Statistics and Its Interface Volume 13 (2020) 221–235

A composite nonparametric product limit approach for estimating the distribution of survival times under length-biased and right-censored data*

Shuqin Fan, Wei Zhao, Alan T. K. Wan†, and Yong Zhou

This paper considers a composite nonparametric product limit estimator for estimating the distribution of survival times when the data are length-biased and right censored. Our method takes into account auxiliary information that frequently arises in survival analysis, and is easier to implement than existing methods for estimating survival functions. We derive a strong representation of the proposed estimator, establish its consistency and asymptotic normality, and derive its convergence rate of approximation. As well, we prove that auxiliary information improves the asymptotic efficiency of the proposed estimator, and provide the values of the composite weights that result in the largest efficiency gain. Our proposed estimator fares well in comparison with other more complex methods in finite samples and offers a clear advantage with respect to computational time.

Keywords and phrases: Biased data, Composite nonparametric estimator, Almost sure representation, Right-censored, Strong consistency, Product-limit estimator.

1. INTRODUCTION

In survival analysis, it is common to use data from prevalent cohorts that exclude individuals who have experienced the failure event before the recruitment time. This results in left truncation of survival times. It is well-known that under the stable disease condition (i.e., the incidence of disease onset is constant over time), the incidence of disease onset follows a stationary Poisson process [20, 6], the truncation time is uniformly distributed, and the sample is length-biased, a term commonly used to refer to the situation where the probability of a survival time being sampled is proportional to its length. The data are also subject to the usual right-censoring as some individuals in the sample may not encounter the failure event before the study ends.

Studies on the estimation of the unbiased survival distribution of length-biased data that utilise the uniform distributional property of the truncation times include [18], [4] and [5]. However, all these methods result in estimators that are computationally intensive as they have no closed-form expressions. [14] proposed an alternative maximum pseudo-partial likelihood estimator (labelled as MPPLE hereafter) that has a closed form expression but at the expense of a slight efficiency loss compared to the established approaches. Their approach is also computationally complex as it entails estimating the distribution function of the censoring time.

[19] developed a truncation product limit estimator (labelled as PLE hereafter) [12] based on a maximisation of the conditional likelihood function. Their approach results in no information loss when the distribution of the truncation time is unspecified, but is less efficient than the full maximum likelihood approach when the parametric form of the truncation time is known. [10] proposed a nonparametric estimator (labelled as HQE hereafter) that takes into account the symmetry between the residual life time and truncation time. Compared to [19]'s PLE, [10]'s method results in a more efficient estimator but has the drawback of being highly complex and computationally challenging. Focusing on the Cox model, [11] developed a composite partial likelihood-based semiparametric approach that yields asymptotically efficient estimators. This approach provides a simple method for estimating the parametric component of the model but is highly complex when applied to the estimation of the baseline hazard function.

In this paper, under length-biased sampling, we develop a simple nonparametric approach of estimating the survival function that is computationally feasible. Our proposed estimator takes the form of the product limit estimator in the spirits of [19], and may be viewed as a composite maximum likelihood estimator within the framework of [11]. We call our estimator the composite nonparametric product limit estimator (CNPLE). We show that the CNPLE is asymptotically efficient. As well, it achieves a convergence approximation rate of $O(n^{-1}\log\log n)$, which is better than the

*Zhou's work was supported by the State Key Program of National Natural Science Foundation of China (71331006), the State Key Program in the Major Research Plan of National Natural Science Foundation of China (91546202). We thank the editor, the associate editor and two referees for helpful comments on an earlier version of the paper. The usual disclaimer applies.
†Corresponding author.

approximation rate of [10]'s HQE [see 16] and identical to that of the estimators of [17] and [21] for left-truncated and right-censored data. We also prove that auxiliary information improves the asymptotic efficiency of the CNPLE, and the biggest efficiency gain is achieved by assigning the composite weight to $w_1 = w_2 = \frac{1}{2}$. We consider the latter a remarkable property of our approach, in addition to the appealing advantage of being simple. In a simulation study based on the same setting as [10], we demonstrate the efficiency gains and computational advantage of our proposed approach over existing approaches in finite samples.

The remainder of the paper is organised as follows. Section 2 discusses the proposed estimator, along with a derivation of its asymptotic properties. Section 3 reports results of a simulation experiment that examines the performance of the proposed estimator in finite samples. An illustration based on data from the Canadian Study of Health and Aging is also presented in the same section. Proofs of theorems are contained in the Appendix.

2. PROPOSED ESTIMATOR AND ITS ASYMPTOTIC PROPERTIES

2.1 Notations and estimators

Let O^0 be the calendar time of the disease onset, T^0 be the time from disease incidence to the event of failure, and $F(\cdot)$ and $f(\cdot)$ be the distribution and density functions of T^0 respectively. Our interest lies in the estimation of $S(\cdot) = 1 - F(\cdot)$, the survival distribution function of T^0, and its corresponding cumulative hazard function $\Lambda(\cdot)$. Let the maximum sampling time $\tau = \inf\{t : Pr(T^0 > t) = 0\}$ be independent of (O^0, T^0). Denote A^0 as the time between disease onset and enrolment, and assume that A^0 and T^0 are independent. An individual can be included in the prevalent cohort only if $T^0 \geq A^0$, meaning that T^0 is left truncated by A^0 with probability $\alpha = P(T^0 \geq A^0) > 0$. Denote A as the observed truncation time, V the residual survival time from the time of enrolment to failure, and $T = A + V$ the failure time. Let the marginal density function of T, A and V be $f_T(t)$, $f_A(t)$ and $f_V(t)$ respectively. It is readily seen that

$$f_T(t) = \frac{tf(t)}{\mu} I(t > 0), \tag{1}$$

where $\mu = \int_0^\infty sf(s)ds$. From [20], the marginal distribution of A and V are identical, i.e.,

$$f_A(t) = f_V(t) = \frac{S(t)}{\mu} I(t > 0). \tag{2}$$

Due to a lack of follow-up and the inevitability of some subjects surviving to the end of the study, observations of survival times in a prevalent cohort are subject to right censoring. Let C be the censoring time with distribution function $G(\cdot)$, and $\delta = I(T \leq A + C)$ be the censoring indicator. Instead of observing the failure time T, we observe the censored failure time $Y = \min(T, A + C) = \min(V, C) + A = \tilde{V} + A$, where $\tilde{V} = \min(V, C)$. The observed data $\{A_i, Y_i, \delta_i\}_{i=1}^n$ are independent and identically distributed (i.i.d.) copies of (A, Y, δ). Clearly, the failure time $T = A + V$ and the total censoring time $A + C$ are dependent, implying that the data are informatively censored.

The proportion of observed failures up to time period t may be represented by the counting process

$$\bar{N}(t) = \frac{1}{n}\sum_{i=1}^n N_i(t) = \frac{1}{n}\sum_{i=1}^n I(Y_i \leq t, \delta_i = 1).$$

As well, the at-risk function $K(t) = E[I(Y \geq t \geq A)]$ can be estimated by

$$\tilde{K}(t) = \frac{1}{n}\sum_{i=1}^n I(Y_i \geq t \geq A_i). \tag{3}$$

[19] proposed the following consistent truncation product-limit estimator (PLE) of the cumulative hazard function $\Lambda(\cdot)$:

$$\tilde{\Lambda}_n(t) = \int_0^t \frac{d\bar{N}(t)}{\tilde{K}(t)}, \tag{4}$$

which results in the consistent survival function estimator

$$\tilde{S}_n(t) = \prod_{u\in[0,t]} \left[1 - d\tilde{\Lambda}_n(u)\right]. \tag{5}$$

One disadvantage of [19]'s approach is that it ignores the auxiliary information (2), the inclusion of which usually leads to efficiency gains in estimation. [10] proposed the following empirical estimator:

$$\begin{aligned}&\tilde{K}_{HQ}(t)\\ =\ &\frac{1}{n}\sum_{i=1}^n I(Y_i \geq t)\\ &- \prod_{u\in[0,t]} \left\{1 - \frac{\sum_{i=1}^n d[I(A_i \geq u) + \delta_i I(\tilde{V}_i \geq u)]}{\sum_{i=1}^n [I(A_i \geq u) + I(\tilde{V}_i \geq u)]}\right\},\end{aligned}$$

which takes into account the auxiliary information of the truncation time A and the residual survival time V having the same distribution. This alternative estimator of $K(t)$ results in

$$\begin{aligned}\tilde{\Lambda}_{HQ}(t) &= \int_0^t \frac{d\bar{N}(t)}{\tilde{K}_{HQ}(t)} \quad \text{and}\\ \tilde{S}_{HQ}(t) &= \prod_{u\in[0,t]} [1 - d\tilde{\Lambda}_{HQ}(t)],\end{aligned}$$

as estimators of the cumulative hazard and survival functions respectively. Because an additional programming step is required to ensure that $\tilde{K}_{HQ}(t) > 0$, $\tilde{K}_{HQ}(t)$ is more difficult to compute than $\tilde{K}(t)$, although it results in estimators of $\Lambda(t)$ and $S(t)$ that are more efficient.

We propose a composite nonparametric product limit estimator (CNPLE) that extends the PLE of [19] to allow for the auxiliary information embedded in (2). Our estimator is not only simpler but also uses auxiliary information in a more natural manner than the HQE of [10]. Now, given that A and V have the same marginal distribution, we have

$$E[I(Y \geq t \geq A)] = E[\delta I(Y \geq t \geq \tilde{V})] = \mu^{-1}S(t)\omega_C(t),$$

where $\omega_C(t) = \int_0^t [1 - G(c)]dc$. Hence we can replace $\tilde{K}(t)$ in (3) by the alternative risk process

$$\hat{K}(t) = \frac{1}{n}\sum_{i=1}^{n}\left[w_1 I(Y_i \geq t \geq A_i) + w_2\delta_i I(Y_i \geq t \geq \tilde{V}_i)\right], \tag{6}$$

that recognises the auxiliary information (2), where w_1 and w_2 are composite weights that satisfy $w_1 + w_2 = 1$. Note that $E[\hat{K}(t)] = E[\tilde{K}(t)] = \mu^{-1}S(t)\omega_C(t)$. The following CNPLE of $\Lambda(\cdot)$ is obtained by replacing $\tilde{K}(t)$ by $\hat{K}(t)$ in (4):

$$\hat{\Lambda}_n(t) = \int_0^t \frac{d\bar{N}(t)}{\hat{K}(t)}. \tag{7}$$

From a computational point of view, our approach has a clear advantage over the HQE of [10] because $\hat{K}(t)$ is, by default, always positive and has a much simpler expression than $\tilde{K}_{HQ}(t)$. The estimator of the survival function based on our approach is given by

$$\hat{S}_n(t) = \prod_{u\in[0,t]}\left\{1 - \sum_{i=1}^{n}\frac{dI(Y_i \leq u, \delta_i = 1)}{n\hat{K}(Y_i)}\right\}. \tag{8}$$

We will show in the subsequent sections that the relatively simplicity of our approach is not offset by any efficiency loss compared to the approach of [10]. In fact, for the special case of the Cox model, there is an equivalence between [11]'s composite maximum likelihood estimator under length-biased and right censored data and the estimator of (7) when $w_1 = w_2 = \frac{1}{2}$. Now, consider the Cox model,

$$\lambda(t) = \lambda_0(t)\exp\{\beta' X\},$$

where $\lambda(t)$ is the derivative of $\Lambda(t)$, $\lambda_0(t)$ is an unspecified baseline hazard function, β is a $p \times 1$ vector of regression parameters, and X is a $p \times 1$ vector of covariates. [11] proposed a composite conditional likelihood equivalent to the product of the truncation likelihood of T conditional on A and the likelihood of T conditional on V, i.e.,

$$\prod_{i=1}^{n}\frac{\{\lambda(Y_i)\exp(\beta' X)\}^{2\delta_i}\exp\{-(1+\delta_i)\Lambda(Y_i)\exp(\beta' X)\}}{\exp[-\{\Lambda(A_i) + \delta_i\lambda(\tilde{V}_i)\exp(\beta' X)\}]}. \tag{9}$$

For any fixed β, (9) attains a maximum at

$$\hat{\Lambda}_{0n}(t) = \sum_{i=1}^{n}\frac{2N_i(t)}{\sum_{i=1}^{n}\exp(\beta' X)K_i(Y_i)},$$

where $K_i(t) = I(Y_i \geq t \geq A_i) + \delta_i I(Y_i \geq t \geq \tilde{V}_i)$. It is readily seen that when $\beta = 0$, $\tilde{\Lambda}_{0n}(t)$ is the same as $\hat{\Lambda}(t)$ when $w_1 = w_2 = \frac{1}{2}$. However, to obtain $\tilde{\Lambda}_{0n}(t)$, one has to compute the composite conditional likelihood, which is intricate, whereas the derivation of $\hat{\Lambda}(t)$ is direct and straightforward.

