac{1}{R(t)}\Big|\nu_n \max_{1\le i\le \nu_n} |H_n(x_{i+1}) - H(x_{i+1}) - H_n(x_i) + H(x_i)| \\
= O(n^{-3/4}\log n) \quad a.s. \qquad (13)
\end{aligned}$$

The next step is to obtain the convergence rate of II.

$$\begin{aligned}
II \le& \max_{1\le i\le \nu_n} \sup_{x_i\le t\le x_{i+1}} \int_{x_i}^{t} \Big|(\frac{1}{R_n(u)} - \frac{1}{R(u)} - \frac{1}{R_n(x_i)} + \frac{1}{R(x_i)})d[H_n(u) - H(u)]\Big| \\
&+ \sup_{0\le t\le b}\Big|\frac{1}{R_n(t)} - \frac{1}{R(t)}\Big| \max_{1\le i\le \nu_n} \sup_{x_i\le t\le x_{i+1}} |H_n(t) - H(t) - H_n(x_i) + H(x_i)| \\
\le& 2 \max_{1\le i\le \nu_n} \sup_{x_i\le t\le x_{i+1}} \Big|\frac{1}{R_n(t)} - \frac{1}{R(t)} - \frac{1}{R_n(x_i)} + \frac{1}{R(x_i)}\Big| + O(n^{-1}\log n), \quad a.s.
\end{aligned}$$

Since $\sup_{0<t<b}(R_n(t) - R(t))^2 = O\left(n^{-1}\log\log n\right)$ a.s., we have

$$\begin{aligned}
III :=& \max_{1\le i\le \nu_n} \sup_{x_i\le t\le x_{i+1}} \Big|\frac{1}{R_n(t)} - \frac{1}{R(t)} - \frac{1}{R_n(x_i)} + \frac{1}{R(x_i)}\Big| \\
&\le \max_{1\le i\le \nu_n} \sup_{x_i\le t\le x_{i+1}} \Big|\frac{R_n(t) - R(t)}{R^2(t)} - \frac{R_n(x_i) - R(x_i)}{R^2(x_i)}\Big| \\
&\quad + O(n^{-1}\log\log n), \quad a.s. \\
&\le \sup_{0\le t\le b} |R_n(t) - R(t)| \max_{1\le i\le \nu_n} \sup_{x_i\le t\le x_{i+1}} \Big|\frac{1}{R^2(t)} - \frac{1}{R^2(x_i)}\Big| \\
&+ \max_{1\le i\le \nu_n} \frac{1}{R^2(x_i)} \max_{1\le i\le \nu_n} \sup_{x_i\le t\le x_{i+1}} |R_n(t) - R(t) - R_n(x_i) + R(x_i)| \\
&+ O(n^{-1}\log\log n), \quad a.s.
\end{aligned}$$

Here we have $\max_i |x_{i+1} - x_i| = O(n^{-1/2}\sqrt{\log\log n})$. An application of the modulus of continuity of the empirical process yield

$$III = O(n^{-3/4}\log n), \quad a.s.$$

Moreover, we conclude

$$II = O(n^{-3/4}\log n), \quad a.s.$$

which together with (13) yields the result.

Proof of Theorem 1 It is easy to check that

$$\begin{aligned}\hat{\Lambda}_n(t) - \Lambda(t) &= \int_0^t \frac{dH_n(u)}{R(u)} - \int_0^t \frac{R_n(u)}{R^2(u)} dH(u) + L_n(t) \\ &= \frac{1}{n}\sum_{i=1}^n \phi_i(t) + L_n(t),\end{aligned}$$

where ϕ_i is defined in (4) and

$$\begin{aligned}L_n(t) &= \int_0^t \left(\frac{1}{R_n(u)} - \frac{1}{R(u)}\right) d[H_n(u) - H(u)] + \int_0^t \frac{(R_n(u) - R(u))^2}{R_n(u)R^2(u)} dH(u) \\ &=: L_{n1}(t) + L_{n2}(t).\end{aligned}$$

In view of Lemma 3, $L_{n1}(t)$ is $O\left(n^{-3/4}\log n\right)$, $a.s.$ Further, by (10) we have

$$\begin{aligned}L_{n2}(t) &\le \sup_{0<u<t} (R_n(u) - R(u))^2 \int_0^t \frac{dH(u)}{R_n(u)R^2(u)} \\ &= O\left(n^{-1}\log\log n\right) \quad a.s.,\end{aligned}$$

This completes the proof of Theorem 1.

Proof of Theorem 2 Applying integration by parts, one can easily get the following

$$\begin{aligned}|\hat{\Lambda}_n(t) - \Lambda(t)| &\le \int_0^t \left|\frac{d[H_n(u) - H(u)]}{R(u)}\right| + \int_0^t \left|\frac{1}{R(u)} - \frac{1}{R_n(u)}\right| dH_n(u) \\ &\le \frac{|H_n(t) - H(t)|}{R(t)} + \int_0^t \frac{|H_n(u) - H(u)|}{R^2(u)} dR(u) \\ &+ \int_0^t \frac{|R_n(u) - R(u)|}{R(u)R_n(u)} dH_n(u)\end{aligned}$$

Thus we obtain the desired result by Eq. (10) and the LIL for the empirical processes $H_n(t) - H(t)$.

