AISM (2006) 58: 647–674
DOI 10.1007/s10463-006-0041-0

Tomoyuki Sugimoto · Toshimitsu Hamasaki

Properties of estimators of baseline hazard functions in a semiparametric cure model

Received: 14 September 2004 / Revised: 28 July 2005 / Published online: 7 September 2006
© The Institute of Statistical Mathematics, Tokyo 2006

Abstract We consider a semiparametric cure model combining the Cox model with the logistic model. There are the two distinct methods for estimating the nonparametric baseline hazard function of the model; one is based on a pseudo partial likelihood and the other is to use an EM algorithm. In this paper, we discuss the consistency and the asymptotic normality of the estimators from the two methods. Then, we show that the estimator from the pseudo partial likelihood can be characterized by the (forward) Volterra integral equation, and the estimator from the EM algorithm by the Fredholm integral equation. These characterizations reveal differences in the properties between the estimators from the two methods. In addition, a simulation study is performed to numerically confirm the results in several finite samples.

Keywords Cox model · Logistic model · Nonparametric baseline hazard · Volterra integral equation · Fredholm integral equation · Asymptotic property

1 Introduction

In statistical analysis of time-to-event data, we may often face with a situation in which some individuals never experience the event of interest on a finite time interval. For example, in survival analysis, if the event of interest is restricted to death by an original cancer, individuals achieving the cure from the cancer do not have the event at all. Then, such cured individuals are observed as censoring at the

T. Sugimoto(✉) · T. Hamasaki
Department of Biomedical Statistics, Graduate School of Medicine,
Osaka University, 2-2 Yamada-oka, Suita,
Osaka 565-0871, Japan
E-mail: sugimoto@medstat.med.osaka-u.ac.jp

end of the follow-up period. The standard Cox model (Cox, 1972) is extremely useful tool in analysis of censored data, while it may be inappropriate to describe a true structure of such data. To overcome this problem, the cure-mixture model can be employed, where the model is formulated by assuming that population is a mixture of susceptible and non-susceptible (cured) individuals.

The parametric cure-mixture models have received much attention earlier. Recently, however, nonparametric modeling approaches with a cure rate have been proposed by several authors (Kuk and Chen, 1992; Lu and Ying, 2004; Tsodikov, 1998). Especially, Kuk and Chen (1992) proposed the semiparametric generalization of the cure-mixture model by Farewell (1986), which used the Cox model for the latency distribution and the logistic model for cure fraction; we call this "the Cox cure model".

Several methods have been available for estimating regression parameters in the Cox cure model. Kuk and Chen (1992) developed a method involving Monte Carlo simulation for approximating a rank-based likelihood. Peng and Dear (2000) and Sy and Taylor (2000) considered a method based on the EM algorithm to compute the joint parametric–nonparametric likelihood. Furthermore, Sugimoto et al. (2005) proposed a method based on the pseudo partial likelihood. Expect the formulation of likelihood, major differences among the methods are caused by how the nonparametric baseline hazard function is estimated or eliminated. Therefore, the estimation of the baseline hazard function is the key in the Cox cure model and it is surely important to know the properties of estimators for the baseline hazard. However, unfortunately the properties of estimators for the baseline hazard have not yet been established.

The two distinct methods have been proposed in estimating the baseline hazard of the Cox cure model. One is based on the pseudo partial likelihood method (referred to as the pseudo-estimation) and the other is to use the EM algorithm (referred to as the EM-estimation). In this paper, we discuss properties of the two estimators from the pseudo-estimation and the EM-estimation. This would throw light on the background of the existence of the two estimators. The main purpose of this paper is to investigate the asymptotic behaviors of the two estimators of the baseline hazard. Especially, for given regression parameters, we establish the strong consistency and asymptotic normality of the two estimators. Concurrently, we formulate an attractive relationship between the two estimators by characterizing them using the (forward) Volterra and Fredholm integral equations, respectively. Then, the Fredholm integral equation is altered into a backward Volterra under a certain additional condition. This fact intuitionally suggests that the two estimators would have the similar relationship to that in the forward and backward system in a stochastic process.

This paper is structured as follows: in Sect. 2 we describe the key notation and formulation of the model and the pseudo- and EM-estimations used in the subsequence sections. In Sect. 3 we establish the large sample properties of the two estimators for the baseline hazard including the strong consistency and asymptotic normality, and then we provide these proofs in Sects. 4 and 5. In Sect. 6 we perform a simulation study to confirm the results obtained in the previous sections. Finally, in Sect. 7 we summarize some findings as concluding remarks.

2 Notation and formulation

2.1 The Cox cure model

Let T_i denote the i-th observable random variable $\min(T_i^*, U_i)$, where T_i^* and U_i are the random variables of true survival and censoring, respectively. Suppose that T_i and U_i are independent. Let Δ_i denote the i-th censoring indicator $I(T_i^* \leq U_i)$, where $I(\cdot)$ is the indicator function. Let X_i and Z_i be the i-th covariate vectors related to cure rate and uncured survival, respectively, where the covariates X_i and Z_i are usually random variables. In the Cox cure model, the data of $(T_i, \Delta_i, X_i, Z_i)$ are supposed to be observed for $i = 1, \ldots, n$. Also, as usual, information on (T_i, Δ_i) can be expressed as $(N_i(t), Y_i(t)), 0 \leq t$ by the counting and at-risk processes, where $N_i(t) = I(T_i \leq t, \Delta_i = 1)$ and $Y_i(t) = I(T_i \geq t)$.

Let λ_i and r_i be the i-th hazard function on given Z_i and the i-th relative risk function, respectively. Suppose that λ_i holds the proportional hazards model if the i-th individual is uncured and λ_i is zero-hazard otherwise. That is, we define

$$\lambda_i(t) = \lambda_i(t, \beta) = G_i \lambda_0(t) r_i(\beta) \quad \text{and} \quad r_i = r_i(\beta) = r(\beta' Z_i) = \exp(\beta' Z_i),$$

where $G_i = I(\text{the } i\text{−th individual is uncured})$, $\lambda_0(t)$ is the baseline hazard function and β is the parameter vector corresponding to Z_i.

Let c_i be the i-th cure rate function defined as

$$c_i = c_i(\alpha) = c(\alpha' X_i) = \Pr(G_i = 0 | X_i),$$

where α is the parameter vector corresponding to X_i and $c(\cdot)$ is a link function in the terminology of generalized linear model, for example, $c(x) = \exp(x)/\{1+\exp(x)\}$ in the logistic model.

Let $w_i(t; \theta, \Lambda_0)$ be the conditional probability that the i-th individual will eventually belong in the uncured group given (θ, Λ_0) and $Y_i(t) = 1$ for t, where $\theta = (\alpha', \beta')'$ and $\Lambda_0(t)$ is the cumulative baseline hazard $\int_0^t \lambda_0(s) \mathrm{d}s$. Suppose that the censoring density $\Pr(U_i | Z_i)$ and the conditional density $\Pr(Z_i | X_i)$ of covariates do not depend on G_i. Then w_i is written as

$$w_i(t; \theta, \Lambda_0) = (1 - c_i(\alpha)) S_i(t; \beta, \Lambda_0) / \{c_i(\alpha) + (1 - c_i(\alpha)) S_i(t; \beta, \Lambda_0)\},$$

where $S_i(t; \beta, \Lambda_0)$ is the i-th survival function $\exp\{-r_i(\beta)\Lambda_0(t)\}$.

For simplicity, we may occasionally omit θ when it is clear that a function depends on θ, for example, $r_i = r_i(\beta)$, $c_i = c_i(\alpha)$, $S_i(t; \Lambda_0) = S_i(t; \beta, \Lambda_0)$ and so on.

2.2 The estimator of baseline hazard from the pseudo-estimation

Here we briefly describe the pseudo-estimation for the estimator of baseline hazard. Let $\hat{\Lambda}_0^\sharp$ be the estimator of cumulative baseline hazard function from the pseudo-estimation, then we can write

$$\hat{\Lambda}_0^\sharp(t; \theta) = \sum_{\{i: T_i \leq t\}} \frac{\Delta_i}{\sum_{j=1}^n Y_j(T_i) r_j(\beta) w_j \left(T_i; \theta, \hat{\Lambda}_0^\sharp\right)}. \tag{1}$$

Suppose that $\mathscr{F}_{t-}$ is the history just prior to t but does not include the information on G_i, $i = 1, \ldots, n$. Given $\mathscr{F}_{t-}$, the conditional expectation of $\mathrm{d}N_i(t)$ is

$$\mathrm{E}[\mathrm{d}N_i(t)|\mathscr{F}_{t-}] = Y_i(t)r_i(\beta)w_i(t;\theta,\Lambda_0)\mathrm{d}\Lambda_0(t).$$

Hence, the processes $M_i(t) = N_i(t) - \int_0^t Y_i(s)r_i(\beta)w_i(s;\theta,\Lambda_0)\mathrm{d}\Lambda_0(s)$, $i = 1, \ldots, n$ are $\mathscr{F}_t$-martingales if (θ, Λ_0) is known. The (pseudo) full likelihood constructed from the counting processes in such filtration $\mathscr{F}_t$ is written as

$$L_{\mathrm{pf}}(\theta,\Lambda_0) = \prod_{i=1}^{n} \{\lambda_0(T_i)r_i(\beta)w_i(T_i;\theta,\Lambda_0)\}^{\Delta_i} \times \exp\left\{-\int_0^{T_i} r_i(\beta)w_i(t;\theta,\Lambda_0)\mathrm{d}\Lambda_0(t)\right\}.$$

Then, we can find $\hat{\Lambda}_0^{\sharp}(\cdot;\theta)$ in (1) as a Breslow's estimate (Breslow, 1972) corresponding to L_{pf}. Substituting $\hat{\Lambda}_0^{\sharp}(\cdot;\theta)$ into Λ_0 in $L_{\mathrm{pf}}(\theta,\Lambda_0)$ leads to a partial likelihood for θ:

$$L_{\mathrm{pp}}\left(\theta,\hat{\Lambda}_0^{\sharp}\right) = \prod_{i=1}^{n}\left\{r_i(\beta)w_i\left(T_i;\theta,\hat{\Lambda}_0^{\sharp}\right) \Big/ \sum_{j=1}^{n} Y_j(T_i)r_j(\beta)w_j\left(T_i;\theta,\hat{\Lambda}_0^{\sharp}\right)\right\}^{\Delta_i},$$

which we call the pseudo partial likelihood. The estimate of θ can be obtained by maximizing $L_{\mathrm{pp}}(\theta,\hat{\Lambda}_0^{\sharp})$ over θ. For further discussions, see Sugimoto et al. (2005).

2.3 The estimator of baseline hazard from the EM-estimation

Here we briefly describe the EM-estimation for the estimator of baseline hazard. Let $\hat{\Lambda}_0^{\flat}$ be the estimator of cumulative baseline hazard function from the EM-estimation, then we can write

$$\hat{\Lambda}_0^{\flat}(t;\theta) = \sum_{\{i:T_i\le t\}} \frac{\Delta_i}{\sum_{j=1}^{n} Y_j(T_i)r_j(\beta)\left\{\Delta_j + (1-\Delta_j)w_j\left(T_j;\theta,\hat{\Lambda}_0^{\flat}\right)\right\}}. \quad (2)$$

The observed full (joint parametric–nonparametric) likelihood in the Cox cure model is

$$L_{\mathrm{mf}}(\theta,\Lambda_0) = \prod_{i=1}^{n}\{\lambda_0(T_i)r_i(\beta)(1-c_i(\alpha))S_i(T_i;\beta,\Lambda_0)\}^{\Delta_i} \times \{c_i(\alpha) + (1-c_i(\alpha))S_i(T_i;\beta,\Lambda_0)\}^{1-\Delta_i}.$$

The E-step transforms an incomplete element contributed as a marginal probability $c_i + (1-c_i)S_i(T_i;\Lambda_0)$ into a complete-data form based on the EM algorithm. The EM-type complete likelihood via the m-th E-step for L_{mf} is

$$L_{\mathrm{f}}^{\mathrm{EM}}(\theta,\Lambda_0|\overline{w}^{(m)}) = \prod_{i=1}^{n} c_i(\alpha)^{1-\overline{w}_i^{(m)}}(1-c_i(\alpha))^{\overline{w}_i^{(m)}}\{\lambda_0(T_i)r_i(\beta)\}^{\Delta_i}\, S_i(T_i;\beta,\Lambda_0)^{\overline{w}_i^{(m)}},$$

where $\overline{w}_i^{(m)} = \Delta_i + (1 - \Delta_i)w_i(T_i; \theta^{(m)}, \Lambda_0^{(m)})$ and $(\theta^{(m)}, \Lambda_0^{(m)})$ is a pair of current estimates in the m-th EM-iteration ($m \geq 0$). In the m-th M-step, we find a Breslow's estimate $\hat{\Lambda}_0^\flat(t|\beta, \overline{w}^{(m)}) = \sum_{\{i:T_i \leq t\}} \Delta_i / \sum_{j=1}^n Y_j(T_i)r_j(\beta)\overline{w}_j^{(m)}$ for Λ_0. Substituting $\hat{\Lambda}_0^\flat(\cdot|\beta, \overline{w}^{(m)})$ into Λ_0 in $L_{\mathrm{f}}^{\mathrm{EM}}(\theta, \Lambda_0|\overline{w}^{(m)})$ leads to a partial likelihood for θ:

$$L_{\mathrm{p}}^{\mathrm{EM}}(\theta|\overline{w}^{(m)}) = \prod_{i=1}^n c_i(\alpha)^{1-\overline{w}_i^{(m)}}(1 - c_i(\alpha))^{\overline{w}_i^{(m)}} \left\{ r_i(\beta) \Big/ \sum_{j=1}^n Y_j(T_i)r_j(\beta)\overline{w}_j^{(m)} \right\}^{\Delta_i}.$$

The M-step of $L_{\mathrm{f}}^{\mathrm{EM}}(\theta, \Lambda_0|\overline{w}^{(m)})$ on (θ, Λ_0) is replaced by maximizing $L_{\mathrm{p}}^{\mathrm{EM}}(\theta|\overline{w}^{(m)})$ over only θ in cases where Λ_0 is nonparametric nuisance, so that we have $\theta^{(m+1)} = \mathrm{argmax}_\theta L_{\mathrm{p}}^{\mathrm{EM}}(\theta|\overline{w}^{(m)})$ and $\Lambda_0^{(m+1)}(\cdot) = \hat{\Lambda}_0^\flat(\cdot|\beta^{(m+1)}, \overline{w}^{(m)})$. For the next $(m+1)$-th E-step, $\overline{w}^{(m)}$ is updated to $\overline{w}^{(m+1)}$. We can search $\theta^{(m)}$ by maximizing $L_{\mathrm{mf}}(\theta^{(m)}, \Lambda_0^{(m)})$ with the above EM-iterations. Also, this EM-estimation is usually equivalent to achieving $(\theta^{(m)}, \Lambda_0^{(m)}) = (\theta^{(m+1)}, \Lambda_0^{(m+1)})$. Then, since $\overline{w}^{(m)} = \overline{w}^{(m+1)}$, we have $\Lambda_0^{(m+1)}(\cdot) = \hat{\Lambda}_0^\flat(\cdot|\beta^{(m+1)}, \overline{w}^{(m+1)}) = \hat{\Lambda}_0^\flat(\cdot|\theta^{(m+1)}, \Lambda_0^{(m+1)})$. For such $\theta = \theta^{(m+1)}$ and $\hat{\Lambda}_0^\flat = \Lambda_0^{(m+1)}$, we can find $\hat{\Lambda}_0^\flat(t; \theta) = \hat{\Lambda}_0^\flat(t|\theta, \hat{\Lambda}_0^\flat)$ described in (2). For further discussions, see Sugimoto and Goto (1999), Peng and Dear (2000) and Sy and Taylor (2000).