2.2 Asymptotic properties

The purpose of this section is two-fold. We first establish the asymptotic properties of the estimators $\hat{\Lambda}_n(t)$ and $\hat{S}_n(t)$. We then obtain the strong representation and the rate of the approximation of the CNPLE. Our proofs of results require the following assumptions:

(A1) (A^0, T^0) and the total censoring time $A + C$ are mutually independent.
(A2) O^0 has a constant density function, which guarantees the length-biasedness of the data.
(A3) F, G, F_V and F_A are continuous, where F_V, F_A and H are distribution functions of V, A and Y respectively.
(A4) For any $0 < b < b_H = \sup\{t : H(t) < 1\}$,

$$\int_0^b \frac{dW_1(t)}{K^3(t)} = \mu^2 \int_0^b \frac{dF(t)}{\omega_c^2(t)S(t)^3} < \infty. \tag{10}$$

Assumptions (A1)-(A3) are regular conditions in survival studies. Assumption (A4) is adopted from [21] and is needed in order to obtain the approximation rate of the remainder terms (see also Remark 1). For notational convenience, write

$$L_n(t) = \int_0^t \frac{d\beta_n(u)}{K(u)} - \int_0^t \frac{\hat{K}(u) - K(u)}{K^2(u)}dW_1(u), \tag{11}$$

where $\beta_n(t) = \bar{N}(t) - W_1(t)$ and $W_1(t) = P(Y \leq t, \delta = 1)$. Theorems 1 and 2 below provide the i.i.d. representations of $\hat{\Lambda}_n$ and $\hat{S}_n$.

Theorem 1. *Let Assumptions (A1)-(A4) hold. Then we have, for* $0 \leq t \leq b < b_H$,

$$\hat{\Lambda}_n(t) - \Lambda(t) = L_n(t) + R_n^0(t),$$

with

$$\sup_{0\leq t\leq b}|\hat{\Lambda}_n(t) - \Lambda(t)| = O(n^{-1/2}(\log\log n)^{1/2}), \quad a.s. \tag{12}$$

and

$$\sup_{0\leq t\leq b}|R_n^0(t)| = O(n^{-1}\log\log n), \quad a.s. \tag{13}$$

where

$$R_n^0(t) = \int_0^t \frac{K(u) - \hat{K}(u)}{K^2(u)}d\beta_n(u)$$

$$+\int_0^t \frac{[\hat{K}(u)-K(u)]^2}{\hat{K}(u)K^2(u)}d\bar{N}(u)$$
$$=: \quad R_{n1}^0(t)+R_{n2}^0(t).$$

Theorem 2. *Let Assumptions (A1)-(A4) hold. Then for* $0 \le t \le b < b_H$*, there exist* $R_n(t)$ *and* $R_n'(t)$ *such that*

$$(14) \quad \hat{S}_n(t)-S(t) = -S(t)[\hat{\Lambda}_n(t)-\Lambda(t)]+R_n(t),$$
$$(15) \quad \hat{S}_n(t)-S(t) = -S(t)L_n(t)+R_n'(t),$$

and

$$(16) \quad \sup_{0\le t\le b}|R_n(t)| = \sup_{0\le t\le b}|R_n'(t)| = O(n^{-1}\log\log n). \quad a.s.$$

Remark 1. *Assumption (A4) is a weak integrability condition that remains valid when one or both of* $C=\infty$ *and* $A=0$ *are satisfied, meaning that it continues to hold when the data are only either length-biased or right-censored, or even complete.*

Remark 2. *By Assumption (A4) and Theorem 2,* $\|\hat{S}_n - S\|$ *is bounded asymptotically by* $O(n^{-1/2}(\log\log n)^{1/2})$.

It can be verified from (7) and (8) that the mappings defined by $\hat{\Lambda}_n$ and $\hat{S}_n$ are compactly differentiable with respect to the supremum norm. Hence it makes sense to study the large sample properties of $\hat{\Lambda}_n$ and $\hat{S}_n$ by the functional delta method [13, Theorem 2.8]. Let us define

$$\phi_i(t) = \int_0^t \frac{dI(Y_i\le u,\delta_i=1)}{K(u)} - \int_0^t \frac{w_1 I(A_i\le u\le Y_i)+w_2\delta_i I(\tilde{V}_i\le u\le Y_i)}{K^2(u)}dW_1(u),$$
$$i=1,\dots,n.$$

We can represent $L_n(t)$ as the sum of i.i.d. $\phi_i(t)$, $i=1,\dots,n$, i.e., $L_n(t)=\sum_{i=1}^n \phi_i(t)$. In Theorem 3 and Corollary 1, for the purpose of developing the asymptotic properties of $\hat{S}_n(t)$, we use a Gaussian process $B(u)$, $0\le u<\infty$, with mean zero and covariance matrix $\Sigma_B(t_1,t_2)$. The proofs are outlined in the Appendix.

Theorem 3. *Let Assumptions (A1)-(A4) hold. Then for* $0<t_1,t_2<b_H$,

(i) there exists a Gaussian process $B(u)$, $0\le u<\infty$, *with* $E[B(u)]=0$ *and covariance function*

$$(17) \quad \Sigma_B(t_1,t_2)=E[B(t_1)B(t_2)] = \int_0^{t_1\wedge t_2}\frac{dW_1(u)}{K^2(u)}+\int_0^{t_2}\int_0^{t_1}\frac{H(u,v)}{K^2(u)K^2(v)}dW_1(u)dW_1(v),$$

where

$$H(u,v) = (w_1^2-w_1)\mu^{-1}S(u\vee v)[\omega_C(u\vee v)-\omega_C(|u-v|)] +(w_2^2-w_2)\mu^{-1}S(u\vee v)\omega_C(u\wedge v) +w_1w_2\int_{u\vee v}^{u+v}\frac{f(t)}{\mu}[\omega_C(u)+\omega_C(v)-\omega_C(t-u)-\omega_C(t-v)]dt,$$

such that

$$\sup_{0<u\le b}|\sqrt{n}(\hat{S}_n(u)-S(u))-S(u)B(u)| = O(n^{-1/2}\log n), \quad a.s.$$

for $b<b_H$*;*

(ii) there exists a sequence of i.i.d. Gaussian processes $B_1(u)$, $B_2(u),\cdots$, *with* $EB_i(u)=0$, $i=1,2,\cdots$, *and covariance function (18), such that*

$$\sup_{0<u\le b}|\sqrt{n}(\hat{S}_n(u)-S(u))-n^{-1/2}S(u)\sum_{i=1}^n B_i(u))| = O(n^{-1/2}\log^2 n), \quad a.s.;$$

and

(iii) there exists a two-parameter Gaussian process $\{G(z,u), 0\le z<\infty, u\ge 0\}$ *with mean zero and covariance function*

$$E[G(t_1,u_1)G(t_2,u_2)] = u_1^{-1/2}u_2^{-1/2}(u_1\wedge u_2)S(t_1)S(t_2)\Sigma_B(t_1,t_2),$$

such that

$$\sup_{0<z\le b}|\sqrt{n}\{\hat{S}_n(z)-S(z)\}-G(z,n)| = O(n^{-1/2}\log^2 n). \quad a.s.$$

Corollaries 1 to 3 stated below, whose proofs are given in the Appendix, are direct consequences of Theorem 3.

Corollary 1. *Let Assumption (A4) hold. Then we have, uniformly in* $0\le t\le b<b_H$,

(i) $\hat{S}_n(t)\longrightarrow S(t)$*, and*
(ii) $\sqrt{n}(\hat{S}_n(t)-S(t))\xrightarrow{\mathcal{D}} N(0,\Sigma_t)$,

where $\xrightarrow{\mathcal{D}}$ *denotes convergence in distribution, and*

$$(18) \quad \Sigma_t = S^2(t)\Big[\int_0^t\frac{dW_1(u)}{K^2(u)} + \int_0^t\int_0^t\frac{H(u,v)}{K^2(u)K^2(v)}dW_1(u)dW_1(v)\Big].$$

Remark 3. *The strong approximation results for* $\hat{\Lambda}_n(t)$ *can be obtained in a similar way.*

Corollary 2. *Denote $\tilde{S}_n(t)$ as the survival distribution estimated by $\tilde{\Lambda}_n(t)$ in (4). Then we have*

$$\sqrt{n}(\tilde{S}_n(t) - S(t)) \xrightarrow{\mathcal{D}} N(0, \tilde{\Sigma}_t), \text{ and}$$
$$\Sigma_t \leq \tilde{\Sigma}_t,$$

where $\tilde{\Sigma}_t = S^2(t) \int_0^t \frac{dW_1(u)}{K^2(u)}$.

Corollary 3. *Let Assumption (A4) hold. Then $\hat{S}_n$ attains maximum efficiency when $w_1 = w_2 = \frac{1}{2}$.*

Remark 4. *The assignment of $w_1 = w_2 = \frac{1}{2}$ is equivalent to assigning the same weight to A and V in the risk process $\hat{K}(t)$ in (6). This is justifiable as there exists no prior information to determine which component should be given more weight. We adhere to the choice of $w_1 = w_2 = \frac{1}{2}$ in the simulation and real data studies.*

3. SIMULATION AND REAL DATA ANALYSIS

3.1 A simulation experiment

In this section we use a simulation experiment to identify and compare the finite sample properties of the CNPLE with the PLE, HQE and MPPLE developed by [14]. Following [10], we set ξ, the calendar time of enrolment, to 100, and generated O^0, the calendar onset time, from a uniform distribution over $[1, 100]$. Observations of the underlying failure time T^0 are generated independently from a Weibull distribution with the survival distribution $S(t) = \exp(-t^2/4)$. To obtain a prevalent cohort, we generate observation pairs of (O^0, T^0) repeatedly until there exist n observation pairs that satisfy $O^0 + T^0 \geq \xi$. The censoring time is generated from a uniform distribution such that the censoring rate is approximately 30%, 50%, 70% and 90%. We set $n = 200$ and $n = 800$ and the number of replications to 1000. Our target is to estimate the survival probabilities at various time points t. Based on the results of Corollary 3, we set $w_1 = w_2 = \frac{1}{2}$. We gauge the estimators' sampling performance in terms of the magnitude of empirical bias (BIAS), empirical standard deviation (SE) and empirical mean square error (MSE). In the case of the CNPLE, we also compute the average standard error estimate (ASE), which should be close to the corresponding SE if the proposed method for calculating the standard error based on equation (18). The time (in seconds) required for calculating the estimates is also reported.

