Lifetime Data Anal
DOI 10.1007/s10985-015-9325-0

Estimating the survival function based on the semi-Markov model for dependent censoring

Ziqiang Zhao[1] · Ming Zheng[1] · Zhezhen Jin[2]

Received: 10 March 2014 / Accepted: 7 March 2015

Abstract In this paper, we study a nonparametric maximum likelihood estimator (NPMLE) of the survival function based on a semi-Markov model under dependent censoring. We show that the NPMLE is asymptotically normal and achieves asymptotic nonparametric efficiency. We also provide a uniformly consistent estimator of the corresponding asymptotic covariance function based on an information operator. The finite-sample performance of the proposed NPMLE is examined with simulation studies, which show that the NPMLE has smaller mean squared error than the existing estimators and its corresponding pointwise confidence intervals have reasonable coverages. A real example is also presented.

Keywords Semi-Markov model · Dependent censoring · NPMLE · Survival function

Electronic supplementary material The online version of this article (doi:10.1007/s10985-015-9325-0) contains supplementary material, which is available to authorized users.

✉ Zhezhen Jin
zj7@columbia.edu

Ziqiang Zhao
johnzhao2835@gmail.com

Ming Zheng
mingzheng@fudan.edu.cn

[1] Department of Statistics, School of Management, Fudan University, 670 Guoshun Road, Shanghai, China

[2] Department of Biostatistics, Mailman School of Public Health, Columbia University, 722 West 168th Street, New York, NY, USA

Published online: 14 March 2015

1 Introduction

In survival analysis, the survival function of the failure time is commonly estimated by the Kaplan-Meier estimator and the Nelson-Aalen estimator. For these two estimators, a key assumption is that the censoring time and the survival time are independent. It is challenging to estimate the survival function under dependent censorship (Tsiatis 1975).

In oncology studies, independent dropout yields independent censoring. In addition, subjects are often censored due to the ending of the study or to the onset of progressive disease (PD). The study-ending censoring is usually independent of the survival time while the PD-related censoring might be dependent on the survival time since the PD could be a precursor of both death and lost of follow-up. In other words, in the presence of PD, the patients have a much higher risk of death and are more likely to leave the study. Ignoring such dependence would yield biased and inconsistent estimation of the survival function. On the other hand, it is possible to improve the estimation if the dependency information is properly used. Datta et al. (2000) considered nonparametric estimation using a three-stage irreversible illness–death model.

In the case of dependent censoring, Lee and Tsai (2005) proposed a semi-Markov model and developed an empirical-type estimator of the survival function. The asymptotic variance of their proposed estimator, however, is too complicated to compute. In this paper, we present a nonparametric maximum likelihood estimator (NPMLE) of the survival function based on the semi-Markov model. We show that our proposed NPMLE converges weakly to a Gaussian process and achieves asymptotic efficiency. In addition, we develop a consistent estimator of its asymptotic covariance function based on an information operator, which can be easily calculated.

The remainder of the paper is organized as follows. In Sect. 2, we will introduce the semi-Markov model. In Sect. 3, we will derive the NPMLE of the survival function. In Sect. 4, we will establish the asymptotic properties of the NPMLE, and construct a consistent estimator of its asymptotic covariance function based on an information operator. In Sect. 5, we will present simulation studies and re-analysis of the example in Lee and Tsai (2005). We will conclude with a short discussion in Sect. 6 and provide sketches of the proofs of theorems in the Appendix.

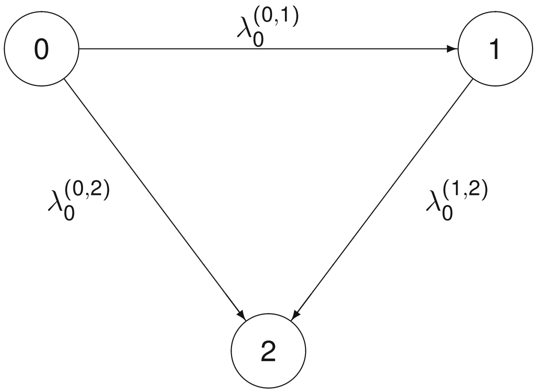
2 The semi-Markov model

Let T be the survival time and U be the PD censoring time. The semi-Markov model proposed by Lee and Tsai (2005) assumes that:

$$\lambda_{\mathsf{T}|\mathsf{U}}(t|u) = \lambda_0^{(0,2)}(t)\,\mathrm{I}\{t \leqslant u\} + \lambda_0^{(1,2)}(t-u)\,\mathrm{I}\{t > u\}, \quad \text{for } t, u \geqslant 0, \tag{1}$$

where $\lambda_{\mathsf{T}|\mathsf{U}}$ is the conditional hazard function of T given U, $\mathrm{I}\{\cdot\}$ is the indicator function taking the value 1 if the condition is satisfied and the value 0 otherwise, $\lambda_0^{(0,2)}$ and $\lambda_0^{(1,2)}$ are unknown hazard functions for death without PD and death with PD respectively.

Fig. 1 Transition



Note that the cumulative form of Model (1) is:

$$\Lambda_{\mathsf{T}|\mathsf{U}}(t|u) = \Lambda_0^{(0,2)}(\min\{t,u\}) + \Lambda_0^{(1,2)}(t-u)\,\mathrm{I}\{t>u\}, \quad \text{for } t,u \geqslant 0, \tag{2}$$

and that for any $t, u \geqslant 0$,

$$\mathsf{S}_{\mathsf{T}|\mathsf{U}}(t|u) = \begin{cases} \mathsf{S}_0^{(0,2)}(t) & t \leqslant u, \\ \mathsf{S}_0^{(0,2)}(u)\,\mathsf{S}_0^{(1,2)}(t-u) & t > u, \end{cases}$$

where $\Lambda_0^{(0,1)}$, $\Lambda_0^{(0,2)}$, $\Lambda_0^{(1,2)}$ and $\mathsf{S}_0^{(0,1)}$, $\mathsf{S}_0^{(0,2)}$, $\mathsf{S}_0^{(1,2)}$ are the corresponding cause specific cumulative hazard functions and the corresponding cause specific survival functions for $\lambda_0^{(0,1)}$, $\lambda_0^{(0,2)}$ and $\lambda_0^{(1,2)}$, respectively. The superscript values 0, 1, 2 indicate three different states, with 0 being the state of alive without PD, 1 being the state of alive with PD and 2 being the state of death.

More precisely, Model (1) corresponds to a non-homogeneous semi-Markov process J with the state space {0, 1, 2}:

$$\mathsf{J}(t) = \begin{cases} 0 \ \mathsf{T} > t, \mathsf{U} > t & \text{(Alive without PD at time } t) \\ 1 \ \mathsf{T} > t, \mathsf{U} \leqslant t & \text{(Alive with PD at time } t), \quad \text{for } t \geqslant 0. \\ 2 \ \mathsf{T} \leqslant t & \text{(Died at time } t) \end{cases}$$

With this specification, if $\lambda_0^{(0,1)}$ denotes the hazard function of U, i.e. the hazard function for PD, it can be shown that $\lambda_0^{(0,1)}$, $\lambda_0^{(0,2)}$ and $\lambda_0^{(1,2)}$ are respectively the cause specific hazard functions for the transition from state 0 to state 1, state 0 to state 2, and state 1 to state 2. Figure 1 provides a plot of transition for illustration.

In addition, we use C to denote independent censoring, which is independent of (T, U). Let $\mathsf{X} = \min\{\mathsf{T}, \mathsf{C}\}$ and $\Delta = \mathrm{I}\{\mathsf{T} \leqslant \mathsf{C}\}$. When $\mathsf{X} \leqslant \mathsf{U}$, the subject is dead or censored before PD, i.e., (X, Δ) is observed while U is not. When $\mathsf{X} > \mathsf{U}$, PD occurs before both death and the independent censoring C and so the subject would leave the study at time U with a probability θ_0, i.e., U is observed and (X, Δ) is observed with probability θ_0, where θ_0 is an unknown parameter.

Let $\mathsf{V} = \min\{\mathsf{X}, \mathsf{U}\}$, $\mathsf{R} = \mathrm{I}\{\mathsf{X} \leqslant \mathsf{U}\}$ and ξ be the indicator of observing (X, Δ). The observed data for a subject is $(\mathsf{V}, \mathsf{R}, \xi, \xi\mathsf{X}, \xi\Delta)$. Throughout the paper, it is assumed that ξ is conditionally independent of $(\mathsf{T}, \mathsf{U}, \mathsf{C})$ given R with $\mathsf{P}(\xi = 1|\mathsf{R} = 1) = 1$ and $\mathsf{P}(\xi = 1|\mathsf{R} = 0) = \theta_0$. For a sample of size n, we observe $(\mathsf{V}_i, \mathsf{R}_i, \xi_i, \xi_i\mathsf{X}_i, \xi_i\Delta_i)$, $i = 1, \ldots, n$.

3 Nonparametric maximum likelihood estimation

In general, without additional information, likelihood based estimation is not available for dependent censoring problems. Using the semi-Markov model specification, it is possible to develop likelihood based estimation.

Next, we present the likelihood for the unknown parameters:

$$\left(\Lambda^{(0,1)}, \Lambda^{(0,2)}, \Lambda^{(1,2)}, \Lambda^{(\mathsf{c})}, \theta\right),$$

where $\Lambda^{(\mathsf{c})}$ is the parameter for $\Lambda_0^{(\mathsf{c})}$ which is the true cumulative hazard function of C. Note that $\Lambda^{(\mathsf{c})}$ and θ are nuisance parameters.

Suppose that $\Lambda^{(0,1)}$, $\Lambda^{(0,2)}$, $\Lambda^{(1,2)}$ and $\Lambda^{(\mathsf{c})}$ are differentiable with corresponding derivatives: $\lambda^{(0,1)}$, $\lambda^{(0,2)}$, $\lambda^{(1,2)}$ and $\lambda^{(\mathsf{c})}$. Let $\mathsf{S}^{(0,1)}$, $\mathsf{S}^{(0,2)}$, $\mathsf{S}^{(1,2)}$ and $\mathsf{S}^{(\mathsf{c})}$ be the corresponding survival functions of $\Lambda^{(0,1)}$, $\Lambda^{(0,2)}$, $\Lambda^{(1,2)}$ and $\Lambda^{(\mathsf{c})}$. With these notations, the likelihood of $\left(\Lambda^{(0,1)}, \Lambda^{(0,2)}, \Lambda^{(1,2)}, \Lambda^{(\mathsf{c})}, \theta\right)$ can be derived.

For subjects with $\xi_i = 1$, $(\mathsf{V}_i, \mathsf{X}_i, \Delta_i, \mathsf{R}_i, \xi_i)$ is observed and its corresponding likelihood can be obtained based on the joint distribution of $(\mathsf{V}, \mathsf{X}, \Delta, \mathsf{R}, \xi)$. Specifically, when $\xi_i = 1$, $\mathsf{R}_i = 1$ and $\Delta_i = 1$, the likelihood is:

$$\lambda^{(0,2)}(\mathsf{V}_i)\,\mathsf{S}^{(0,2)}(\mathsf{V}_i)\,\mathsf{S}^{(\mathsf{c})}(\mathsf{V}_i)\,\mathsf{S}^{(0,1)}(\mathsf{V}_i);$$

When $\xi_i = 1$, $\mathsf{R}_i = 1$ and $\Delta_i = 0$, the likelihood is:

$$\lambda^{(\mathsf{c})}(\mathsf{V}_i)\,\mathsf{S}^{(\mathsf{c})}(\mathsf{V}_i)\,\mathsf{S}^{(0,2)}(\mathsf{V}_i)\,\mathsf{S}^{(0,1)}(\mathsf{V}_i);$$

When $\xi_i = 1$, $\mathsf{R}_i = 0$ and $\Delta_i = 1$, the likelihood is:

$$\lambda^{(1,2)}(\mathsf{X}_i - \mathsf{V}_i)\,\mathsf{S}^{(1,2)}(\mathsf{X}_i - \mathsf{V}_i)\,\mathsf{S}^{(\mathsf{c})}(\mathsf{X}_i)\,\lambda^{(0,1)}(\mathsf{V}_i)\,\mathsf{S}^{(0,1)}(\mathsf{V}_i)\,\mathsf{S}^{(0,2)}(\mathsf{V}_i)\,\theta;$$

When $\xi_i = 1$, $\mathsf{R}_i = 0$ and $\Delta_i = 0$, the likelihood is:

$$\lambda^{(\mathsf{c})}(\mathsf{X}_i)\,\mathsf{S}^{(\mathsf{c})}(\mathsf{X}_i)\,\mathsf{S}^{(1,2)}(\mathsf{X}_i - \mathsf{V}_i)\,\lambda^{(0,1)}(\mathsf{V}_i)\,\mathsf{S}^{(0,1)}(\mathsf{V}_i)\,\mathsf{S}^{(0,2)}(\mathsf{V}_i)\,\theta.$$

For subjects with $\xi_i = 0$, $(\mathsf{V}_i, \mathsf{R}_i, \xi_i)$ is observed and its corresponding likelihood is:

$$\lambda^{(0,1)}(\mathsf{V}_i)\,\mathsf{S}^{(0,1)}(\mathsf{V}_i)\,\mathsf{S}^{(0,2)}(\mathsf{V}_i)\,\mathsf{S}^{(\mathsf{c})}(\mathsf{V}_i)\,(1 - \theta).$$

Hence, the likelihood based on the observed data $(\mathsf{V}_i, \mathsf{R}_i, \xi_i, \xi_i \mathsf{X}_i, \xi_i \Delta_i)$, $i = 1, \ldots, n$, is proportional to

$$\prod_{i=1}^{n} \left[\left[\lambda^{(0,1)} (\mathsf{V}_i) \right]^{(1-\mathsf{R}_i)} \prod_{y \in [0, \mathsf{V}_i)} \left(1 - \Lambda^{(0,1)} (dy) \right) \right]$$
$$\times \prod_{i=1}^{n} \left[\left[\lambda^{(0,2)} (\mathsf{V}_i) \right]^{\Delta_i \mathsf{R}_i} \prod_{y \in [0, \mathsf{V}_i)} \left(1 - \Lambda^{(0,2)} (dy) \right) \right]$$
$$\times \prod_{i=1}^{n} \left[\left[\lambda^{(1,2)} (\mathsf{X}_i - \mathsf{V}_i) \right]^{\Delta_i} \prod_{y \in [0, \mathsf{X}_i - \mathsf{V}_i)} \left(1 - \Lambda^{(1,2)} (dy) \right) \right]^{\xi_i (1-\mathsf{R}_i)} .$$

Unfortunately, this function is unbounded from above and the usual maximum likelihood estimator (MLE) does not exist when $\Lambda_0^{(0,1)}$, $\Lambda_0^{(0,2)}$ and $\Lambda_0^{(1,2)}$ are restricted to continuous functions. With discretized extensions by allowing $\Lambda_0^{(0,1)}$, $\Lambda_0^{(0,2)}$ and $\Lambda_0^{(1,2)}$ to be discontinuous, the NPMLE is well defined in the sense of Kiefer and Wolfowitz (1956) and Scholz (1980). Specifically, we assume that $\Lambda_0^{(0,1)}$, $\Lambda_0^{(0,2)}$ and $\Lambda_0^{(1,2)}$ are cadlag, piecewise constant and right continuous with left limits.