3 Main results

We shall observe the survival time on the interval $[0, \tau_e]$, where τ_e is finite and the largest follow-up time. Let $\lambda_0^*(\cdot)$ be the true function of $\lambda_0(\cdot)$. Also, let α^* and β^* be the true parameters of α and β. We define the subdistribution function $\mathbb{N}(t) = \Pr(T \leq t, \Delta = 1)$, then we can write

$$\begin{aligned}\mathbb{N}(t) &= \mathrm{E}\left[I(T \leq t, \Delta = 1|X, Z)\right] \\ &= \mathrm{E}\left[\int_0^t \lambda_0^*(T)r^*(1 - c^*)S^*(T)S_C^*(T; X, Z)\mathrm{d}T\right],\end{aligned}$$

where $c^* = c(\alpha^{*\prime}X)$, $r^* = r(\beta^{*\prime}Z)$, $S^*(t) = \exp\{-r^*\Lambda_0^*(t)\}$, $\Lambda_0^*(t) = \int_0^t \lambda_0^*(s)\mathrm{d}s$ and $S_C^*(t; X, Z)$ is the true survival function of censoring. We assume $S_C^*(t; X, Z)$ is Lipschitz continuous on $t \in [0, \tau_e)$ and, for simplicity, we write $S_C^*(t) = S_C^*(t; X, Z)$. Differentiating $\mathbb{N}(t)$ with respect to t, we have the relationship

$$\lambda_0^*(t)\mathrm{d}t = \mathrm{d}\mathbb{N}(t) \big/ \mathrm{E}\left[r^*(1 - c^*)S^*(t)S_C^*(t)\right]. \tag{3}$$

Convergence almost surely ($\rightarrow_{a.s.}$), convergence in probability ($\rightarrow_p$) and convergence in distribution ($\rightarrow_D$) are relative to the probability measures parameterized by λ_0^*, α^*, β^*, and S_C^*.

Let $\mathbb{N}_n(t)$ be the empirical estimate of $\mathbb{N}(t)$. Then we can find the following expressions

$$\mathbb{N}_n(t) = \mathrm{E}_n \left[\int_0^t \mathrm{d}N_i(s) \right] = \mathrm{E}_n \left[(1 - Y_i(t))\Delta_i \right] = \mathrm{E}_n \left[I(T_i \le t, \Delta_i = 1) \right],$$

where $\mathrm{E}_n[f(R_i)] = n^{-1} \sum_{i=1}^n f(R_i)$ is the empirical expectation corresponding to the theoretical expectation $\mathrm{E}[f(R)]$ of a random function $f(R)$. In this paper, we consistently use this notation.

3.1 On the estimator $\hat{\Lambda}_0^\sharp$ from the pseudo-estimation

First, we define the important notations:

$$\mathbb{S}_n^\sharp(t; \theta, \Lambda_0) = \mathrm{E}_n[r_j(\beta) w_j(t; \theta, \Lambda_0) Y_j(t)] \quad \text{and}$$
$$\mathbb{S}^\sharp(t; \theta, \Lambda_0) = \mathrm{E}[r(\beta) w(t; \theta, \Lambda_0) Y(t)].$$

The empirical expectation expression of (1) is

$$\hat{\Lambda}_0^\sharp(t; \theta) = \int_0^t \mathrm{d}\mathbb{N}_n(s) \Big/ \mathbb{S}_n^\sharp \left(s; \theta, \hat{\Lambda}_0^\sharp \right). \tag{4}$$

To compute $\hat{\Lambda}_0^\sharp(\cdot)$, we can solve (4) as follows: given $\hat{\Lambda}_0^\sharp(T_{(i-1)})$, the next $\hat{\Lambda}_0^\sharp(T_{(i)})$ is found from $x = \hat{\Lambda}_0^\sharp(T_{(i)}) - \hat{\Lambda}_0^\sharp(T_{(i-1)})$ such that $x - \Delta_{(i)} / \sum_j Y_j(T_{(i)}) r_j w_j(T_{(i)}; \hat{\Lambda}_0^\sharp(T_{(i-1)}) + x) = 0$, where $(T_{(1)}, \ldots, T_{(k)})$ is the order statistic of observed failure times of $(T_1, \ldots, T_n)$ and $T_{(0)} = 0$.

The heuristic limit of (4) is

$$\Lambda_0^\sharp(t; \theta) = \int_0^t \mathrm{d}\mathbb{N}(s) \Big/ \mathbb{S}^\sharp \left(s; \theta, \Lambda_0^\sharp \right). \tag{5}$$

Note that we can write $\mathbb{S}^\sharp(s; \theta, \Lambda_0) = \mathrm{E}\big[r(\beta) w(s; \theta, \Lambda_0) \{c^* + (1 - c^*) S^*(s)\} S_C^*(s) \big]$ since the conditional expectation of $Y(s)$ given (X, Z) is $\{c^* + (1 - c^*) S^*(s)\} S_C^*(s)$. So, when $\theta = \theta^*$, one solution of $\Lambda_0^\sharp(\cdot; \theta^*)$ is $\Lambda_0^*(\cdot)$ using (3) since we have $\mathbb{S}^\sharp(t; \theta^*, \Lambda_0^*) = \mathrm{E}[r^* w^*(s) \{c^* + (1 - c^*) S^*(s)\} S_C^*(s)] = \mathrm{E}[r^* (1 - c^*) S^*(t) S_C^*(t)]$. To see this uniqueness, we consider the following equation derived from a Taylor expansion of (5) around Λ_0^*:

$$\Lambda_0^\sharp(t; \theta^*) - \Lambda_0^*(t) = \int_0^t \left(\Lambda_0^\sharp(s; \theta^*) - \Lambda_0^*(s) \right) \tilde{B}^{*\sharp}(s) \mathrm{d}\Lambda_0^*(s) \Big/ \mathbb{S}^\sharp \left(s; \theta^*, \Lambda_0^\sharp \right),$$

where $\tilde{B}^{*\sharp}(s) = \mathrm{E}[r^{*2} w(s; \theta^*, \tilde{\Lambda}_0)(1 - w(s; \theta^*, \tilde{\Lambda}_0)) Y(s)]$ and $\tilde{\Lambda}_0(s)$ is on the line segment between $\Lambda_0^\sharp(s; \theta^*)$ and $\Lambda_0^*(s)$ for all $s \in [0, \tau_e]$. For a similar detailed discussion, see Sect. 4.1. This is a Volterra integral equation, so that it can be uniquely solved by Theorem A.1 when $\tau_e < \infty$ and $\sup_{s \in [0, \tau_e]} |\tilde{B}^{*\sharp}(s) \mathrm{d}\Lambda_0^*(s) / \mathbb{S}^\sharp(s; \theta^*, \Lambda_0^\sharp)| < \infty$.

Now we shall summarize the results that $\hat{\Lambda}_0^{\sharp}(\cdot;\theta)$ converges almost surely to the true functional form $\Lambda_0^{\sharp}(\cdot;\theta)$ and the asymptotic distribution of $\sqrt{n}(\hat{\Lambda}_0^{\sharp}(\cdot;\theta)-\Lambda_0^{\sharp}(\cdot;\theta))$ is a Gaussian martingale under the true parameter θ^*.

Conditions ($\sharp$) Let Θ be a neighborhood of θ^*. $\Lambda_0^{\sharp}$ is a cumulative hazard defined in (5).

(0) $\mathrm{E}\left[r(\beta)^2\right] < \infty$ on $\beta \in \Theta$.

(I$^\sharp$) $\mathbb{S}^{\sharp}(t;\theta,\Lambda_0^{\sharp})$ is bounded away from zero on $(\theta,t) \in \Theta \times [0,\tau_e]$.

(II$^\sharp$) $\sup_{\theta\in\Theta,t\in[0,\tau_e]} \left|\mathbb{S}_n^{\sharp}(t;\theta,\Lambda_0^{\sharp}) - \mathbb{S}^{\sharp}(t;\theta,\Lambda_0^{\sharp})]\right| \to_{a.s.} 0$ (as $n\to\infty$).

(III$^\sharp$) $\sqrt{n}\left\{\mathbb{S}_n^{\sharp}(t;\theta^*,\Lambda_0^{\sharp}) - \mathbb{S}^{\sharp}(t;\theta^*,\Lambda_0^{\sharp})]\right\}$ converges weakly to a zero-mean Gaussian process on $t \in [0,\tau_e]$ (as $n\to\infty$).

Lemma 1 *Suppose that Conditions (0), (I$^\sharp$), and (II$^\sharp$) are satisfied. Then, as $n\to\infty$*

$$\sup_{\theta\in\Theta,t\in[0,\tau_e]} |\hat{\Lambda}_0^{\sharp}(t;\theta) - \Lambda_0^{\sharp}(t;\theta)| \to_{a.s.} 0.$$

Lemma 2 *Suppose that Conditions (0), (I$^\sharp$), and (III$^\sharp$) are satisfied at $\theta=\theta^*$. Then, as $n\to\infty$, $\sqrt{n}(\hat{\Lambda}_0^{\sharp}(t;\theta^*)-\Lambda_0^{*}(t))$ converges weakly to a zero-mean Gaussian martingale process on $t\in[0,\tau_e]$ with the covariance function given in (12).*

The proofs of Lemmas 1 and 2 are given in Sects. 4.2 and 4.3, respectively. Also, the derivations of the covariance function and the martingale property are discussed in Sect. 4.4.

3.2 On the estimator $\hat{\Lambda}_0^{\flat}$ from the EM-estimation

First, we define the important notations:

$$\mathbb{S}_n^{\flat}(t;\theta,\Lambda_0) = \mathrm{E}_n\left[r_j(\beta)\{\Delta_j + (1-\Delta_j)w_j(T_j;\theta,\Lambda_0)\}Y_j(t)\right] \quad \text{and}$$
$$\mathbb{S}^{\flat}(t;\theta,\Lambda_0) = \mathrm{E}\left[r(\beta)\{\Delta + (1-\Delta)w(T;\theta,\Lambda_0)\}Y(t)\right].$$

The empirical expectation expression of (2) is

$$\hat{\Lambda}_0^{\flat}(t;\theta) = \int_0^t \mathrm{d}\mathbb{N}_n(s) \Big/ \mathbb{S}_n^{\flat}\left(s;\theta,\hat{\Lambda}_0^{\flat}\right). \tag{6}$$

Also, by the process expression, $\mathbb{S}_n^{\flat}$ can be written as

$$\mathbb{S}_n^{\flat}\left(s;\theta,\hat{\Lambda}_0^{\flat}\right) = \mathrm{E}_n\left[\int_s^{\tau_e} r_j(\beta)\mathrm{d}N_j(x)\right] + \mathrm{E}_n\left[\int_s^{\tau_e} r_j(\beta)w_j\left(x;\theta,\hat{\Lambda}_0^{\flat}\right)\mathrm{d}N_j^C(x)\right],$$

where $N_j^C(x) = I(T_j \le x, \Delta_j = 0)$. From these expressions, we can find that $\hat{\Lambda}_0^{\flat}(t;\theta)$ depends on future values $\hat{\Lambda}_0^{\flat}(x;\theta)$ for all $x > t$ included in $w(x;\theta,\hat{\Lambda}_0^{\flat})$.

To avoid such dependencies, and to obtain an easier expression to compute the estimate, using a backward procedure, we can rewrite $\hat{\Lambda}_0^{\flat}(t;\theta)$ by

$$\hat{\Lambda}_0^{\flat}(t;\theta) = \hat{\Lambda}_0^{\flat}(\tau_e;\theta) - \int_{t+}^{\tau_e} \mathrm{d}\mathbb{N}_n(s) \Bigg/ \Bigg\{ \mathrm{E}_n\left[\int_s^{\tau_e} r_j(\beta)\mathrm{d}N_j(x)\right] + \mathrm{E}_n\left[\int_s^{\tau_e} r_j(\beta) w_j\left(x;\theta,\hat{\Lambda}_0^{\flat}\right)\mathrm{d}N_j^C(x)\right]\Bigg\}.$$

Therefore, given a value of $\hat{\Lambda}_0^{\flat}(T_{(k)};\theta)$, the use of this new expression can automatically provide a sequence of $\hat{\Lambda}_0^{\flat}(T_{(k-1)};\theta),\ldots,\hat{\Lambda}_0^{\flat}(T_{(1)};\theta)$ and $\hat{\Lambda}_0^{\flat}(T_{(0)};\theta)$ in turn. Then, the sequence from $\hat{\Lambda}_0^{\flat}(T_{(k)};\theta)$ which gives $\hat{\Lambda}_0^{\flat}(T_{(0)};\theta) = 0$ is the solution of (6).

The heuristic limit of (6) is

$$\Lambda_0^{\flat}(t;\theta) = \int_0^t \mathrm{d}\mathbb{N}(s) \Big/ \mathbb{S}^{\flat}\left(s;\theta,\Lambda_0^{\flat}\right). \tag{7}$$

Also, using a process expression as well as $\mathbb{S}_n^{\flat}$, we rewrite

$$\mathbb{S}^{\flat}\left(s;\theta,\Lambda_0^{\flat}\right) = \mathrm{E}\left[\int_s^{\tau_e} r(\beta)\mathrm{d}N(x)\right] + \mathrm{E}\left[\int_s^{\tau_e} r(\beta)w\left(x;\theta,\Lambda_0^{\flat}\right)\mathrm{d}N^C(x)\right].$$

Note that $\mathrm{E}[\mathrm{d}N(x)|X,Z] = (1-c^*)S^*(x)S_C^*(x)r^*\mathrm{d}\Lambda_0^*(x)$ and $\mathrm{E}[\mathrm{d}N^C(x)|X,Z] = \{c^* + (1-c^*)S^*(x)\}\,S_C^*(x)\mathrm{d}\Lambda_C^*(x)$, where Λ_C^* is the cumulative hazard corresponding to S_C^*. So, when $\theta = \theta^*$, one solution of $\Lambda_0^{\flat}(\cdot;\theta^*)$ is $\Lambda_0^*(\cdot)$ using (3) since we have

$$\begin{aligned}\mathbb{S}^{\flat}(s;\theta^*,\Lambda_0^*) &= \mathrm{E}\left[r^*\int_s^{\tau_e}(1-c^*)S^*(x)S_C^*(x)r^*\mathrm{d}\Lambda_0^*(x)\right] \\ &\quad + \mathrm{E}\left[r^*\int_s^{\tau_e} w^*(x)\left\{c^* + (1-c^*)S^*(x)\right\}S_C^*(x)\mathrm{d}\Lambda_C^*(x)\right] \\ &= \mathrm{E}\left[r^*(1-c^*)\int_s^{\tau_e} S^*(x)S_C^*(x)\{r^*\mathrm{d}\Lambda_0^*(x) + \mathrm{d}\Lambda_C^*(x)\}\right] \\ &= \mathrm{E}\left[r^*(1-c^*)S^*(s)S_C^*(s)\right].\end{aligned}$$

To see this uniqueness, we consider the following equation derived from a Taylor expansion of (7) around Λ_0^*:

$$\Lambda_0^\flat(t;\theta^*) - \Lambda_0^*(t) = \int_0^{\tau_e} \left(\Lambda_0^\flat(s;\theta^*) - \Lambda_0^*(s)\right) \mathrm{d}\tilde{B}^{*\flat}(s) \times \left\{ \int_0^{\min(s,t)} \mathrm{d}\Lambda_0^*(u) \Big/ \mathbb{S}^\flat\left(u;\theta^*,\Lambda_0^\flat\right) \right\},$$

where $\mathrm{d}\tilde{B}^{*\flat}(s) = \mathrm{E}[r^{*2} w(s;\theta^*,\tilde{\Lambda}_0)(1 - w(s;\theta^*,\tilde{\Lambda}_0))\mathrm{d}N^C(s)]$ and $\tilde{\Lambda}_0(s)$ is on the line segment between $\Lambda_0^\flat(s;\theta^*)$ and $\Lambda_0^*(s)$ for all $s \in [0,\tau_e]$. For a similar detailed discussion, see Sect. 5.1. This is a Fredholm integral equation, so that it is solved by Theorem A.2. For the Fredholm determinant $D(\xi)$ defined in Theorem A.2 with $K^\flat(t,s)\mathrm{d}s = \mathrm{d}\tilde{B}^{*\flat}(s) \left\{\int_0^{\min(s,t)} \mathrm{d}\Lambda_0^*(u)/\mathbb{S}^\flat(u;\theta^*,\Lambda_0^\flat)\right\}$, let $D_{\theta^*}^\flat = D(1)$. When we can assume $D_{\theta^*}^\flat \neq 0$, the solution is unique by Theorem A.2. Similarly, this fact is held in the Fredholm determinant $D(1)$ corresponding to cases of some finite n-sample or $\theta \neq \theta^*$. For simplicity, suppose that the uniquenesses of $\hat{\Lambda}_0^\flat(\cdot;\theta)$ and $\Lambda_0^\flat(\cdot;\theta)$ described by such Fredholm determinant are held without further investigation in this paper.