Table 1 presents the results for $n = 200$ at the time points $t = 0.94, 1.43, 1.91, 2.54$ that correspond to the survival probabilities of $S(t) = 0.8, 0.6, 0.4, 0.2$ respectively. The relatively small empirical bias obtained in all cases suggest that all four estimators produce accurate estimates. For the CNPLE, the ASEs are all very close to their corresponding SEs, indicating that our standard error estimator performs satisfactorily. With few exceptions, other things being equal, the BIAS, SE, ASE and MSE of estimators generally worsen as the censoring rate increases. As expected, the PLE, which ignores auxiliary information, invariably produces the least efficient estimates in terms of SE and MSE. The MPPLE dominates the PLE in terms of SE and MSE, but it is inferior to the HQE and the CNPLE by the same yardsticks in a large region of the parameter space. Overall, the difference in sampling performance between the HQE and CNPLE are small, with the HQE being marginally better in terms of BIAS and the CNPLE being slightly superior in terms of MSE; in most cases, however, there is little to choose numerically between the two estimators with respect to their sampling performance. This similarity in performance of the HQE and CNPLE is likely attributable to the fact that both estimators incorporate the same auxiliary information, namely, the distributions of truncation time A and the residual survival time V being identical. Having said that, in terms of computing time, the CNPLE is the superior alternative over its more complex rival by a long shot. In most cases, the calculation of the CNPLE requires less than half the time for computing the HQE.

The general comments above also apply in broad terms to the case where $n = 800$, the results for which are reported in Table 2. While there are exceptions, the BIAS, SE, ASE and MSE generally decrease as n increases. In summary, our simulation results demonstrate that our proposed estimator performs well, even under high censoring rates. It also fares well in comparison with other more complex methods with respect to estimator efficiency and computational time.

3.2 A case study

In this section, we apply the four methods considered in Section 3.1 to data from the first wave of the Canadian Study of Health and Aging (CSHA-1). The CSHA-1 is a study on the prevalence of dementia among elderly Canadians conducted in 1991. The original sample contains 1132 elderly subjects (aged 65 or above) identified with dementia by the Modified Mini-Mental State Examination in Canada. Our analysis is based on a subsample of 807 subjects after excluding those with missing data on dementia type or disease onset date, and those who survived more than 20 years after dementia onset due to their low likelihood of developing Alzheimer's disease or vascular dementia [4]. The purpose of our study is to estimate the survival probabilities of subjects at various time points after disease onset. The death rate of these subjects within a ten-year follow-up period is 78% (or 627 deaths). Our observed survival times are left truncated because the sample excludes subjects who died before the study began, length-biased because those who survived longer had a higher chance of being included, and right-censored because some subjects were still alive at the end of the study or lost to follow-up. We applied [1]'s test of the stationarity assumption to the data. The test result confirmed that the incidence of dementia was constant

Table 1. Simulation results of estimators for $n = 200$

t	$S(t)$	PLE Bias	PLE SE	PLE MSE	PLE Time	HQE Bias	HQE SE	HQE MSE	HQE Time
				CensoringRate = 30%					
0.94	0.8	3	513	26	0.8	18	486	24	2.9
1.43	0.6	1	506	26	0.8	18	486	22	2.9
1.91	0.4	5	450	20	0.8	7	399	16	2.9
2.54	0.2	10	336	11	0.8	7	297	9	2.9
				Censoring Rate = 50%					
0.94	0.8	25	580	34	0.8	39	540	29	2.8
1.43	0.6	15	596	36	0.8	29	528	28	2.8
1.91	0.4	13	536	29	0.8	21	462	21	2.8
2.54	0.2	2	416	17	0.8	0	356	13	2.8
				Censoring Rate = 70%					
0.94	0.8	10	560	31	0.9	−1	480	23	3.1
1.43	0.6	26	684	47	0.9	25	594	35	3.1
1.91	0.4	34	703	49	0.9	46	640	41	3.1
2.54	0.2	6	610	37	0.9	12	576	33	3.1
				Censoring Rate = 90%					
0.94	0.8	9	938	88	0.9	47	821	68	3.1
1.43	0.6	37	1051	111	0.9	58	959	92	3.1
1.91	0.4	88	1029	106	0.9	95	962	93	3.1
2.54	0.2	14	842	71	0.9	12	819	67	3.1

t	$S(t)$	MPPLE Bias	MPPLE SE	MPPLE MSE	MPPLE Time	CNPLE Bias	CNPLE SE	CNPLE MSE	CNPLE ASE	CNPLE Time
				Censoring Rate = 30%						
0.94	0.8	21	488	24	129.4	18	486	475	24	1.2
1.43	0.6	24	478	23	129.4	17	468	465	22	1.2
1.91	0.4	15	424	18	129.4	9	400	393	16	1.2
2.54	0.2	6	319	10	129.4	0	296	280	9	1.2
				Censoring Rate = 50%						
0.94	0.8	45	540	29	126.1	42	539	504	29	1.3
1.43	0.6	32	545	30	126.1	35	532	519	28	1.3
1.91	0.4	21	483	24	126.1	30	465	454	22	1.3
2.54	0.2	-6	368	14	126.1	17	352	331	12	1.3
				Censoring Rate = 70%						
0.94	0.8	47	514	27	130.2	48	475	432	23	1.5
1.43	0.6	67	631	40	130.2	69	604	578	37	1.5
1.91	0.4	67	659	44	130.2	67	651	619	43	1.5
2.54	0.2	16	584	34	130.2	18	576	558	33	1.5
				Censoring Rate = 90%						
0.94	0.8	92	880	78	130.8	82	730	611	54	1.5
1.43	0.6	45	994	99	130.8	57	904	836	82	1.5
1.91	0.4	15	990	98	130.8	9	930	891	86	1.5
2.54	0.2	19	808	65	130.8	36	781	749	61	1.5

[1] PLE is the truncation product-limit estimator; HQE is the nonparametric estimator proposed by Huang and Qin (2011); MPPLE is the maximum pseudo-partial-likelihood estimator; CNPLE is our proposed composite nonparametric estimator;

[2] Bias and SE are the empirical bias ($\times 10^4$) and empirical standard deviation ($\times 10^4$) based on 1000 replications respectively; ASE is the average of 1000 standard error estimates based on (18); MSE is the mean square error ($\times 10^4$) based on 1000 replications; Time is the time (seconds) required for computing 1000 estimates by a Mac Pro i5 PC.

Table 2. Simulation results of estimators for $n = 800$

t	$S(t)$	PLE Bias	PLE SE	PLE MSE	PLE Time	HQE Bias	HQE SE	HQE MSE	HQE Time
				Censoring Rate = 30%					
0.94	0.8	5	277	8	87.7	10	276	8	374.5
1.43	0.6	10	265	7	87.7	18	253	6	374.5
1.91	0.4	15	230	5	87.7	22	210	5	374.5
2.54	0.2	7	159	2	87.7	11	142	2	374.5
				Censoring Rate = 50%					
0.94	0.8	7	290	8	96.7	11	274	8	411.4
1.43	0.6	9	293	9	96.7	11	266	7	411.4
1.91	0.4	12	262	7	96.7	12	231	5	411.4
2.54	0.2	3	193	4	96.7	1	166	3	411.4
				Censoring Rate = 70%					
0.94	0.8	3	269	7	105.4	8	233	5	450.9
1.43	0.6	13	341	12	105.4	22	301	9	450.9
1.91	0.4	7	352	12	105.4	17	322	10	450.9
2.54	0.2	10	323	10	105.4	17	313	10	450.9
				Censoring Rate = 90%					
0.94	0.8	18	436	19	113.9	12	370	14	487.6
1.43	0.6	30	517	27	113.9	30	448	20	487.6
1.91	0.4	5	526	28	113.9	8	480	23	487.6
2.54	0.2	3	436	19	113.9	7	416	17	487.6

t	$S(t)$	MPPLE Bias	MPPLE SE	MPPLE MSE	MPPLE Time	CNPLE Bias	CNPLE SE	CNPLE ASE	CNPLE MSE	CNPLE Time
				Censoring Rate = 30%						
0.94	0.8	11	270	7	63030.5	10	276	269	8	139.8
1.43	0.6	18	256	7	63030.5	18	253	240	6	139.8
1.91	0.4	21	223	5	63030.5	23	211	201	5	139.8
2.54	0.2	10	151	2	63030.5	11	142	141	2	139.8
				Censoring Rate = 50%						
0.94	0.8	12	273	8	64655.8	11	275	268	8	183.9
1.43	0.6	13	268	7	64655.8	12	269	268	7	183.9
1.91	0.4	16	239	6	64655.8	15	235	233	6	183.9
2.54	0.2	6	175	3	64655.8	5	169	169	3	183.9
				Censoring Rate = 70%						
0.94	0.8	11	249	6	66235.9	16	230	224	5	229.3
1.43	0.6	22	322	10	66235.9	29	311	297	10	229.3
1.91	0.4	17	335	11	66235.9	24	331	320	11	229.3
2.54	0.2	16	313	10	66235.9	20	317	294	10	229.3
				Censoring Rate = 90%						
0.94	0.8	44	421	18	67489.3	45	363	342	13	271.1
1.43	0.6	51	500	25	67489.3	55	456	452	21	271.1
1.91	0.4	22	510	26	67489.3	27	494	475	24	271.1
2.54	0.2	10	429	18	67489.3	16	423	406	18	271.1

[1] PLE is the truncation product-limit estimator; HQE is the nonparametric estimator proposed by Huang and Qin (2011); MPPLE is the maximum pseudo-partial-likelihood estimator; CNPLE is our proposed composite nonparametric estimator;

[2] Bias and SE are the empirical bias ($\times 10^4$) and empirical standard deviation ($\times 10^4$) based on 1000 replications respectively; ASE is the average of 1000 standard error estimates based on (18); MSE is the mean square error ($\times 10^4$) based on 1000 replications; Time is the time (seconds) required for computing 1000 estimates by a Mac Pro i5 PC.

over time. As in the simulation study, we choose the weight $w_1 = w_2 = \frac{1}{2}$ following the result given in Corollary 3.

Table 3 presents results on the estimated survival probabilities and their standard errors by the PLE, HQE, MPPLE and CNPLE at time points ranging from 12 to 120 months after disease onset. At a given t, the PLE invariably produces the lowest survival probabilities, while HQE generally yields the largest among the four estimators. Generally speaking, the longer is the survival time, the smaller are the standard errors of the probability estimates. Our proposed CNPLE and the MPPLE deliver very similar results, and both approaches yield smaller standard errors than the PLE.

Table 3. Results of real data study: estimated survival probabilities at selected time-points

	PLE		HQE	
Month	Est	SE	Est	SE
12	0.889	0.054	0.927	0.033
24	0.722	0.051	0.781	0.035
36	0.573	0.045	0.633	0.033
48	0.437	0.037	0.495	0.028
72	0.247	0.024	0.297	0.019
96	0.136	0.016	0.174	0.014
120	0.065	0.010	0.095	0.021
	MPPLE		**CNPLE**	
Month	Est	SE	Est	SE
12	0.935	0.028	0.922	0.035
24	0.780	0.034	0.765	0.038
36	0.616	0.033	0.608	0.035
48	0.468	0.029	0.465	0.030
72	0.258	0.020	0.266	0.019
96	0.137	0.014	0.151	0.014
120	0.065	0.009	0.082	0.009

[1] PLE is the truncation product-limit estimator; HQE is the nonparametric estimator proposed by Huang and Qin (2011); MPPLE is the maximum pseudo-partial-likelihood estimator; CNPLE is the proposed composite product-limit estimator; Est and SE are the estimate and its standard error respectively.