To obtain the NPMLE, we rewrite the likelihood function with the discretized $\Lambda^{(0,1)}$, $\Lambda^{(0,2)}$ and $\Lambda^{(1,2)}$:

$$\overline{\mathcal{L}}_n \left(\Lambda^{(0,1)}, \Lambda^{(0,2)}, \Lambda^{(1,2)} \right) = \overline{\mathcal{L}}_n^{(0,1)} \left(\Lambda^{(0,1)} \right) \overline{\mathcal{L}}_n^{(0,2)} \left(\Lambda^{(0,2)} \right) \overline{\mathcal{L}}_n^{(1,2)} \left(\Lambda^{(1,2)} \right),$$

where for any cumulative hazard function Λ,

$$\overline{\mathcal{L}}_n^{(0,1)} (\Lambda) = \prod_{i=1}^{n} \left[\left[\Lambda \{\mathsf{V}_i\} \right]^{(1-\mathsf{R}_i)} \prod_{y \in [0, \mathsf{V}_i)} (1 - \Lambda (dy)) \right],$$
$$\overline{\mathcal{L}}_n^{(0,2)} (\Lambda) = \prod_{i=1}^{n} \left[\left[\Lambda \{\mathsf{V}_i\} \right]^{\Delta_i \mathsf{R}_i} (1 - \Lambda \{\mathsf{V}_i\})^{1-\Delta_i \mathsf{R}_i} \prod_{y \in [0, \mathsf{V}_i)} (1 - \Lambda (dy)) \right],$$
$$\overline{\mathcal{L}}_n^{(1,2)} (\Lambda)$$
$$= \prod_{i=1}^{n} \left[\left[\Lambda \{\mathsf{X}_i - \mathsf{V}_i\} \right]^{\Delta_i} (1 - \Lambda \{\mathsf{X}_i - \mathsf{V}_i\})^{1-\Delta_i} \prod_{y \in [0, \mathsf{X}_i - \mathsf{V}_i)} (1 - \Lambda (dy)) \right]^{\xi_i (1-\mathsf{R}_i)},$$

with $\Lambda \{t\} = \Lambda (t) - \Lambda (t-)$ for all $t \geqslant 0$.

It is also assumed that $\Lambda_0^{(0,1)}$ does not share jump points with $\Lambda_0^{(0,2)}$ and $\Lambda_0^{(\mathsf{c})}$, which means that the time of PD censoring occurrence is different from the time of death or the time of independent censoring. Then, for any cumulative hazard function Λ,

$$\log \overline{\mathcal{L}}_n^{(0,1)}(\Lambda) \leqslant \int_0^{+\infty} \log [\Lambda\{t\}]\, \overline{\mathsf{N}}_n^{(0,1)}(dt) + \sum_{t \geqslant 0} \overline{\mathsf{Y}}_n^{(0)}(t+) \log [1 - \Lambda\{t\}] \tag{3}$$

$$\leqslant \int_0^{+\infty} \log \left[\hat{\Lambda}_n^{(0,1)}\{t\}\right] \overline{\mathsf{N}}_n^{(0,1)}(dt) + \sum_{t \geqslant 0} \overline{\mathsf{Y}}_n^{(0)}(t+) \log \left[1 - \hat{\Lambda}_n^{(0,1)}\{t\}\right] \tag{4}$$

$$= \log \overline{\mathcal{L}}_n^{(0,1)}\left(\hat{\Lambda}_n^{(0,1)}\right),$$

$$\log \overline{\mathcal{L}}_n^{(0,2)}(\Lambda) \leqslant \int_0^{+\infty} \log [\Lambda\{t\}]\, \overline{\mathsf{N}}_n^{(0,2)}(dt) + \sum_{t \geqslant 0} \overline{\mathsf{Y}}_n^{(0)}(t+) \log [1 - \Lambda\{t\}] \tag{5}$$

$$\leqslant \int_0^{+\infty} \log \left[\hat{\Lambda}_n^{(0,2)}\{t\}\right] \overline{\mathsf{N}}_n^{(0,2)}(dt) + \sum_{t \geqslant 0} \overline{\mathsf{Y}}_n^{(0)}(t+) \log \left[1 - \hat{\Lambda}_n^{(0,2)}\{t\}\right] \tag{6}$$

$$= \log \overline{\mathcal{L}}_n^{(0,2)}\left(\hat{\Lambda}_n^{(0,2)}\right),$$

$$\log \overline{\mathcal{L}}_n^{(1,2)}(\Lambda) \leqslant \int_0^{+\infty} \log [\Lambda\{t\}]\, \overline{\mathsf{N}}_n^{(1,2)}(dt) + \sum_{t \geqslant 0} \overline{\mathsf{Y}}_n^{(1)}(t+) \log [1 - \Lambda\{t\}] \tag{7}$$

$$\leqslant \int_0^{+\infty} \log \left[\hat{\Lambda}_n^{(1,2)}\{t\}\right] \overline{\mathsf{N}}_n^{(1,2)}(dt) + \sum_{t \geqslant 0} \overline{\mathsf{Y}}_n^{(1)}(t+) \log \left[1 - \hat{\Lambda}_n^{(1,2)}\{t\}\right]$$

$$= \log \overline{\mathcal{L}}_n^{(1,2)}\left(\hat{\Lambda}_n^{(1,2)}\right), \tag{8}$$

where for any $t \geqslant 0$,

$$\overline{\mathsf{Y}}_n^{(0)}(t) = \sum_{i=1}^n \mathrm{I}\{\mathsf{V}_i \geqslant t\},\ \overline{\mathsf{Y}}_n^{(1)}(t) = \sum_{i=1}^n \xi_i (1 - \mathsf{R}_i)\, \mathrm{I}\{\mathsf{X}_i - \mathsf{V}_i \geqslant t\},$$

$$\overline{\mathsf{N}}_n^{(0,1)}(t) = \sum_{i=1}^n (1 - \mathsf{R}_i)\, \mathrm{I}\{\mathsf{V}_i \leqslant t\},\ \overline{\mathsf{N}}_n^{(0,2)}(t) = \sum_{i=1}^n \Delta_i \mathsf{R}_i \mathrm{I}\{\mathsf{V}_i \leqslant t\},$$

$$\overline{\mathsf{N}}_n^{(1,2)}(t) = \sum_{i=1}^n \xi_i \Delta_i (1 - \mathsf{R}_i)\, \mathrm{I}\{\mathsf{X}_i - \mathsf{V}_i \leqslant t\},$$

$$\hat{\Lambda}_n^{(0,1)}(t) = \int_{[0,t]} \left(\overline{\mathsf{Y}}_n^{(0)}(y)\right)^{-1} \mathrm{I}\left\{\overline{\mathsf{Y}}_n^{(0)}(y) > 0\right\} \overline{\mathsf{N}}_n^{(0,1)}(dy),$$

$$\hat{\Lambda}_n^{(0,2)}(t) = \int_{[0,t]} \left(\overline{\mathsf{Y}}_n^{(0)}(y)\right)^{-1} \mathrm{I}\left\{\overline{\mathsf{Y}}_n^{(0)}(y) > 0\right\} \overline{\mathsf{N}}_n^{(0,2)}(dy),$$

$$\hat{\Lambda}_n^{(1,2)}(t) = \int_{[0,t]} \left(\overline{\mathsf{Y}}_n^{(1)}(y)\right)^{-1} \mathrm{I}\left\{\overline{\mathsf{Y}}_n^{(1)}(y) > 0\right\} \overline{\mathsf{N}}_n^{(1,2)}(dy).$$

The equalities for (3), (5), and (7) hold if and only if Λ is a pure jump function. The inequalities in (4), (6), and (8) follow from the fact that for any $\lambda \in (0, 1)$ and any $\alpha, \beta > 0$,

$$\alpha \log \lambda + \beta \log (1 - \lambda) \leqslant \alpha \log (\alpha/(\alpha + \beta)) + \beta \log (\beta/(\alpha + \beta)) .$$

According to Criteria (B) of Scholz (1980), $\left(\hat{\Lambda}_n^{(0,1)}, \hat{\Lambda}_n^{(0,2)}, \hat{\Lambda}_n^{(1,2)}\right)$ is the NPMLE of $\left(\Lambda_0^{(0,1)}, \Lambda_0^{(0,2)}, \Lambda_0^{(1,2)}\right)$, since $\overline{\mathcal{L}}_n \left(\hat{\Lambda}_n^{(0,1)}, \hat{\Lambda}_n^{(0,2)}, \hat{\Lambda}_n^{(1,2)}\right) > 0$.

Let S_0 be the marginal survival function of T. Note that for any $t \geqslant 0$,

$$\begin{aligned} \mathsf{S}_0 (t) &= \mathsf{P} (\mathsf{T} > t) = - \int_{\mathbb{R}_+} \mathsf{S}_{\mathsf{T}|\mathsf{U}} (t|u) \; \mathsf{S}_0^{(0,1)} (du) \\ &= \mathsf{S}_0^{(0,1)} (t) \, \mathsf{S}_0^{(0,2)} (t) - \int_{[0,t]} \mathsf{S}_0^{(0,2)} (u) \, \mathsf{S}_0^{(1,2)} (t-u) \, \mathsf{S}_0^{(0,1)} (du) . \end{aligned}$$

Correspondingly, for any $t \geqslant 0$, define

$$\hat{\mathsf{S}}_n (t) = \hat{\mathsf{S}}_n^{(0,1)} (t) \, \hat{\mathsf{S}}_n^{(0,2)} (t) - \int_{[0,t]} \hat{\mathsf{S}}_n^{(0,2)} (u) \, \hat{\mathsf{S}}_n^{(1,2)} (t-u) \, \hat{\mathsf{S}}_n^{(0,1)} (du) ,$$

where for any $t \geqslant 0$,

$$\begin{aligned} \hat{\mathsf{S}}_n^{(0,1)} (t) &= \prod_{y \in [0,t]} \left[1 - \hat{\Lambda}_n^{(0,1)} (dy)\right], \; \hat{\mathsf{S}}_n^{(0,2)} (t) = \prod_{y \in [0,t]} \left[1 - \hat{\Lambda}_n^{(0,2)} (dy)\right], \\ \hat{\mathsf{S}}_n^{(1,2)} (t) &= \prod_{y \in [0,t]} \left[1 - \hat{\Lambda}_n^{(1,2)} (dy)\right]. \end{aligned}$$

By the invariance property, it follows that $\left(\hat{\mathsf{S}}_n^{(0,1)}, \hat{\mathsf{S}}_n^{(0,2)}, \hat{\mathsf{S}}_n^{(1,2)}\right)$ is the NPMLE of $\left(\mathsf{S}_0^{(0,1)}, \mathsf{S}_0^{(0,2)}, \mathsf{S}_0^{(1,2)}\right)$ and $\hat{\mathsf{S}}_n$ is the NPMLE of S_0.

4 Asymptotic results

In this section, we present the asymptotic properties of the NPMLE.

4.1 Regularity conditions

We list regularity conditions in this subsection.

For any $t \geqslant 0$, define $\mathsf{L}_0^{(0)} (t) = \mathsf{S}_0^{(0,1)} (t) \, \mathsf{S}_0^{(0,2)} (t) \, \mathsf{S}_0^{(\mathsf{c})} (t)$ and

$$\mathsf{L}_0^{(1)} (t) = -\theta_0 \mathsf{S}_0^{(1,2)} (t) \int_{\mathbb{R}_+} \mathsf{S}_0^{(0,2)} (u) \, \mathsf{S}_0^{(\mathsf{c})} (u+t) \; \mathsf{S}_0^{(0,1)} (du) .$$

The regularity conditions are as follows:

A.1 For a sample of size n, $(\mathsf{T}_i, \mathsf{U}_i, \mathsf{C}_i, \mathsf{X}_i, \Delta_i, \mathsf{V}_i, \mathsf{R}_i, \xi_i)$, $i = 1, \ldots, n$ are n independent copies of $(\mathsf{T}, \mathsf{U}, \mathsf{C}, \mathsf{X}, \Delta, \mathsf{V}, \mathsf{R}, \xi)$.