Proof of Theorem 4 By integration by parts, Theorem 3 and assumption of bounded variation for kernel function K, we have

$$\begin{aligned}\hat{f}_n(t) - \bar{f}_n(t) &= -h_n^{-1}\int_{B_n}[\hat{F}_n(x) - F(x)]dK\left(\frac{t-x}{h_n}\right)\\ &= -h_n^{-1}\int_{B_n}(1-F(x))[\hat{\Lambda}_n(x) - \Lambda(x)]dK\left(\frac{t-x}{h_n}\right)\\ &\quad + O\left(\frac{\log\log n}{nh_n}\right) \quad a.s., \end{aligned} \tag{14}$$

where $B_n = \{x \geq 0;\ (t-x)/h_n \in (-1, 1)\}$. According to the definition of hazard function and its estimator and using integration by parts, the first term of the right-hand side of the above equation can be viewed as

$$\begin{aligned} J :=& - h_n^{-1}\int_{B_n}(1-F(x))[\hat{\Lambda}_n(x) - \Lambda(x)]dK\left(\frac{t-x}{h_n}\right)\\ =& - h_n^{-1}\int_{B_n}(1-F(x))\int_0^x \frac{d[H_n(u) - H(u)]}{R(u)}dK\left(\frac{t-x}{h_n}\right)\\ &- h_n^{-1}\int_{B_n}(1-F(x))\int_0^x\left(\frac{1}{R_n(u)} - \frac{1}{R(u)}\right)dH_n(u)dK\left(\frac{t-x}{h_n}\right)\\ =&h_n^{-1}\int_{B_n}\frac{H(x) - H_n(x)}{\mu^{-1}w(x)}dK\left(\frac{t-x}{h_n}\right)\\ &+ h_n^{-1}\int_{B_n}(1-F(x))\int_0^x\frac{H_n(u) - H(u)}{R^2(u)}dR(u)dK\left(\frac{t-x}{h_n}\right)\\ &- h_n^{-1}\int_{B_n}(1-F(x))\int_0^x\frac{R_n(u) - R(u)}{R(u)R_n(u)}dH_n(u)dK\left(\frac{t-x}{h_n}\right)\\ =:& J_1 + J_2 + J_3 \end{aligned}$$

Use change of variables to get

$$\begin{aligned} J_1 =&h_n^{-1}\mu\int_{B_n}\frac{H(x) - H_n(x)}{w(t)}dK\left(\frac{t-x}{h_n}\right)\\ &+ h_n^{-1}\mu\int_{B_n}\frac{H_n(x) - H(x)}{w(t)w(x)}(w(x) - w(t))dK\left(\frac{t-x}{h_n}\right) \end{aligned}$$

By LIL for empirical processes and noticing that $\inf_{a\leq t\leq b} w(t) > 0$, the second term of the above expression is of order $\sqrt{\log\log n/n}$, uniformly in $a \leq t \leq b$. Thus, by

Taylor's expansion, we have uniformly in $0 < a \le t \le b$

$$J_1 = \frac{\mu}{w(t)}(g_n(t) - \mathrm{E}[g_n(t)]) + O\left(\sqrt{\frac{\log\log n}{n}}\right) \quad a.s. \tag{15}$$

Denote $C_n(x) = \int_0^x \frac{H_n(u)-H(u)}{R^2(u)} dR(u)$ and $V(K)$ as the total variation of K. As to J_2, put $u = (t-x)/h_n$, so that

$$\begin{aligned} J_2 \le& h_n^{-1}\left|\int_{-1}^{1}(1-F(t-uh_n))\,(C_n(t-uh_n) - C_n(t))dK(u)\right| \\ &+ h_n^{-1}|C_n(t)|\int_{-1}^{1}\left|K'\left(\frac{t-x}{h_n}\right)\right|du \\ \le& \sup_{0<x<\infty}|H_n(x) - H(x)|\left\{\sup_{a\le t\le b}\left|\frac{R'(t)}{R^2(t)}\right|\int_{-1}^{1}|K'(u)|du \right. \\ &\left. + \int_0^\infty \left|\frac{dR(u)}{R^2(u)}\right| V(K)\right\} \\ =& O\left(\sqrt{\frac{\log\log n}{n}}\right) \quad a.s. \end{aligned} \tag{16}$$