Now we shall summarize the results that $\hat{\Lambda}_0^\flat(\cdot;\theta)$ converges almost surely to the true functional form $\Lambda_0^\flat(\cdot;\theta)$ and the asymptotic distribution of $\sqrt{n}(\hat{\Lambda}_0^\flat(\cdot;\theta) - \Lambda_0^\flat(\cdot;\theta))$ is Gaussian under the true parameter θ^*.

Conditions ($\flat$) Let Θ be a neighborhood of θ^*. $\Lambda_0^\flat$ is a cumulative hazard defined in (7).

- (0) $\mathrm{E}\left[r(\beta)^2\right] < \infty$ on $\beta \in \Theta$.
- (I$^\flat$) $\mathbb{S}^\flat(t;\theta,\Lambda_0^\flat)$ is bounded away from zero on $(\theta,t) \in \Theta \times [0,\tau_e]$.
- (II$^\flat$) $\sup_{\theta\in\Theta,t\in[0,\tau_e]} \left|\mathbb{S}_n^\flat(t;\theta,\Lambda_0^\flat) - \mathbb{S}^\flat(t;\theta,\Lambda_0^\flat)\right| \to_{a.s.} 0$ (as $n \to \infty$).
- (III$^\flat$) $\sqrt{n}\left\{\mathbb{S}_n^\flat(t;\theta^*,\Lambda_0^\flat) - \mathbb{S}^\flat(t;\theta^*,\Lambda_0^\flat)\right\}$ converges weakly to a zero-mean Gaussian process on $t \in [0,\tau_e]$ (as $n \to \infty$).

Lemma 3 *Suppose that Conditions (0), (I$^\flat$), and (II$^\flat$) are satisfied. Let $\hat{\Lambda}_0^\flat(\cdot;\theta)$ and $\Lambda_0^\flat(\cdot;\theta)$ be unique. Then, as $n \to \infty$,*

$$\sup_{\theta\in\Theta,t\in[0,\tau_e]} \left|\hat{\Lambda}_0^\flat(t;\theta) - \Lambda_0^\flat(t;\theta)\right| \to_{a.s.} 0.$$

Lemma 4 *Suppose that Conditions (0), (I$^\flat$), and (III$^\flat$) are satisfied at $\theta = \theta^*$. Let $\hat{\Lambda}_0^\flat(\cdot;\theta^*)$ and $\Lambda_0^\flat(\cdot;\theta^*)$ be unique. Then, as $n \to \infty$, $\sqrt{n}(\hat{\Lambda}_0^\flat(t;\theta^*) - \Lambda_0^*(t))$ converges weakly to a zero-mean Gaussian process on $t \in [0,\tau_e]$ with the covariance function given in (16)*

The proofs of Lemmas 3 and 4 are given in Sects. 5.2 and 5.3, respectively. The derivation of the covariance function is discussed in Sect. 5.4.

3.3 Identically distributed covariates

Conditions (♯) and Conditions (♭) given in the previous sections are formed to conveniently use in the proof of Lemmas 1, 2, 3, and 4. However, the Conditions may appear to be not so direct for the random variables included in the model. In this section, we consider the i.i.d. case of covariates and provide elementary conditions corresponding to Conditions (♯) and Conditions (♭).

Conditions (i.i.d.) In the i.i.d. case with finite dimensional vectors $(X_i, Z_i, T_i, \Delta_i)$ left continuous with right hand limits, Conditions (♯) and Conditions (♭) are satisfied if $\Pr(Y(\tau_e) = 1) > 0$ and there exists a neighborhood Θ of θ^* such that Θ is compact and

$$\Lambda_0^\sharp(\tau_e; \theta) < \infty, \quad \Lambda_0^\flat(\tau_e; \theta) < \infty, \quad 0 < \mathrm{E}[r(\beta)^2] < \infty, \text{and} \quad \mathrm{E}[c(\alpha)] < 1.$$

Proof Let $\mathcal{S}_i^\sharp(t, \theta) = r_i(\beta) w_i(t; \theta, \Lambda_0^\sharp) Y_i(t)$ and $\mathcal{S}_i^\flat(t, \theta) = r_i(\beta)\{\Delta_i + (1 - \Delta_i) w_i(T_i; \theta, \Lambda_0^\flat)\} Y_i(t)$, $i = 1, \ldots, n$, which are also i.i.d. elements of the processes indexed by $[0, \tau_e] \times \Theta$.

By $\mathrm{E}[c(\alpha)] < 1$, $\mathrm{E}[r(\beta)] < \infty$ and $\Lambda_0^\jmath(\tau_e; \theta) < \infty$, we have $\mathrm{E}[w(\tau_e; \theta, \Lambda_0^\jmath)] > 0$ ($\jmath = \sharp, \flat$). Also, by $\Pr(Y(\tau_e) = 1) > 0$ and $\mathrm{E}[r(\beta)^2] > 0$, we have $\mathrm{E}[Y(\tau_e)] > 0$ and $\mathrm{E}[r(\beta)] > 0$, respectively. Hence, since the probability of $\mathcal{S}_i^\jmath(t, \theta) = 0$ ($\jmath = \sharp, \flat$) is not zero at each $t \in [0, \tau_e]$, ($\mathrm{I}^\sharp$) and ($\mathrm{I}^\flat$) are satisfied.

For ($\mathrm{II}^\sharp$) and ($\mathrm{II}^\flat$), we use a strong law of large numbers (SLLN) brought from the Glivenko–Cantelli. Here, we can refer the SLLN on space of right-continuous functions on $[0, \tau_e]$ with left limits taking values in a separable Banach space (Andersen and Gill, 1982, Theorem III.1). Since $\mathcal{S}_i^\sharp(t, \theta)$ and $\mathcal{S}_i^\flat(t, \theta)$ are continuous functions on compact Θ and we have $\mathrm{E} \sup_{t \in [0, \tau_e], \theta \in \Theta} |\mathcal{S}_i^\jmath(t, \theta)| \leq \mathrm{E} \sup_{\theta \in \Theta} |\mathcal{S}_i^\jmath(0, \theta)| < \infty$ ($\jmath = \sharp, \flat$), ($\mathrm{II}^\sharp$) and ($\mathrm{II}^\flat$) are proved by direct use of the SLLN.

For ($\mathrm{III}^\sharp$) and ($\mathrm{III}^\flat$), we use a central limit theorem (CLT) brought from the Donsker. Here, we can refer the CLT in van der Vaart and Wellner (1996, Theorem 2.11.9). Since sample paths $t \mapsto \mathcal{S}_i^\jmath(t, \theta^*)$ of the processes $\mathcal{S}_i^\jmath(\cdot, \theta^*)$ are non-increasing ($\jmath = \sharp, \flat$), the envelop functions of $\mathcal{S}_i^\jmath(\cdot, \theta^*)$ are $\mathcal{S}_i^\jmath(0, \theta^*)$ and are square integrable by $\mathrm{E}[r(\beta)^2] < \infty$, so that Lindeberg's condition for the envelops in the CLT is satisfied. Also, the entropy condition in the CLT is satisfied using Example 2.11.16 in van der Vaart and Wellner (1996) because of the right-continuity with left limits and the monotoneity of sample paths $t \mapsto \mathcal{S}_i^\jmath(t, \theta^*)$ ($\jmath = \sharp, \flat$). Therefore, ($\mathrm{III}^\sharp$) and ($\mathrm{III}^\flat$) are proved by use of the CLT. □

4 The asymptotic behavior of $\hat{\Lambda}_0^\sharp$ from the pseudo-estimation

For simplicity, for some $\theta \in \Theta$, we write $\Lambda_0^\sharp(t) = \Lambda_0^\sharp(t; \theta)$ and $\hat{\Lambda}_0^\sharp(t) = \hat{\Lambda}_0^\sharp(t; \theta)$ by dropping θ. Also, we write $\hat{w}^\sharp(t) = w(t; \theta, \hat{\Lambda}_0^\sharp)$ and $w^\sharp(t) = w(t; \theta, \Lambda_0^\sharp)$.

4.1 A derivation of the Volterra integral equation

In this section, we show that the estimator from the pseudo-estimation is characterized by the (forward) Volterra integral equation. This fact can be used for the proofs of Lemmas 1 and 2.

The difference $\hat{\Lambda}_0^\sharp(t) - \Lambda_0^\sharp(t)$ can be divided into the two terms as follows:

$$\hat{\Lambda}_0^\sharp(t) - \Lambda_0^\sharp(t) = A_n^\sharp(t) + \int_0^t \left(\mathrm{d}\mathbb{N}_n(s)\Big/\mathbb{S}_n^\sharp\left(s;\theta,\hat{\Lambda}_0^\sharp\right) - \mathrm{d}\mathbb{N}_n(s)\Big/\mathbb{S}_n^\sharp\left(s;\theta,\Lambda_0^\sharp\right)\right),$$

where

$$A_n^\sharp(t) = \int_0^t \mathrm{d}\mathbb{N}_n(s)\Big/\mathbb{S}_n^\sharp\left(s;\theta,\Lambda_0^\sharp\right) - \int_0^t \mathrm{d}\mathbb{N}(s)\Big/\mathbb{S}^\sharp\left(s;\theta,\Lambda_0^\sharp\right).$$

Then, a first-order Taylor expansion of $\hat{w}^\sharp(s)$ in $\hat{\Lambda}_0^\sharp(s)$ around $\Lambda_0^\sharp(s)$ gives $\hat{w}^\sharp(s) = w^\sharp(s) - r(\beta)\tilde{w}^\sharp(s)(1-\tilde{w}^\sharp(s))(\hat{\Lambda}_0^\sharp(s) - \Lambda_0^\sharp(s))$, where $\tilde{w}^\sharp(s) = w(s;\theta,\tilde{\Lambda}_0^\sharp)$ and $\tilde{\Lambda}_0^\sharp(s)$ is on the line segment between $\hat{\Lambda}_0^\sharp(s)$ and $\Lambda_0^\sharp(s)$ for all $s \in [0,\tau_e]$. So, let $\tilde{B}_n^\sharp(s) = \mathrm{E}_n[r_j(\beta)^2\tilde{w}_j^\sharp(s)(1-\tilde{w}_j^\sharp(s))Y_j(s)]$, then we have

$$\begin{aligned}&1\Big/\mathbb{S}_n^\sharp\left(s;\theta,\hat{\Lambda}_0^\sharp\right) - 1\Big/\mathbb{S}_n^\sharp\left(s;\theta,\Lambda_0^\sharp\right)\\ &\quad= \tilde{B}_n^\sharp(s)\left(\hat{\Lambda}_0^\sharp(s) - \Lambda_0^\sharp(s)\right)\Big/\mathbb{S}_n^\sharp\left(s;\theta,\hat{\Lambda}_0^\sharp\right)\mathbb{S}_n^\sharp\left(s;\theta,\Lambda_0^\sharp\right).\end{aligned}$$

Using this, we have a (forward) Volterra integral equation

$$\hat{\Lambda}_0^\sharp(t) - \Lambda_0^\sharp(t) = A_n^\sharp(t) + \int_0^t \left(\hat{\Lambda}_0^\sharp(s) - \Lambda_0^\sharp(s)\right)\tilde{B}_n^\sharp(s)\mathrm{d}\hat{C}_n^\sharp(s), \tag{8}$$

where $\mathrm{d}\hat{C}_n^\sharp(s) = \mathrm{d}\bar{\Lambda}_0^\sharp(s)/\mathbb{S}_n^\sharp(s;\theta,\hat{\Lambda}_0^\sharp)$ and $\mathrm{d}\bar{\Lambda}_0^\sharp(s) = \mathrm{d}\mathbb{N}_n(s)/\mathbb{S}_n^\sharp(s;\theta,\Lambda_0^\sharp)$.

4.2 Consistency : proof of Lemma 1

The Volterra integral equation of (8) can be uniquely solved by Theorem A.1. For $\Gamma^\sharp(t,s,\xi)$ defined with the kernel $K^\sharp(t,s)\mathrm{d}s = \tilde{B}_n^\sharp(s)\mathrm{d}\hat{C}_n^\sharp(s)$ in Theorem A.1, let $\Gamma_n^\sharp(t,s) = \Gamma^\sharp(t,s,1)$ and $K_n^\sharp(t,s) = K^\sharp(t,s)$. Since we can apply Proposition A.1 in a discrete Volterra integral equation, we have

$$\Gamma_n^\sharp(t,s)\mathrm{d}s = \frac{\tilde{B}_n^\sharp(s)\mathrm{d}\hat{C}_n^\sharp(s)}{\pi_{u\in[s,t]}\left(1 - \tilde{B}_n^\sharp(u)\mathrm{d}\hat{C}_n^\sharp(u)\right)},$$

where π denotes the product integral. Then, the solution of (8) is

$$\hat{\Lambda}_0^\sharp(t) - \Lambda_0^\sharp(t) = A_n^\sharp(t) + \int_0^t A_n^\sharp(s)\Gamma_n^\sharp(t,s)\mathrm{d}s. \tag{9}$$

The discussion of the limit of $A_n^\sharp(\cdot)$ is the same as that of the Cox model with time-dependent covariates. That is, by Conditions ($\mathrm{I}^\sharp$) and ($\mathrm{II}^\sharp$), we have $\sup_{\theta\in\Theta,t\in[0,\tau_e]}|A_n^\sharp(t)| = \sup_{\theta\in\Theta,t\in[0,\tau_e]}|\bar{\Lambda}_0^\sharp(t) - \Lambda_0^\sharp(t)| \to_{a.s.} 0$.

We easily show $\Lambda_0^\sharp(\tau_e) < \infty$ by Condition (I$^\sharp$), so that we have almost surely $\mathbb{S}_n^\sharp(\tau_e; \theta, \Lambda_0^\sharp) > 0$ on Θ from Condition (II$^\sharp$). This gives almost surely $\mathrm{E}_n[Y_j(\tau_e)r_j(\beta)] > 0$, so that $\mathbb{S}_n^\sharp(t; \theta, \hat{\Lambda}_0^\sharp) > 0$ is almost surely held under $\hat{\Lambda}_0^\sharp(t) < \infty$. Using the mathematical induction to the estimation process of $\hat{\Lambda}_0^\sharp(\cdot)$, we have almost surely $\hat{\Lambda}_0^\sharp(\tau_e) < \infty$ on Θ.