APPENDIX A. OUTLINE OF PROOFS

Proof of Theorem 1. Note that

$$\begin{aligned}
&\hat{\Lambda}_n(t) - \Lambda(t)\\
=\ & \int_0^t \frac{d\bar{N}(u) - dW_1(u)}{K(u)} + \int_0^t \frac{d\bar{N}(u)}{\hat{K}(u)} - \int_0^t \frac{d\bar{N}(u)}{K(u)}\\
=\ & \int_0^t \frac{d\beta_n(u)}{K(u)} - \int_0^t \frac{\hat{K}(u) - K(u)}{K^2(u)}\\
& \cdot[dW_1(u) - dW_1(u) + \frac{K(u)}{\hat{K}(u)} d\bar{N}(u)]\\
=\ & L_n(t) - \int_0^t \frac{K(u) - \hat{K}(u)}{K^2(u)} dW_1(u)\\
& - \int_0^t \frac{\hat{K}(u) - K(u)}{K^2(u)} \frac{K(u)}{\hat{K}(u)} d\bar{N}(u)\\
=\ & L_n(t) + \int_0^t \frac{K(u) - \hat{K}(u)}{K^2(u)} [d\bar{N}(u) - dW_1(u)]\\
& - \int_0^t \frac{\hat{K}(u) - K(u)}{K^2(u)} \left[\frac{K(u)}{\hat{K}(u)} - 1\right] d\bar{N}(u)\\
=\ & L_n(t) + \int_0^t \frac{K(u) - \hat{K}(u)}{K^2(u)} d\beta_n(u)\\
& + \int_0^t \frac{[\hat{K}(u) - K(u)]^2}{K^2(u)\hat{K}(u)} d\bar{N}(u)\\
=\ & L_n(t) + R^0_{n1}(t) + R^0_{n2}(t). \qquad (19)
\end{aligned}$$

We divide the proof of Theorem 1 into three parts to focus on the derivations of the supremum norms of $R^0_{n1}(t)$, $R^0_{n2}(t)$ and $\hat{\Lambda}_n(t) - \Lambda(t)$ respectively.

Part 1: Note that

$$\begin{aligned}
& R^0_{n1}(t)\\
=\ & w_1 \int_0^t \frac{K(u) - \tilde{K}(u)}{K^2(u)} d\beta_n(u)\\
& + w_2 \int_0^t \frac{K(u) - \hat{K}_2(u)}{K^2(u)} d\beta_n(u)\\
=:\ & w_1 R^0_{n11}(t) + w_2 R^0_{n12}(t),
\end{aligned}$$

where $\hat{K}_2(t) = \frac{1}{n}\sum_{i=1}^n \delta_i I(\tilde{V}_i \le t \le Y_i)$. We apply the Hoeffding Decomposition for U-statistic from [9] to obtain the asymptotic properties of R^0_{n11} and R^0_{n12}. Let T and S be two independent failure time variables and A be a truncation variable. Given a subset $Q \subset \{1, \ldots, n\}$ with $|Q| = 3$, let H_Q denote the set of all square-integrable random variables of type $g_Q(A_i, T_i, S_i : i \in Q)$, for measurable functions g_Q of $|Q|$ argument such that

$$E(g_Q(A_i, T_i, S_i : i \in Q)|A_j, T_j, S_j, j \in B) = 0, \quad \text{for any } B : |B| < |Q|,$$

where $|\cdot|$ is the cardinality of the corresponding set. Let $\pi_Q f$ denote the projection of f onto H_Q. Then by the orthogonality of the H_Q, the Hájek projection is

$$\begin{aligned}
\pi_Q f(H_1, H_2) &= -f(H_1, H_2) + E_{H_1}[f(H_1, H_2)]\\
&\quad + E_{H_2}[f(H_1, H_2)] - E[f(H_1, H_2)],
\end{aligned}$$

where $H_1, H_2 \in \mathbb{K}^3$. It is readily seen that

$$R^0_{n12}(t) = \frac{1}{n^2} \sum_{i=1}^n \sum_{i=1}^n \pi_Q f_t^{(1)}((A_i, T_i, S_i), (A_j, T_j, S_j)), \qquad (20)$$

where

$$f_t^{(1)}((a_1, t_1, s_1), (a_2, t_2, s_2))$$

$$= \quad -\frac{1}{K^2(t_2)} I\{a_1 \le t_2 \le (t_1 \wedge s_1), a_2 \le t_2 \le (s_2 \wedge t)\}.$$

The subset of $\mathbb{K}^3 \times \mathbb{K}^3 : \{D_t = I\{a_1 \le t_2 \le (t_1 \wedge s_1), a_2 \le t_2 \le (s_2 \wedge t)\}, t \in \mathbb{K}\}$ is a measurable Vapnik-Červonenkis (VC) class of sets, with each set being the intersection of at most six half spaces of $\mathbb{K}^6$ [7]. By the definition of VC subgraph, the class of functions $\{\frac{1}{K^2(t)} I_{D_t} : 0 \le t \le b\}$ is a measurable VC subgraph, and hence a measurable VC class of functions from the definition of $K(t)$ (by a simple modification of Lemma 2.5 II.5 in [15]). By the property of VC class and Assumption (A4), it can be seen that the envelope $H(a_1, t_1, s_1, a_2, t_2, s_2) = \frac{1}{K^2(t_2)} I_{D_\infty}$ has a finite integral, i.e.,

$$EH^2 = \int_0^\infty \frac{dW_1(t)}{K^3(t)S^2(t)} \le \frac{1}{S^2(b)} \int_0^\infty \frac{dW_1(t)}{K^3(t)} < \infty.$$

The Law of the Iterated Logarithm (LIL) according to [3] implies that

$$\sup_{0 \le t \le b} \left| \frac{1}{n^2} \sum_{i=1}^n \sum_{i \ne j} \pi_Q f_t^{(1)}((A_i, T_i, S_i), (A_j, T_j, S_j)) \right| = O(n^{-1} \log\log n). \quad a.s.$$

Now, the sum of the diagonal terms of the U-statistic in (20) is

$$(21) \qquad \frac{1}{n^2} \sum_{i=1}^n \left[\left\{ \frac{1}{K^2(T_i)} + \frac{1}{K(T_i)} \right\} I(A_i \le T_i \le S_i \le t) + \int_0^t \frac{I(A_i \le u \le T_i \le S_i)}{K(u)S(u)} d\bar{N}(u) - \Lambda(t) \right],$$

which is a biased empirical process over a VC class of functions. The Law of Large Numbers [see Theorem 8.3 of 8] implies that the supremum norm of (21) is almost surely $O(n^{-1})$. Therefore, $\sup_{0 \le t \le b} |R^0_{n12}(t)| = O(n^{-1} \log\log n)$.

Similarly, the supremum norm of $R^0_{n11}(t)$ has the same order as $O(n^{-1} \log\log n)$. Hence we have

$$(22) \qquad \sup_{0 \le t \le b} |R^0_{n1}(t)| = O(n^{-1} \log\log n).$$

Part 2: Write

$$(23) \qquad \begin{aligned} &\sup_{0 < t \le b} |R^0_{n2}(t)| \\ = &\sup_{0 < t \le b} \left| \int_0^t \frac{[K(u) - \hat{K}(u)]^2}{K^2(u)\hat{K}(u)} d\bar{N}(u) \right| \\ = &\sup_{0 < t \le b} \left| \int_0^t \left\{ \frac{[K(u) - \hat{K}(u)]^3}{K^3(u)\hat{K}(u)} + \frac{[K(u) - \hat{K}(u)]^2}{K^3(u)} \right\} d\bar{N}(u) \right| \\ \le &\sup_{0 < t \le b} \left| \int_0^t \frac{[K(u) - \hat{K}(u)]^3}{K^3(u)\hat{K}(u)} d\bar{N}(u) \right| + \sup_{0 < t \le b} \left| \int_0^t \frac{[K(u) - \hat{K}(u)]^2}{K^3(u)} d\bar{N}(u) \right| \\ =: &\; J_1 + J_2. \end{aligned}$$

For J_1, we have

$$(24) \qquad \begin{aligned} &J_1 \\ \le &\sup_{0 < t \le b} \left| \frac{K(t) - \hat{K}(t)}{\sqrt{K(t)}} \right|^2 \sup_{0 < t \le b} \int_0^t \frac{|K(u) - \hat{K}(u)|}{K^2(u)\hat{K}(u)} d\bar{N}(u) \\ \le &\sup_{0 < t \le b} \left| \frac{K(t) - \hat{K}(t)}{\sqrt{K(t)}} \right|^2 \sup_{0 < t \le b} |K(t) - \hat{K}(t)| \\ &\cdot \sup_{i: Y_i \le b} \frac{K(Y_i)}{\hat{K}(Y_i)} \int_0^b \frac{d\bar{N}(t)}{K^3(t)} \end{aligned}$$

For the purpose of obtaining the supremum norm of J_1, we first give the order of each term in (24).

(i) Write

$$\hat{K}_2(t) = \frac{1}{n} \sum_{i=1}^n \delta_i I(\tilde{V}_i \le t \le Y_i).$$

By the definition of $\hat{K}(t)$,

$$(25) \qquad \begin{aligned} &\sup_{0 < t \le b} \left| \frac{\hat{K}(t) - K(t)}{\sqrt{K(t)}} \right| \\ \le &\sup_{0 < t \le b} w_1 \left| \frac{\tilde{K}(t) - K(t)}{\sqrt{K(t)}} \right| + \sup_{0 < t \le b} w_2 \left| \frac{\hat{K}_2(t) - K(t)}{\sqrt{K(t)}} \right| \\ =: &\; w_1 M_1 + w_2 M_2. \end{aligned}$$

From Equation (2.12) of [21], for any $\varepsilon > 0$,

$$(26) \qquad M_1 = o(n^{-1/2} (\log n)^{(1+\varepsilon)/2}). \ a.s.$$

For M_2, as

$$\begin{aligned} P(I(A \le t \le Y) = 1) &= E[I(A \le t \le Y)] \\ &= \mu^{-1} S(t) \omega_C(t) \\ &= E[\delta I(\tilde{V} \le t \le Y)) \\ &= P(\delta I(\tilde{V} \le t \le Y) = 1) \end{aligned}$$

and

$$\begin{aligned} P(I(A \le t \le Y) = 0) &= 1 - P(I(A \le t \le Y) = 1) \\ &= 1 - P(\delta I(\tilde{V} \le t \le Y)) \\ &= P(\delta I(\tilde{V} \le t \le Y) = 0), \end{aligned}$$

so $I(A \le t \le Y)$ and $\delta I(\tilde{V} \le t \le Y)$ are identically distributed. Hence $\tilde{K}(t)$ and $\hat{K}_2(t)$ are also identically distributed. Based on this observation, it is easy to show that

$$(27) \quad M_2 = \sup_{0 \le t \le b} \left| \frac{\hat{K}_2(t) - K(t)}{\sqrt{K(t)}} \right| = o(n^{-1/2} (\log n)^{(1+\varepsilon)/2}).$$

Substituting (26) and (27) into (25), we have, for $b < b_H$ and any $\varepsilon > 0$,

$$\sup_{0<t\le b} \frac{\hat{K}(t)-K(t)}{\sqrt{K(t)}} = o(n^{-1/2}(\log n)^{(1+\varepsilon)/2}). \quad a.s. \tag{28}$$

(ii) For $b < b_H$,

$$\begin{aligned}
&\sup_{i:Y_i\le b} \frac{K(Y_i)}{\hat{K}(Y_i)} \qquad (29)\\
= &\sup_{i:Y_i\le b} (E[I(A\le Y_i\le Y)])[n^{-1}\sum_{j=1}^{n}[w_1 I(A_j\le Y_i\le Y_j)\\
&+w_2\delta_j I(\tilde{V}_j\le Y_i\le Y_j)]]^{-1}\\
\le &\sup_{i:Y_i\le b} \frac{E[I(A\le Y_i\le Y)]}{n^{-1}w_1\sum_{j=1}^{n} I(A_j\le Y_i\le Y_j)}\\
= &\sup_{i:Y_i\le b} \frac{K(Y_i)}{w_1\tilde{K}(Y_i)}\\
= &O(\log n).
\end{aligned}$$