Introduce $D_n(x) = \int_0^x \frac{R_n(u)-R(u)}{R(u)R_n(u)} dH_n(u)$. Now, by change of variables

$$\begin{aligned} J_3 \le& h_n^{-1}\left|\int_{-1}^{1}(1-F(t-uh_n))(D_n(t-uh_n) - D_n(t))dK(u)\right| \\ &+ |D_n(t)|\left|\int_{-1}^{1}(1-F(t-uh_n))dK(u)\right| \\ \le& h_n^{-1}\sup_{a\le t\le b}\left|\frac{R_n(t)-R(t)}{R_nR(t)}\right|\left|\left\{2\sup_{0<t\le b}|H_n(t)-H(t)|\right|\int_{-1}^{1}dK(u)\right| \\ &\left. + h_nH'(t^*)\left|\int_{-1}^{1}u\,dK(u)\right| + |H_n(t)|\int_{-1}^{1}\left|K'\left(\frac{t-x}{h_n}\right)\right|du\right\} \end{aligned}$$

where t^* is between t and $t - uh_n$. Thus by (10) we can get uniformly in $0 \le t \le b < \tau$

$$J_3 = O\left(\frac{\log\log n}{nh_n}\right) + O\left(\sqrt{\frac{\log\log n}{n}}\right) \tag{17}$$

Finally the proof of the theorem is completed by Eqs. (14)–(17).

References

Addona, V., & Wolfson, D. B. (2006). A formal test for the stationarity of the incidence rate using data from a prevalent cohort study with follow-up. *Lifetime Data Analysis, 12*, 267–284.

Asgharian, M., Carone, M., & Fakoor, V. (2012). Large-sample study of the kernel density estimators under multiplicative censoring. *The Annals of Statistics, 40*(1), 159–187.

Asgharian, M., M'Lan, C. E., & Wolfson, D. B. (2002). Length-biased sampling with right censoring: An unconditional approach. *Journal of the American Statistical Association, 97*(457), 201–209.

Asgharian, M., & Wolfson, D. B. (2005). Asymptotic behavior of the unconditional NPMLE of the length-biased survivor function from right censored prevalent cohort data. *The Annals of Statistics, 33*(5), 2109–2131.

Brunel, E., Comte, F., & Guilloux, A. (2009). Nonparametric density estimation in presence of bias and censoring. *Test, 18*(1), 166–194.

Devroye, L. (1987). *A course in density estimation*. Birkhauser Boston Inc.

Devroye, L., & Györfi, L. (1985). *Nonparametric density estimation: The L_1 view.* Wiley.

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

Han, X., Small, D. S., Foster, D. P., & Patel, V. (2011). The effect of winning an Oscar award on survival: Correcting for healthy performer survivor bias with a rank preserving structural accelerated failure time model. *The Annals of Applied Statistics, 5*(2A), 746–772.

He, S., & Yang, G. L. (1998). Estimation of the truncation probability in the random truncation model. *The Annals of Statistics, 26*(3), 1011–1027.

Huang, C.-Y., & Qin, J. (2011). Nonparametric estimation for length-biased and right-censored data. *Biometrika, 98*(1), 177–186.

Lo, S., Mack, Y., & Wang, J. (1989). Density and hazard rate estimation for censored data via strong representation of the Kaplan-Meier estimator. *Probability Theory and Related Fields, 80*(3), 461–473.

Luo, X., & Tsai, W. Y. (2009). Nonparametric estimation for right-censored length-biased data: A pseudo-partial likelihood approach. *Biometrika, 96*(4), 873–886.

Marron, J. S., & Padgett, W. J. (1987). Asymptotically optimal bandwidth selection for kernel density estimators from randomly right-censored samples. *The Annals of Statistics, 15*(4), 1520–1535.

Mielniczuk, J. (1986). Some asymptotic properties of kernel estimators of a density function in case of censored data. *The Annals of Statistics, 14*(2), 766–773.

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics, 33*(3), 1065–1076.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics, 27*, 832–837.

Sánchez-Sellero, C., González-Manteiga, W., & Cao, R. (1999). Bandwidth selection in density estimation with truncated and censored data. *Annals of the Institute of Statistical Mathematics, 51*(1), 51–70.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Routledge.

Stute, W., & Wang, J.-L. (1993). The strong law under random censorship. *The Annals of Statistics, 21*(3), 1591–1607.

Tsai, W.-Y., Jewell, N. P., & Wang, M.-C. (1987). A note on the product-limit estimator under right censoring and left truncation. *Biometrika, 74*(4), 883–886.

Vardi, Y. (1989). Multiplicative censoring, renewal processes, deconvolution and decreasing density: Nonparametric estimation. *Biometrika, 76*(4), 751–761.

Wang, M.-C. (1991). Nonparametric estimation from cross-sectional survival data. *Journal of the American Statistical Association, 86*(413), 130–143.

Wang, Y., Zhou, Z., Zhou, X.-H., & Zhou, Y. (2017). Nonparametric and semiparametric estimation of quantile residual lifetime for length-biased and right-censored data. *Canadian Journal of Statistics, 45*(2), 220–250.

Winter, B., & Földes, A. (1988). A product-limit estimator for use with length-biased data. *Canadian Journal of Statistics, 16*(4), 337–355.