Ann Inst Stat Math (2013) 65:859–888
DOI 10.1007/s10463-012-0395-4

Asymptotic distribution of the nonparametric distribution estimator based on a martingale approach in doubly censored data

Tomoyuki Sugimoto

Received: 23 March 2012 / Revised: 12 September 2012 / Published online: 18 January 2013

Abstract For analysis of time-to-event data with incomplete information beyond right-censoring, many generalizations of the inference of the distribution and regression model have been proposed. However, the development of martingale approaches in this area has not progressed greatly, while for right-censored data such an approach has spread widely to study the asymptotic properties of estimators and to derive regression diagnosis methods. In this paper, focusing on doubly censored data, we discuss a martingale approach for inference of the nonparametric maximum likelihood estimator (NPMLE). We formulate a martingale structure of the NPMLE using a score function of the semiparametric profile likelihood. Finally, an expression of the asymptotic distribution of the NPMLE is derived more conveniently without depending on an infinite matrix expression as in previous research. A further useful point is that a variance-covariance formula of the NPMLE computable in a larger sample is obtained as an empirical version of the limit form presented here.

Keywords Counting process · Forward-backward intensities · Nonparametric maximum likelihood estimator · Semiparametric profile likelihood · Weak convergence

This study was supported by Grant-in-Aid for Scientific Research Young Scientists (B), 23700336, from JSPS.

T. Sugimoto (✉)
Department of Mathematical Sciences, Graduate School of Science and Technology,
Hirosaki University, 3 Bunkyocho, Hirosaki, Aomori 036-8561, Japan
e-mail: tomoyuki@cc.hirosaki-u.ac.jp

1 Introduction

In statistical analysis for right-censored data, the Kaplan–Meier estimate and Cox regression are typical tools. A large sample study of their estimators and several methods of regression diagnosis are elegantly constructed by the counting processes and the corresponding martingale theories (Andersen and Gill 1982; Fleming and Harrington 1991; Andersen et al. 1993; Therneau and Grambsch 2000). Many researchers predict that the counting processes and their associated martingales will certainly continue to play an important role in this area (see, e.g., Oakes 2000). For analysis of time-to-event data with incomplete information beyond right-censoring, such as doubly or interval censored data, generalizations of inference procedures of the nonparametric distribution estimators (e.g., Turnbull 1974, 1976; Chang 1990; Gu and Zhang 1993; Gentleman and Geyer 1994; Mykland and Ren 1996; Wellner and Zhang 1997) and the Cox regression model (e.g., Kim 2003; Cai and Cheng 2004) have been proposed by many authors. However, in cases with such incomplete data, it remains uncertain whether to characterize the properties of their estimators using a martingale approach; in right-censored data, by contrast, the approach has become widespread as a main tool for such a purpose. Our interest here is how such an approach can be developed in the case of double-censoring, which consists of complete, right-censored or left-censored observations.

We discuss a martingale approach for inference of the nonparametric maximum likelihood estimator (NPMLE) in doubly censored data. First, as a natural request, it is necessary to consider both forward and backward aspects of the counting processes to record doubly censored data, since the processes and the corresponding martingales play central roles in right- and left-censored-only data, respectively. Therefore, as fundamental results, we provide martingale properties for the forward and backward counting processes and then formulate the correlation structure of their martingales. The analogy of right-censoring case will often work for such studies. As a study concerned in the topics, Patilea and Rolin (2006) proposes latent variable models related to doubly censored data, where a backward martingale approach is implied to derive the product-limit estimators of survival function. On the other hand, the NPMLE is a solution of an integral equation, termed the self-consistent equation, which cannot be expressed by a closed form in doubly or interval censored data. For such a reason, the asymptotic properties of the NPMLE have been studied using some infinite matrix or operator expression taken from the self-consistent equations (Tsai and Crowley 1985; Chang and Yang 1987; Gu and Zhang 1993; Yu and Li 2001). Similarly, even if the martingale properties presented here are incorporated into the self-consistent equations, it is difficult to discuss the asymptotic distribution of the NPMLE without some infinite matrix expression in doubly censored data. Hence, such a manner does not lead to an elegant expansion of a martingale approach as in right- or left-censored-only data.