Substituting (28) and (29) in (24), using Assumption (A4) and the Law of the Iterated Logarithm for empirical process, we obtain

$$J_1 \le o(n^{-3/2}(\log n)^{2+\epsilon}(\log\log n)^{1/2}). \quad a.s. \tag{30}$$

On the other hand,

$$\begin{aligned}
J_2 &\le \sup_{0<t\le b}|K(t)-\hat{K}(t)|^2 \sup_{0<t\le b}\int_0^t \frac{d\bar{N}(u)}{K^3(u)}\\
&= \sup_{0<t\le b}|K(t)-\hat{K}(t)|^2 \int_0^b \frac{d\bar{N}(t)}{K^3(t)}\\
&= O(n^{-1}\log\log n). \quad a.s. \qquad (31)
\end{aligned}$$

Hence, by (23), (30) and (31), we have

$$\sup_{0<t\le b}|R^0_{n2}(t)| = O(n^{-1}\log\log n). \tag{32}$$

Part 3: Write

$$I_n(t) = \int_0^t \frac{d\bar{N}(u)}{K(u)} - \int_0^t \frac{dW_1(u)}{K(u)}.$$

The process $I_n(t)$ is an empirical process over a VC class that has square integral envelope. Hence for $b < b_H$, $\sup_{0<t\le b}|I_n(t)| = O(n^{-1/2}(\log\log n)^{1/2})$ a.s. [2]. By (29), (31), Assumption (A4) and the Law of the iterated logarithm for empirical processes, for $b < b_H$,

$$\begin{aligned}
\Delta &:= \sup_{0<t\le b}\left|\int_0^t \frac{d\bar{N}(u)}{\hat{K}(u)} - \int_0^t \frac{d\bar{N}(u)}{K(u)}\right|\\
&= \sup_{0<t\le b}\left|\sum_{i=1}^{n}\frac{I(Y_i\le t)}{n\hat{K}(Y_i)} - \sum_{i=1}^{n}\frac{I(Y_i\le t)}{nK(Y_i)}\right|\\
&\le \sup_{i:0<Y_i\le b}\frac{K(Y_i)}{\hat{K}(Y_i)}\left|\int_0^b \frac{[\hat{K}(u)-K(u)]^2}{K^3(u)}d\bar{N}(u)\right|\\
&\quad+\left|\int_0^b \frac{\hat{K}(u)-K(u)}{K^2(u)}d\bar{N}(u)\right|\\
&\le \sup_{i:0<Y_i\le b}\frac{K(Y_i)}{\hat{K}(Y_i)}\left|\int_0^b \frac{[\hat{K}(u)-K(u)]^2}{K^3(u)}d\bar{N}(u)\right|\\
&\quad+\sup_{0\le u\le b}|\hat{K}(u)-K(u)|\left|\int_0^b \frac{d\bar{N}(u)}{K^2(u)}\right|\\
&= O(n^{-1/2}(\log\log n)^{1/2}). \quad a.s.
\end{aligned}$$

Hence we obtain

$$\begin{aligned}
&\sup_{0<t\le b}|\Lambda_n(t)-\Lambda(t)|\\
\le &\sup_{0<t\le b}\left|\int_0^t \frac{d\bar{N}(u)}{\hat{K}(u)} - \int_0^t \frac{d\bar{N}(u)}{K(u)}\right|\\
&+\sup_{0<t\le b}\left|\int_0^t \frac{d\bar{N}(u)}{K(u)} - \int_0^t \frac{dW_1(u)}{K(u)}\right|\\
(33) = &\Delta + \sup_{0<t\le b}|I_n(t)| = O(n^{-1/2}(\log\log n)^{1/2}). \quad a.s.
\end{aligned}$$

Combining (19), (22), (32) and (33), Theorem 1 is proved. □

Proof of Theorem 2. Consider the Taylor series' expansion of $e^{-\hat{\Lambda}_n(t)} - e^{-\Lambda(t)}$,

$$\begin{aligned}
&e^{-\hat{\Lambda}_n(t)} - e^{-\Lambda(t)}\\
= &e^{-\Lambda(t)} + \sum_{i=1}^{\infty}\frac{(-1)^i e^{-\Lambda(t)}}{i!}[\hat{\Lambda}_n(t)-\Lambda(t)]^i - e^{-\Lambda(t)}\\
= &-e^{-\Lambda(t)}[\hat{\Lambda}_n(t)-\Lambda(t)] + \frac{1}{2}e^{-\Lambda^*(t)}[\hat{\Lambda}_n(t)-\Lambda(t)]^2,
\end{aligned}$$

where $|\Lambda^*_n(t)-\Lambda(t)| \le |\Lambda_n(t)-\Lambda(t)|$. Write

$$\bar{S}_n(t) = \prod_{i:Y_i\le t}\left(\frac{n\hat{K}(Y_i)}{n\hat{K}(Y_i)+1}\right)^{\delta_i}$$

in lieu of $\hat{S}_n(t)$ to avoid the logarithm of $\hat{S}_n(t)$'s becoming $\log 0$. It follows that

$$\begin{aligned}
&\hat{S}_n(t) - S(t)\\
= &e^{-\hat{\Lambda}_n(t)} - e^{-\Lambda(t)} + \hat{S}_n(t) - \bar{S}_n(t) + \bar{S}_n(t) - e^{-\hat{\Lambda}_n(t)}\\
= &-e^{-\Lambda(t)}[\hat{\Lambda}_n(t)-\Lambda(t)] + R_n(t),
\end{aligned}$$

with

$$\begin{aligned}
&\sup_{0\le t\le b}|R_n(t)| \qquad (34)\\
\le &\sup_{0\le t\le b}\left|\frac{1}{2}e^{-\Lambda^*(t)}[\hat{\Lambda}_n(t)-\Lambda(t)]^2\right| + \sup_{0\le t\le b}|\hat{S}_n(t)-\bar{S}_n(t)|
\end{aligned}$$

$$
\begin{aligned}
&+\sup_{0\le t\le b}|\bar{S}_n(t)-e^{-\hat{\Lambda}_n(t)}| \\
\le\ & \left(\sup_{0\le t\le b}|\hat{\Lambda}_n(t)-\Lambda(t)|\right)^2+\sup_{0\le t\le b}|\hat{S}_n(t)-\bar{S}_n(t)| \\
&+\sup_{0\le t\le b}|\bar{S}_n(t)-e^{-\hat{\Lambda}_n(t)}| \\
=:\ & R_1+R_2+R_3.
\end{aligned}
$$

It follows straightforwardly from (33) that

$$
R_1=O(n^{-1}\log\log n). \tag{35}
$$

For R_2, given that $\left|\prod_{j=1}^n a_j-\prod_{j=1}^n b_j\right|\le\sum_{j=1}^n|a_j-b_j|$ and $|a_j|,|b_j|\le 1$, we obtain

$$
\begin{aligned}
& R_2 \\
=\ & \sup_{0\le t\le b}|\hat{S}_n(t)-\bar{S}_n(t)| \\
\le\ & \sup_{0\le t\le b}\Big|\prod_{i:Y_i\le t}\left(\frac{n\hat{K}(Y_i)-1}{n\hat{K}(Y_i)}\right)^{\delta_i} \\
& -\prod_{i:Y_i\le t}\left(\frac{n\hat{K}(Y_i)}{n\hat{K}(Y_i)+1}\right)^{\delta_i}\Big| \\
=\ & \sup_{0\le t\le b}\sum_{i:Y_i\le t}\frac{I(\delta_i=1)}{n\hat{K}(Y_i)[n\hat{K}(Y_i)+1]} \\
=\ & \sup_{0\le t\le b}\int_0^t\frac{d\bar{N}(t)}{\hat{K}(u)[n\hat{K}(u)+1]} \\
=\ & \sup_{0\le t\le b}\int_0^t\frac{\hat{K}(u)K(u)+K(u)[K(u)-\hat{K}(u)]}{nK^2(u)\hat{K}(u)[\hat{K}(u)+n^{-1}]}d\bar{N}(u) \\
\le\ & \frac{1}{n}\sup_{0\le t\le b}\left|\int_0^t\frac{d\bar{N}(u)}{K(u)[\hat{K}(u)+n^{-1}]}\right| \\
& +\frac{1}{n}\sup_{0<t\le b}\left|\frac{K(t)-\hat{K}(t)}{\sqrt{K(t)}}\right|\sup_{i:Y_i\le b}\frac{K(Y_i)}{\hat{K}(Y_i)} \\
& \cdot\int_0^b\frac{d\bar{N}(u)}{\sqrt{K^3(u)}[\hat{K}(u)+n^{-1}]} \\
=\ & O(n^{-1}).\quad a.s.
\end{aligned}
\tag{36}
$$

For R_3, applying the Taylor series' expansion to $\log(\bar{S}_n(t))$, we have

$$
\begin{aligned}
\log(\bar{S}_n(t)) &= -\sum_{i:Y_i\le t}\log\left(1+\frac{\delta_i}{n\hat{K}(Y_i)}\right) \\
&= -\sum_{i:Y_i\le t}\frac{\delta_i}{n\hat{K}(Y_i)}-\sum_{j=2}^{\infty}\sum_{i:Y_i\le t}\frac{1}{j}\left(\frac{\delta_i}{n\hat{K}(Y_i)}\right)^j \\
&=: -\hat{\Lambda}_n(t)+D_n(t).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
& \sup_{0\le t\le b}|D_n(t)| \\
=\ & \sup_{0\le t\le b}|\log(\bar{S}_n(t))+\hat{\Lambda}_n(t)| \\
=\ & \sup_{0\le t\le b}\Big|\sum_{i:Y_i\le t}\sum_{m=1}^{\infty}\left\{-\frac{\delta_i}{m[n\hat{K}(Y_i)+1]^m}\right\} \\
& +\sum_{i=1}^{n}\frac{I(Y_i\le t,\delta_i=1)}{n\hat{K}(Y_i)}\Big| \\
=\ & \sup_{0\le t\le b}\sum_{i:Y_i\le t}\Big|\sum_{m=2}^{\infty}\left\{-\frac{\delta_i}{m[n\hat{K}(Y_i)+1]^m}\right\} \\
& +\frac{\delta_i}{n\hat{K}(Y_i)}-\frac{\delta_i}{n\hat{K}(Y_i)+1}\Big| \\
=\ & \sup_{0\le t\le b}\sum_{i:Y_i\le t}\frac{\delta_i}{n\hat{K}(Y_i)[n\hat{K}(Y_i)+1]} \\
& \left|1-\sum_{m=2}^{\infty}\frac{n\hat{K}(Y_i)}{m[n\hat{K}(Y_i)+1]^{m+1}}\right|.
\end{aligned}
\tag{37}
$$

Note that

$$
\begin{aligned}
& \sum_{m=2}^{\infty}\frac{n\hat{K}(Y_i)}{m[n\hat{K}(Y_i)+1]^{m+1}} \\
=\ & \frac{n\hat{K}(Y_i)}{n\hat{K}(Y_i)+1}\sum_{m=2}^{\infty}\frac{[n\hat{K}(Y_i)+1]^{-m}}{m} \\
=\ & \frac{n\hat{K}(Y_i)}{n\hat{K}(Y_i)+1}\left[\log\left(\frac{n\hat{K}(Y_i)+1}{n\hat{K}(Y_i)}\right)-\frac{1}{n\hat{K}(Y_i)+1}\right] \\
=\ & O(1).
\end{aligned}
\tag{38}
$$

Substituting (38) back into (37), we have

$$
\sup_{0\le t\le b}|D_n(t)| = \sup_{0\le t\le b}|\log(\bar{S}_n(t))+\hat{\Lambda}_n(t)|=O(1)
$$