A.2 Model (2) holds with $\Lambda_0^{(1,2)}(0) = 0$.
A.3 The censoring time C is independent of (T, U).
A.4 The indicator ξ is conditionally independent of $(\mathsf{T}, \mathsf{U}, \mathsf{C})$ given R, with

$$\mathsf{P}(\xi = 1|\mathsf{R} = 1) = 1 \text{ and } \mathsf{P}(\xi = 1|\mathsf{R} = 0) = \theta_0 \in (0, 1).$$

A.5 There exists $\tau > 0$ such that $\mathsf{L}_0^{(0)}(\tau-) > 0$ and $\mathsf{L}_0^{(1)}(\tau-) > 0$.
A.6 For any $t \in [0, \tau]$, $\Lambda_0^{(0,1)}\{t\}\Lambda_0^{(0,2)}\{t\} = \Lambda_0^{(0,1)}\{t\}\Lambda_0^{(\mathsf{c})}\{t\} = 0$.

Assumption A.1 is commonly satisfied. Assumption A.2 is a technical condition for model specification. Assumption A.3 indicates that C is independent censoring. Assumption A.4 assumes that there is a subgroup of the potentially dependent censored subjects whose (X, Δ) are observed. Assumption A.5 is equivalent to $\mathsf{L}_0^{(0)}(0) > 0$ and $\mathsf{L}_0^{(1)}(0) > 0$. The required τ is generally not unique and could be chosen as large as possible. Assumption A.6 is a mild technical condition which is weaker than continuity.

4.2 Asymptotic normality

In this subsection, we establish the asymptotic normality and the asymptotic efficiency of the NPMLE estimators. The asymptotic normality is established by convergence to a tight Gaussian process in the space of uniformly bounded functions on $[0, \tau]$. Specifically, $\hat{\mathsf{S}}_n$ is viewed as a random element in $\ell^\infty([0, \tau])$, and $\left(\hat{\Lambda}_n^{(0,1)}, \hat{\Lambda}_n^{(0,2)}, \hat{\Lambda}_n^{(1,2)}\right)$ and $\left(\hat{\mathsf{S}}_n^{(0,1)}, \hat{\mathsf{S}}_n^{(0,2)}, \hat{\mathsf{S}}_n^{(1,2)}\right)$ are viewed as random elements in $\ell_3^\infty([0, \tau]) = \ell^\infty([0, \tau]) \times \ell^\infty([0, \tau]) \times \ell^\infty([0, \tau])$, where $\ell^\infty([0, \tau])$ denotes the space of all uniformly bounded real valued functions defined on $[0, \tau]$ equipped with the uniform norm. The asymptotic efficiency is shown by convolution theorem.

The first theorem gives the asymptotic normality for $\left(\hat{\Lambda}_n^{(0,1)}, \hat{\Lambda}_n^{(0,2)}, \hat{\Lambda}_n^{(1,2)}\right)$.

Theorem 1 *Under assumptions A.1–A.6,*

$$n^{1/2}\left(\hat{\Lambda}_n^{(0,1)} - \Lambda_0^{(0,1)}, \hat{\Lambda}_n^{(0,2)} - \Lambda_0^{(0,2)}, \hat{\Lambda}_n^{(1,2)} - \Lambda_0^{(1,2)}\right)$$

weakly converges to a tight zero-mean Gaussian process in $\ell_3^\infty([0, \tau])$ *with covariance function* $\mathcal{U}_0$, *as* $n \to \infty$, *where for any* $t, s \in [0, \tau]$,

$$\mathcal{U}_0(t, s) = \operatorname{diag}\left(\mathcal{U}_0^{(0,1)}(t, s), \mathcal{U}_0^{(0,2)}(t, s), \mathcal{U}_0^{(1,2)}(t, s)\right),$$

in which, for any $(i, j) \in \{(0, 1), (0, 2), (1, 2)\}$ *and any* $t, s \in [0, \tau]$,

$$\mathcal{U}_0^{(i,j)}(t, s) = \int_{[0,\min\{t,s\}]} \left(\mathsf{L}_0^{(i)}(y-)\right)^{-1}\left(1 - \Lambda_0^{(i,j)}\{y\}\right)\Lambda_0^{(i,j)}(dy).$$

The second theorem gives the asymptotic normality for $\left(\hat{\mathsf{S}}_n^{(0,1)}, \hat{\mathsf{S}}_n^{(0,2)}, \hat{\mathsf{S}}_n^{(1,2)}\right)$.

Theorem 2 *Under assumptions A.1–A.6,*

$$n^{1/2}\left(\hat{\mathsf{S}}_n^{(0,1)}-\mathsf{S}_0^{(0,1)},\hat{\mathsf{S}}_n^{(0,2)}-\mathsf{S}_0^{(0,2)},\hat{\mathsf{S}}_n^{(1,2)}-\mathsf{S}_0^{(1,2)}\right)$$

weakly converges to a tight zero-mean Gaussian process in $\ell_3^\infty([0,\tau])$ *with covariance function* $\mathcal{V}_0$*, as* $n\to\infty$*, where for any* $t,s\in[0,\tau]$*,*

$$\mathcal{V}_0(t,s)=\operatorname{diag}\left(\mathcal{V}_0^{(0,1)}(t,s),\mathcal{V}_0^{(0,2)}(t,s),\mathcal{V}_0^{(1,2)}(t,s)\right),$$

in which, for any $(i,j)\in\{(0,1),(0,2),(1,2)\}$ *and any* $t,s\in[0,\tau]$*,*

$$\begin{aligned}\mathcal{V}_0^{(i,j)}(t,s)&=\mathsf{S}_0^{(i,j)}(t)\,\mathsf{S}_0^{(i,j)}(s)\,\mathcal{W}_0^{(i,j)}(t,s),\\ \mathcal{W}_0^{(i,j)}(t,s)&=\int_{[0,\min\{t,s\}]}\left(\mathsf{L}_0^{(i)}(y-)\right)^{-1}\left(1-\Lambda_0^{(i,j)}\{y\}\right)^{-1}\Lambda_0^{(i,j)}(dy).\end{aligned}$$

The next theorem establishes the asymptotic normality for $\hat{S}_n$.

Theorem 3 *Under assumptions A.1–A.6,* $n^{1/2}(\hat{\mathsf{S}}_n-\mathsf{S}_0)$ *weakly converges to a tight zero-mean Gaussian process in* $\ell^\infty([0,\tau])$ *with covariance function* Ω_0*, as* $n\to\infty$*, where for any* $t,s\in[0,\tau]$*,*

$$\begin{aligned}\Omega_0(t,s)&=\Omega_0^{(0,1)}(t,s)+\Omega_0^{(0,2)}(t,s)+\Omega_0^{(1,2)}(t,s),\\ \Omega_0^{(0,1)}(t,s)&=\int_{(0,t]}\int_{(0,s]}\mathcal{V}_0^{(0,1)}(x-,y-)\ \mathsf{S}_0^{(0,2)}(dy)\ \mathsf{S}_0^{(0,2)}(dx)\\ &\quad-\mathsf{S}_0^{(0,2)}(0)\int_{(0,t]}\int_{(0,s]}\mathcal{V}_0^{(0,1)}(x-,s-y)\ \mathsf{S}_0^{(1,2)}(dy)\ \mathsf{S}_0^{(0,2)}(dx)\\ &\quad-\mathsf{S}_0^{(0,2)}(0)\int_{(0,t]}\int_{(0,s]}\mathcal{V}_0^{(0,1)}(t-x,y-)\ \mathsf{S}_0^{(0,2)}(dy)\ \mathsf{S}_0^{(1,2)}(dx)\\ &\quad+\left[\mathsf{S}_0^{(0,2)}(0)\right]^2\int_{(0,t]}\int_{(0,s]}\mathcal{V}_0^{(0,1)}(t-x,s-y)\ \mathsf{S}_0^{(1,2)}(dy)\ \mathsf{S}_0^{(1,2)}(dx),\\ \Omega_0^{(0,2)}(t,s)&=\mathsf{S}_0^{(0,1)}(t)\,\mathsf{S}_0^{(0,1)}(s)\,\mathcal{V}_0^{(0,2)}(t,s)\\ &\quad-\mathsf{S}_0^{(0,1)}(t)\int_{[0,s]}\mathcal{V}_0^{(0,2)}(t,y)\,\mathsf{S}_0^{(1,2)}(s-y)\ \mathsf{S}_0^{(0,1)}(dy)\\ &\quad-\mathsf{S}_0^{(0,1)}(s)\int_{[0,t]}\mathcal{V}_0^{(0,2)}(x,s)\,\mathsf{S}_0^{(1,2)}(t-x)\ \mathsf{S}_0^{(0,1)}(dx)\\ &\quad+\int_{[0,t]}\int_{[0,s]}\mathcal{V}_0^{(0,2)}(x,y)\,\mathsf{S}_0^{(1,2)}(t-x)\,\mathsf{S}_0^{(1,2)}(s-y)\,\mathsf{S}_0^{(0,1)}(dy)\ \mathsf{S}_0^{(0,1)}(dx),\\ \Omega_0^{(1,2)}(t,s)&=\int_{[0,t]}\int_{[0,s]}\mathcal{V}_0^{(1,2)}(t-x,s-y)\,\mathsf{S}_0^{(0,2)}(x)\,\mathsf{S}_0^{(0,2)}(y)\,\mathsf{S}_0^{(0,1)}(dy)\,\mathsf{S}_0^{(0,1)}(dx).\end{aligned}$$

It is worth noting that, although $\hat{\Lambda}_n^{(0,1)}-\Lambda_0^{(0,1)}$, $\hat{\Lambda}_n^{(0,2)}-\Lambda_0^{(0,2)}$ and $\hat{\Lambda}_n^{(1,2)}-\Lambda_0^{(1,2)}$ themselves are martingales, they do not form a joint martingale, since no common σ-field is available. Hence, the martingale limit theory cannot be directly used here. The proof of the above theorems are based on empirical process theory and are provided in the Appendix.

4.3 Asymptotic efficiency

In this subsection, we turn to the asymptotic nonparametric efficiency, which is formally defined by the convolution theorem (Theorem VIII.3.1 of Andersen et al. 1993).

We explicitly specify the Hilbert space required by the convolution theorem. Define $\mathbb{H} = \mathbb{H}^{(0,1)} \times \mathbb{H}^{(0,2)} \times \mathbb{H}^{(1,2)}$, where for $(i, j) \in \{(0,1), (0,2), (1,2)\}$,

$$\mathbb{H}^{(i,j)} = \left\{ h : \int_{[0,\tau]} [h(y)]^2 \mathsf{L}_0^{(i)}(y-) \left(1 - \Lambda_0^{(i,j)}\{y\}\right)^{-1} \Lambda_0^{(i,j)}(dy) < +\infty \right\}.$$

For any $g = \left(g^{(0,1)}, g^{(0,2)}, g^{(1,2)}\right), h = \left(h^{(0,1)}, h^{(0,2)}, h^{(1,2)}\right) \in \mathbb{H}$, define

$$\langle g, h \rangle_{\mathbb{H}} = \sum_{(i,j)} \int_{[0,\tau]} g^{(i,j)}(y) h^{(i,j)}(y) \mathsf{L}_0^{(i)}(y-) \left(1 - \Lambda_0^{(i,j)}\{y\}\right)^{-1} \Lambda_0^{(i,j)}(dy),$$

where the summation is taken over $\{(0,1), (0,2), (1,2)\}$. For any $h \in \mathbb{H}$, define $\|h\|_{\mathbb{H}} = \langle h, h \rangle_{\mathbb{H}}^{1/2}$. It can be shown that $\mathbb{H}$ is a Hilbert space with the inner product $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ and the norm $\|\cdot\|_{\mathbb{H}}$.

For any $(i, j) \in \{(0,1), (0,2), (1,2)\}$ and any $h = \left(h^{(0,1)}, h^{(0,2)}, h^{(1,2)}\right) \in \mathbb{H}$, define

$$\Lambda_{n,h}^{(i,j)}(t) = \int_{[0,t]} \left(1 + n^{-1/2} h^{(i,j)}(y)\right) \Lambda_0^{(i,j)}(dy),$$

the likelihood ratio of $\left(\Lambda_{n,h}^{(0,1)}, \Lambda_{n,h}^{(0,2)}, \Lambda_{n,h}^{(1,2)}\right)$ to $\left(\Lambda_0^{(0,1)}, \Lambda_0^{(0,2)}, \Lambda_0^{(1,2)}\right)$ is:

$$\mathcal{R}_n(h) = \mathcal{R}_n^{(0,1)}(h) \mathcal{R}_n^{(0,2)}(h) \mathcal{R}_n^{(1,2)}(h),$$

where for any $h = \left(h^{(0,1)}, h^{(0,2)}, h^{(1,2)}\right) \in \mathbb{H}$,

$$\log \mathcal{R}_n^{(0,1)}(h) = \sum_{i=1}^n (1 - \mathsf{R}_i) \log\left[1 + n^{-1/2} h^{(0,1)}(\mathsf{V}_i)\right]$$
$$+ \sum_{i=1}^n \left[\log \prod_{y \in [0,\mathsf{V}_i)} \left(1 - \Lambda_{n,h}^{(0,1)}(dy)\right) - \log \prod_{y \in [0,\mathsf{V}_i)} \left(1 - \Lambda_0^{(0,1)}(dy)\right)\right],$$

$$\log \mathcal{R}_n^{(0,2)}(h) = \sum_{i=1}^n \Delta_i \mathsf{R}_i \log\left[1 + n^{-1/2} h^{(0,2)}(\mathsf{V}_i)\right]$$
$$+ \sum_{i=1}^n (1 - \Delta_i \mathsf{R}_i) \left[\log\left(1 - \Lambda_{n,h}^{(0,2)}\{\mathsf{V}_i\}\right) - \log\left(1 - \Lambda_0^{(0,2)}\{\mathsf{V}_i\}\right)\right]$$
$$+ \sum_{i=1}^n \left[\log \prod_{y \in [0,\mathsf{V}_i)} \left(1 - \Lambda_{n,h}^{(0,2)}(dy)\right) - \log \prod_{y \in [0,\mathsf{V}_i)} \left(1 - \Lambda_0^{(0,2)}(dy)\right)\right],$$

$$\log \mathcal{R}_n^{(1,2)}(h) = \sum_{i=1}^{n} \xi_i (1-\mathsf{R}_i) \Delta_i \log\left[1 + n^{-1/2} h^{(1,2)} (\mathsf{X}_i - \mathsf{V}_i)\right]$$
$$+ \sum_{i=1}^{n} \xi_i (1-\mathsf{R}_i)(1-\Delta_i) \log\left(1 - \Lambda_{n,h}^{(1,2)} \{\mathsf{X}_i - \mathsf{V}_i\}\right)$$
$$- \sum_{i=1}^{n} \xi_i (1-\mathsf{R}_i)(1-\Delta_i) \log\left(1 - \Lambda_0^{(1,2)} \{\mathsf{X}_i - \mathsf{V}_i\}\right)$$
$$+ \sum_{i=1}^{n} \xi_i (1-\mathsf{R}_i) \log \prod_{y\in[0,\mathsf{X}_i-\mathsf{V}_i)} \left(1 - \Lambda_{n,h}^{(1,2)} (dy)\right)$$
$$- \sum_{i=1}^{n} \xi_i (1-\mathsf{R}_i) \log \prod_{y\in[0,\mathsf{X}_i-\mathsf{V}_i)} \left(1 - \Lambda_0^{(1,2)} (dy)\right).$$