In this paper, to overcome this difficulty, we characterize the martingale structure of the NPMLE using a score function of the log profile likelihood (Murphy and van der Vaart 1997). The asymptotic distribution of the NPMLE is discussed based on the martingale structure in such an expansion. Finally, we show that the limit distribution of the NPMLE converges weakly to a Gaussian process freed from some infinite matrix or operator expression. A further useful point of this result is that the variance-covariance formula of the NPMLE proposed by Turnbull (1974), which is worked out computationally and theoretically by Sugimoto (2011), can be captured as a natural estimate (empirical version) of the limit form of the variance-covariance derived here. This variance-covariance formula is iteration-free and computable in a larger sample and reduces to the Greenwood formula in right-censored data.

In Sect. 2 we briefly review the semiparametric profile likelihood inference and its structure of derivatives in doubly censored data. In Sect. 3 we formulate forward and backward martingale properties for the counting processes and their correlation structure. In Sect. 4 we derive a martingale structure of the NPMLE using a score function of the profile likelihood; we then show that the asymptotic distribution of the NPMLE is a superposition of two Gaussian martingale processes, as more convenient expression without depending on an infinite matrix expression as in previous research (Chang and Yang 1987; Gu and Zhang 1993).

2 Preliminary

2.1 Empirical likelihood and the NPMLE

In a doubly censored sample (Gehan 1965; Turnbull 1974) of size n, the ith observation T_i and censoring indicator Δ_i $(i = 1, \ldots, n)$ are available as

$$\begin{aligned} T_i &= \max[\min(T_i^*, C_i^R), C_i^L] = \min[\max(T_i^*, C_i^L), C_i^R], \\ \Delta_i &= \begin{cases} 1 & \text{if } C_i^L < T_i^* \le C_i^R \quad \text{(no censoring)} \\ 2 & \text{if } C_i^R < T_i^* \qquad\quad\;\; \text{(right-censoring)} \\ 3 & \text{if } T_i^* \le C_i^L \qquad\quad\;\; \text{(left-censoring)} \end{cases}. \end{aligned} \tag{1}$$

Here $T_1^*, \ldots, T_n^*$ are independent and identically distributed (i.i.d.) random variables following a true distribution function $F^*(t) = 1 - S^*(t)$, and $(C_1^L, C_1^R), \ldots, (C_n^L, C_n^R)$ are i.i.d. vectors of left- and right-censoring times independent of T_i^*'s with $C_i^L \le C_i^R$. Let $F^L(t) = 1 - S^L(t)$ and $F^R(t) = 1 - S^R(t)$ denote true marginal distribution functions of C_i^L and C_i^R, respectively. For the sake of simplicity, we assume throughout this paper that F^*, F^L and F^R are continuous functions.

Example Consider a study of the age-of-onset distributions for a disease. Some persons had already suffered from the disease at the time of registration to this study, so their ages were left-censored observations. The others were followed-up and received periodical medical examination. The ages of onset for some of persons are observed exactly during the time-span of the study. The ages of the rest persons who had not yet suffered from the disease until the end of the study are treated as right-censored observations.

Let $F(t)\,(0 \le t)$ be a discretized parameter function to estimate the unknown $F^*(\cdot)$ nonparametrically, where $F(\cdot)$ is equivalently parameterized by the individual vector expression $\boldsymbol{F} = (F_{[1]}, F_{[2]}, \ldots, F_{[n-1]}, F_{[n]})$ such that $F_{[i]} = F(T_{(i)})$ for the order statistics $T_{(1)} < T_{(2)} < \cdots < T_{(n)}$ of T_i's. The NPMLE $\widehat{F}(t)$ $(0 \le t)$ of $F^*(t)$ maximizes the log empirical likelihood

$$\begin{aligned} l_n(\boldsymbol{F}) &= l_n(F_{[1]}, \ldots, F_{[n]}) \\ &= \log \prod\nolimits_{i=1}^{n} (F_{[i]} - F_{[i-1]})^{\mathbb{1}(\Delta_{(i)}=1)} (1 - F_{[i]})^{\mathbb{1}(\Delta_{(i)}=2)} F_{[i]}^{\mathbb{1}(\Delta_{(i)}=3)} \end{aligned}$$

with the constraint condition

$$\begin{aligned} &0 = F_0 \le F_{[1]} \le F_{[2]} \le \cdots \le F_{[n-1]} \le F_{[n]} = 1, \\ &F_{[i-1]} < F_{[i]} \ \text{ if } \ \Delta_{(i)} = 1 \ \text{ for } \ i = 1, \ldots, n, \end{aligned} \tag{2}$$