It then follows, for any $b<b_H$, that

$$
\begin{aligned}
& R_3 \\
=\ & \sup_{0\le t\le b}|\bar{S}_n(t)-e^{-\hat{\Lambda}_n(t)}| \\
=\ & \sup_{0\le t\le b}|e^{-\hat{\Lambda}_n(t)}(e^{D_n(t)}-1)| \\
=\ & \sup_{0\le t\le b}\left|e^{-\hat{\Lambda}_n(t)}\sum_{j=1}^{\infty}\frac{[D_n(t)]^j}{j!}\right| \\
\le\ & \sup_{0\le t\le b}|D_n(t)|\sup_{0\le t\le b}\left|\sum_{j=0}^{\infty}\frac{[D_n(t)]^j}{(j+1)!}\right|=O(n^{-1}).
\end{aligned}
\tag{39}
$$

Combining (34), (35), (36) and (39) yields

$$
\begin{aligned}
\sup_{0\le t\le b}|R_n(t)| &\le O(n^{-1}\log\log n)+O(n^{-1})+O(n^{-1}) \\
&= O(n^{-1}\log\log n).
\end{aligned}
$$

This completes the proof of Theorem 2. □

Proof of Theorem 3. By the definition of $L_n(t)$, we have

$$\begin{aligned}
& L_n(t) \\
= \; & \int_0^t \frac{d\bar{N}(u)}{K(u)} - \int_0^t \frac{\hat{K}(u)}{K^2(u)} dW_1(u) \\
= \; & \int_0^t \frac{d\big[\sum_{i=1}^n I(Y_i \le u, \delta_i = 1)\big]}{nK(u)} \\
& - \int_0^t \frac{\sum_{i=1}^n [w_1 I(A_i \le u \le Y_i) + w_2\delta_i I(\tilde{V}_i \le u \le Y_i)]}{nK^2(u)} \\
& \cdot dW_1(u) \\
= \; & \frac{1}{n}\sum_{i=1}^n \int_0^t \frac{dI(Y_i \le u, \delta_i = 1)}{K(u)} \\
& - \frac{1}{n}\sum_{i=1}^n \int_0^t \frac{w_1 I(A_i \le u \le Y_i) + w_2\delta_i I(\tilde{V}_i \le u \le Y_i)}{K^2(u)} \\
& \quad \cdot dW_1(u) \\
=: \; & \frac{1}{n}\sum_{i=1}^n \phi_i(t).
\end{aligned}$$

By the functional delta method [13] and Central Limit Theorem in empirical process, for any $0 < t_1$ and $t_2 < b$, $\sqrt{n}(\hat{S}_n(t) - S(t))$ follows approximately a Gaussian process with mean zero and covariance matrix $\Sigma(t_1, t_2) = S(t_1)S(t_2)E[\phi_1(t_1)\phi_1(t_2)]$. This provides a basis for obtaining the representation of $E[\phi_1(t_1)\phi_1(t_2)]$. Without loss of generality, let $t_1 \le t_2$. For notational convenience, write $\hat{I}_1(t) = w_1 I(A_1 \le t \le Y_1) + w_2\delta_1 I(\tilde{V}_1 \le t \le Y_1)$. Then we have

$$\begin{aligned}
(40) \quad & E[\phi_1(t_1)\phi_1(t_2)] \\
= \; & E\Bigg[\Bigg\{\int_0^{t_1} \frac{dI(Y_1 \le u, \delta_1 = 1)}{K(u)} - \int_0^{t_1} \frac{\hat{I}_1(u)}{K^2(u)} dW_1(u)\Bigg\} \\
& \Bigg\{\int_0^{t_2} \frac{dI(Y_1 \le u, \delta_1 = 1)}{K(u)} - \int_0^{t_2} \frac{\hat{I}_1(u)}{K^2(u)} dW_1(u)\Bigg\}\Bigg] \\
= \; & E\Bigg[\frac{I(Y_1 \le t_1, \delta_1 = 1)}{K(Y_1)} \frac{I(Y_1 \le t_2, \delta_1 = 1)}{K(Y_1)}\Bigg] \\
& - E\Bigg[\frac{I(Y_1 \le t_1, \delta_1 = 1)}{K(Y_1)} \int_0^{t_2} \frac{\hat{I}_1(u)}{K^2(u)} dW_1(u)\Bigg] \\
& - E\Bigg[\frac{I(Y_1 \le t_2, \delta_1 = 1)}{K(Y_1)} \int_0^{t_1} \frac{\hat{I}_1(u)}{K^2(u)} dW_1(u) \\
& + E\Bigg[\int_0^{t_1} \frac{\hat{I}_1(u)}{K^2(u)} dW_1(u) \int_0^{t_2} \frac{\hat{I}_1(u)}{K^2(u)} dW_1(u)\Bigg] \\
= \; & \int_0^{t_1} \frac{dW_1(u)}{K^2(u)} + \int_0^{t_2}\int_0^{t_1} \frac{E[\hat{I}_1(u)\hat{I}_1(v)]}{K^2(u)K^2(v)} dW_1(u)dW_1(v) \\
& - \int_0^{t_1} E\Bigg[\frac{w_1\delta_1 I(A_1 \le u \le Y_1 \le t_2)}{K(Y_1)} \\
& + \frac{w_2\delta_1 I(\tilde{V}_1 \le u \le Y_1 \le t_2)}{K(Y_1)}\Bigg] \frac{dW_1(u)}{K^2(u)} \\
& - \int_0^{t_2} E\Bigg[\frac{w_1\delta_1 I(A_1 \le u \le Y_1 \le t_1)}{K(Y_1)} \\
& \frac{w_2\delta_1 I(\tilde{V}_1 \le u \le Y_1 \le t_1)}{K(Y_1)}\Bigg] \frac{dW_1(u)}{K^2(u)} \\
=: \; & \int_0^{t_1} \frac{dW_1(u)}{K^2(u)} + A(t_1, t_2) - B(t_1, t_2).
\end{aligned}$$

In order to give the exact expression of (40), we first provide the exact expressions of $A(t_1, t_2)$ and $B(t_1, t_2)$.

(i) Note that

$$\begin{aligned}
& E\Bigg[\frac{I(A_1 \le u \le Y_1 \le t_1, \delta_1 = 1)}{K(Y_1)}\Bigg] \\
= \; & E\Bigg[\frac{I(u \le T_1 \le t_1, 0 \le A_1 \le u, T_1 - A_1 \le C_1)}{K(T_1)}\Bigg] \\
= \; & \int_u^{t_1}\int_0^u\int_{t-a}^\infty \frac{f(t)}{\mu K(t)} dG(c)dadt \\
= \; & \int_u^{t_1} \frac{f(t)}{S(t)\omega_C(t)} [\omega_C(t) - \omega_C(t-u)]dt,
\end{aligned}$$

and

$$\begin{aligned}
& E\Bigg[\frac{I(\tilde{V}_1 \le u \le Y_1 \le t_1, \delta_1 = 1)}{K(Y_1)}\Bigg] \\
= \; & E\Bigg[\frac{I(u \le T_1 \le t_1, T_1 - u \le A_1 \le T_1, T_1 - A_1 \le C_1)}{K(T_1)}\Bigg] \\
= \; & \int_u^{t_1} \frac{f(t)}{S(t)\omega_C(t)} \omega_C(u)dt.
\end{aligned}$$

It follows immediately that

$$\begin{aligned}
& E\Bigg[\frac{w_1\delta_1 I(A_1 \le u \le Y_1 \le t_1) + w_2\delta_1 I(\tilde{V}_1 \le u \le Y_1 \le t_1)}{K(Y_1)}\Bigg] \\
= \; & \int_u^{t_1} \frac{f(t)}{S(t)\omega_C(t)} [w_1(\omega_C(t) - \omega_C(t-u)) + w_2\omega_C(u)]dt.
\end{aligned}$$

Therefore,

$$\begin{aligned}
& B(t_1, t_2) \\
= \; & \int_0^{t_1}\int_u^{t_2} \frac{f(t)f(u)[w_1(\omega_C(t) - \omega_C(t-u)) + w_2\omega_C(u)]}{\mu^{-1}S(t)\omega_C(t)S^2(u)\omega_C(u)} dtdu \\
& + \int_0^{t_2}\int_u^{t_1} \frac{f(t)f(u)[w_1(\omega_C(t) - \omega_C(t-u)) + w_2\omega_C(u)]}{\mu^{-1}S(t)\omega_C(t)S^2(u)\omega_C(u)} dtdu.
\end{aligned}$$

(ii) Note that

$$\begin{aligned}
(41) \quad & E[\hat{I}_1(u)\hat{I}_1(v)] \\
= \; & E[\{w_1 I(A_1 \le u \le Y_1) + w_2\delta_1 I(\tilde{V}_1 \le u \le Y_1)\} \\
& \{w_1 I(A_1 \le v \le Y_1) + w_2\delta_1 I(\tilde{V}_1 \le v \le Y_1)\}] \\
= \; & w_1^2 E[I(A_1 \le u, v \le Y_1)] + w_2^2 E[\delta_1 I(\tilde{V}_1 \le u, v \le Y_1)] \\
& + w_1 w_2 E[\delta_1 I(A_1 \le u \le Y_1, \tilde{V}_1 \le v \le Y_1)] \\
& + w_1 w_2 E[\delta_1 I(A_1 \le v \le Y_1, \tilde{V}_1 \le u \le Y_1)] \\
=: \; & I_{11} + I_{12} + I_{13} + I_{14}.
\end{aligned}$$

We can write

$$I_{11} = w_1^2 E[I(A_1 \le u, v \le Y_1)]$$
$$= w_1^2 E[I(A_1 \le u \wedge v, Y_1 \ge u \vee v)]$$
$$= w_1^2 E[I(A_1 \le u \wedge v, T_1 \ge u \vee v, C_1 \ge u \vee v - A_1)]$$
$$= w_1^2 \int_{u\vee v}^{\infty} \int_0^{u\wedge v} \frac{f(t)}{\mu} S_C(u \vee v - a) da dt$$
$$= w_1^2 \int_{u\vee v}^{\infty} \frac{f(t)}{\mu} dt \int_{u\vee v}^{|u-v|} S(y) d(-y)$$
$$(42) \quad = w_1^2 \mu^{-1} S(u \vee v)[\omega_C(u \vee v) - \omega_C(|u-v|)],$$

$$I_{12} = w_2^2 E[\delta_1 I(\tilde{V}_1 \le u, v \le Y_1)]$$
$$= w_2^2 E[I(T_1 - A_1 \le u \wedge v, u \vee v \le T_1 \le A_1 + C_1)]$$
$$= w_2^2 \int_{u\vee v}^{\infty} \int_{t-u\wedge v}^{t} \int_{t-a}^{\infty} \frac{f(t)}{\mu} dG(c) da dt$$
$$= w_2^2 \int_{u\vee v}^{\infty} \frac{f(t)}{\mu} \omega_C(u \wedge v) dt$$
$$(43) \quad = w_2^2 \mu^{-1} S(u \vee v) \omega_C(u \wedge v),$$

and

$$\begin{aligned} I_{13} &= w_1 w_2 E[\delta_1 I(A_1 \le u \le Y_1, \tilde{V}_1 \le v \le Y_1)] \\ &= w_1 w_2 E[I(A_1 \le u \le T_1 \le A_1 + C_1, \\ &\quad T_1 - A_1 \le v \le T_1 \le A_1 + C_1)] \\ &= w_1 w_2 E[I(T_1 - v \le A_1 \le u, T_1 - v \le u, \\ &\quad u \vee v \le T_1, T_1 - A_1 \le C_1)] \\ &= w_1 w_2 \int_{u\vee v}^{u+v} \int_{t-v}^{u} \int_{t-a}^{\infty} \frac{f(t)}{\mu} dG(c) da dt \\ &= w_1 w_2 \int_{u\vee v}^{u+v} \int_{t-v}^{u} \frac{f(t)}{\mu} S_C(t-a) da dt \\ &= w_1 w_2 \int_{u\vee v}^{u+v} \frac{f(t)}{\mu} \int_{v}^{t-u} S(y) d(-y) \end{aligned}$$
$$(44) \quad = w_1 w_2 \int_{u\vee v}^{u+v} \frac{f(t)}{\mu} [\omega_C(v) - \omega_C(t-u)] dt.$$