Theorem 4 *Under assumptions A.1–A.6, for any $m > 0, h_1, \ldots, h_m \in \mathbb{H}$,*

$$(\log \mathcal{R}_n (h_1), \ldots, \log \mathcal{R}_n (h_m))$$

weakly converges to a Gaussian random vector in $\mathbb{R}^m$ with mean

$$-\frac{1}{2}\left(\|h_1\|_{\mathbb{H}}^2, \ldots, \|h_m\|_{\mathbb{H}}^2\right)$$

and covariance matrix

$$\begin{pmatrix} \langle h_1, h_1\rangle_{\mathbb{H}} & \cdots & \langle h_1, h_m\rangle_{\mathbb{H}} \\ \vdots & \ddots & \vdots \\ \langle h_m, h_1\rangle_{\mathbb{H}} & \cdots & \langle h_m, h_m\rangle_{\mathbb{H}} \end{pmatrix}.$$

Hence, the information operator $\mathcal{I}_0 : \mathbb{H} \times \mathbb{H} \to \mathbb{R}$ is: for any $g, h \in \mathbb{H}$,

$$\mathcal{I}_0 (g, h) = \langle g, h\rangle_{\mathbb{H}}.$$

For any $h = \left(h^{(0,1)}, h^{(0,2)}, h^{(1,2)}\right) \in \mathbb{H}$, define

$$\hat{\kappa}_n (h) = \int_{[0,\tau]} \left(h^{(0,1)}(y)\ \hat{\Lambda}_n^{(0,1)}(dy) + h^{(0,2)}(y)\ \hat{\Lambda}_n^{(0,2)}(dy) + h^{(1,2)}(y)\ \hat{\Lambda}_n^{(1,2)}(dy)\right).$$

Note that for any $g, h \in \mathbb{H}$, the asymptotic covariance of $\hat{\kappa}_n(g)$ and $\hat{\kappa}_n(h)$ is

$$\sum_{(i,j)} \int_{[0,\tau]} g^{(i,j)}(y)\, h^{(i,j)}(y)\, \mathsf{L}_0^{(i)}(y-) \left(1 - \Lambda_0^{(i,j)}\{y\}\right)^{-1} \Lambda_0^{(i,j)}(dy) \tag{9}$$

and can be interpreted as the "inverse" of the information operator $\mathcal{I}_0$, where the summation is taken over $\{(0,1),(0,2),(1,2)\}$. The efficiency result for $\left(\hat{\Lambda}_n^{(0,1)}, \hat{\Lambda}_n^{(0,2)}, \hat{\Lambda}_n^{(1,2)}\right)$ and $\hat{\mathsf{S}}_n$ is stated in the following theorem.

Theorem 5 *Under assumptions A.1–A.6,* $\left(\hat{\Lambda}_n^{(0,1)}, \hat{\Lambda}_n^{(0,2)}, \hat{\Lambda}_n^{(1,2)}\right)$ *and* $\hat{\mathsf{S}}_n$ *are efficient.*

The proof of the above theorems are given in the Appendix.

4.4 Asymptotic covariance function estimation

To carry out statistical inference, it is necessary to estimate the asymptotic covariance function. In this subsection, we provide uniform consistent estimators for the asymptotic covariance functions.

For any $g = \left(g^{(0,1)}, g^{(0,2)}, g^{(1,2)}\right), h = \left(h^{(0,1)}, h^{(0,2)}, h^{(1,2)}\right) \in \mathbb{H}$, define

$$\hat{\mathcal{I}}_n(g,h) = \sum_{(i,j)} \int_{[0,\tau]} g^{(i,j)}(y)\, h^{(i,j)}(y) \left(1 - \hat{\Lambda}_n^{(i,j)}\{y\}\right)^{-1} \overline{\mathsf{N}}_n^{(i,j)}(dy)$$

where the summation is taken over $\{(0,1),(0,2),(1,2)\}$.

For any $g = \left(g^{(0,1)}, g^{(0,2)}, g^{(1,2)}\right), h = \left(h^{(0,1)}, h^{(0,2)}, h^{(1,2)}\right) \in \mathbb{H}$, it can be shown that $\hat{\mathcal{I}}_n(g,h)$ is the negative second order directional derivative of $\log \overline{\mathcal{L}}_n$ at $\left(\hat{\Lambda}_n^{(0,1)}, \hat{\Lambda}_n^{(0,2)}, \hat{\Lambda}_n^{(1,2)}\right)$ along the direction $\left[\mathsf{G}_n, \mathsf{H}_n\right]$, where

$$\mathsf{G}_n = \left(\int g^{(0,1)} d\hat{\Lambda}_n^{(0,1)}, \int g^{(0,2)} d\hat{\Lambda}_n^{(0,2)}, \int g^{(1,2)} d\hat{\Lambda}_n^{(1,2)}\right),$$
$$\mathsf{H}_n = \left(\int h^{(0,1)} d\hat{\Lambda}_n^{(0,1)}, \int h^{(0,2)} d\hat{\Lambda}_n^{(0,2)}, \int h^{(1,2)} d\hat{\Lambda}_n^{(1,2)}\right).$$

Since the function $\overline{\mathcal{L}}_n$ plays the role of the likelihood, $\hat{\mathcal{I}}_n$ should be a reasonable estimator for $\mathcal{I}_0$. The "inverse" operation in (9) leads to an estimator for the asymptotic covariance function of $\hat{\kappa}_n$:

$$\sum_{(i,j)} \int_{[0,\tau]} g^{(i,j)}(y)\, h^{(i,j)}(y) \left(1 - \hat{\Lambda}_n^{(i,j)}\{y\}\right)^{-1} \overline{\mathsf{Y}}_n^{(0)}(y)\ \hat{\Lambda}_n^{(i,j)}(dy)$$

for all $g = \left(g^{(0,1)}, g^{(0,2)}, g^{(1,2)}\right), h = \left(h^{(0,1)}, h^{(0,2)}, h^{(1,2)}\right) \in \mathbb{H}$, where the summation is taken over $\{(0,1),(0,2),(1,2)\}$.

Hence, the estimators for $\mathcal{U}_0$ and $\mathcal{V}_0$ are given as follows: for any $t, s \in [0,\tau]$,

$$\hat{\mathcal{U}}_n(t,s) = \operatorname{diag}\left(\hat{\mathcal{U}}_n^{(0,1)}(t,s), \hat{\mathcal{U}}_n^{(0,2)}(t,s), \hat{\mathcal{U}}_n^{(1,2)}(t,s)\right),$$
$$\hat{\mathcal{V}}_n(t,s) = \operatorname{diag}\left(\hat{\mathcal{V}}_n^{(0,1)}(t,s), \hat{\mathcal{V}}_n^{(0,2)}(t,s), \hat{\mathcal{V}}_n^{(1,2)}(t,s)\right),$$

where for any $(i, j) \in \{(0,1), (0,2), (1,2)\}$ and any $t, s \in [0, \tau]$,

$$\hat{\mathcal{V}}_n^{(i,j)}(t,s) = \hat{\mathsf{S}}_n^{(i,j)}(t)\,\hat{\mathsf{S}}_n^{(i,j)}(s)\,\hat{\mathcal{W}}_n^{(i,j)}(t,s),$$

$$\hat{\mathcal{U}}_n^{(i,j)}(t,s) = n\int_{[0,\min\{t,s\}]}\left(1-\hat{\Lambda}_n^{(i,j)}\{y\}\right)\left(\overline{\mathsf{Y}}_n^{(i)}(y)\right)^{-1}\ \hat{\Lambda}_n^{(i,j)}(dy),$$

$$\hat{\mathcal{W}}_n^{(i,j)}(t,s) = n\int_{[0,\min\{t,s\}]}\left(\overline{\mathsf{Y}}_n^{(i)}(y)\right)^{-1}\left(1-\hat{\Lambda}_n^{(i,j)}\{y\}\right)^{-1}\ \hat{\Lambda}_n^{(i,j)}(dy).$$

The estimator for Ω_0 is given as $\hat{\Omega}_n$: for any $t, s \in [0, \tau]$,

$$\hat{\Omega}_n(t,s) = \hat{\Omega}_n^{(0,1)}(t,s) + \hat{\Omega}_n^{(0,2)}(t,s) + \hat{\Omega}_n^{(1,2)}(t,s),$$

where for any $(i, j) \in \{(0,1), (0,2), (1,2)\}$ and any $t, s \in [0, \tau]$,

$$\begin{aligned}
\hat{\Omega}_n^{(0,1)}(t,s) = &\int_{(0,t]}\int_{(0,s]}\hat{\mathcal{V}}_n^{(0,1)}(x-,y-)\ \hat{\mathsf{S}}_n^{(0,2)}(dy)\ \hat{\mathsf{S}}_n^{(0,2)}(dx)\\
&-\hat{\mathsf{S}}_n^{(0,2)}(0)\int_{(0,t]}\int_{(0,s]}\hat{\mathcal{V}}_n^{(0,1)}(x-,s-y)\ \hat{\mathsf{S}}_n^{(1,2)}(dy)\ \hat{\mathsf{S}}_n^{(0,2)}(dx)\\
&-\hat{\mathsf{S}}_n^{(0,2)}(0)\int_{(0,t]}\int_{(0,s]}\hat{\mathcal{V}}_n^{(0,1)}(t-x,y-)\ \hat{\mathsf{S}}_n^{(0,2)}(dy)\ \hat{\mathsf{S}}_n^{(1,2)}(dx)\\
&+\left[\hat{\mathsf{S}}_n^{(0,2)}(0)\right]^2\int_{(0,t]}\int_{(0,s]}\hat{\mathcal{V}}_n^{(0,1)}(t-x,s-y)\ \hat{\mathsf{S}}_n^{(1,2)}(dy)\ \hat{\mathsf{S}}_n^{(1,2)}(dx),\\
&\hat{\Omega}_n^{(0,2)}(t,s)\\
=\ &\hat{\mathsf{S}}_n^{(0,1)}(t)\,\hat{\mathsf{S}}_n^{(0,1)}(s)\,\hat{\mathcal{V}}_n^{(0,2)}(t,s)\\
&-\hat{\mathsf{S}}_n^{(0,1)}(t)\int_{[0,s]}\hat{\mathcal{V}}_n^{(0,2)}(t,y)\,\hat{\mathsf{S}}_n^{(1,2)}(s-y)\ \hat{\mathsf{S}}_n^{(0,1)}(dy)\\
&-\hat{\mathsf{S}}_n^{(0,1)}(s)\int_{[0,t]}\hat{\mathcal{V}}_n^{(0,2)}(x,s)\,\hat{\mathsf{S}}_n^{(1,2)}(t-x)\ \hat{\mathsf{S}}_n^{(0,1)}(dx)\\
&+\int_{[0,t]}\int_{[0,s]}\hat{\mathcal{V}}_n^{(0,2)}(x,y)\,\hat{\mathsf{S}}_n^{(1,2)}(t-x)\,\hat{\mathsf{S}}_n^{(1,2)}(s-y)\,\hat{\mathsf{S}}_n^{(0,1)}(dy)\ \hat{\mathsf{S}}_n^{(0,1)}(dx),
\end{aligned}$$

$$\begin{aligned}
\hat{\Omega}_n^{(1,2)}(t,s) = &\int_{[0,t]}\int_{[0,s]}\hat{\mathcal{V}}_n^{(1,2)}(t-x,s-y)\,\hat{\mathsf{S}}_n^{(0,2)}(x)\,\hat{\mathsf{S}}_n^{(0,2)}(y)\ \hat{\mathsf{S}}_n^{(0,1)}(dy)\\
&\times\hat{\mathsf{S}}_n^{(0,1)}(dx).
\end{aligned}$$

The following theorem states the uniform consistency for $\hat{\mathcal{U}}_n$, $\hat{\mathcal{V}}_n$ and $\hat{\Omega}_n$.

Theorem 6 *Under assumptions A.1–A.6, as $n \to \infty$,*

$$\sup_{t,s\in[0,\tau]}\left|\hat{\mathcal{U}}_n(t,s)-\mathcal{U}_0(t,s)\right|,\ \sup_{t,s\in[0,\tau]}\left|\hat{\mathcal{V}}_n(t,s)-\mathcal{V}_0(t,s)\right|,$$
$$\sup_{t,s\in[0,\tau]}\left|\hat{\Omega}_n(t,s)-\Omega_0(t,s)\right|$$

converge to zero in probability.

The proof of the theorem is given in the Appendix.