We finally discuss the boundedness of kernel $K_n^\sharp(\cdot)$. $\tilde{B}_n^\sharp(\cdot)$ is almost surely bounded on $[0, \tau_e] \times \Theta$ by Condition (0). $\mathbb{S}_n^\sharp(s; \theta, \hat{\Lambda}_0^\sharp)$ is almost surely bounded away from zero on $[0, \tau_e] \times \Theta$ under $\hat{\Lambda}_0^\sharp(\tau_e) < \infty$. $\mathrm{d}\bar{\Lambda}_0^\sharp(s)/\mathrm{d}s$ is almost surely bounded on $[0, \tau_e]$ in each $\theta \in \Theta$ by the Lipschitz property of $\Lambda_0^\sharp(\cdot)$, where the notation $\mathrm{d}\bar{\Lambda}_0^\sharp(s)/\mathrm{d}s$ includes $\{\bar{\Lambda}_0^\sharp(T_{(i)}) - \bar{\Lambda}_0^\sharp(T_{(i-1)})\}/(T_{(i)} - T_{(i-1)})$ under a finite n $(i = 1, \ldots, k)$. Therefore, $K_n^\sharp(\cdot)$ is almost surely bounded on $[0, \tau_e] \times \Theta$, so that $\Gamma_n^\sharp(\cdot, \cdot)$ is almost surely bounded on $[0, \tau_e]^2 \times \Theta$ by applying Proposition A.2. Consequently, since we have almost surely $|\hat{\Lambda}_0^\sharp(t) - \Lambda_0^\sharp(t)| \leq M \sup_t |A_n^\sharp(t)|$ for some $M < \infty$ from (9), it follows that $\sup_{\theta\in\Theta, t\in[0,\tau_e]} |\hat{\Lambda}_0^\sharp(t) - \Lambda_0^\sharp(t)| \to_{a.s.} 0$. □

4.3 Asymptotic normality: proof of Lemma 2

Note that $\Lambda_0^*(t) = \Lambda_0^\sharp(t; \theta^*)$, so that we write $w^*(t) = w(t; \theta^*, \Lambda_0^\sharp)$, $\mathbb{S}_n^{*\sharp}(t) = \mathbb{S}_n^\sharp(t; \theta^*, \Lambda_0^\sharp)$ and $\mathbb{S}^{*\sharp}(t) = \mathbb{S}^\sharp(t; \theta^*, \Lambda_0^\sharp)$. Also, to specify the case of $\theta = \theta^*$ considered here, we write $\hat{\Lambda}_0^\sharp(t; \theta^*) = \hat{\Lambda}_0^\sharp(t)|_{\theta=\theta^*}$, $\bar{\Lambda}_0^\sharp(t; \theta^*) = \bar{\Lambda}_0^\sharp(t)|_{\theta=\theta^*}$, and $\tilde{w}^\sharp(s; \theta^*) = \tilde{w}^\sharp(s)|_{\theta=\theta^*}$ without dropping θ. Also, let Condition (II$_m^\sharp$): $\sup_{t\in[0,\tau_e]} |\mathbb{S}_n^{*\sharp}(t) - \mathbb{S}^{*\sharp}(t)| \to_p 0$, then (III$^\sharp$) includes (II$_m^\sharp$). So the results obtained in Sect. 4.2 are held in a version of convergence in probability.

From (9) with $\theta = \theta^*$, we have

$$\sqrt{n}\left(\hat{\Lambda}_0^\sharp(t; \theta^*) - \Lambda_0^*(t)\right) = \sqrt{n}A_n^{*\sharp}(t) + \int_0^t \sqrt{n}A_n^{*\sharp}(s)\Gamma_n^{*\sharp}(t, s)\mathrm{d}s, \qquad (10)$$

where $A_n^{*\sharp}(\cdot) = A_n^\sharp(\cdot)|_{\theta=\theta^*}$ and $\Gamma_n^{*\sharp}(\cdot, \cdot) = \Gamma_n^\sharp(\cdot, \cdot)|_{\theta=\theta^*}$. For a proof to converge to a Gaussian process, we discuss that $\sqrt{n}A_n^{*\sharp}(\cdot)$ converges weakly to a Gaussian process and $\Gamma_n^\sharp(\cdot, \cdot)$ converges in probability to a deterministic function.

The discussion on the asymptotic normality of $\sqrt{n}A_n^{*\sharp}(\cdot)$ is similar to that of the Cox model with time-dependent covariates. From Conditions (I$^\sharp$) and (III$^\sharp$) and a delta method or martingale approach, we have $\sqrt{n}A_n^{*\sharp}(\cdot) \to_D \mathbb{G}_A^\sharp(\cdot)$, where $\mathbb{G}_A^\sharp(\cdot)$ is a zero-mean Gaussian martingale process.

Let $\tilde{B}_n^{*\sharp}(\cdot) = \tilde{B}_n^\sharp(\cdot)|_{\theta=\theta^*}$, $\hat{C}_n^{*\sharp}(\cdot) = \hat{C}_n^\sharp(\cdot)|_{\theta=\theta^*}$, and $K_n^{*\sharp}(t, s)\mathrm{d}s = \tilde{B}_n^{*\sharp}(s)\mathrm{d}\hat{C}_n^{*\sharp}(s)$. Here, to prove the result on $\Gamma_n^{*\sharp}(\cdot, \cdot)$, we discuss $\sup_{t\in[0,\tau_e]} |\int_0^t K_n^\sharp(t, s)\mathrm{d}s - \int_0^t K^{*\sharp}(t, s)\mathrm{d}s| \to_p 0$, where $K^{*\sharp}(t, s)\mathrm{d}s = B^{*\sharp}(s)\mathrm{d}C^{*\sharp}(s)$, $B^{*\sharp}(s) = \mathrm{E}[r^{*2}w^*(s)(1 - w^*(s))Y(s)]$ and $\mathrm{d}C^{*\sharp}(s) = \mathrm{d}\Lambda_0^*(s)/\mathbb{S}^{*\sharp}(s)$.

We consider $|\tilde{B}_n^{*\sharp}(\cdot) - B^{*\sharp}(\cdot)| \leq b_1^\sharp(\cdot) + b_2^\sharp(\cdot)$ obtained via the triangle inequality, where

$$\begin{aligned}
b_1^\sharp(t) = &\Big|\mathrm{E}_n\left[r_j^{*2}\tilde{w}_j^\sharp(t;\theta^*)\left(1-\tilde{w}_j^\sharp(t;\theta^*)\right)Y_j(t)\right] \\
&- \mathrm{E}_n\left[r_j^{*2}w_j^*(t)\left(1-w_j^*(t)\right)Y_j(t)\right]\Big|, \\
b_2^\sharp(t) = &\left|\mathrm{E}_n\left[r_j^{*2}w_j^*(t)\left(1-w_j^*(t)\right)Y_j(t)\right] - \mathrm{E}\left[r^{*2}w^*(t)\left(1-w^*(t)\right)Y(t)\right]\right|.
\end{aligned}$$

A Taylor expansion provides $|\tilde{w}^\sharp(t;\theta^*)^m - w^*(t)^m| \le r^*\check{w}(t;\theta^*)|\tilde{\Lambda}_0^\sharp(t;\theta^*) - \Lambda_0^*(t)|(m=1,2)$, where $\check{w}(t;\theta^*) = \tilde{w}^\sharp(t;\theta^*)|_{\tilde{\Lambda}_0^\sharp=\check{\Lambda}_0}$ and $\min(\tilde{\Lambda}_0^\sharp(t;\theta^*), \Lambda_0^*(t)) < \check{\Lambda}_0(t) < \max(\tilde{\Lambda}_0^\sharp(t;\theta), \Lambda_0^*(t))$. We have $r^*\check{w}(t;\theta^*) < \check{\Lambda}_0(t)^{-1}$ using $x\mathrm{e}^{-xy} \le \mathrm{e}^{-1}y^{-1}(x,y \ge 0)$. For some η_n such that $\sup_t|\check{\Lambda}_0(t;\theta^*) - \Lambda_0^*(t)| \le \eta_n^2$, let $t_{\eta_n} = \inf\{t : \check{\Lambda}_0(t) \ge \eta_n\}$, then we have $b_1^\sharp(t) \le \eta_n\mathrm{E}_n[r_j^{*2}]$ in $t \in [t_{\eta_n}, \tau_e]$. Hence, $\sup_t b_1^\sharp(t) \to_p 0$ is shown by Condition (0) and $\eta_n \to_p 0$ and $t_{\eta_n} \to_p 0$ obtained from ($\mathrm{II}_m^\sharp$). Next, in the discussion on $b_2^\sharp$, we use Condition ($\mathrm{III}^\sharp$). From the existence of covariance function of the Gaussian process in ($\mathrm{III}^\sharp$), we have uniformly $\mathrm{E}_n[r_j^{*2}w_j^*(\cdot)^2Y_j(\cdot)] \to_p \mathrm{E}[r^{*2}w^*(\cdot)^2Y(\cdot)]$ and $\mathrm{E}_n[r_j^{*2}w_j^*(\cdot)(1-c_j^*)Y_j(\cdot)] \to_p \mathrm{E}[r^{*2}w^*(\cdot)(1-c_j^*)Y(\cdot)]$. Also, the latter result and the Glivenko–Cantelli property provide uniformly $\mathrm{E}_n[r_j^{*2}w_j^*(\cdot)Y_j(\cdot)] \to_p \mathrm{E}[r^{*2}w^*(\cdot)Y(\cdot)]$. Hence, $\sup_t b_2^\sharp(t) \to_p 0$ is obtained from these results, so that $\sup_{t\in[0,\tau_e]}|\tilde{B}_n^{*\sharp}(t) - B^{*\sharp}(t)| \to_p 0$ is shown. Also, we have uniformly $\mathbb{S}_n^\sharp(\cdot;\theta^*,\hat{\Lambda}_0^\sharp) \to_p \mathbb{S}_n^\sharp(\cdot;\theta^*,\Lambda_0^\sharp)$ similarly to the discussion of $\sup b_1^\sharp \to_p 0$ and obtain $\sup_t|\bar{\Lambda}_0^\sharp(t;\theta^*) - \Lambda_0^*(t)| \to_p 0$ and $\Lambda_0^*(\tau_e) < \infty$ from results in Sect. 4.2 with ($\mathrm{II}_m^\sharp$). Hence $\sup_t|\hat{C}_n^{*\sharp}(t) - C^{*\sharp}(t)| \to_p 0$ is shown by these results, Conditions ($\mathrm{I}^\sharp$) and ($\mathrm{II}_m^\sharp$) and the continuous mapping property.

From the above discussion, we conclude $\sup_t|\int_0^t K_n^\sharp(t,s)\mathrm{d}s - \int_0^t K^{*\sharp}(t,s)\mathrm{d}s| \to_p 0$ via the triangle inequality. Therefore, by applying Proposition A.4, we have $\sup_{t,s\in[0,\tau_e]}|\Gamma_n^{*\sharp}(t,s) - \Gamma^{*\sharp}(t,s)| \to_p 0$, where $\Gamma^{*\sharp}(\cdot,\cdot)$ is a bounded deterministic function such that

$$\Gamma^{*\sharp}(t,s) = \frac{B^{*\sharp}(s)\mathrm{d}C^{*\sharp}(s)}{\prod_{x\in[s,t]}\left(1 - B^{*\sharp}(x)\mathrm{d}C^{*\sharp}(x)\right)}.$$

Therefore, from Slutsky's lemma and asymptotically tightness of $\sqrt{n}A_n^{*\sharp}(\cdot)\Gamma_n^{*\sharp}(\cdot,\cdot)$, we show that $\sqrt{n}A_n^{*\sharp}(s_1)\Gamma_n^{*\sharp}(t,s_1),\ldots,\sqrt{n}A_n^{*\sharp}(s_m)\Gamma_n^{*\sharp}(t,s_m)$ converge weakly to zero-mean normal distributions $\mathbb{G}_A^\sharp(s_1)\Gamma^{*\sharp}(t,s_1),\ldots,\mathbb{G}_A^\sharp(s_m)\Gamma^{*\sharp}(t,s_m)$ for every finite set $s_1,\ldots,s_m \in [0,t](0 \le t \le \tau_e)$. Since a tight, Borel measurable Gaussian process is transformed to some Gaussian process by every continuous linear map into a Banach space, $\sqrt{n}(\hat{\Lambda}_0^\sharp(\cdot;\theta^*) - \Lambda_0^*(\cdot))$ converges weakly to a zero-mean Gaussian process $\mathbb{G}_A^\sharp(\cdot) + \int_0^\cdot \mathbb{G}_A^\sharp(s)\Gamma^{*\sharp}(\cdot,s)\mathrm{d}s$. □

4.4 Covariance function

Here, we discuss the derivation of the covariance function and martingale property in Lemma 2.

The asymptotic covariance function of $\sqrt{n}(\hat{\Lambda}_0^{\sharp}(\cdot;\theta^*) - \Lambda_0^*(\cdot))$ is conveniently derived by using a martingale calculus. This is one of the advantages of the pseudo-estimation. From Sect. 2.2, the first step is to find that $A_n^{*\sharp}(t)$ is an $\mathscr{F}_t$-martingale such that

$$A_n^{*\sharp}(t) = \int_0^t \mathrm{d}\mathbb{N}_n(s) \big/ \mathbb{S}_n^{*\sharp}(s) - \int_0^t \mathrm{E}_n\left[Y_i(s) r_i^* w_i^*(s) \mathrm{d}\Lambda_0^*(s)\right] \big/ \mathbb{S}_n^{*\sharp}(s)$$
$$= \int_0^t \mathrm{d}\bar{M}(s) \big/ n\mathbb{S}_n^{*\sharp}(s),$$

where $\bar{M}(s) = \sum_{i=1}^n M_i(s)$ and $\mathrm{d}M_i(s) = \mathrm{d}N_i(s) - Y_i(s) r_i^* w_i^*(s) \mathrm{d}\Lambda_0^*(s)$.