Similarly, we can show that

$$I_{14} = w_1 w_2 E[\delta_1 I(A_1 \le v \le Y_1, \tilde{V}_1 \le u \le Y_1)]$$
$$(45) \quad = w_1 w_2 \int_{u\vee v}^{u+v} \frac{f(t)}{\mu} [\omega_C(u) - \omega_C(t-v)] dt.$$

Substituting (42), (43), (44) and (45) into (41), we have

$$\begin{aligned} & E[\hat{I}_1(u)\hat{I}_1(v)] \\ = \; & w_1^2 \mu^{-1} S(u \vee v)[\omega_C(u \vee v) - \omega_C(|u-v|)] \\ & + w_2^2 \mu^{-1} S(u \vee v) \omega_C(u \wedge v) \\ & + w_1 w_2 \mu^{-1} \int_{u\vee v}^{u+v} f(t)[\omega_C(v) + \omega_C(u) \\ & \quad - \omega_C(t-u) - \omega_C(t-v)] dt \\ =: \; & H_1(u,v) + H(u,v), \end{aligned}$$

where

$$\begin{aligned} H_1(u,v) &= w_1 \mu^{-1} S(u \vee v)[\omega_C(u \vee v) - \omega_C(|u-v|)] \\ &\quad + w_2 \mu^{-1} S(u \vee v) \omega_C(u \wedge v), \text{and} \\ H(u,v) &= (w_1^2 - w_1) \mu^{-1} S(u \vee v) \\ &\quad [\omega_C(u \vee v) - \omega_C(|u-v|)] \\ &\quad + (w_2^2 - w_2) \mu^{-1} S(u \vee v) \omega_C(u \wedge v) \\ &\quad + w_1 w_2 \mu^{-1} \int_{u\vee v}^{u+v} f(t)[\omega_C(v) + \omega_C(u) \\ &\quad - \omega_C(t-u) - \omega_C(t-v)] dt. \end{aligned}$$

From the definition of $A(t_1, t_2)$, it follows that

$$\begin{aligned} A(t_1,t_2) &= \int_0^{t_1} \int_0^{t_2} \frac{H_1(u,v)}{K^2(u)K^2(v)} dW_1(u) dW_1(v) \\ (46) \quad &\quad + \int_0^{t_1} \int_0^{t_2} \frac{H(u,v)}{K^2(u)K^2(v)} dW_1(u) dW_1(v). \end{aligned}$$

By the exact expression of $B(t_1, t_2)$ in (41), the first term of $A(t_1, t_2)$ in (46) can be expressed as

$$\begin{aligned} & \int_0^{t_1} \int_0^{t_2} \frac{H_1(u,v)}{K^2(u)K^2(v)} dW_1(v) dW_1(u) \\ = \; & \int_0^{t_1} \int_0^{t_2} [\frac{w_1 \mu^{-1} S(u \vee v)[\omega_C(u \vee v) - \omega_C(|u-v|)]}{K^2(u)K^2} \\ & + \frac{w_2 \mu^{-1} S(u \vee v) \omega_C(u \wedge v)}{K^2(u)K^2(v)}] dW_1(v) dW_1(u) \\ = \; & \int_0^{t_1} \int_0^{u} \frac{[w_1(\omega_C(u) - \omega_C(u-v)) + w_2 \omega_C(v)]}{\mu^{-2} S^2(u) \omega_C(u)} \\ & \mu^{-1} S(u) f(u) f(v) dv du \\ & + \int_0^{t_1} \int_u^{t_2} \frac{[w_1(\omega_C(v) - \omega_C(v-u)) + w_2 \omega_C(u)]}{\mu^{-2} S^2(u) \omega_C(u) S^2(v) \omega_C(v)} \\ & \mu^{-1} S(v) f(u) f(v) dv du \\ = \; & \int_0^{t_2} \int_v^{t_1} \frac{[w_1(\omega_C(u) - \omega_C(u-v)) + w_2 \omega_C(v)]}{\mu^{-1} S(u) \omega_C(u) S^2(v) \omega_C(v)} \\ & f(u) f(v) dv du \\ & + \int_0^{t_1} \int_u^{t_2} \frac{[w_1(\omega_C(v) - \omega_C(v-u)) + w_2 \omega_C(u)]}{\mu^{-1} S^2(u) \omega_C(u) S(v) \omega_C(v)} \\ & f(u) f(v) dv du \\ = \; & B(t_1, t_2). \end{aligned}$$

Therefore,

$$\begin{aligned} (47) \quad & A(t_1, t_2) \\ = \; & B(t_1, t_2) + \int_0^{t_1} \int_0^{t_2} \frac{H(u,v)}{K^2(u)K^2(v)} dW_1(u) dW_1(v). \end{aligned}$$

Combining (40) with (47), it is readily seen that

$$E[\phi_1(t_1)\phi_1(t_2)] = \int_0^{t_1} \frac{dW_1(u)}{K^2(u)} + \int_0^{t_2}\int_0^{t_1} \frac{H(u,v)}{K^2(u)K^2(v)} dW_1(u)dW_1(v),$$

where

$$\begin{aligned} & H(u,v) \\ = \; & (w_1^2 - w_1)\mu^{-1}S(u \vee v)[\omega_C(u \vee v) - \omega_C(|u-v|)] \\ & +(w_2^2 - w_2)\mu^{-1}S(u \vee v)\omega_C(u \wedge v) \\ & +w_1 w_2 \mu^{-1}\int_{u\vee v}^{u+v} f(t) \\ & \cdot[\omega_C(v) + \omega_C(u) - \omega_C(t-u) - \omega_C(t-v)]dt. \end{aligned} \tag{48}$$

This completes the proof of Theorem 3. □

Proof of Corollary 1. (i) By Theorem 2 and recognising that $E[L_n(t)] = 0$, we can show that $\hat{S}_n(t) - S(t) \longrightarrow 0$; (ii) It follows directly from part (i) of Theorem 3 by setting $t_1 = t_2 = t$. □

Proof of Corollary 2. (i) We first prove that $\sqrt{n}(\tilde{S}_n(t) - S(t)) \xrightarrow{\mathcal{D}} N(0, \tilde{\Sigma}_t)$. Write

$$\tilde{L}_n(t) = \int_0^t \frac{d\beta_n(u)}{K(u)} - \int_0^t \frac{\tilde{K}(u) - K(u)}{K^2(u)} dW_1(u).$$

Similar to the proofs of Theorem 1 and 2, for $0 \le t \le b \le b_H$, we obtain

$$\begin{aligned} \tilde{\Lambda}_n(t) - \Lambda(t) &= \tilde{L}_n(t) + \tilde{R}_n^0(t), \\ \tilde{S}_n(t) - S(t) &= -S(t)[\tilde{\Lambda}_n(t) - \Lambda(t)] + \tilde{R}_n(t), \quad \text{and} \\ \tilde{S}_n(t) - S(t) &= -S(t)\tilde{L}_n(t) + \tilde{R}_n'(t), \end{aligned}$$

with

$$\begin{aligned} \sup_{0\le t\le b} |\tilde{\Lambda}_n(t) - \Lambda(t)| &= O(n^{-1/2}(\log\log n)^{1/2}), \quad a.s. \\ \sup_{0\le t\le b} |\tilde{R}_n^0(t)| &= \sup_{0\le t\le b} |\tilde{R}_n(t)| \\ &= \sup_{0\le t\le b} |\tilde{R}_n'(t)| = O(n^{-1}\log\log n). \end{aligned}$$

By the definition of $\tilde{L}_n(t)$, we have

$$\begin{aligned} \tilde{L}_n(t) &= \int_0^t \frac{d\bar{N}(u)}{K(u)} - \int_0^t \frac{\tilde{K}(u)}{K^2(u)} dW_1(u) \\ &= \int_0^t \frac{d\big[\sum_{i=1}^n I(Y_i \le u, \delta_i = 1)\big]}{nK(u)} \\ &\quad - \int_0^t \frac{\sum_{i=1}^n I(A_i \le u \le Y_i)}{nK^2(u)} dW_1(u) \\ &= \frac{1}{n}\sum_{i=1}^n \Big[\int_0^t \frac{dI(Y_i \le u, \delta_i = 1)}{K(u)} \\ &\quad - \int_0^t \frac{I(A_i \le u \le Y_i)}{K^2(u)} dW_1(u)\Big] \\ &=: \frac{1}{n}\sum_{i=1}^n \tilde{\phi}_i(t), \end{aligned}$$

where $\tilde{\phi}_i(t) = \int_0^t \frac{dI(Y_i \le u, \delta_i=1)}{K(u)} - \int_0^t \frac{I(A_i\le u\le Y_i)}{K^2(u)} dW_1(u)$. By the functional delta method [13] and Central Limit Theorem in empirical process, for any $0 < t_1$ and $t_2 < b$, $\sqrt{n}(\tilde{S}_n(t) - S(t))$ follows approximately a Gaussian process with mean zero and covariance matrix $\tilde{\Sigma}(t_1,t_2) = S(t_1)S(t_2)E[\tilde{\phi}_1(t_1)\tilde{\phi}_1(t_2)]$. This provides a basis for obtaining the representation of $E[\tilde{\phi}_1(t_1)\tilde{\phi}_1(t_2)]$. Without loss of generality, let $t_1 \le t_2$. Then we have

$$\begin{aligned} & E[\tilde{\phi}_1(t_1)\tilde{\phi}_1(t_2)] \\ =\; & E\Big[\Big\{\int_0^{t_1} \frac{dI(Y_1 \le u, \delta_1 = 1)}{K(u)} - \int_0^{t_1} \frac{I(A_1 \le u \le Y_1)}{K^2(u)} dW_1(u)\Big\} \\ & \Big\{\int_0^{t_2} \frac{dI(Y_1 \le u, \delta_1 = 1)}{K(u)} - \int_0^{t_2} \frac{I(A_1 \le u \le Y_1)}{K^2(u)} dW_1(u)\Big\}\Big] \\ =\; & E\Big[\frac{I(Y_1 \le t_1, \delta_1 = 1)}{K(Y_1)}\frac{I(Y_1 \le t_2, \delta_1 = 1)}{K(Y_1)} \\ & - \frac{I(Y_1 \le t_1, \delta_1 = 1)}{K(Y_1)}\int_0^{t_2} \frac{I(A_1 \le u \le Y_1)}{K^2(u)} dW_1(u) \\ & - \frac{I(Y_1 \le t_2, \delta_1 = 1)}{K(Y_1)}\int_0^{t_1} \frac{I(A_1 \le u \le Y_1)}{K^2(u)} dW_1(u) \\ & + \int_0^{t_1} \frac{I(A_1 \le u \le Y1)}{K^2(u)} dW_1(u) \\ & \int_0^{t_2} \frac{I(A_1 \le u \le Y_1)}{K^2(u)} dW_1(u)\Big] \\ =:\; & \int_0^{t_1} \frac{dW_1(u)}{K^2(u)} + \tilde{A}(t_1,t_2) - \tilde{B}(t_1,t_2). \end{aligned} \tag{49}$$

In order to give the exact expression of (49), we next calculate exact expression of $\tilde{A}(t_1,t_2)$ and $\tilde{B}(t_1,t_2)$.