Based on the asymptotic theory and the estimation of the asymptotic covariance function, various types of inference can be conducted. For example, for any fixed $t_0 \in [0, \tau]$, a 95 % pointwise confidence interval can be constructed as:

$$\hat{\mathsf{S}}_n(t_0) \pm 1.96\hat{\Omega}_n^{1/2}(t_0, t_0).$$

5 Numerical studies

5.1 Simulation studies

Four sets of simulation studies were carried out to examine the finite-sample performance of the proposed NPMLE.

In the first and the third sets of the simulation studies, the independent censoring time C was specified as $+\infty$. In the second and the fourth sets, the independent censoring time C was generated from a uniform distribution such that $\mathsf{P}(\mathsf{T} \geqslant \mathsf{C}) = 0.15$.

In the first and the second sets of the simulation studies, the PD censoring time U was generated from Uniform $(0, \eta)$, while in the third and the fourth sets, the PD censoring time U was generated from the exponential distribution with hazard rate η.

In all simulation studies, the survival time T was generated as

$$\mathsf{T} = \begin{cases} \mathsf{T}^{(1)} & \text{if } \mathsf{T}^{(1)} \leqslant \mathsf{U} \\ \mathsf{U} + \mathsf{T}^{(2)} & \text{if } \mathsf{T}^{(1)} > \mathsf{U} \end{cases},$$

where $\mathsf{T}^{(1)}$ was generated from the exponential distribution with hazard rate 1 and $\mathsf{T}^{(2)}$ was generated from the exponential distribution with hazard rate λ. The parameters were set to be $\lambda \in \{1, 2\}$, $\mathsf{P}(\mathsf{U} < \mathsf{T}) \in \{0.3, 0.7\}$ and $\theta_0 \in \{0.3, 0.6, 1\}$. For each scenario, we generated 1000 samples of size $n = 100$.

The simulation results were summarized in Tables 1, 2, 3, 4, 5, 6, 7 and 8, which report the empirical bias (Bias), the empirical standard deviation (Std), the square root of the empirical mean squared error (sqrt[MSE]) and the average estimated standard deviation (AES) of the proposed estimator, as well as the empirical coverage rate (CR) of the corresponding 95 % pointwise confidence interval, at the 0.3th, 0.5th and 0.7th quantiles of T. The values of Bias, Std, sqrt(MSE) and AES were multiplied by 10,000. For comparison, the results from the Kaplan–Meier approach were also included in the tables. In addition, the empirical bias, the empirical standard deviation and the square root of the empirical mean squared error of Lee and Tsai's estimator were included in Tables 1, 2, 5 and 6. Because there is no valid estimator of the variance of Lee and Tsai's estimator, the AES and the CR are not reported for the Lee and Tsai's estimator. The Lee and Tsai's estimator was not included in Tables 3, 4, 7 and 8, because it does not incorporate the additional independent censoring.

For the cases with $\mathsf{C} = +\infty$, in which the Lee and Tsai's approach is applicable, the proposed NPMLE and the Lee and Tsai's estimator were nearly unbiased and consistent. The NPMLE had smaller sqrt(MSE) than the Lee and Tsai's estimator for all three quantiles, the improvement ranged from 0.6 to 7.9 %. The coverage of the 95 % pointwise confidence interval for the proposed NPMLE were close to the nominal level.

Table 1 Simulation results when $C = +\infty$ and U is generated from the uniform distribution with $P(T > U) = 0.3$

λ	θ	τ	Lee and Tsai's estimator			Kaplan–Meier estimator					NPMLE				
			Bias	Std	sqrt(MSE)	Bias	Std	sqrt(MSE)	AES	CR	Bias	Std	sqrt(MSE)	AES	CR
1	0.3	0.3	16.322	449.049	449.346	17.504	452.963	453.301	464.366	0.946	15.095	444.900	445.156	453.465	0.948
1	0.3	0.5	20.922	514.706	515.131	19.302	519.672	520.030	519.932	0.952	18.855	507.437	507.788	508.858	0.950
1	0.3	0.7	15.460	493.339	493.581	4.205	483.332	483.350	497.773	0.954	20.245	487.241	487.662	511.333	0.966
1	0.6	0.3	11.145	477.190	477.320	9.971	487.063	487.165	460.309	0.932	9.911	473.493	473.596	447.447	0.928
1	0.6	0.5	11.651	505.172	505.306	9.105	517.079	517.160	510.235	0.945	10.361	497.356	497.464	489.322	0.942
1	0.6	0.7	29.273	472.164	473.071	30.350	486.644	487.589	479.243	0.935	25.349	461.293	461.989	467.520	0.946
1	1.0	0.3	19.700	456.914	457.339	19.700	456.914	457.339	454.647	0.949	17.716	444.797	445.150	444.825	0.940
1	1.0	0.5	−8.600	507.468	507.541	−8.600	507.468	507.541	497.407	0.947	−9.648	481.701	481.798	480.689	0.953
1	1.0	0.7	−9.100	457.686	457.777	−9.100	457.686	457.777	455.098	0.948	−10.352	437.898	438.020	446.506	0.951
2	0.3	0.3	9.379	467.485	467.579	89.851	471.065	479.558	460.533	0.935	12.260	464.082	464.244	453.541	0.944
2	0.3	0.5	−3.479	507.639	507.651	183.352	510.943	542.845	518.393	0.932	−2.961	502.958	502.966	512.860	0.946
2	0.3	0.7	−11.359	480.614	480.749	282.399	502.682	576.575	505.049	0.927	−12.246	477.239	477.396	518.828	0.965
2	0.6	0.3	−7.299	457.866	457.924	36.817	463.476	464.936	459.002	0.940	−5.320	448.094	448.126	446.429	0.943
2	0.6	0.5	1.588	523.249	523.251	103.364	533.389	543.312	508.829	0.931	8.289	507.991	508.059	493.961	0.943
2	0.6	0.7	−16.654	478.814	479.104	133.208	504.190	521.490	481.332	0.935	−15.314	464.329	464.582	488.198	0.961
2	1.0	0.3	−9.700	440.306	440.413	−9.700	440.306	440.413	456.149	0.956	−10.120	426.256	426.376	443.491	0.952
2	1.0	0.5	−11.000	473.924	474.051	−11.000	473.924	474.051	497.740	0.949	−8.285	460.149	460.224	485.975	0.957
2	1.0	0.7	9.900	449.504	449.613	9.900	449.504	449.613	456.035	0.960	9.806	430.754	430.865	476.470	0.970

The AES and the CR are not reported for the Lee and Tsai's estimator, since the corresponding variance estimator has not been developed. The values of Bias, Std, sqrt(MSE) and AES have been multiplied by 10,000

Bias the empirical bias, *Std* the empirical standard deviation, *sqrt(MSE)* the square root of the empirical mean squared error, *AES* the average estimated standard deviation, *CR* the empirical coverage rate of the 95 % confidence interval

Table 2 Simulation results when $C = +\infty$ and U is generated from the uniform distribution with $P(T > U) = 0.7$

λ	θ	τ	Lee and Tsai's estimator			Kaplan–Meier estimator					NPMLE				
			Bias	Std	sqrt(MSE)	Bias	Std	sqrt(MSE)	AES	CR	Bias	Std	sqrt(MSE)	AES	CR
1	0.3	0.3	−11.337	499.437	499.565	−0.496	520.377	520.377	503.210	0.930	−14.576	492.276	492.492	474.698	0.932
1	0.3	0.5	−31.311	658.791	659.534	−33.695	675.060	675.901	644.222	0.925	−30.850	647.774	648.508	631.552	0.939
1	0.3	0.7	−2.828	716.866	716.871	0.973	760.752	760.753	728.592	0.918	−3.802	710.683	710.694	697.288	0.920
1	0.6	0.3	4.242	442.553	442.574	1.272	469.836	469.838	480.906	0.946	2.315	418.693	418.699	437.047	0.959
1	0.6	0.5	5.565	525.029	525.059	3.813	558.605	558.618	562.140	0.950	4.559	496.985	497.006	511.491	0.952
1	0.6	0.7	−5.090	531.379	531.404	−6.285	560.956	560.991	556.616	0.946	−5.814	510.060	510.093	521.131	0.951
1	1.0	0.3	1.900	467.809	467.813	1.900	467.809	467.813	455.301	0.941	−9.390	430.119	430.221	420.918	0.932
1	1.0	0.5	−2.100	489.127	489.131	−2.100	489.127	489.131	497.592	0.945	−5.547	447.852	447.886	454.722	0.950
1	1.0	0.7	−6.200	448.980	449.023	−6.200	448.980	449.023	455.336	0.950	−6.642	414.769	414.822	427.758	0.956
2	0.3	0.3	−5.362	495.129	495.158	245.676	499.341	556.506	483.087	0.895	−5.022	479.782	479.808	467.545	0.945
2	0.3	0.5	12.071	602.643	602.763	513.093	595.947	786.395	593.462	0.851	8.185	589.272	589.329	581.017	0.935
2	0.3	0.7	1.222	645.529	645.530	466.958	739.829	874.869	721.418	0.881	6.178	634.232	634.262	623.757	0.932
2	0.6	0.3	7.979	460.885	460.954	138.658	474.635	494.474	470.687	0.931	16.016	438.841	439.133	428.777	0.943
2	0.6	0.5	11.827	527.470	527.603	247.354	563.289	615.206	546.354	0.914	11.581	494.984	495.120	491.471	0.950
2	0.6	0.7	−6.385	511.517	511.557	193.050	566.862	598.833	559.222	0.938	−1.412	485.327	485.329	479.425	0.944
2	1.0	0.3	−8.200	457.627	457.700	−8.200	457.627	457.700	455.865	0.953	−15.557	409.854	410.149	413.332	0.956
2	1.0	0.5	−23.800	483.329	483.915	−23.800	483.329	483.915	497.645	0.956	−24.953	436.142	436.856	449.685	0.958
2	1.0	0.7	−14.800	445.847	446.093	−14.800	445.847	446.093	454.984	0.960	−18.720	392.816	393.261	406.233	0.951

The AES and the CR are not reported for the Lee and Tsai's estimator, since the corresponding variance estimator has not been developed. The values of Bias, Std, sqrt(MSE) and AES have been multiplied by 10,000

Bias the empirical bias, *Std* the empirical standard deviation, *sqrt(MSE)* the square root of the empirical mean squared error, *AES* the average estimated standard deviation, *CR* the empirical coverage rate of the 95 % confidence interval

Table 3 Simulation results when $\mathsf{P}(\mathsf{T} > \mathsf{C}) = 0.15$ and U is generated from the uniform distribution with $\mathsf{P}(\mathsf{T} > \mathsf{U}) = 0.3$

λ	θ	τ	Kaplan–Meier estimator					NPMLE				
			Bias	Std	sqrt(MSE)	AES	CR	Bias	Std	sqrt(MSE)	AES	CR
1	0.3	0.3	3.157	475.857	475.868	465.098	0.943	1.603	464.227	464.229	454.039	0.944
1	0.3	0.5	−8.948	518.809	518.886	520.695	0.949	−12.346	515.592	515.740	510.107	0.952
1	0.3	0.7	−24.535	480.116	480.742	497.576	0.949	−26.005	488.030	488.722	512.812	0.961
1	0.6	0.3	−55.071	457.719	461.020	463.386	0.952	−53.657	444.683	447.908	450.794	0.943
1	0.6	0.5	−27.605	521.312	522.042	509.740	0.944	−23.875	496.274	496.848	489.593	0.939
1	0.6	0.7	−4.198	477.638	477.657	477.467	0.952	−6.434	444.951	444.998	466.468	0.960
1	1.0	0.3	5.800	456.393	456.430	455.269	0.957	2.850	450.296	450.305	445.537	0.951
1	1.0	0.5	−1.200	517.326	517.327	497.304	0.938	−0.957	504.409	504.410	480.853	0.945
1	1.0	0.7	−9.700	457.368	457.470	455.068	0.949	−4.476	439.057	439.080	446.865	0.947
2	0.3	0.3	73.204	466.208	471.920	461.281	0.932	−3.787	461.198	461.213	454.278	0.936
2	0.3	0.5	186.877	519.638	552.220	518.041	0.932	5.280	508.992	509.020	513.616	0.955
2	0.3	0.7	285.552	520.929	594.060	504.759	0.911	9.750	494.252	494.348	519.513	0.964
2	0.6	0.3	54.006	466.143	469.262	458.288	0.947	11.512	451.738	451.884	445.745	0.937
2	0.6	0.5	124.852	533.052	547.478	509.020	0.923	18.940	508.319	508.671	494.004	0.936
2	0.6	0.7	168.701	481.221	509.935	483.245	0.939	16.348	447.174	447.473	489.082	0.963
2	1.0	0.3	−4.200	473.568	473.587	455.499	0.945	−4.680	459.618	459.642	442.950	0.934
2	1.0	0.5	8.000	498.583	498.647	497.498	0.948	4.254	478.715	478.734	485.907	0.955
2	1.0	0.7	8.600	444.391	444.474	456.052	0.961	5.087	423.137	423.167	476.635	0.974

The Lee and Tsai's estimator is not included, since the additional independent censoring is not incorporated in this approach. The values of Bias, Std, sqrt(MSE) and AES have been multiplied by 10,000

Bias the empirical bias, *Std* the empirical standard deviation, *sqrt(MSE)* the square root of the empirical mean squared error, *AES* the average estimated standard deviation, *CR* the empirical coverage rate of the 95 % confidence interval

Table 4 Simulation results when $P(T > C) = 0.15$ and U is generated from the uniform distribution with $P(T > U) = 0.7$