Exchanging the order of integration of $\int_0^t \sqrt{n} A_n^{*\sharp}(s) \Gamma_n^{*\sharp}(t,s)\mathrm{d}s$, we can write

$$\int_0^t \sqrt{n} A_n^{*\sharp}(s) \Gamma_n^{*\sharp}(t,s)\mathrm{d}s = \int_0^t \left[\int_0^s \mathrm{d}\bar{M}(u) \big/ \sqrt{n}\mathbb{S}_n^{*\sharp}(u)\right] \Gamma_n^{*\sharp}(t,s)\mathrm{d}s$$
$$= \int_0^t \Phi^{\sharp}(s,t) \mathrm{d}\bar{M}(s) \big/ \sqrt{n}\mathbb{S}_n^{*\sharp}(s)$$
$$+ \int_0^t \left[\int_s^t \Gamma_n^{*\sharp}(t,u)\mathrm{d}u - \Phi^{\sharp}(s,t)\right] \mathrm{d}\bar{M}(s) / \sqrt{n}\mathbb{S}_n^{*\sharp}(s), \tag{11}$$

where $\Phi^{\sharp}(s,t) (s \le t)$ denotes the deterministic function (predictable process) such that

$$\Phi^{\sharp}(s,t) = \int_s^t \Gamma^{*\sharp}(t,u)\mathrm{d}u, \qquad \sup_{0 \le s \le t \le \tau_e} \left| \Phi^{\sharp}(s,t) - \int_s^t \Gamma_n^{*\sharp}(t,u)\mathrm{d}u \right| \to_p 0.$$

Also, when $B^{*\sharp}$ and $\mathrm{d}C^{*\sharp}$ are continuous, we have

$$\Phi^{\sharp}(s,t) = \frac{\int_s^t B^{*\sharp}(u)\mathrm{d}C^{*\sharp}(u)}{\prod_{[u,t]} \left(1 - B^{*\sharp}(x)\mathrm{d}C^{*\sharp}(x)\right)}$$
$$= -1 + \exp\left\{\int_s^t B^{*\sharp}(x)\mathrm{d}C^{*\sharp}(x)\right\}.$$

As for the second term of (11), we have

$$|\text{the second term of (11)}|$$
$$\le \sup_{s',t'} \left| \int_{s'}^{t'} \Gamma^{*\sharp}(t',u)\mathrm{d}u - \Phi^{\sharp}(s',t') \right| \left| \int_0^t \mathrm{d}\bar{M}(s) \big/ \sqrt{n}\mathbb{S}_n^{*\sharp}(s) \right|.$$

Applying Lenglart's inequality to the right-hand side of this inequality, we have

$$\Pr\left(\sup_{t\in[0,\tau_e]}\left|\int_0^t \mathrm{d}\bar{M}(s)\Big/\sqrt{n}\mathbb{S}_n^{*\sharp}(s)\right| > \eta\right)$$
$$\leq \delta/\eta^2 + \Pr\left(\int_0^t \lambda_0^*(s)\mathrm{d}s\Big/\mathbb{S}_n^{*\sharp}(s) > \delta\right).$$

Then, the probability of the left-hand side tends to zero as $n \to \infty$ by letting $\delta < \eta \to \infty$ for sufficiently large δ. Hence, since $\left|\int_0^t\left[\int_s^t \Gamma_n^{*\sharp}(u)\mathrm{d}u - \Phi^\sharp(s,t)\right]\mathrm{d}\bar{M}(s)/\sqrt{n}\mathbb{S}_n^{*\sharp}(s)\right| = o_P(1)$, we have

$$\sqrt{n}\left(\hat{\Lambda}_0^\sharp(t;\theta^*) - \Lambda_0^*(t)\right) = \int_0^t (\Phi^\sharp(s,t)+1)\mathrm{d}\bar{M}(s)\Big/\sqrt{n}\mathbb{S}_n^{*\sharp}(s) + o_P(1)\ .$$

Therefore, we conclude that the limit process of $\sqrt{n}(\hat{\Lambda}_0^\sharp(t;\theta^*) - \Lambda_0^\sharp(t;\theta^*))$ is the Gaussian martingale process $\mathbb{G}_\Lambda^\sharp(t) = \mathbb{G}_A^\sharp(t) + \int_0^t \mathbb{G}_A^\sharp(s)\Gamma^{*\sharp}(t,s)\mathrm{d}s$ and has the covariance function

$$\mathrm{Cov}\left(\mathbb{G}_\Lambda^\sharp(t), \mathbb{G}_\Lambda^\sharp(u)\right) = \int_0^{t\wedge u} \exp\left\{2\int_s^{t\wedge u} B^{*\sharp}(x)\mathrm{d}\Lambda_0^*(x)\Big/\mathbb{S}^{*\sharp}(x)\right\}\mathrm{d}\Lambda_0^*(s)\Big/\mathbb{S}^{*\sharp}(s)\ . \tag{12}$$

5 The asymptotic behavior of $\hat{\Lambda}_0^\flat$ from the EM-estimation

For simplicity, for some $\theta \in \Theta$, we write $\Lambda_0^\flat(t) = \Lambda_0^\flat(t;\theta)$ and $\hat{\Lambda}_0^\flat(t) = \hat{\Lambda}_0^\flat(t;\theta)$ by dropping θ. Also, we write $\hat{w}^\flat(t) = w(t;\theta,\hat{\Lambda}_0^\flat)$ and $w^\flat(t) = w(t;\theta,\Lambda_0^\flat)$.

5.1 A derivation of the Fredholm integral equation

In this section, we show that the estimator from the EM-estimation is characterized by the Fredholm integral equation. This fact can be used for the proofs of Lemmas 3 and 4. Also, we provide that the estimator can be described by a backward Volterra integral equation under a certain condition.

The difference $\hat{\Lambda}_0^\flat(t) - \Lambda_0^\flat(t)$ can be divided into the two terms as follows:

$$\hat{\Lambda}_0^\flat(t) - \Lambda_0^\flat(t) = A_n^\flat(t) + \int_0^t \left(\mathrm{d}\mathbb{N}_n(s)\Big/\mathbb{S}_n^\flat\left(s;\theta,\hat{\Lambda}_0^\flat\right) - \mathrm{d}\mathbb{N}_n(s)\Big/\mathbb{S}_n^\flat\left(s;\theta,\Lambda_0^\flat\right)\right),$$

where

$$A_n^\flat(t) = \int_0^t \mathrm{d}\mathbb{N}_n(s) \Big/ \mathbb{S}_n^\flat\left(s;\theta,\Lambda_0^\flat\right) - \int_0^t \mathrm{d}\mathbb{N}(s) \Big/ \mathbb{S}^\flat\left(s;\theta,\Lambda_0^\flat\right).$$

A first-order Taylor expansion of $\hat{w}^\flat(s)$ in $\hat{\Lambda}_0^\flat(s)$ around $\Lambda_0^\flat(s)$ gives $\hat{w}^\flat(s) = w^\flat(s) - r(\beta)\tilde{w}^\flat(s)(1-\tilde{w}^\flat(s))(\hat{\Lambda}_0^\flat(s) - \Lambda_0^\flat(s))$, where $\tilde{w}^\flat(s) = w(s;\theta,\tilde{\Lambda}_0^\flat)$ and $\tilde{\Lambda}_0^\flat(s)$ is on the line segment between $\hat{\Lambda}_0^\flat(s)$ and $\Lambda_0^\flat(s)$ for all $s \in [0,\tau_e]$. So, let $\mathrm{d}\tilde{B}_n^\flat(s) = \mathrm{E}_n[r_j(\beta)^2 \tilde{w}_j^\flat(s)(1-\tilde{w}_j^\flat(s))\mathrm{d}N_j^C(s)]$, then we have

$$\begin{aligned}
&1/\mathbb{S}_n^\flat\left(s;\theta,\hat{\Lambda}_0^\flat\right) - 1/\mathbb{S}_n^\flat\left(s;\theta,\Lambda_0^\flat\right) \\
&= \mathrm{E}_n\left[\int_s^{\tau_e} r_j(\beta)\{w_j^\flat(u) - \hat{w}_j^\flat(u)\}\mathrm{d}N_j^C(u)\right] \Big/ \mathbb{S}_n^\flat\left(s;\theta,\hat{\Lambda}_0^\flat\right)\mathbb{S}_n^\flat\left(s;\theta,\Lambda_0^\flat\right) \\
&= \int_s^{\tau_e} (\hat{\Lambda}_0^\flat(u) - \Lambda_0^\flat(u))\mathrm{d}\tilde{B}_n^\flat(u) \Big/ \mathbb{S}_n^\flat\left(s;\theta,\hat{\Lambda}_0^\flat\right)\mathbb{S}_n^\flat\left(s;\theta,\Lambda_0^\flat\right).
\end{aligned}$$

Using this, we have

$$\hat{\Lambda}_0^\flat(t) - \Lambda_0^\flat(t) = A_n^\flat(t) + \int_0^t \left\{\int_s^{\tau_e} \left(\hat{\Lambda}_0^\flat(u) - \Lambda_0^\flat(u)\right)\mathrm{d}\tilde{B}_n^\flat(u)\right\}\mathrm{d}\hat{C}_n^\flat(s),$$

where $\mathrm{d}\hat{C}_n^\flat(s) = \mathrm{d}\bar{\Lambda}_0^\flat(s)/\mathbb{S}_n^\flat(s;\theta,\hat{\Lambda}_0^\flat)$ and $\mathrm{d}\bar{\Lambda}_0^\flat(s) = \mathrm{d}\mathbb{N}_n(s)/\mathbb{S}_n^\flat(s;\theta,\Lambda_0^\flat)$. Further, this can be rewritten by exchanging the order of integration as follows:

$$\begin{aligned}
\hat{\Lambda}_0^\flat(t) - \Lambda_0^\flat(t) &= A_n^\flat(t) \\
&\quad + \int_0^{\tau_e} I(s \le t)\mathrm{d}\hat{C}_n^\flat(s) \int_0^{\tau_e} I(s \le u)\left(\hat{\Lambda}_0^\flat(u) - \Lambda_0^\flat(u)\right)\mathrm{d}\tilde{B}_n^\flat(u) \\
&= A_n^\flat(t) + \int_0^{\tau_e} \left(\hat{\Lambda}_0^\flat(s) - \Lambda_0^\flat(s)\right)\hat{C}_n^\flat(s \wedge t)\mathrm{d}\tilde{B}_n^\flat(s). \qquad (13)
\end{aligned}$$

This is a Fredholm integral equation.

We investigate a viewpoint as a Volterra integral equation of (13). Providing $t = 0$ in (13), we have

$$0 = \hat{\Lambda}_0^\flat(0) - \Lambda_0^\flat(0) = \int_0^{\tau_e} \left(\hat{\Lambda}_0^\flat(s) - \Lambda_0^\flat(s)\right)\left\{\int_0^s \mathrm{d}\hat{C}_n^\flat(u)\right\}\mathrm{d}\tilde{B}_n^\flat(s).$$

Using this, we transform (13) into

$$\hat{\Lambda}_0^\flat(t) - \Lambda_0^\flat(t) = A_n^-(t,\tau_e) - \int_{t_+}^{\tau_e} \left(\hat{\Lambda}_0^\flat(s) - \Lambda_0^\flat(s)\right)\left\{\int_{t_+}^s \mathrm{d}\hat{C}_n^\flat(u)\right\}\mathrm{d}\tilde{B}_n^\flat(s).$$

This is a backward Volterra integral equation if $\hat{\Lambda}_0^\flat(\tau_e) - \Lambda_0^\flat(\tau_e)$ is fixed, where $A_n^-(t, \tau_e) = \hat{\Lambda}_0^\flat(\tau_e) - \Lambda_0^\flat(\tau_e) - \{A_n^\flat(\tau_e) - A_n^\flat(t)\}$. However, since we cannot usually fix $\hat{\Lambda}_0^\flat(\tau_e) - \Lambda_0^\flat(\tau_e)$, a solution of (13) follows the manner of the Fredholm integral equation.

5.2 Consistency: proof of Lemma 3

The Fredholm integral equation of (13) can be uniquely solved under the Fredholm determinant $\neq 0$ by Theorem A.2. For $\Gamma^\flat(t, s, \xi)$, $D(\xi)$ and $D(t, s, \xi)$ defined in Theorem A.2 with $K^\flat(t, s)\mathrm{d}s = \mathrm{d}\tilde{B}_n^\flat(s)\hat{C}_n^\flat(s \wedge t)$ and $\xi = 1$, let $\Gamma_n^\flat(t, s) = \Gamma^\flat(t, s, 1)$, $D_n^\flat = D(1)$, $D_n^\flat(t, s) = D(t, s, 1)$ and $K_n^\flat(t, s) = K^\flat(t, s)$. The uniqueness condition provides $D_n^\flat \neq 0$. Then, the solution of (13) is

$$\hat{\Lambda}_0^\flat(t) - \Lambda_0^\flat(t) = A_n^\flat(t) + \int_0^{\tau_e} A_n^\flat(s)\Gamma_n^\flat(t, s)\mathrm{d}s. \tag{14}$$

The limit of $A_n^\flat(\cdot)$ can be discussed within the framework of the standard empirical processes. By results on stochastic integrals and Conditions (I$^\flat$) and (II$^\flat$), we have $\sup_{\theta\in\Theta, t\in[0,\tau_e]} |A_n^\flat(t)| = \sup_{\theta\in\Theta, t\in[0,\tau_e]} |\bar{\Lambda}_0^\flat(t) - \Lambda_0^\flat(t)| \to_{a.s.} 0$.

We easily have $\Lambda_0^\flat(\tau_e) < \infty$ by Condition (I$^\flat$), so that we have almost surely $\mathbb{S}_n^\flat(\tau_e; \theta, \Lambda_0^\flat) > 0$ on Θ from Condition (II$^\flat$). This gives almost surely $\mathrm{E}_n[Y_j(\tau_e)r_j(\beta)] > 0$, so that $\mathbb{S}_n^\flat(t; \theta, \hat{\Lambda}_0^\flat) > 0$ is almost surely held under $\hat{\Lambda}_0^\flat(t) < \infty$. Using the reductive absurdity to the estimation process of $\hat{\Lambda}_0^\flat(\cdot)$, we have almost surely $\hat{\Lambda}_0^\flat(\tau_e) < \infty$ on Θ.

Here we finally discuss the boundedness of kernel $K_n^\flat(\cdot, \cdot)$. $\mathrm{d}\tilde{B}_n^\flat(s)/\mathrm{d}s$ is almost surely bounded on $[0, \tau_e]$ in each $\theta \in \Theta$ by Condition (0) and the Lipschitz property of $S^C(\cdot)$, where $\mathrm{d}\tilde{B}_n^\flat(s)/\mathrm{d}s$ includes $\mathrm{E}_n[r_j^2\tilde{w}_j^\flat(T_{(i)})(1-\tilde{w}_j^\flat(T_{(i)}))\{N_j^C(T_{(i+1)}) - N_j^C(T_{(i)})\}/(T_{(i+1)} - T_{(i)})]$ under a finite n ($i = 1, \ldots, k$) and $T_{(k+1)} = \tau_e + \varepsilon$ for some $\varepsilon > 0$. $\mathbb{S}^\flat\left(s; \theta, \hat{\Lambda}_0^\flat\right)$ is almost surely bounded away from zero on $[0, \tau_e] \times \Theta$ by Condition (I$^\flat$) under $\hat{\Lambda}_0^\flat(\tau_e) < \infty$. $\bar{\Lambda}_0^\flat(\cdot)$ is almost surely bounded on $[0, \tau_e] \times \Theta$ by $\bar{\Lambda}_0^\flat(\tau_e) \to_{a.s.} \Lambda_0^\flat(\tau_e)$. Therefore, $K_n^\flat(\cdot, \cdot)$ is almost surely bounded on $[0, \tau_e]^2 \times \Theta$, so that $\Gamma_n^\flat(\cdot, \cdot)$ is almost surely bounded by applying Proposition A.3 and $D_n^\flat \neq 0$. Consequently, since we have almost surely $|\hat{\Lambda}_0^\flat(t) - \Lambda_0^\flat(t)| \leq M \sup_t |A_n^\flat(t)|$ for some $M < \infty$ from (14), it follows that $\sup_{\theta\in\Theta, t\in[0,\tau_e]} |\hat{\Lambda}_0^\flat(t) - \Lambda_0^\flat(t)| \to_{a.s.} 0$. □

5.3 Asymptotic normality: proof of Lemma 4

Note that $\Lambda_0^*(t) = \Lambda_0^\flat(t; \theta^*)$, so that we write $w^*(t) = w(t; \theta^*, \Lambda_0^\flat)$, $\mathbb{S}_n^{*\flat}(t) = \mathbb{S}_n^\flat(t; \theta^*, \Lambda_0^\flat)$ and $\mathbb{S}^{*\flat}(t) = \mathbb{S}^\flat(t; \theta^*, \Lambda_0^\flat)$. To specify the case of $\theta = \theta^*$ considered here, we write $\hat{\Lambda}_0^\flat(t; \theta^*) = \hat{\Lambda}_0^\flat(t)|_{\theta=\theta^*}$, $\bar{\Lambda}_0^\flat(t; \theta^*) = \bar{\Lambda}_0^\flat(t)|_{\theta=\theta^*}$ and $\tilde{w}^\flat(s; \theta^*) =$

$\tilde{w}^{\flat}(s)|_{\theta=\theta^*}$ without dropping θ. Also, let Condition ($\mathrm{II}_m^{\flat}$): $\sup_{t\in[0,\tau_e]} |\mathbb{S}_n^{*\flat}(t)-\mathbb{S}^{*\flat}(t)| \to_p 0$, then ($\mathrm{III}^{\flat}$) includes ($\mathrm{II}_m^{\flat}$). So the results obtained in Sect. 5.2 are held in a version of convergence in probability.