Note that

$$\begin{aligned} & E\Big[\frac{I(A_1 \le u \le Y_1 \le t_1, \delta_1 = 1)}{K(Y_1)}\Big] \\ =\; & E\Big[\frac{I(u \le T_1 \le t_1, 0 \le A_1 \le u, T_1 - A_1 \le C_1)}{K(T_1)}\Big] \\ =\; & \int_u^{t_1}\int_0^u\int_{t-a}^\infty \frac{f(t)}{\mu K(t)} dG(c)dadt \\ =\; & \int_u^{t_1} \frac{f(t)}{S(t)\omega_C(t)}[\omega_C(t) - \omega_C(t-u)]dt. \end{aligned}$$

Therefore,

$$\text{(50)}\quad \begin{aligned} &\tilde{B}(t_1,t_2)\\ =\ & \int_0^{t_1}\int_u^{t_2} \frac{f(t)f(u)[\omega_C(t)-\omega_C(t-u)]}{\mu^{-1}S(t)\omega_C(t)S^2(u)\omega_C(u)}dtdu\\ &+\int_0^{t_2}\int_u^{t_1} \frac{f(t)f(u)[\omega_C(t)-\omega_C(t-u))]}{\mu^{-1}S(t)\omega_C(t)S^2(u)\omega_C(u)}dtdu. \end{aligned}$$

On the other hand,

$$\text{(51)}\quad \begin{aligned} & E[I(A_1\le u\le Y_1)I(A_1\le v\le Y_1)]\\ =\ & E[I(A_1\le u,v\le Y_1)]\\ =\ & E[I(A_1\le u\wedge v, Y_1\ge u\vee v)]\\ =\ & E[I(A_1\le u\wedge v, T_1\ge u\vee v, C_1\ge u\vee v-A_1)]\\ =\ & \int_{u\vee v}^{\infty}\int_0^{u\wedge v}\frac{f(t)}{\mu}S_C(u\vee v-a)dadt\\ =\ & \int_{u\vee v}^{\infty}\frac{f(t)}{\mu}dt\int_{u\vee v}^{|u-v|}S(y)d(-y)\\ =\ & \mu^{-1}S(u\vee v)[\omega_C(u\vee v)-\omega_C(|u-v|)]. \end{aligned}$$

From the definition of $\tilde{A}(t_1,t_2)$, it can be shown that

$$\text{(52)}\quad \begin{aligned} &\tilde{A}(t_1,t_2)\\ =\ & \int_0^{t_1}\int_0^{t_2}\frac{I(A_1\le u\le Y1)I(A_1\le v\le Y1)}{K^2(u)K^2(v)}\\ & dW_1(u)dW_1(v)\\ =\ & \int_0^{t_1}\int_0^{t_2}\frac{\mu^{-1}S(u\vee v)[\omega_C(u\vee v)-\omega_C(|u-v|)]}{K^2(u)K^2(v)}\\ & dW_1(v)dW_1(u)\\ =\ & \int_0^{t_1}\int_0^{u}\frac{\mu^{-1}S(u)[\omega_C(u)-\omega_C(u-v)]}{\mu^{-2}S^2(u)\omega_C(u)S^2(v)\omega_C(v)}dvdu\\ & +\int_0^{t_1}\int_u^{t_2}\frac{\mu^{-1}S(v)[\omega_C(v)-\omega_C(v-u)]}{\mu^{-2}S^2(u)\omega_C(u)S^2(v)\omega_C(v)}dvdu\\ =\ & \tilde{B}(t_1,t_2). \end{aligned}$$

Combining (52) with (49), it is readily seen that

$$E[\phi_1(t_1)\phi_1(t_2)] = \int_0^{t_1}\frac{dW_1(u)}{K^2(u)}.$$

Hence we have

$$\sqrt{n}(\tilde{S}_n(t)-S(t)) \xrightarrow{\mathcal{D}} N(0,\tilde{\Sigma}_t),$$

where $\tilde{\Sigma}_t = S^2(t)\int_0^t \frac{dW_1(u)}{K^2(u)}$.

(ii) In order to prove that $\Sigma_t\le\tilde{\Sigma}_t$, we only need to verify that $H(u,v)$ in (48) is less equal than 0 for any u and v. Now, note that

$$\begin{aligned} H(u,v) &= (w_1^2-w_1)\mu^{-1}S(u\vee v)[\omega_C(u\vee v)-\omega_C(|u-v|)]\\ &\quad + (w_2^2-w_2)\mu^{-1}S(u\vee v)\omega_C(u\wedge v)\\ &\quad + w_1w_2\mu^{-1}\int_{u\vee v}^{u+v}f(t)\\ &\quad [\omega_C(v)+\omega_C(u)-\omega_C(t-u)-\omega_C(t-v)]dt.\\ &\le (w_1^2-w_1)\mu^{-1}S(u\vee v)[\omega_C(u\vee v)-\omega_C(|u-v|)]\\ &\quad + (w_2^2-w_2)\mu^{-1}S(u\vee v)\omega_C(u\wedge v)\\ &\quad + w_1w_2\mu^{-1}(\omega_C(u)+\omega_C(v))(S(u+v)-S(v))\\ &= \mu^{-1}w_1(w_1+w_2-1)S(v)\omega_C(v)\\ &\quad + \mu^{-1}w_2(w_1+w_2-1)\omega_C(u)S(v)\\ &\quad + \mu^{-1}w_1(w_1+w_2-1)[S(v)\omega_C(u-v)\\ &\quad + (\omega_C(u)+\omega_C(v))S(u+v)]\\ &= 0. \end{aligned}$$

This completes the proof of Corollary 2. ☐

Proof of Corollary 3. By Corollary 1, $\hat{S}_n$ is most efficient at the values of w_1 and w_2 that minimise

$$M(t;w_1,w_2) =: \int_0^t\int_0^t \frac{H(u,v)}{K^2(u)K^2(v)}dW_1(u)dW_1(v),$$

where

$$\begin{aligned} & H(u,v)\\ =\ & (w_1^2-w_1)\mu^{-1}S(u\vee v)[\omega_C(u\vee v)-\omega_C(|u-v|)]\\ & +(w_2^2-w_2)\mu^{-1}S(u\vee v)\omega_C(u\wedge v)\\ & +w_1w_2\mu^{-1}\int_{u\vee v}^{u+v}f(t)[\omega_C(v)+\omega_C(u)\\ & -\omega_C(t-u)-\omega_C(t-v)]dt. \end{aligned}$$

Since $w_1+w_2=1$, we can rewrite $M(t;w_1,w_2)$ as

$$\text{(53)}\quad \begin{aligned} \tilde{M}(t,w_1) &= (w_1^2-w_1)\mu^{-1}\{S(u\wedge v)\\ &\quad [\omega_C(u)+\omega_C(v)-\omega_C(|u-v|)]\\ &\quad -\int_{u\wedge v}^{u+v}f(t)[\omega_C(u)+\omega_C(v)\\ &\quad -\omega_C(t-u)-\omega_C(t-v)]dt\}. \end{aligned}$$

Differentiating (53) with respect to w_1 yields

$$\begin{aligned} \frac{\partial\tilde{M}(t;w_1)}{\partial w_1} &= (2w_1-1)\mu^{-1}\{S(u\wedge v)\\ &\quad [\omega_C(u)+\omega_C(v)-\omega_C(|u-v|)]\\ &\quad -\int_{u\wedge v}^{u+v}f(t)[\omega_C(u)+\omega_C(v)\\ &\quad -\omega_C(t-u)-\omega_C(t-v)]dt\}\\ &= 0 \end{aligned}$$

when $w_1=\frac{1}{2}$. This proves Corollary 3. ☐

Received 8 February 2019

REFERENCES

[1] Addona, V. and Wolfson, D. B. (2006). A formal test for stationary of the incidence rate using data from a prevalent cohort study with follow-up. *Lifetime Data Analysis* **12** 267–284. MR2328577

[2] Alexander, K. S. and Talagrand, M. (1989). The law of the iterated logarithm for empirical processes on Vapnik-Červonenkis classes. *Journal of Multivariate Analysis* **30** 155–166. MR1003714

[3] Arcones, M. A. and Giné, E. (1995). On the law of the iterated logarithm for canonical U-statistics and processes. *Stochastic processes and their applications* **58** 217–245. MR1348376

[4] Asgharian, M., M'Lan, C. M. and Wolfson, D. B. (2002). Length-biased sampling with right censoring: an unconditional approach. *Journal of the American Statistical Association* **97** 207–209. MR1947280

[5] Asgharian, M. and Wolfson, D. B. (2005). Asymptotic behaviour of the unconditional NPMLE of the length-biased survival function from right censored prevalent cohort data. *Annals of Statistics* **33** 2109–2131. MR2211081

[6] de Uña Álvarez, J. (2004). Nonparametric estimation under length-biased sampling and Type I censoring: a moment based approach. *Annals of the Institute of Statistical Mathematics* **56** 667–681. MR2126804

[7] Dudley, R. M. (1984). A course on empirical processes. In Hennequin, P. L. (ed.), *Ecole d'été de Probabilités de Saint-Flour XII-1982* 1–142. Springer, Berlin, Heidelberg. MR0876079

[8] Giné, E. and Zinn, J. (1984). Some limit theorems for empirical processes. *The Annals of Probability* **12** 929–989. MR0757767

[9] Hoeffding, W. (1992). A class of statistics with asymptotically normal distribution. In Kotz, S., Johnson, N. L. (eds.), *Breakthroughs in Statistics. Springer Series in Statistics (Perspectives in Statistics)* 308–334. Springer, New York, NY. MR1182794

[10] Huang, C. Y. and Qin, J. (2011). Nonparametric estimation for length-biased and right-censored data. *Biometrika* **98** 177–186. MR2804218

[11] Huang, C. Y. and Qin, J. (2012). Composite partial likelihood estimation under length-biased sampling, with application to a prevalent cohort study of dementia. *Journal of the American Statistical Association* **107** 946–957. MR3010882

[12] Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53** 457–481. MR0093867

[13] Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer, New York. MR2724368

[14] Luo, X. D. and Tsai, W. Y. (2009). Nonparametric estimation for right-censored length-biased data: a pseudo-partial likelihood approach. *Biometrika* **96** 873–886. MR2767276

[15] Pollard, D. (1990). Empirical processes: theory and applications. *NSF-CBMS Regional Conference Series in Probability and Statistics* **2** i–86. MR1089429

[16] Shi, J. H., Chen, X. R. and Zhou, Y. (2015). The strong representation for the nonparametric estimator of length-biased and right-censored data. *Statistics and Probability Letters* **104** 49–57. MR3360704

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

[18] Vardi, Y. (1989). Multiplicative censoring, renewal processes, deconvolution and decreasing density: nonparametric estimation. *Biometrika* **76** 751–761. MR1041420

[19] Wang, M. C. (1991). Nonparametric estimation from cross-sectional survival data. *Journal of the American Statistical Association* **86** 343–354. MR1137104

[20] Winter, B. B. and Foldes, A. (1988). A product-limit estimator for use with length-biased data. *Canadian Journal of Statistics* **16** 337–355. MR1007165

[21] Zhou, Y. and Yip, P. 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. MR1703375

Shuqin Fan
School of Mathematical Sciences
University of Chinese Academy of Sciences
Beijing
China
Department of Management Sciences
City University of Hong Kong
Kowloon, Hong Kong
Hong Kong
E-mail address: fanshuqin1990@126.com

Wei Zhao
Department of Biostatistics and Bioinformatics
Emory University
Atlanta
U.S.A.
School of Mathematical Sciences
University of Chinese Academy of Sciences
Beijing
China
Department of Management Sciences
City University of Hong Kong
Kowloon, Hong Kong
Hong Kong
E-mail address: wei.zhao@emory.edu

Alan T. K. Wan
Department of Management Sciences
City University of Hong Kong
Kowloon, Hong Kong
Hong Kong
E-mail address: msawan@cityu.edu.hk

Yong Zhou
Academy of Mathematics and Systems Science
Chinese Academy of Sciences
Beijing
China
E-mail address: yzhou@amss.ac.cn