λ	θ	τ	Kaplan–Meier estimator					NPMLE				
			Bias	Std	sqrt(MSE)	AES	CR	Bias	Std	sqrt(MSE)	AES	CR
1	0.3	0.3	−11.321	508.927	509.053	504.172	0.952	−7.032	491.383	491.433	473.164	0.938
1	0.3	0.5	−2.953	674.774	674.780	642.816	0.938	4.681	649.068	649.085	629.126	0.930
1	0.3	0.7	35.088	748.811	749.633	731.977	0.936	21.350	714.376	714.695	698.724	0.933
1	0.6	0.3	−1.934	492.202	492.205	480.297	0.941	−1.255	440.665	440.666	437.101	0.948
1	0.6	0.5	−1.948	557.711	557.714	561.540	0.950	11.016	501.484	501.605	511.521	0.945
1	0.6	0.7	21.931	556.991	557.423	556.910	0.947	21.264	510.031	510.474	521.362	0.935
1	1.0	0.3	−5.000	475.711	475.738	455.495	0.943	−5.471	432.805	432.840	420.981	0.937
1	1.0	0.5	−15.700	510.455	510.696	497.374	0.938	−12.767	461.219	461.395	454.286	0.945
1	1.0	0.7	−12.200	468.413	468.572	454.838	0.946	−21.023	433.071	433.581	427.045	0.938
2	0.3	0.3	268.558	468.726	540.211	482.080	0.889	9.309	466.171	466.264	466.745	0.940
2	0.3	0.5	500.114	593.135	775.837	594.952	0.851	−1.386	587.409	587.410	581.482	0.945
2	0.3	0.7	420.009	739.105	850.108	722.839	0.900	−13.806	641.460	641.608	621.044	0.924
2	0.6	0.3	97.707	453.296	463.707	473.168	0.945	−27.117	413.581	414.469	430.981	0.960
2	0.6	0.5	217.812	555.667	596.831	546.970	0.916	−17.538	483.365	483.683	492.237	0.950
2	0.6	0.7	190.462	561.836	593.242	559.105	0.942	−7.811	482.644	482.708	479.955	0.937
2	1.0	0.3	0.100	453.780	453.780	455.556	0.957	1.700	407.523	407.527	412.883	0.948
2	1.0	0.5	−18.300	514.198	514.523	497.333	0.936	−7.882	450.739	450.807	449.654	0.938
2	1.0	0.7	−20.600	465.653	466.108	454.485	0.943	−19.324	415.083	415.533	405.605	0.937

The Lee and Tsai's estimator is not included, since the additional independent censoring is not incorporated in this approach. The values of Bias, Std, sqrt(MSE) and AES have been multiplied by 10,000

Bias the empirical bias, *Std* the empirical standard deviation, *sqrt(MSE)* the square root of the empirical mean squared error, *AES* the average estimated standard deviation, *CR* the empirical coverage rate of the 95 % confidence interval

Table 5 Simulation results when $C = +\infty$ and U is generated from the exponential distribution with $P(T > U) = 0.3$

λ	θ	τ	Lee and Tsai's estimator			Kaplan–Meier estimator					NPMLE				
			Bias	Std	sqrt(MSE)	Bias	Std	sqrt(MSE)	AES	CR	Bias	Std	sqrt(MSE)	AES	CR
1	0.3	0.3	10.296	456.820	456.936	11.247	469.847	469.981	468.018	0.941	10.442	450.538	450.659	456.574	0.944
1	0.3	0.5	4.645	522.233	522.254	10.906	527.338	527.451	526.662	0.955	3.852	517.105	517.120	521.405	0.953
1	0.3	0.7	−2.288	513.833	513.838	−2.456	503.665	503.671	505.226	0.943	−3.613	510.711	510.724	529.396	0.952
1	0.6	0.3	21.891	464.160	464.676	19.899	470.593	471.014	461.643	0.940	20.407	457.762	458.216	446.718	0.942
1	0.6	0.5	14.251	495.721	495.925	14.596	503.023	503.235	513.288	0.951	12.513	489.975	490.135	492.027	0.947
1	0.6	0.7	−0.230	475.155	475.155	0.286	482.683	482.683	481.632	0.944	1.168	464.456	464.457	473.894	0.954
1	1.0	0.3	19.800	456.154	456.583	19.800	456.154	456.583	454.665	0.958	19.156	446.730	447.141	442.640	0.952
1	1.0	0.5	5.600	485.494	485.526	5.600	485.494	485.526	497.628	0.947	4.393	469.216	469.236	479.710	0.944
1	1.0	0.7	6.200	453.594	453.636	6.200	453.594	453.636	455.838	0.956	7.415	429.566	429.630	448.339	0.963
2	0.3	0.3	18.762	440.773	441.172	116.418	447.944	462.825	462.360	0.938	19.215	433.321	433.747	456.426	0.953
2	0.3	0.5	2.862	514.297	514.305	222.344	511.420	557.662	523.404	0.934	2.159	508.523	508.527	522.428	0.950
2	0.3	0.7	−1.447	496.163	496.165	318.350	506.289	598.059	512.735	0.915	−3.142	489.412	489.422	526.006	0.961
2	0.6	0.3	−12.466	454.920	455.091	43.213	460.479	462.502	460.518	0.934	−15.996	438.648	438.939	446.454	0.944
2	0.6	0.5	−8.363	503.876	503.945	113.093	513.398	525.707	512.080	0.943	−10.435	488.754	488.865	496.818	0.955
2	0.6	0.7	−0.668	472.414	472.414	167.725	498.620	526.074	486.202	0.943	−1.017	458.369	458.371	491.470	0.971
2	1.0	0.3	−22.600	457.667	458.224	−22.600	457.667	458.224	456.493	0.948	−18.267	440.063	440.442	441.278	0.939
2	1.0	0.5	−22.500	517.350	517.839	−22.500	517.350	517.839	497.298	0.941	−16.422	489.006	489.282	484.956	0.944
2	1.0	0.7	1.400	452.751	452.753	1.400	452.751	452.753	455.626	0.951	−2.116	435.387	435.392	476.523	0.955

The AES and the CR are not reported for the Lee and Tsai's estimator, since the corresponding variance estimator has not been developed. The values of Bias, Std, sqrt(MSE) and AES have been multiplied by 10,000

Bias the empirical bias, *Std* the empirical standard deviation, *sqrt(MSE)* the square root of the empirical mean squared error, *AES* the average estimated standard deviation, *CR* the empirical coverage rate of the 95 % confidence interval

Table 6 Simulation results when $\mathsf{C} = +\infty$ and U is generated from the exponential distribution with $\mathsf{P}(\mathsf{T} > \mathsf{U}) = 0.7$

λ	θ	τ	Lee and Tsai's estimator			Kaplan–Meier estimator					NPMLE				
			Bias	Std	sqrt(MSE)	Bias	Std	sqrt(MSE)	AES	CR	Bias	Std	sqrt(MSE)	AES	CR
1	0.3	0.3	−5.225	538.632	538.657	−3.818	550.361	550.374	526.103	0.935	−4.202	524.809	524.825	507.769	0.936
1	0.3	0.5	24.438	685.975	686.410	29.837	672.759	673.420	649.392	0.932	22.835	677.260	677.645	655.729	0.931
1	0.3	0.7	20.986	710.008	710.318	35.836	701.711	702.626	683.782	0.933	14.405	703.195	703.343	685.009	0.928
1	0.6	0.3	0.793	458.252	458.253	−0.823	479.782	479.783	490.888	0.956	0.680	435.084	435.085	445.071	0.956
1	0.6	0.5	−17.919	536.533	536.832	−19.713	560.278	560.625	566.159	0.939	−18.499	514.169	514.501	521.047	0.946
1	0.6	0.7	−17.088	526.254	526.532	−13.209	553.737	553.894	545.974	0.929	−15.800	505.719	505.966	516.219	0.946
1	1.0	0.3	−17.500	450.740	451.080	−17.500	450.740	451.080	456.367	0.956	−12.883	413.698	413.899	417.697	0.960
1	1.0	0.5	−11.100	484.336	484.463	−11.100	484.336	484.463	497.638	0.945	−16.014	445.981	446.268	457.365	0.959
1	1.0	0.7	−13.400	443.302	443.505	−13.400	443.302	443.505	455.083	0.957	−10.796	411.697	411.839	430.601	0.955
2	0.3	0.3	−24.212	523.877	524.437	289.561	503.138	580.511	497.892	0.881	−23.854	509.650	510.208	498.870	0.934
2	0.3	0.5	−24.450	618.911	619.393	503.502	612.361	792.780	612.976	0.854	−18.987	606.850	607.147	613.808	0.945
2	0.3	0.7	−14.467	622.933	623.101	598.470	678.103	904.428	666.906	0.856	−12.684	614.458	614.588	619.218	0.939
2	0.6	0.3	−5.620	447.090	447.125	153.853	465.542	490.307	478.634	0.929	4.358	422.673	422.695	435.396	0.946
2	0.6	0.5	0.120	511.504	511.504	259.435	540.600	599.629	554.792	0.926	1.101	489.888	489.889	503.871	0.964
2	0.6	0.7	−3.514	498.036	498.048	266.248	542.172	604.018	549.819	0.930	−10.181	469.893	470.004	488.709	0.955
2	1.0	0.3	21.600	469.163	469.660	21.600	469.163	469.660	454.390	0.942	21.138	417.597	418.132	406.461	0.933
2	1.0	0.5	31.400	501.483	502.465	31.400	501.483	502.465	497.459	0.938	23.507	455.782	456.388	451.001	0.943
2	1.0	0.7	6.100	457.320	457.361	6.100	457.320	457.361	455.782	0.954	15.701	421.121	421.414	424.771	0.955

The AES and the CR are not reported for the Lee and Tsai's estimator, since the corresponding variance estimator has not been developed. The values of Bias, Std, sqrt(MSE) and AES have been multiplied by 10,000

Bias the empirical bias, *Std* the empirical standard deviation, *sqrt(MSE)* the square root of the empirical mean squared error, *AES* the average estimated standard deviation, *CR* the empirical coverage rate of the 95 % confidence interval

Table 7 Simulation results when $P(T > C) = 0.15$ and U is generated from the exponential distribution with $P(T > U) = 0.3$

λ	θ	τ	Kaplan–Meier estimator					NPMLE				
			Bias	Std	sqrt(MSE)	AES	CR	Bias	Std	sqrt(MSE)	AES	CR
1	0.3	0.3	−37.142	486.215	487.632	469.845	0.948	−32.751	468.945	470.088	458.227	0.946
1	0.3	0.5	−9.926	527.004	527.098	526.192	0.944	−7.799	521.589	521.647	520.417	0.948
1	0.3	0.7	15.360	508.382	508.614	505.214	0.940	14.140	511.564	511.759	528.501	0.956
1	0.6	0.3	−1.646	440.079	440.082	463.025	0.960	−0.956	421.928	421.929	448.100	0.965
1	0.6	0.5	16.974	487.732	488.027	513.509	0.961	19.143	461.533	461.930	492.530	0.963
1	0.6	0.7	8.971	464.171	464.258	482.019	0.957	9.506	438.781	438.884	473.189	0.960
1	1.0	0.3	−5.500	473.206	473.238	455.534	0.948	−2.166	459.658	459.663	443.548	0.939
1	1.0	0.5	2.600	517.978	517.985	497.297	0.930	7.841	491.710	491.773	479.562	0.935
1	1.0	0.7	−8.700	451.950	452.033	455.201	0.950	−7.242	431.096	431.157	448.312	0.950
2	0.3	0.3	106.272	444.514	457.041	462.601	0.942	4.028	445.565	445.583	457.066	0.954
2	0.3	0.5	218.342	499.220	544.879	523.112	0.941	2.863	501.590	501.598	522.948	0.955
2	0.3	0.7	320.370	500.009	593.840	512.668	0.919	1.890	481.911	481.914	528.818	0.972
2	0.6	0.3	57.458	475.318	478.778	459.332	0.936	6.531	458.627	458.673	444.893	0.933
2	0.6	0.5	148.054	524.993	545.470	511.501	0.931	27.558	500.385	501.143	496.141	0.949
2	0.6	0.7	192.218	481.920	518.840	487.073	0.941	22.617	447.304	447.875	492.088	0.973
2	1.0	0.3	−24.800	474.120	474.768	456.404	0.944	−24.034	449.961	450.602	441.301	0.947
2	1.0	0.5	−12.100	491.289	491.438	497.570	0.943	−21.879	480.151	480.650	485.176	0.948
2	1.0	0.7	−23.100	457.478	458.060	454.476	0.961	−25.766	429.097	429.870	476.058	0.970

The Lee and Tsai's estimator is not included, since the additional independent censoring is not incorporated in this approach. The values of Bias, Std, sqrt(MSE) and AES have been multiplied by 10,000

Bias the empirical bias, *Std* the empirical standard deviation, *sqrt(MSE)* the square root of the empirical mean squared error, *AES* the average estimated standard deviation, *CR* the empirical coverage rate of the 95 % confidence interval

Table 8 Simulation results when $\mathsf{P}(\mathsf{T} > \mathsf{C}) = 0.15$ and U is generated from the exponential distribution with $\mathsf{P}(\mathsf{T} > \mathsf{U}) = 0.7$