From (14) with $\theta = \theta^*$, we have

$$\sqrt{n}\left(\hat{\Lambda}_0^{\flat}(t;\theta^*) - \Lambda_0^*(t)\right) = \sqrt{n}A_n^{*\flat}(t) + \int_0^{\tau_e} \sqrt{n}A_n^{*\flat}(s)\Gamma_n^{*\flat}(t,s)\mathrm{d}s, \tag{15}$$

where $A_n^{*\flat}(\cdot) = A_n^{\flat}(\cdot)|_{\theta=\theta^*}$ and $\Gamma_n^{*\flat}(\cdot,\cdot) = \Gamma_n^{\flat}(\cdot,\cdot)|_{\theta=\theta^*}$. For a proof to converge to a Gaussian process, we discuss that $\sqrt{n}A_n^{*\flat}(\cdot)$ converges weakly to a Gaussian process and $\Gamma_n^{\flat}(\cdot,\cdot)$ converges in probability to a deterministic function.

The discussion on the asymptotic normality of $\sqrt{n}A_n^{*\flat}(\cdot)$ is based on a delta method. Recall that $A_n^{*\flat}(t) = \int_0^t \mathrm{d}\mathbb{N}_n(s)\big/\mathbb{S}_n^{*\flat}(s) - \int_0^t \mathrm{d}\mathbb{N}(s)\big/\mathbb{S}^{*\flat}(s)$. Using a functional delta method (see van der Vaart and Wellner (1996, p. 384)) under Conditions ($\mathrm{I}^{\flat}$) and ($\mathrm{III}^{\flat}$), we have

$$\begin{aligned}\sqrt{n}A_n^{*\flat}(\cdot) &= \sqrt{n}\left(\bar{\Lambda}_0^{\flat}(\cdot;\theta^*) - \Lambda_0^*(\cdot)\right)\\ \to_D \mathbb{G}_A^{\flat}(\cdot) &= \int_0^{\cdot} \mathrm{d}\mathbb{G}_{\mathbb{N}}(s)\big/\mathbb{S}^{*\flat}(s) - \int_0^{\cdot} \mathbb{G}_{\mathbb{S}}^{\flat}(s)\mathrm{d}\mathbb{N}(s)\big/\mathbb{S}^{*\flat}(s)^2,\end{aligned}$$

where $\mathbb{G}_{\mathbb{N}}$ and $\mathbb{G}_{\mathbb{S}}^{\flat}$ are zero-mean Gaussian processes such that

$$\sqrt{n}\left(\mathbb{N}_n(\cdot) - \mathbb{N}(\cdot), \mathbb{S}_n^{*\flat}(\cdot) - \mathbb{S}^{*\flat}(\cdot)\right) \to_D \left(\mathbb{G}_{\mathbb{N}}(\cdot), \mathbb{G}_{\mathbb{S}}^{\flat}(\cdot)\right),$$

which are guaranteed by the Donsker theorem and ($\mathrm{III}^{\flat}$). Therefore, $\sqrt{n}A_n^{*\flat}(\cdot)$ converges weakly to a zero-mean Gaussian $\mathbb{G}_A^{\flat}(\cdot)$.

Let $\tilde{B}_n^{*\flat}(\cdot) = \tilde{B}_n^{\flat}(\cdot)|_{\theta=\theta^*}$, $\hat{C}_n^{*\flat}(\cdot) = \hat{C}_n^{\flat}(\cdot)|_{\theta=\theta^*}$ and $K_n^{*\flat}(t,s)\mathrm{d}s = \mathrm{d}\tilde{B}_n^{*\flat}(s)\hat{C}_n^{*\flat}(s\wedge t)$. Here, to prove the result on $\Gamma_n^{*\flat}(\cdot,\cdot)$, we discuss $\sup_{t\in[0,\tau_e]} | \int_0^t K_n^{\flat}(t,s)\mathrm{d}s - \int_0^t K^{*\flat}(t,s)\mathrm{d}s| \to_p 0$, where $K^{*\flat}(t,s)\mathrm{d}s = \mathrm{d}B^{*\flat}(s)C^{*\flat}(s\wedge t)$, $\mathrm{d}B^{*\flat}(s) = \mathrm{E}[r^{*2}w^*(s)(1-w^*(s))\mathrm{d}N^C(s)]$ and $C^{*\flat}(s\wedge t) = \int_0^{s\wedge t} \mathrm{d}\Lambda_0^*(u)/\mathbb{S}^{*\flat}(u)$.

We consider $| \int_t^{\tau_e} \mathrm{d}\tilde{B}_n^{*\flat}(s) - \int_t^{\tau_e} \mathrm{d}B^{*\flat}(s)| \le b_1^{\flat}(t) + b_2^{\flat}(t)$ obtained via the triangle inequality, where (recall that $\mathrm{E}_n[\int_t^{\tau_e} f_j(s)\mathrm{d}N_j^C(s)] = \mathrm{E}_n[f_j(T_j)(1-\Delta_j)Y_j(t)]$)

$$\begin{aligned}b_1^{\flat}(t) &= \Bigg|\mathrm{E}_n\left[\int_t^{\tau_e} r_j^{*2}\tilde{w}_j^{\flat}(s;\theta^*)\left(1-\tilde{w}_j^{\flat}(s;\theta^*)\right)\mathrm{d}N_j^C(s)\right]\\ &\quad - \mathrm{E}_n\left[\int_t^{\tau_e} r_j^{*2}w_j^*(s)\left(1-w_j^*(s)\right)\mathrm{d}N_j^C(s)\right]\Bigg|,\\ b_2^{\flat}(t) &= \big|\mathrm{E}_n\left[r_j^{*2}w_j^*(T_j)\left(1-w_j^*(T_j)\right)(1-\Delta_j)Y_j(t)\right]\\ &\quad - \mathrm{E}\left[r^{*2}w^*(T)(1-w^*(T))(1-\Delta)Y(t)\right]\big|.\end{aligned}$$

For $b_1^\flat$, we can use the similar discussion to that to evaluate $\sup_t b_1^\sharp(t) \to_p 0$ in Sect. 4.3, so that $\sup_t b_1^\flat(t) \to_p 0$ is shown by Conditions (0) and ($\mathrm{II}_m^\flat$). From the existence of covariance function of the Gaussian process in Condition ($\mathrm{III}^\flat$), we have uniformly $\mathrm{E}_n[r_j^{*2} w_j^*(T_j)^2(1-\Delta_j)Y_j(\cdot)] \to_p \mathrm{E}[r^{*2}w^*(T)^2(1-\Delta)Y(\cdot)]$. This result and the Glivenko–Cantelli property provide $\sup_t b_2^\flat(t) \to_p 0$. Hence, $\sup_{t\in[0,\tau_e]} |\int_t^{\tau_e} \mathrm{d}\tilde{B}_n^{*\flat}(s) - \int_t^{\tau_e} \mathrm{d}B^{*\flat}(s)| \to_p 0$ is shown. Also, we have uniformly $\mathbb{S}_n^\sharp(\cdot;\theta^*, \hat{\Lambda}_0^\flat) \to_p \mathbb{S}_n^\flat(\cdot;\theta^*, \Lambda_0^\flat)$ similarly to the discussion of $\sup b_1^\flat \to_p 0$ and obtain $\sup_t |\bar{\Lambda}_0^\flat(t;\theta^*) - \Lambda_0^*(t)| \to_p 0$ and $\Lambda_0^*(\tau_e) < \infty$ from results in Sect. 5.2 with ($\mathrm{II}_m^\flat$). Hence $\sup_t |\hat{C}_n^{*\flat}(t) - C^{*\flat}(t)| \to_p 0$ is shown by these results, Conditions ($\mathrm{I}^\flat$) and ($\mathrm{II}_m^\flat$) and the continuous mapping property.

From the above discussion, we conclude $\sup_t |\int_0^t K_n^\flat(t,s)\mathrm{d}s - \int_0^t K^{*\flat}(t,s)\mathrm{d}s| \to_p 0$ via the triangle inequality. Therefore, by applying Proposition A.4, we have $\sup_{t,s\in[0,\tau_e]} |\Gamma_n^{*\flat}(t,s) - \Gamma^{*\flat}(t,s)| \to_p 0$, where $\Gamma^{*\flat}(\cdot,\cdot)$ is a bounded deterministic function.

Therefore, by Slutsky's lemma and asymptotically tightness of $\sqrt{n}A_n^{*\flat}(\cdot)\Gamma_n^{*\flat}(\cdot,\cdot)$, we show that $\sqrt{n}A_n^{*\flat}(s_1)\Gamma_n^{*\flat}(t,s_1), \ldots, \sqrt{n}A_n^{*\flat}(s_m)\Gamma_n^{*\flat}(t,s_m)$ converge weakly to zero-mean normal distributions $\mathbb{G}_A^\flat(s_1)\Gamma^{*\flat}(t,s_1), \ldots, \mathbb{G}_A^\flat(s_m)\Gamma^{*\flat}(t,s_m)$ for every finite set $s_1, \ldots, s_m \in [0,\tau_e] (0 \le t \le \tau_e)$. Since a tight, Borel measurable Gaussian process is transformed to some Gaussian process by every continuous linear map into a Banach space, $\sqrt{n}(\hat{\Lambda}_0^\flat(\cdot;\theta^*) - \Lambda_0^*(\cdot))$ converges weakly to a zero-mean Gaussian process $\mathbb{G}_A^\flat(\cdot) + \int_0^{\tau_e} \mathbb{G}_A^\flat(s)\Gamma^{*\flat}(\cdot,s)\mathrm{d}s$. □

5.4 Covariance function

Here, we discuss the derivation of the covariance function in Lemma 4.

The derivation of the asymptotic covariance function of $\sqrt{n}(\hat{\Lambda}_0^\flat(\cdot;\theta^*) - \Lambda_0^*(\cdot))$ is not easy so as in case of $\hat{\Lambda}_0^\sharp(\cdot;\theta^*)$. One of the reasons is complexity of operator $\Gamma^{*\flat}(\cdot,\cdot)$, another is intricacy of covariance of $\sqrt{n}A_n^{*\flat}(\cdot)$. To describe the estimated covariance easily, we explain the covariance function using an infinite matrix.

Let $\hat{\mathbb{Z}}_i = \sqrt{n}(\hat{\Lambda}_0^\flat(T_{(i)};\theta^*) - \Lambda_0^*(T_{(i)}))$, $\hat{\mathbb{A}}_i = \sqrt{n}A_n^{*\flat}(T_{(i)})$, $\mathbb{B}_i = B^{*\flat}(T_{(i+1)}) - B^{*\flat}(T_{(i)})$, and $\mathbb{C}_i = C^{*\flat}(T_{(i)})$ $(i = 1, \ldots, k)$. A limit form of (13) is written as the discrete type of Fredholm integral equation $\hat{\mathbb{Z}}_i = \hat{\mathbb{A}}_i + \sum_{j=1}^k \hat{\mathbb{Z}}_j \mathbb{C}_{\min(i,j)}\mathbb{B}_j$. For $\hat{\mathbb{Z}}^k = (\hat{\mathbb{Z}}_1, \ldots, \hat{\mathbb{Z}}_k)'$ and $\hat{\mathbb{A}}^k = (\hat{\mathbb{A}}_1, \ldots, \hat{\mathbb{A}}_k)'$, this matrix representation is

$$\hat{\mathbb{Z}}^k = \hat{\mathbb{A}}^k + \begin{pmatrix} \mathbb{B}_1\mathbb{C}_1 & \mathbb{B}_2\mathbb{C}_1 & \mathbb{B}_3\mathbb{C}_1 & \cdots & \mathbb{B}_k\mathbb{C}_1 \\ \mathbb{B}_1\mathbb{C}_1 & \mathbb{B}_2\mathbb{C}_2 & \mathbb{B}_3\mathbb{C}_2 & \cdots & \mathbb{B}_k\mathbb{C}_2 \\ \mathbb{B}_1\mathbb{C}_1 & \mathbb{B}_2\mathbb{C}_2 & \mathbb{B}_3\mathbb{C}_3 & \cdots & \mathbb{B}_k\mathbb{C}_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbb{B}_1\mathbb{C}_1 & \mathbb{B}_2\mathbb{C}_2 & \mathbb{B}_3\mathbb{C}_3 & \cdots & \mathbb{B}_k\mathbb{C}_k \end{pmatrix} \hat{\mathbb{Z}}^k = \hat{\mathbb{A}}^k + \mathbb{K}_k\hat{\mathbb{Z}}^k.$$

Hence, this equation can be solved as $\hat{\mathbb{Z}}^k = (\mathrm{I}_k - \mathbb{K}_k)^{-1}\hat{\mathbb{A}}^k$, so that the covariance of $\hat{\mathbb{Z}}_i$ and $\hat{\mathbb{Z}}_j$ is

$$\mathrm{E}[\hat{\mathbb{Z}}_i\hat{\mathbb{Z}}_j] = (\mathrm{I}_k - \mathbb{K}_k)^{-1}_{(i)} \begin{pmatrix} \mathrm{E}[\hat{\mathbb{A}}_1^2] & \mathrm{E}[\hat{\mathbb{A}}_2\hat{\mathbb{A}}_1] & \cdots & \mathrm{E}[\hat{\mathbb{A}}_k\hat{\mathbb{A}}_1] \\ \vdots & & & \\ \mathrm{E}[\hat{\mathbb{A}}_1\hat{\mathbb{A}}_k] & \mathrm{E}[\hat{\mathbb{A}}_2\hat{\mathbb{A}}_k] & \cdots & \mathrm{E}[\hat{\mathbb{A}}_k^2] \end{pmatrix} \left\{(\mathrm{I}_k - \mathbb{K}_k)^{-1}_{(j)}\right\}',$$

where $(\mathrm{I}_k - \mathbb{K}_k)^{-1}_{(i)}$ is the i-th row vector of the matrix $(\mathrm{I}_k - \mathbb{K}_k)^{-1}$. Since $\hat{\mathbb{A}}_i = \sqrt{n}A_n^{*\flat}(t_i)$ converges weakly to a zero-mean normal distribution $\mathbb{G}_A^\flat(t_i)$, we have $\mathrm{E}[\hat{\mathbb{A}}_i\hat{\mathbb{A}}_j] \to_p \mathrm{E}[\mathbb{G}_A^\flat(t_i)\mathbb{G}_A^\flat(t_j)]$.

In the following sentence, we derive the concrete form of $\mathrm{E}[\mathbb{G}_A^\flat(t_i)\mathbb{G}_A^\flat(t_j)]$. We write

$$\mathbb{G}_A^\flat(t) = \int_0^t H(x)\mathrm{d}\mathbb{G}_a^\flat(x),$$

where

$$\mathbb{G}_a^\flat(x) = \mathbb{G}_\mathbb{N}(x) - \int_a^x \mathbb{G}_\mathbb{S}^\flat(y)\mathrm{d}\Lambda_0^*(y) \quad \text{and} \quad H(x) = 1/\mathbb{S}^{*\flat}(x).$$

Then, we have

$$\begin{aligned}\mathrm{E}\left[\mathbb{G}_A^\flat(s)\mathbb{G}_A^\flat(t)\right] &= H(s)H(t)\mathrm{E}\left[\mathbb{G}_a^\flat(s)\mathbb{G}_a^\flat(t)\right] \\ &\quad + \int_0^t\int_0^s \mathrm{E}\left[\mathbb{G}_a^\flat(x)\mathbb{G}_a^\flat(y)\right]\mathrm{d}H(x)\mathrm{d}H(y) \\ &\quad - H(t)\int_0^s \mathrm{E}\left[\mathbb{G}_a^\flat(t)\mathbb{G}_a^\flat(x)\right]\mathrm{d}H(x) \\ &\quad - H(s)\int_0^t \mathrm{E}\left[\mathbb{G}_a^\flat(s)\mathbb{G}_a^\flat(y)\right]\mathrm{d}H(y).\end{aligned}$$

Each term in the above equation includes $\mathrm{E}[\mathbb{G}_a^\flat(\cdot)\mathbb{G}_a^\flat(\cdot)]$, which is divided such that

$$\begin{aligned}\mathrm{E}\left[\mathbb{G}_a^\flat(s)\mathbb{G}_a^\flat(t)\right] &= \mathrm{E}\left[\mathbb{G}_\mathbb{N}(s)\mathbb{G}_\mathbb{N}(t)\right] + \int_0^t\int_0^s \mathrm{E}\left[\mathbb{G}_\mathbb{S}^\flat(x)\mathbb{G}_\mathbb{S}^\flat(y)\right]\mathrm{d}\Lambda_0^*(x)\mathrm{d}\Lambda_0^*(y) \\ &\quad - \int_0^s \mathrm{E}\left[\mathbb{G}_\mathbb{N}(t)\mathbb{G}_\mathbb{S}^\flat(x)\right]\mathrm{d}\Lambda_0^*(x) - \int_0^t \mathrm{E}\left[\mathbb{G}_\mathbb{N}(s)\mathbb{G}_\mathbb{S}^\flat(y)\right]\mathrm{d}\Lambda_0^*(y),\end{aligned}$$

whose elements include the following expectations

$$\begin{aligned}
\mathrm{E}\left[\mathbb{G}_{\mathbb{N}}(s)\mathbb{G}_{\mathbb{N}}(t)\right] &= \mu_{\mathbb{N}}(1, s \wedge t) - \mu_{\mathbb{N}}(1, s)\mu_{\mathbb{N}}(1, t) \\
\mathrm{E}\left[\mathbb{G}^{\flat}_{\mathbb{S}}(s)\mathbb{G}^{\flat}_{\mathbb{S}}(t)\right] &= \mu_{\mathbb{S}}(\mathrm{e}^{\beta^{*\prime}Z}, s \vee t) - \mu_{\mathbb{S}}(1, s)\mu_{\mathbb{S}}(1, t) - \mu_B(s \vee t) \\
\mathrm{E}\left[\mathbb{G}^{\flat}_{\mathbb{S}}(s)\mathbb{G}_{\mathbb{N}}(t)\right] &= -\mu_{\mathbb{N}}(\mathrm{e}^{\beta^{*\prime}Z}, s \wedge t) + \mu_{\mathbb{N}}(\mathrm{e}^{\beta^{*\prime}Z}, t) - \mu_{\mathbb{N}}(1, s)\mu_{\mathbb{S}}(1, t),
\end{aligned}$$

where $\mu_{\mathbb{N}}(g(Z), t) = \mathrm{E}[g(Z)\Delta(1-Y(t))]$, $\mu_B(t) = \mathrm{E}[r^{*2}w^*(T)(1-w^*(T))(1-\Delta)Y(t)]$, and $\mu_{\mathbb{S}}(g(Z), t) = \mathrm{E}[g(Z)r^*\{\Delta+(1-\Delta)w^*(T)\}Y(t)]$. Then, using these notations and the relation (19) between $\mu_{\mathbb{N}}$ and $\mu_{\mathbb{S}}$, we have

$$\begin{aligned}
\mathrm{E}\left[\mathbb{G}^{\flat}_{a}(s)\mathbb{G}^{\flat}_{a}(t)\right] = \mu_{\mathbb{N}}(s \wedge t) - \int_0^{s\wedge t} \mu_B(y)\Lambda_0^*(y)\mathrm{d}\Lambda_0^*(y) \\
- \int_0^{s\vee t} \mu_B(y)\Lambda_0^*(y)\mathrm{d}\Lambda_0^*(y).
\end{aligned}$$

To express $\mathrm{E}[\mathbb{G}^{\flat}_{A}(\cdot)\mathbb{G}^{\flat}_{A}(\cdot)]$ more simply, let $\mathrm{d}\mu(x) = \mathrm{d}\mu_{\mathbb{N}}(x) - \mu_B(x)\Lambda_0^*(x)\mathrm{d}\Lambda_0^*(x)$ and $\mathrm{d}\epsilon(x) = \mu_B(x)\Lambda_0^*(x)\mathrm{d}\Lambda_0^*(x)$, then $\mathrm{E}[\mathbb{G}^{\flat}_{a}(s)\mathbb{G}^{\flat}_{a}(t)] = \mu(s \wedge t) - \epsilon(s \vee t)$. Using this, we can write

$$\begin{aligned}
\mathrm{E}\left[\mathbb{G}^{\flat}_{A}(s)\mathbb{G}^{\flat}_{A}(t)\right] = &\int_0^{s\wedge t} H^2(x)\mathrm{d}\mu(x) - H(s)H(t)\epsilon(s \vee t) \\
&- \int_0^{s\wedge t} \epsilon(y)H(y)\mathrm{d}H(y) - \int_0^{s\vee t} \epsilon(y)H(y)\mathrm{d}H(y) \\
&+ H(t)\int_0^{s} \epsilon(t \vee y)\mathrm{d}H(y) + H(s)\int_0^{t} \epsilon(s \vee x)\mathrm{d}H(x).
\end{aligned}$$

Further, the other expression under $t \geq s$ is

$$\begin{aligned}
\mathrm{E}\left[\mathbb{G}^{\flat}_{A}(s)\mathbb{G}^{\flat}_{A}(t)\right] = &\int_0^{s} H^2(x)\mathrm{d}\mu(x) + H(s)H(t)\epsilon(t) - \frac{1}{2}\epsilon(s)H^2(s) - \frac{1}{2}\epsilon(t)H^2(t) \\
&+ \frac{1}{2}\int_0^{s} H^2(x)\mathrm{d}\epsilon(x) + \frac{1}{2}\int_0^{t} H^2(x)\mathrm{d}\epsilon(x) \\
&- H(s)\int_s^{t} H(x)\mathrm{d}\epsilon(x) - H(0)\{H(t)\epsilon(t) + H(s)\epsilon(s)\}.
\end{aligned}$$

Therefore, we conclude that the limit process of $\sqrt{n}(\hat{\Lambda}_0^{\flat}(t;\theta^*) - \Lambda_0^*(t))$ is the Gaussian process $\mathbb{G}_{\Lambda}^{\flat}(t) = \mathbb{G}_A^{\flat}(t) + \int_0^{\tau_e} \mathbb{G}_A^{\flat}(s)\Gamma^{*\flat}(t,s)\mathrm{d}s$ and has the covariance function for $t = t_i$ and $u = t_j$

$$\begin{aligned}
&\mathrm{Cov}\left(\mathbb{G}_{\Lambda}^{\flat}(t), \mathbb{G}_{\Lambda}^{\flat}(u)\right) \\
&= \lim_{n\to\infty} (\mathrm{I}_k - \mathbb{K}_k)^{-1}_{(i)} \begin{pmatrix} \mathrm{E}\left[\mathbb{G}_A^{\flat}(t_1)^2\right] & \mathrm{E}\left[\mathbb{G}_A^{\flat}(t_2)\mathbb{G}_A^{\flat}(t_1)\right] & \cdots & \mathrm{E}\left[\mathbb{G}_A^{\flat}(t_k)\mathbb{G}_A^{\flat}(t_1)\right] \\ & \vdots & & \\ \mathrm{E}\left[\mathbb{G}_A^{\flat}(t_1)\mathbb{G}_A^{\flat}(t_k)\right] & \mathrm{E}\left[\mathbb{G}_A^{\flat}(t_2)\mathbb{G}_A^{\flat}(t_k)\right] & \cdots & \mathrm{E}\left[\mathbb{G}_A^{\flat}(t_k)^2\right] \end{pmatrix} \\
&\times \left\{(\mathrm{I}_k - \mathbb{K}_k)^{-1}_{(j)}\right\}'. \qquad (16)
\end{aligned}$$

6 Simulation

We conduct the simulation study to examine how well a large sample theory for the two estimators of the baseline hazard derived in this paper works under some situations of finite-sample. For the simulation design, we use the same configuration as discussed by Kuk and Chen (1992) and Peng and Dear (2000), where the datasets of two groups with the same size of sample in each for $n = 100, 300, 500, 1{,}000$, and 2,000 are generated. We set $(\alpha_0^*, \alpha_1^*) = (-0.5, 0.5)$, which provides the cure rate of 38 and 50% in each group. We use the standard exponential distribution as the baseline survival distribution for uncured individuals and $\beta^* = 0.5\log(0.5)$. The censoring times are generated according to a uniform distribution on $[0, \tau_e]$ with $\tau_e = 4$. Under these settings, we generate 1,000 set of simulated data.

We provide the values of two biases (Bias_1,Bias_2) and MSE of $\hat{\Lambda}_0^{\sharp}(\cdot)$ and $\hat{\Lambda}_0^{\flat}(\cdot)$ from the simulated data in Table 1, where the Bias_1 denotes $\mathrm{E}[\sup_{t\in[0,\tau_e]}|\hat{\Lambda}_0^J(t;\theta^*) - \Lambda_0^*(t)|]$, the Bias_2 denotes $\sup_{t\in[0,\tau_e]}|\mathrm{E}[\hat{\Lambda}_0^J(t;\theta^*)] - \Lambda_0^*(t)|$ and the MSE denotes $\mathrm{E}[\int_0^{\tau_e}(\hat{\Lambda}_0^J(t;\theta^*) - \Lambda_0^*(t))^2\mathrm{d}t]$ ($J = \sharp, \flat$). The Bias_1 is the important measure for Lemmas 1 and 3 and the Bias_2 for the mean of the asymptotic distribution in Lemmas 2 and 4. In addition, we provide the values of the PCI_a which measures the number of the true $\Lambda_0^*(\cdot)$ falling into an a% confidence intervals for $\hat{\Lambda}_0^{\sharp}(\cdot;\theta^*)$ and $\hat{\Lambda}_0^{\flat}(\cdot;\theta^*)$. The intervals are constructed by substituting empirical estimates and estimated values corresponding to theoretical values in (12) and (16), based on Lemmas 2, and 4. For example, the substitutions of $\Lambda_0^*(\cdot)$, $B^{*J}(\cdot)$, and $\mathrm{E}[f(w^*(\cdot), N(\cdot), Y(\cdot), \Delta, Z, X)]$ are $\hat{\Lambda}_0^J(\cdot;\theta^*)$, $\hat{B}_n^{*J}(\cdot)$, and $\mathrm{E}_n[f(w_i(\cdot;\theta^*, \hat{\Lambda}_0^J)$, $N_i(\cdot), Y_i(\cdot), \Delta_i, Z_i, X_i)]$, respectively ($J = \sharp, \flat$). For simplicity, the intervals constructed here are the pointwise confidence intervals using the form only of variance functions, so that the PCI_a counts the average rate of $\Lambda_0^*(T_1), \ldots, \Lambda_0^*(T_n)$ falling into the intervals constructed marginally at each time.

The result of the biases shows that the values of Bias_1 for $\hat{\Lambda}_0^{\sharp}$ are larger than those for $\hat{\Lambda}_0^{\flat}$ for all cases of sample size n, while the values of Bias_2 for $\hat{\Lambda}_0^{\sharp}$ are smaller than those for $\hat{\Lambda}_0^{\flat}$ ignoring the sign of values and all values of Bias_2 for $\hat{\Lambda}_0^{\flat}$ are negative. The differences in the values of Bias_1 and Bias_2 between the two

Table 1 Two biases, MSE, and PCI_{90} of $\hat{\Lambda}_0^{\sharp}(\cdot)$ and $\hat{\Lambda}_0^{\flat}(\cdot)$

	Estimation	$n = 100$	$n = 300$	$n = 500$	$n = 1{,}000$	$n = 2{,}000$
Bias_1	Pseudo	1.3047	1.0746	0.9824	0.8175	0.6530
	EM	0.6945	0.6364	0.6113	0.6104	0.5428
Bias_2	Pseudo	−0.1067	−0.0118	0.1565	0.0880	0.0362
	EM	−0.6372	−0.4773	−0.3252	−0.2544	−0.1838
MSE	Pseudo	0.2075	0.0621	0.0396	0.0201	0.0095
	EM	0.0697	0.0319	0.0198	0.0126	0.0070
PCI_{90} (%)	Pseudo	92.24	91.93	91.80	91.19	90.43
	EM	89.57	–	–	–	–

estimators become smaller as n is larger. The result of MSE shows that the values for $\hat{\Lambda}_0^{\sharp}$ are larger than those for $\hat{\Lambda}_0^{\flat}$ and the differences between the two estimators become smaller as n is larger. The result of PCI_{90} shows that the values for the pseudo-estimation are sightly larger than the true value of 90% and become closely toward 90% as n is larger. However, in the EM-estimation, it is not possible to compute the inverse of higher dimensional matrix with larger n in the standard manner.

7 Concluding remarks

We discuss the two distinct estimators of the nonparametric baseline hazard function in the Cox cure model. We show that the estimator from the pseudo-estimation can be characterized by the (forward) Volterra integral equation, and the estimator from the EM-estimation by the redholm integral equation. Then, we establish, for given regression parameters, the strong consistency and asymptotic normality of the two estimators. Such characterizations by the integral equations reveal the background of existence of the two estimators. The results of simulation study not only ensure the asymptotic results but also draw the differences of behaviors between the two estimators in small and moderate samples.

The asymptotic results in this paper will provide the fundamental and important step to establish the large sample theory on the estimators $\hat{\theta}$ of θ in the Cox cure model. In fact, the nuisances $\hat{\Lambda}_0^{\sharp}$ and $\hat{\Lambda}_0^{\flat}$ are included in the criterions L_{pp} and $L_{\mathrm{p}}^{\mathrm{EM}}$, respectively, though this problem did not occur in the standard Cox model since the nuisance is perfectly eliminated in Cox's partial likelihood. If the asymptotic distributions of the estimators $\hat{\theta}$ are established, the asymptotic distributions of $\hat{\Lambda}_0^{\sharp}(\cdot;\hat{\theta})$ and $\hat{\Lambda}_0^{\flat}(\cdot;\hat{\theta})$ and these corresponding survival functions with $\hat{\theta}$ will be easily obtained by using a delta method and the results in this paper, such as

$$\sqrt{n}\left(\hat{\Lambda}_0^{\jmath}(t;\hat{\theta}) - \Lambda_0^{*}(t)\right) \approx \sqrt{n}\left(\hat{\theta} - \theta^{*}\right)\partial\hat{\Lambda}_0^{\jmath}(t;\theta)/\partial\theta$$
$$+\sqrt{n}\left(\hat{\Lambda}_0^{\jmath}(t;\theta^{*}) - \Lambda_0^{*}(t)\right) \quad (\jmath = \sharp, \flat).$$

Furthermore, the twin relationship between the two estimators of the baseline hazard may be useful as a common framework included in other emiparametric counting process model with incomplete data, such as a forward and backward system in a stochastic process.