λ	θ	τ	Kaplan–Meier estimator					NPMLE				
			Bias	Std	sqrt(MSE)	AES	CR	Bias	Std	sqrt(MSE)	AES	CR
1	0.3	0.3	11.840	530.033	530.165	525.895	0.928	15.547	504.538	504.778	506.972	0.947
1	0.3	0.5	15.475	636.189	636.378	650.307	0.947	15.571	625.898	626.092	654.501	0.955
1	0.3	0.7	54.060	674.713	676.876	684.240	0.942	34.744	662.199	663.110	685.833	0.949
1	0.6	0.3	−7.869	501.497	501.559	490.966	0.931	−10.250	451.606	451.723	446.259	0.937
1	0.6	0.5	−7.043	577.838	577.881	565.730	0.944	−15.676	525.635	525.869	521.541	0.944
1	0.6	0.7	−19.481	547.570	547.916	546.076	0.943	−21.699	510.310	510.771	515.834	0.938
1	1.0	0.3	−20.200	448.635	449.089	456.502	0.965	−22.370	400.871	401.494	418.539	0.960
1	1.0	0.5	−13.100	500.169	500.340	497.480	0.938	−20.281	457.657	458.107	457.652	0.957
1	1.0	0.7	−11.600	465.061	465.206	454.898	0.944	−4.168	428.946	428.966	430.929	0.941
2	0.3	0.3	317.395	496.285	589.100	496.722	0.873	−10.674	501.787	501.900	497.296	0.945
2	0.3	0.5	559.884	620.296	835.606	612.199	0.826	−10.036	621.620	621.701	613.743	0.943
2	0.3	0.7	658.430	663.681	934.881	670.392	0.849	4.927	621.533	621.552	618.646	0.932
2	0.6	0.3	163.339	486.410	513.103	478.177	0.917	5.585	437.106	437.142	435.284	0.949
2	0.6	0.5	288.481	550.581	621.579	555.227	0.919	14.533	495.363	495.576	504.191	0.954
2	0.6	0.7	286.414	549.552	619.710	550.980	0.927	16.611	472.103	472.395	490.128	0.952
2	1.0	0.3	36.000	460.734	462.139	453.857	0.951	27.417	398.346	399.288	406.952	0.945
2	1.0	0.5	45.700	495.830	497.932	497.505	0.954	35.270	434.118	435.548	451.011	0.956
2	1.0	0.7	32.400	458.891	460.033	456.942	0.951	31.639	410.367	411.585	425.044	0.959

The Lee and Tsai's estimator is not included, since the additional independent censoring is not incorporated in this approach. The values of Bias, Std, sqrt(MSE) and AES have been multiplied by 10,000

Bias the empirical bias, *Std* the empirical standard deviation, *sqrt(MSE)* the square root of the empirical mean squared error, *AES* the average estimated standard deviation, *CR* the empirical coverage rate of the 95 % confidence interval

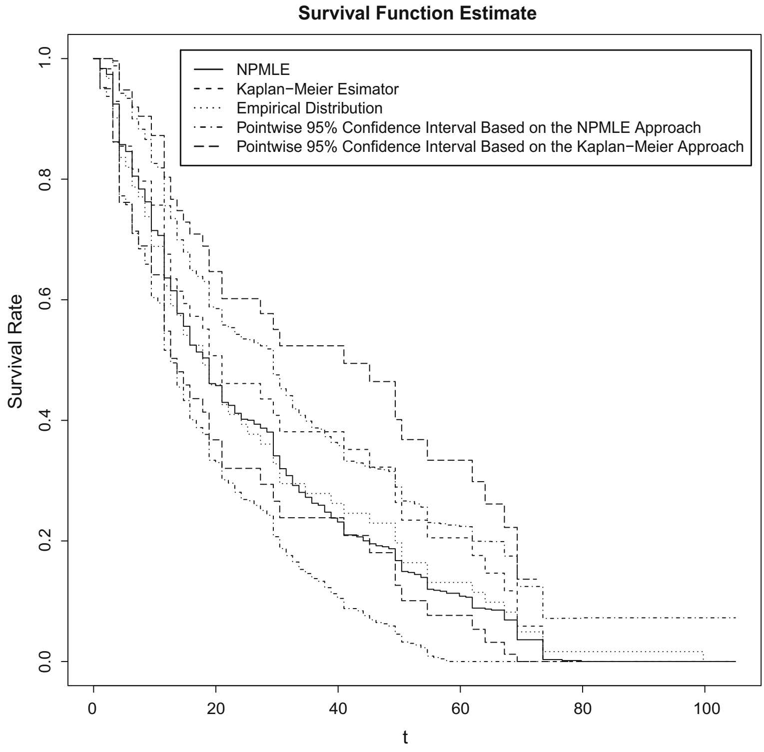


Fig. 2 Kaplan–Meier estimate and NPMLE

For the cases with $\lambda = 1$ when T was independent of U and the cases with $\theta = 1$ when (X, Δ) was fully observed, the Kaplan–Meier estimator was nearly unbiased and consistent and the corresponding 95 % confidence intervals provided reasonable coverage rates; the proposed NPMLE and the Lee and Tsai's estimator were also nearly unbiased and consistent, and the coverage rate of the proposed 95 % confidence intervals were close to the nominal level. Nevertheless, the proposed NPMLE had smaller sqrt(MSE) than the Kaplan–Meier estimator.

For the cases with $\lambda = 2$ and $\theta < 1$, the Kaplan–Meier estimator had larger bias and appeared inconsistent, while the other two estimators performed well.

5.2 An example

We illustrate our method with the data from a clinical trial of lung cancer conducted by the Eastern Cooperative Oncology Group, which was initially analyzed by Lagakos and Williams (1978). The data were also used by Lee and Wolfe (1998) and Lee and Tsai (2005) for the illustration of their methods. The complete data can be found in Lee and Tsai (2005). The study consists of 61 patients with inoperable carcinoma of the lung who were treated with the drug cyclophosphamide. Among the 61 patients, 28 patients experienced metastatic disease or a significant increase in the size of their

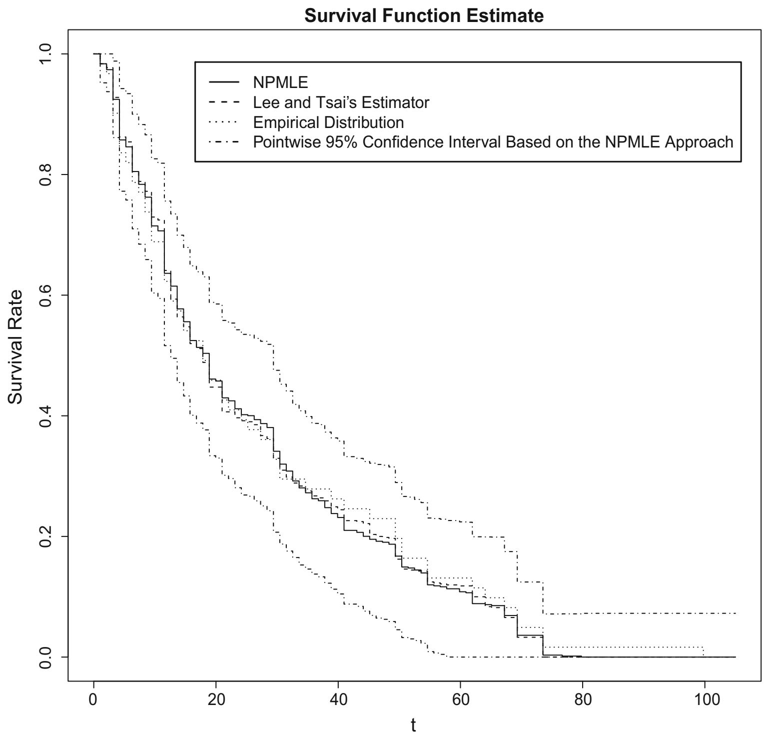


Fig. 3 Lee and Tsai's estimator and NPMLE

primary lesion, which are certain types of PD. In addition, Lee and Tsai (2005) did a pseudo second stage sampling, in which they selected 10 of these 28 patients as a random sample and pretended that the death time of the rest of the 18 patients were unknown.

For illustration, we also pretended that these 18 patients' death time were unknown. We treated the 10 selected patients as if they did not leave the study while the rest of the 18 did. Figure 2 shows the NPMLE and its pointwise 95 % confidence intervals along with the Kaplan–Meier estimator and its pointwise 95 % confidence intervals and the empirical distribution based on the complete data. Figure 3 presents the NPMLE and its pointwise 95 % confidence intervals along with the Lee and Tsai's estimator and the empirical distribution based on the complete data.

Figure 2 clearly shows that the Kaplan–Meier estimator is very different from the empirical distribution, and far away from the other estimators. Figures 2 and 3 show that both NPMLE and the Lee and Tsai's estimator are very close to the empirical distribution. The proposed pointwise 95 % confidence interval contains the Lee and Tsai's estimator and the empirical distribution.

6 Discussion

In oncology studies, in addition to the usual independent censoring, the censoring caused by the PD is a certain type of dependent censoring. In this paper, we have adopted the semi-Markov model provided by Lee and Tsai (2005) with an additional independent censoring and studied the NPMLE of the cause specific cumulative hazard functions and the marginal survival function, where we have shown the asymptotic normality of the NPMLE in the space of uniformly bounded functions and established its asymptotic efficiency. We have also developed uniformly consistent estimators for the covariance function based on the information operator.

A limitation of the model (1) is that it assumes that the risk of death after disease progression only depends on the time since progression, but not the duration of the progression. The limitation could be addressed by regression type models with the inclusion of the time to progression as a covariate. It is of interest to develop diagnostic methods for the model and to investigate regression type models in future studies.

Acknowledgments The research is supported by the National Natural Science Foundation of China (11271081) and the Student Growth Fund Scholarship of School of Management, Fudan University.

Appendix

Preliminary

Lemmas 1 and 2 provide related Donsker properties.

Lemma 1 *The classes* $\{(x,\delta)\mapsto \delta \mathrm{I}\{x\leqslant t\} : t\geqslant 0\}$ *and* $\{(x,\delta)\mapsto \delta \mathrm{I}\{x\geqslant t\} : t\geqslant 0\}$ *are Donsker classes of functions on* $\mathbb{R}_+\times\{0,1\}$.

Proof Combining Lemma 2.6.15 and Lemma 2.6.18(iii, vi) of Vaart and Wellner (1996), it can be shown that $\{(x,\delta)\mapsto \delta \mathrm{I}\{x\leqslant t\} : t\geqslant 0\}$ and $\{(x,\delta)\mapsto \delta \mathrm{I}\{x\geqslant t\} : t\geqslant 0\}$ are VC-classes on $\mathbb{R}_+\times\{0,1\}$. By Theorem 2.6.8 of Vaart and Wellner (1996), they are Donsker classes, where the measurability conditions could be verified via the denseness of the rational numbers. □

Lemma 2 *Let* η *be a positive real number and* Λ *be a nondecreasing cadlag function on* $[0,\eta]$. *The class* $\left\{(x,\delta)\mapsto \int_0^t \delta \mathrm{I}\{x\geqslant y\}\ \Lambda(dy) : t\in[0,\eta]\right\}$ *is a Donsker class of functions on* $\mathbb{R}_+\times\{0,1\}$.

Proof The result is easy to see by combining Example 2.6.21 and Example 2.10.8 of Vaart and Wellner (1996). □

Lemma 3 *Let* η *be a positive real number and* $\mathcal{BV}([0,\eta])$ *be the space of all cadlag functions defined on* $[0,\eta]$ *whose total variation are bounded by* 2. *For any* $(\mathsf{F},\mathsf{G},\mathsf{H})\in\mathcal{BV}([0,\eta])\times\mathcal{BV}([0,\eta])\times\mathcal{BV}([0,\eta])$ *and any* $t\in[0,\eta]$, *let*

$$\phi(\mathsf{F},\mathsf{G},\mathsf{H})[t]=\mathsf{F}(t)\,\mathsf{G}(t)-\int_{[0,t]}\mathsf{G}(u)\,\mathsf{H}(t-u)\ \mathsf{F}(du).$$

For any $\mathsf{F}_0, \mathsf{G}_0, \mathsf{H}_0 \in \mathcal{BV}([0,\eta])$, ϕ *is Hadamard differentiable at* $(\mathsf{F}_0, \mathsf{G}_0, \mathsf{H}_0)$ *with derivative* $\phi'(\mathsf{F}_0, \mathsf{G}_0, \mathsf{H}_0)$*, where for any* $t \in [0,\eta]$*,*

$$\phi'(\beta_0)[\beta][t] = \mathsf{F}_0(t)\,\mathsf{G}(t) + \mathsf{F}(t)\,\mathsf{G}_0(t) - \int_{[0,t]} \mathsf{G}_0(x)\,\mathsf{H}_0(t-x)\,\mathsf{F}(dx) - \int_{[0,t]} \mathsf{G}_0(x)\,\mathsf{H}(t-x)\,\mathsf{F}_0(dx) - \int_{[0,t]} \mathsf{G}(x)\,\mathsf{H}_0(t-x)\,\mathsf{F}_0(dx),$$

where $\beta = (\mathsf{F}, \mathsf{G}, \mathsf{H})$ *and* $\beta_0 = (\mathsf{F}_0, \mathsf{G}_0, \mathsf{H}_0)$.

Proof Let $\mathsf{F}_0, \mathsf{G}_0, \mathsf{H}_0, \mathsf{F}, \mathsf{G}, \mathsf{H} \in \mathcal{BV}([0,\eta])$, $\{(\mathsf{F}_m, \mathsf{G}_m, \mathsf{H}_m) : m \in \mathbb{N}\}$ be a sequence converging to $(\mathsf{F}, \mathsf{G}, \mathsf{H})$ in $\mathcal{BV}([0,\eta]) \times \mathcal{BV}([0,\eta]) \times \mathcal{BV}([0,\eta])$ and $\{h_m : m \in \mathbb{N}\}$ be a sequence of real numbers converging to 0.

Denote $\beta_0 = (\mathsf{F}_0, \mathsf{G}_0, \mathsf{H}_0)$ and $\beta_m = \beta_0 + h_m(\mathsf{F}, \mathsf{G}, \mathsf{H})$.