Appendix A

A.1 The integral equations

Theorem A.1 *(Volterra integral equation). Let* ξ *be a given parameter* ($\xi < \infty$). *Let* $f(t)$ *and* $K^\sharp(t,s)$ *be given functions for* $t, s \in [a,b]$ *with* $\sup_t |f(t)| < \infty$ *and* $\sup_{t,s} |K^\sharp(t,s)| < \infty$ *(*$|a|, |b| < \infty$*). The Volterra equation of the form*

$$\phi(t) = f(t) + \xi \int_a^t K^\sharp(t,s)\phi(s)\mathrm{d}s \tag{17}$$

is solved uniquely as

$$\phi(t) = f(t) + \xi \int_a^t \Gamma^\sharp(t,s,\xi) f(s)\mathrm{d}s,$$

where

$$\Gamma^\sharp(t,s,\xi) = \sum_{m=0}^{\infty} \xi^m K_{m+1}(t,s), \quad K_m(t,s) = \int_s^t K^\sharp(t,u)K_{m-1}(u,s)\mathrm{d}u.$$

Comment There are many books on the Volterra integral equation. In settings of this theorem, K, f, and ϕ are defined as elements in space $L^\infty[a,b]$. Also, this theorem is held in discrete cases such that the integration becomes the summation, which is allowed as a Legesgue–Stieltjes's integration. In more larger space, for example $L^2[a,b]$, the same result is held. See Hochstadt (1973, p. 33), etc.

Proposition A.1 *(The solution for a discrete Volterra integral equation). A discrete Volterra integral equation* $\phi_i = f_i + \sum_{j=1}^{i} K_j^\sharp \phi_j \,(i = 1, \ldots, n)$ *uniquely gives the solution* $\phi_i = f_i + \sum_{j=1}^{i} f_j K_j^\sharp \Big/ \prod_{l=j}^{i} \left(1 - K_l^\sharp\right)$.

Proof This is easily proved from mathematical induction and Theorem A.1. □

For $\Gamma^\sharp$ defined in Theorem A.1, we draw a result found in a proof of Theorem A.1 for this paper.

Proposition A.2 *For* $\Gamma^\sharp$ *and* $K^\sharp$ *defined in Theorem A.1, if* $\sup_{t,s\in[a,b]} |K^\sharp(t,s)| < \infty$ *and* $|b-a| < \infty$, *then* $\sup_{t,s\in[a,b]} |\Gamma^\sharp(t,s,\xi)| < \infty$.

Proof Let $\|K^\sharp\|_\infty = \sup_{t,s\in[a,b]} |K^\sharp(t,s)|$. By the induction $|K_m(t,s)| \le (t-s)^m \|K^\sharp\|_\infty^m / m!$. So, we have $|\Gamma^\sharp(t,s,\xi)| \le \xi^{-1} \sum_{m=1}^{\infty} \xi^m |K_m(t,s)| \le \xi^{-1}[\exp\{\xi \|K^\sharp\|_\infty (t-s)\} - 1]$. □

Theorem A.2 *(Fredholm integral equation). Let* ξ *be a given parameter* ($\xi < \infty$). *Let* $f(t)$ *and* $K^\flat(t,s)$ *be given functions for* $t, s \in [a,b]$ *with* $\sup_t |f(t)| < \infty$ *and* $\sup_{t,s} |K^\flat(t,s)| < \infty$. *The Fredholm equation of the form*

$$\phi(t) = f(t) + \xi \int_a^b K^\flat(t,s)\phi(s)\mathrm{d}s \tag{18}$$

is solved uniquely as

$$\phi(t) = f(t) + \xi \int_a^b \Gamma^\flat(t, s, \xi) f(s) \mathrm{d}s$$

if $D(\xi) \neq 0$*, where* $\Gamma^\flat(t, s, \xi) = D(t, s, \xi)/D(\xi)$*, when* $K^\flat$ *is a continuous kernel or* f *and* $K^\flat$ *are finite-dimensional discrete functions,* $D(t, s, \xi)$ *and* $D(\xi)$ *are written as*

$$D(\xi) = \sum_{m=0}^{\infty} \xi^m D_m,\ D(t, s, \xi) = \sum_{m=0}^{\infty} \xi^m D_m(t, s),$$

$$D_m = \frac{(-1)^m}{m!} \int_a^b \cdots \int_a^b K\begin{pmatrix} x_1 & \cdots & x_m \\ x_1 & \cdots & x_m \end{pmatrix} \mathrm{d}x_1 \cdots \mathrm{d}x_m \ (D_0 = 1),$$

$$D_m(t, s) = \frac{(-1)^m}{m!} \int_a^b \cdots \int_a^b K\begin{pmatrix} t & x_1 & \cdots & x_m \\ s & x_1 & \cdots & x_m \end{pmatrix} \mathrm{d}x_1 \cdots \mathrm{d}x_m$$

$$(D_0(t, s) = K^\flat(t, s)), \quad \text{and}$$

$$K\begin{pmatrix} t_1 & t_2 & \cdots & t_m \\ s_1 & s_2 & \cdots & s_m \end{pmatrix} = \begin{vmatrix} K^\flat(t_1, s_1) & K^\flat(t_1, s_2) & \cdots & K^\flat(t_1, s_m) \\ K^\flat(t_2, s_1) & K^\flat(t_2, s_2) & \cdots & K^\flat(t_2, s_m) \\ \vdots & \ddots & \ddots & \vdots \\ K^\flat(t_m, s_1) & K^\flat(t_m, s_2) & \cdots & K^\flat(t_m, s_m) \end{vmatrix}.$$

Comment There are many books on the Fredholm integral equation as well as the Volttera. Here K, f, and ϕ are defined as elements in space $L^\infty[a, b]$ (Pipkin, 1991, p. 41-2). $D(\xi)$ and $D(t, s, \xi)$ are called the Fredholm determinat and the first Fredholm minor, respectively. Also, this theorem is allowed as the framework of a Legesgue-Stieltjes's integration. In more general $L^2[a, b]$, D_m, and $D_m(t, s)$ are slightly modified (see Hochstadt, 1973, p. 108, p. 241; Smithies, 1958). However, we do not have to use such modifications in this paper.

For $D(\xi)$ and $D(t, s, \xi)$ defined in Theorem A.2, we draw a result found in a proof of Theorem A.2 for this paper.

Proposition A.3 *For* $\Gamma^\flat$ *and* $K^\flat$ *defined in Theorem A.2, if* $\sup_{t,s\in[a,b]} |K^\flat(t, s)| < \infty$ *and* $|b - a| < \infty$*, then* $|D(\xi)| < \infty$ *and* $\sup_{t,s\in[a,b]} |D(t, s, \xi)| < \infty$.

Proof Let $\|K^\flat\|_\infty = \sup_{t,s} |K^\flat(t, s)| < \infty$. Then by Hadamard's inequality, $|D_m| \leq \|K^\flat\|_\infty^m (b-a)^m m^{1/2m}/m!$ and $|D_{m-1}(t, s)| \leq \|K^\flat\|_\infty^m (b-a)^m m^{1/2m}/(m-1)!$. For $c_m = \|K^\flat\|_\infty^m (b-a)^m m^{1/2m}/m!$ or $c_m = \|K^\flat\|_\infty^m (b-a)^m m^{1/2m}/(m-1)!$, $\sum_{m=0}^{\infty} c_m \xi^m$ converges to some finite quantity on $\mathbb{R}$ by $\lim_m c_{m+1}/c_m \to 0$. □

We give an original result for the Volterra and Fredholm integral equations with a kernel which converges in probability to a deterministic continuous function.

Proposition A.4 *Let $(K_n^\sharp(t,s), K^\sharp(t,s))$ and $(K_n^\flat(t,s), K^\flat(t,s))$ be pairs of kernels of the Volterra and Fredholm integral equations defined in $t, s \in [a,b]$, respectively. Suppose that $K^\jmath(t,s)$ ($\jmath = \sharp, \flat$) is continuous for all $s, t \in [a,b]$ and $|b-a| < \infty$. If* $\sup_t |\int_a^t \{K_n^\jmath(t,s) - K^\jmath(t,s)\}\mathrm{d}s| \to_p 0$ *(as $n \to \infty$) under $D(\xi) \neq 0$ in case of a Fredholm integral equation, then* $\sup_{t,s} |\Gamma_n^\jmath(t,s) - \Gamma^\jmath(t,s)| \to_p 0$ *(as $n \to \infty$) ($\jmath = \sharp, \flat$), where $\Gamma_n^\sharp$ and $\Gamma^\sharp$ are $\Gamma^\sharp$ defined in Theorem A.1 in which $K^\sharp$ is substituted by $K_n^\sharp$ and $K^\sharp$, and $\Gamma_n^\flat$ and $\Gamma^\flat$ are $\Gamma^\flat$ defined in Theorem A.2 in which $K^\flat$ is substituted by $K_n^\flat$ and $K^\flat$, respectively.*

Proof For every $\sup_t |f(t)| < \infty$, we define ϕ_n^f and ϕ^f by an integral equation $\phi_n^f(t) = f(t) + \xi \int_a^{b(t)} \phi_n^f(s) K_n^\jmath(t,s)\mathrm{d}s$ and $\phi^f(t) = f(t) + \xi \int_a^{b(t)} \phi^f(s) K^\jmath(t,s)\mathrm{d}s$, respectively ($\jmath = \sharp, \flat$), where $b(t) = t$ in case of a Volterra integral equation and $b(t) = b$ in case of a Fredholm. These solutions can be given by $\phi_n^f(t) = f(t) + \xi \int_a^{b(t)} \Gamma_n^\jmath(t,s,\xi) f(s)\mathrm{d}s$ and $\phi^f(t) = f(t) + \xi \int_a^{b(t)} \Gamma^\jmath(t,s,\xi) f(s)\mathrm{d}s$ from Theorem A.1 or A.2. By Proposition A.2 or A.3, we have $\sup_t |\phi^f(t)| < \infty$ and $\sup_t |\phi_n^f(t)| < \infty$ in probability because of $\sup_{t,s} |K_n^\jmath(t,s)| < \infty$ in probability. The difference $\phi_n^f(t) - \phi^f(t)$ can be divided as

$$\phi_n^f(t) - \phi^f(t) = \xi \int_a^{b(t)} \phi_n^f(t) \left(K_n^\jmath(t,s) - K^\jmath(t,s)\right) \mathrm{d}s$$
$$+\xi \int_a^{b(t)} K^\jmath(t,s) \left(\phi_n^f(s) - \phi^f(s)\right) \mathrm{d}s.$$

We can treat this expression as another integral equation $\phi_n^f(t) - \phi^f(t) = o_p(1)(b(t)) + \xi \int_a^{b(t)} K^\jmath(t,s)(\phi_n^f(s) - \phi^f(s))\mathrm{d}s$, so that we have its solution $\phi_n^f(t) - \phi^f(t) = o_p(1)(b(t)) + \xi \int_a^{b(t)} \Gamma^\jmath(t,s,\xi) o_p(1)(b(s))\mathrm{d}s$. Therefore, $\sup_t |\phi_n^f(t) - \phi^f(t)| \to_p 0$. By combination this result with solutions of $\phi_n^f(t)$ and $\phi^f(t)$, we have $\sup_t |\int_a^{b(t)} \{\Gamma_n^\jmath(t,s,\xi) - \Gamma^\jmath(t,s,\xi)\} f(s)\mathrm{d}s| \to_p 0$. Here consider $f(s) = \pm 1_{[\Gamma_n^\jmath(t,s,\xi) - \Gamma^\jmath(t,s,\xi)]}$, where $\pm 1_{[a]}$ is 1 if $a > 0$ and -1 otherwise. Such $f(\cdot)$'s lead $\sup_{t,s} |\Gamma_n^\jmath(t,s,\xi) - \Gamma^\jmath(t,s,\xi)| \to_p 0$. □

A.2 The relationship between $\mathbb{N}$ and Λ^*

For a function $g(T, X, Z)$ into R^1 from (T, X, Z), we define

$$\mathbb{N}(g,t) = \mathrm{E}\left[\int_0^t g(T,X,Z) r^*(1-c^*) S^*(T) S_C^*(T) \mathrm{d}\Lambda_0^*(T)\right].$$

Then, we obtain a relation

$$\mathrm{d}\Lambda_0^*(t) = \mathrm{d}\mathbb{N}(g,t) \big/ \mathrm{E}\left[g(t,X,Z) r^*(1-c^*) S^*(t) S_C^*(t)\right],$$

which is an extension for $\mathbb{N}(t)$ and (3). As the other expression of $\mathbb{N}(g,t)$, we have

$$\mathbb{N}(g,t) = \mathrm{E}\left[g(T,Z,X)I(T \le t, \Delta = 1)\right] = \mathrm{E}\left[g(T,Z,X)\{1 - Y(t)\}\Delta\right].$$

Hence, we find $\mu_{\mathbb{N}}(g(T,Z,X),t) = \mathbb{N}(g,t)$ in the notation defined in Sect. 5.4. On the other hand, $\mu_{\mathbb{S}}(g(t,Z,X),t)$ is computed as follows

$$\begin{aligned}
\mu_{\mathbb{S}}(g(t,Z,X),t) &= \mathrm{E}\left[g(t,Z,X)r^*Y(t)\Delta\right] + \mathrm{E}\left[g(t,Z,X)r^*Y(t)(1-\Delta)w^*(T)\right] \\
&= \mathrm{E}\left[g(t,Z,X)r^*\int_t^\infty (1-c^*)r^*S^*(T)S_C^*(T)\mathrm{d}\Lambda_0^*(T)\right] \\
&\quad + \mathrm{E}\left[g(t,Z,X)r^*\int_t^\infty w^*(T)\{c^*+(1-c^*)S^*(T)\}S_C^*(T)\mathrm{d}\Lambda_C^*(T)\right] \\
&= \mathrm{E}\left[g(t,Z,X)r^*(1-c^*)S^*(t)S_C^*(t)\right].
\end{aligned}$$

Therefore, for $\mu_{\mathbb{N}}(g(X,Z),t)$ and $\mu_{\mathbb{S}}(g(X,Z),t)$, we have a relationship

$$\mathrm{d}\mu_{\mathbb{N}}(g(X,Z),t) = \mu_{\mathbb{S}}(g(X,Z),t)\mathrm{d}\Lambda_0^*(t). \tag{19}$$

Acknowledgements The authors are grateful to two referees for their constructive comments and helpful suggestions.

References

Andersen, P. K., Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *The Annals of Statistics*, *10*, 1100–1120.
Breslow, N. E. (1972). Contribution to discussion of paper by D. R. Cox. *Journal of the Royal Statistical Society, Series B*, *34*, 216–217.
Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B*, *34*, 187–220.
Farewell, V. T. (1986). Mixture models in survival analysis: Are they worth the risk?. *Canadian Journal of Statistics*, *14*, 257–262.
Hochstadt, H. (1973). *Integral equations*. New York: John Wiley & Sons.
Kuk, A. Y. C., Chen, C. H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, *79*, 531–541.
Lu, W., Ying, Z. (2004). On semiparametric transformation cure models. *Biometrika*, *91*, 331–343.
Peng, Y., Dear, K. B. G. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics*, *56*, 237–243.
Pipkin, A. C. (1991). *A course on integral equations*. New York: Springer-Verlag.
Smithies, F. (1958). *Integral equations*. New York: Cambridge Tracts in Mathematics and Mathematical Physics, Cambridge University Press.
Sugimoto, T., Goto, M. (1999). Adjustment of the proportional hazards model via characteristics of cases and its inference (in Japanese). *Bulletin of the Computational Statistics of Japan*, *12*, 85–106.
Sugimoto, T., Goto, M. (2003). The proportional hazards model of time-varying type and a class of rank test (in Japanese). *Journal of the Japan Statistical Society*, *33*, 365–382.
Sugimoto, T., Hamasaki, T., Goto, M. (2005) Estimation from pseudo partial likelihood in a semiparametric cure model. *Journal of the Japanese Society of Computational Statistics*. (in press).
Sy, J. P., Taylor, J. M. G. (2000). Estimation in a cox proportional hazards cure model. *Biometrics*, *56*, 227–236.

Tsodikov, A. (1998). A proportional hazard model taking account of long-term surviours. *Biometrics*, *54*, 1508–1516.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes*, Springer-Verlag, New York.