Note that for any $t \in [0,\eta]$, $\phi(\beta_m)[t] = \phi(\beta_0)[t] + h_m \Gamma_m(t) + h_m^2 \mathsf{W}_m(t)$, where

$$\begin{aligned}
\Gamma_m(t) &= \mathsf{F}_0(t)\,\mathsf{G}_m(t) + \mathsf{F}_m(t)\,\mathsf{G}_0(t) - \int_{[0,t]} \mathsf{G}_0(x)\,\mathsf{H}_0(t-x)\,\mathsf{F}_m(dx) \\
&\quad - \int_{[0,t]} \mathsf{G}_0(x)\,\mathsf{H}_m(t-x)\,\mathsf{F}_0(dx) - \int_{[0,t]} \mathsf{G}_m(x)\,\mathsf{H}_0(t-x)\,\mathsf{F}_0(dx), \\
\mathsf{W}_m(t) &= \mathsf{F}_m(t)\,\mathsf{G}_m(t) - \int_{[0,t]} \mathsf{H}_0(t-x)\,\mathsf{G}_m(x)\,\mathsf{F}_m(dx) \\
&\quad - \int_{[0,t]} \mathsf{G}_0(x)\,\mathsf{H}_m(t-x)\,\mathsf{F}_m(dx) - \int_{[0,t]} \mathsf{G}_m(x)\,\mathsf{H}_m(t-x)\,\mathsf{F}_0(dx) \\
&\quad - h_m \int_{[0,t]} \mathsf{G}_m(x)\,\mathsf{H}_m(t-x)\,\mathsf{F}_m(dx).
\end{aligned}$$

For any $t \in [0,\eta]$, let

$$\Gamma_0(t) = \mathsf{F}_0(t)\,\mathsf{G}(t) + \mathsf{F}(t)\,\mathsf{G}_0(t) - \int_{[0,t]} \mathsf{G}_0(x)\,\mathsf{H}_0(t-x)\,\mathsf{F}(dx) - \int_{[0,t]} \mathsf{G}_0(x)\,\mathsf{H}(t-x)\,\mathsf{F}_0(dx) - \int_{[0,t]} \mathsf{G}(x)\,\mathsf{H}_0(t-x)\,\mathsf{F}_0(dx).$$

Note that for any $t \in [0,\eta]$, $|\mathsf{W}_m(t)| \leqslant 28 + 8h_m$ and

$$|\Gamma_m(t) - \Gamma_0(t)| \leqslant 6 \sup_{s\in[0,\eta]} |\mathsf{F}_m(s) - \mathsf{F}(s)| + 6 \sup_{s\in[0,\eta]} |\mathsf{G}_m(s) - \mathsf{G}(s)| + 4 \sup_{s\in[0,\eta]} |\mathsf{H}_m(s) - \mathsf{H}(s)|.$$

Hence, as $m \to \infty$, $\sup_{t\in[0,\eta]} |\mathsf{W}_m(t)| = O(1)$ and $\sup_{t\in[0,\eta]} |\Gamma_m(t) - \Gamma_0(t)| = o(1)$.

Therefore, ϕ is Hadamard differentiable at β_0 and for any $t \in [0, \eta]$,

$$\phi'(\beta_0)[\beta][t] = \mathsf{F}_0(t)\,\mathsf{G}(t) + \mathsf{F}(t)\,\mathsf{G}_0(t) - \int_{[0,t]} \mathsf{G}_0(x)\,\mathsf{H}_0(t-x)\,\mathsf{F}(dx) - \int_{[0,t]} \mathsf{G}_0(x)\,\mathsf{H}(t-x)\,\mathsf{F}_0(dx) - \int_{[0,t]} \mathsf{G}(x)\,\mathsf{H}_0(t-x)\,\mathsf{F}_0(dx),$$

where $\beta = (\mathsf{F}, \mathsf{G}, \mathsf{H})$. □

Proof of Theorems 1 to 3

For any $(i, j) \in \{(0, 1), (0, 2), (1, 2)\}$ and any $t \geqslant 0$, define

$$\overline{\mathsf{M}}_n^{(i,j)}(t) = \overline{\mathsf{N}}_n^{(i,j)}(t) - \int_{[0,t]} \overline{\mathsf{Y}}_n^{(i)}(y)\ \Lambda_0^{(i,j)}(dy).$$

Proof Combining Theorem 2.10.6 of Vaart and Wellner (1996) with Lemmas 1 and 2, it can be shown that, as $n \to \infty$,

$$n^{-1/2}\left(\overline{\mathsf{M}}_n^{(0,1)}, \overline{\mathsf{M}}_n^{(0,2)}, \overline{\mathsf{M}}_n^{(1,2)}, \overline{\mathsf{Y}}_n^{(0)} - \mathsf{E}\overline{\mathsf{Y}}_n^{(0)}, \overline{\mathsf{Y}}_n^{(1)} - \mathsf{E}\overline{\mathsf{Y}}_n^{(1)}\right)$$

weakly converges to a tight zero mean Gaussian process in

$$\ell_5^\infty([0, \tau]) = \ell^\infty([0, \tau]) \times \ell^\infty([0, \tau]) \times \ell^\infty([0, \tau]) \times \ell^\infty([0, \tau]) \times \ell^\infty([0, \tau]).$$

Combining this with Lemma 3.9.17, Lemma 3.9.25, and Theorem 3.9.4 of Vaart and Wellner (1996), for any $(i, j) \in \{(0, 1), (0, 2), (1, 2)\}$,

$$\sup_{t\in[0,\tau]} \left| \hat{\Lambda}_n^{(i,j)}(t) - \Lambda_0^{(i,j)}(t) - \int_{[0,t]} \left(\mathsf{L}_0^{(i)}(y-)\right)^{-1} n^{-1}\overline{\mathsf{M}}_n^{(i,j)}(dy) \right| = o_p\left(n^{-1/2}\right), \tag{10}$$

By the continuity and the linearity of the integral operator, as $n \to \infty$,

$$n^{1/2}\left(\hat{\Lambda}_n^{(0,1)} - \Lambda_0^{(0,1)}, \hat{\Lambda}_n^{(0,2)} - \Lambda_0^{(0,2)}, \hat{\Lambda}_n^{(1,2)} - \Lambda_0^{(1,2)}\right)$$

weakly converges to a tight Gaussian process in $\ell_3^\infty([0, \tau])$ with covariance function $\mathcal{U}_0$.

By Lemma 3.9.30 and Theorem 3.9.4 of Vaart and Wellner (1996), as $n \to \infty$,

$$n^{1/2}\left(\hat{\mathsf{S}}_n^{(0,1)} - \mathsf{S}_0^{(0,1)}, \hat{\mathsf{S}}_n^{(0,2)} - \mathsf{S}_0^{(0,2)}, \hat{\mathsf{S}}_n^{(1,2)} - \mathsf{S}_0^{(1,2)}\right)$$

weakly converges to a tight Gaussian process in $\ell_3^\infty([0, \tau])$ with covariance function $\mathcal{V}_0$.

Combining Lemma 3 with Theorem 3.9.4 of Vaart and Wellner (1996),

$$\sup_{t\in[0,\tau]}\left|\hat{\mathsf{S}}_n(t)-\mathsf{S}_0(t)-\phi'(\beta_0)\left[\hat{\beta}_n-\beta_0\right][t]\right|=o_p\left(n^{-1/2}\right),$$

where $\beta_0=\left(\mathsf{S}_0^{(0,1)},\mathsf{S}_0^{(0,2)},\mathsf{S}_0^{(1,2)}\right)$ and $\hat{\beta}_n=\left(\hat{\mathsf{S}}_n^{(0,1)},\hat{\mathsf{S}}_n^{(0,2)},\hat{\mathsf{S}}_n^{(1,2)}\right)$.

Therefore, as $n\to\infty$, $n^{1/2}(\hat{\mathsf{S}}_n-\mathsf{S}_0)$ weakly converges to a tight zero mean Gaussian process in $\ell^\infty([0,\tau])$ with covariance function Ω_0. □

Proof of Theorems 4 and 5

Proof To obtain the efficiency, we will verify the conditions of Theorem VIII.3.2 and Theorem VIII.3.3 of Andersen et al. (1993).

First, we verify the local asymptotic normality (LAN) Assumption (Assumption VIII.3.1 of Andersen et al. 1993).

By the Taylor expansion, it is easy to show that for any $h=(h^{(0,1)},h^{(0,2)},h^{(1,2)})\in\mathbb{H}$,

$$\log\mathcal{R}_n(h)=\mathcal{S}_n^{(0,1)}(h)+\mathcal{S}_n^{(0,2)}(h)+\mathcal{S}_n^{(1,2)}(h)-\frac{1}{2}\|h\|_{\mathbb{H}}^2+o_p(1),\qquad(11)$$

where for any $(i,j)\in\{(0,1),(0,2),(1,2)\}$ and any $h=(h^{(0,1)},h^{(0,2)},h^{(1,2)})\in\mathbb{H}$,

$$\mathcal{S}_n^{(i,j)}(h)=n^{-1/2}\int_{[0,\tau]}h^{(i,j)}(y)\left(1-\Lambda_0^{(i,j)}\{y\}\right)^{-1}\overline{\mathsf{M}}_n^{(i,j)}(dy).$$

Hence, for any $m\in\mathbb{N}^+$ and any $h_1,\ldots,h_m\in\mathbb{H}$,

$$(\log\mathcal{R}_n(h_1),\ldots,\log\mathcal{R}_n(h_m))$$

weakly converges to a Gaussian random vector in $\mathbb{R}^m$ with mean

$$-\frac{1}{2}(\|h_1\|_{\mathbb{H}},\ldots,\|h_m\|_{\mathbb{H}})$$

and covariance matrix

$$\begin{pmatrix}\langle h_1,h_1\rangle_{\mathbb{H}} & \cdots & \langle h_1,h_m\rangle_{\mathbb{H}}\\ \vdots & \ddots & \vdots\\ \langle h_m,h_1\rangle_{\mathbb{H}} & \cdots & \langle h_m,h_m\rangle_{\mathbb{H}}\end{pmatrix}.$$

Next, we verify the Differentiability Assumption (Assumption VIII.3.2 of Andersen et al. 1993).

For any $(i, j) \in \{(0,1), (0,2), (1,2)\}$, any $h = (h^{(0,1)}, h^{(0,2)}, h^{(1,2)}) \in \mathbb{H}$, any $t \in [0, \tau]$,

$$n^{1/2}\left(\Lambda_{n,h}^{(i,j)}(t) - \Lambda_0^{(i,j)}(t)\right) = \int_{[0,t]} h^{(i,j)}(y)\ \Lambda_0^{(i,j)}(dy).$$

Note that, for any $(i, j) \in \{(0,1), (0,2), (1,2)\}$, any $h = (h^{(0,1)}, h^{(0,2)}, h^{(1,2)}) \in \mathbb{H}$, and any $t \in [0, \tau]$,

$$\left|\int_{[0,t]} h^{(i,j)}(y)\ \Lambda_0^{(i,j)}(dy)\right| \leqslant \|h\|_{\mathbb{H}}\, \Lambda_0^{(i,j)}(\tau)\left(\mathsf{L}_0^{(i)}(\tau-)\right)^{-1}.$$

Hence, the Differentiability Assumption is verified.

Next, we show that $\left(\hat{\Lambda}_n^{(0,1)}, \hat{\Lambda}_n^{(0,2)}, \hat{\Lambda}_n^{(1,2)}\right)$ is regular.

Recall the weak convergence of the log-likelihood ratio $\log \mathcal{R}_n(h)$. The continuity follows from the Le Cam's third Lemma and the weak convergence can be shown via similar arguments in the previous subsection. Hence, the regularity follows.

It remains to check Equation (8.3.5) of Andersen et al. (1993) for each coordinate.

For any $t \in [0, \tau]$, define

$$\gamma_t^{(0,1)} = \left(\varphi_t^{(0,1)}, 0, 0\right),\ \gamma_t^{(0,2)} = \left(0, \varphi_t^{(0,2)}, 0\right),\ \gamma_t^{(1,2)} = \left(0, 0, \varphi_t^{(1,2)}\right)$$

where for any $(i, j) \in \{(0,1), (0,2), (1,2)\}$ and any $s \in [0, \tau]$,

$$\varphi_t^{(i,j)}(s) = \mathrm{I}\{s \leqslant t\}\, \mathsf{L}_0^{(i)}(s-)^{-1}\left(1 - \Lambda_0^{(i,j)}\{y\}\right).$$

By (10) and (11), for any $(i, j) \in \{(0,1), (0,2), (1,2)\}$ and any $t \in [0, \tau]$,

$$\hat{\Lambda}_n^{(i,j)}(t) - \Lambda_0^{(i,j)}(t) - n^{-1/2}\left(\log \mathcal{R}_n\left(\gamma_t^{(i,j)}\right) + \frac{1}{2}\|h\|_{\mathbb{H}}^2\right) = o_p\left(n^{-1/2}\right).$$

Therefore, combining all the above results with Theorem VIII.3.2 and Theorem VIII.3.3 of Andersen et al. (1993), it follows that $\left(\hat{\Lambda}_n^{(0,1)}, \hat{\Lambda}_n^{(0,2)}, \hat{\Lambda}_n^{(1,2)}\right)$ is asymptotically efficient.

Combining Theorem VIII.3.4 of Andersen et al. (1993) with Lemma 3, the efficiency of $\hat{\mathsf{S}}_n$ follows. □

References

Andersen PK, Borgan O, Gill RD, Keiding N (1993) Statistical models based on counting processes. Springer, New York

Datta S, Satten GA, Datta S (2000) Nonparametric estimation for the three-stage irreversible illness-death model. Biometrics 56:841–847

Kiefer J, Wolfowitz J (1956) Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. Ann Math Stat 27(4):887–906

Lagakos SW, Williams JS (1978) Models for censored survival analysis: a cone class of variable-sum models. Biometrika 65(1):181–189

Lee SY, Tsai WY (2005) An estimator of the survival function based on the semi-markov model under dependent censorship. Lifetime Data Anal 11:193–211

Lee SY, Wolfe RA (1998) A simple test for independent censoring under the proportional hazards model. Biometrics 54:1176–1182

Scholz FW (1980) Towards a unified definition of maximum likelihood. Can J Stat 8:193–203

Tsiatis AA (1975) A nonidentifiability aspect of the problem of competing risks. Proc Natl Acad Sci USA 72:20–22

Van der Vaart AW, Wellner JA (1996) Weak convergence and empirical processes: with applications to statistics. Springer, New York