where $\mathbb{1}(\cdot)$ is the indicator function. The necessary and sufficient condition to find the NPMLE has been formulated in several ways (e.g., Turnbull 1974; Mykland and Ren 1996; Wellner and Zhang 1997, etc.), but here we introduce the condition based on the score functions to discuss derivatives between parameters later. To identify $F_{[i]}$ and $F_{[i-1]}$ as the same parameter if $F_{[i]} = F_{[i-1]}$, let $[i]$ be an index number j satisfying $F_j = F_{[i]}$ of $\boldsymbol{F}$ such that distinct $F_{[i]}$'s are denoted by $F_1 < F_2 < \cdots < F_{[n]-1} < F_{[n]}$. Similarly, denote the individual vector expression of $\widehat{F}(\cdot)$ by $\widehat{\boldsymbol{F}} = (\widehat{F}_{[1]}, \widehat{F}_{[2]}, \ldots, \widehat{F}_{[n-1]}, \widehat{F}_{[n]})$, so that we can represent the distinct $\widehat{F}_{[i]}$'s as $\widehat{F}_1 < \widehat{F}_2 < \cdots < \widehat{F}_{[n]-1} < \widehat{F}_{[n]} = 1$. Let $u_j(\boldsymbol{F})$ denote the first derivative of $l_n(\boldsymbol{F})$ w.r.t. F_j,

$$u_j(\boldsymbol{F}) = u_j(F_{j-1}, F_j, F_{j+1}) = \frac{\alpha_j}{F_j - F_{j-1}} - \frac{\alpha_{j+1}}{F_{j+1} - F_j} - \frac{\beta_j}{1 - F_j} + \frac{\gamma_j}{F_j},$$

where α_j, β_j and γ_j are

$$\begin{aligned} \alpha_j = \sum\nolimits_{i=1}^{n} \mathbb{1}(T_i = J_j, \Delta_i = 1), \quad \beta_j = \sum\nolimits_{i=1}^{n} \mathbb{1}(T_i \in [J_j, J_{j+1}), \Delta_i = 2), \\ \gamma_j = \sum\nolimits_{i=1}^{n} \mathbb{1}(T_i \in [J_j, J_{j+1}), \Delta_i = 3), \end{aligned}$$

and J_j is the jth time point at which $F(\cdot)$ jumps. Note that $F(t) = F_j$ if $t \in [J_j, J_{j+1})$ in a time and $F_{[i]} = F_j$ if $T_{(i)} \in [J_j, J_{j+1})$ for the (i)th individual. Let $\mathcal{J}$ be the collection of all possible $\boldsymbol{J} = (J_1, \ldots, J_{[n]})$'s considered under (2) and let $\widehat{\boldsymbol{J}}$ be $\boldsymbol{J}$ accompanied by the NPMLE. The NPMLE $\widehat{\boldsymbol{F}}$ satisfies

$$u_j(\widehat{\boldsymbol{F}}) = u_j(\boldsymbol{F})\big|_{\boldsymbol{F}=\widehat{\boldsymbol{F}}} = 0, \quad j = 1, \ldots, [n] - 1 \ \text{ for } \widehat{\boldsymbol{J}} \tag{3}$$

and the dimension $[n]$ of $\widehat{\boldsymbol{J}}$ is the largest among every $\boldsymbol{J} \in \mathcal{J}$ for which F satisfies $u_j(\widehat{\boldsymbol{F}}) = 0$, $j = 1, \ldots, [n] - 1$. For details, see Sugimoto (2011, Lemma 1). A linear transformation of (3) leads to the self-consistent equations (see Turnbull 1974, Lemma A1 and Sugimoto 2011, Lemma 2).

2.2 Profile likelihood and its derivatives

Further analysis such as test or confidence interval about a particular $F^*(t)$ is based on the log profile likelihood for $\theta_t = F(t)$ constructed as

$$pl_n(\theta_t) = \max_{0 \le F_{[1]} \le \cdots \le F_{[i]} \le \theta_t \le F_{[i+1]} \le \cdots \le F_{[n]} \le 1} l_n(\boldsymbol{F}) \quad \text{for } T_{(i)} \le t < T_{(i+1)} \tag{4}$$

To compute $pl_n(\theta_t)$, Chen and Zhou (2003) provided a self-consistent algorithm, while Sugimoto (2011) proposed a Newton–Raphson algorithm. The analysis based on the semiparametric profile likelihood ratio performs better; however, the computation potentially requires a heavy load. Therefore, in a larger sample it will be desirable to replace it with the Wald- or score-type if possible. In the following, we review results of Sugimoto (2011) required for this paper. Let $U_n(\theta_t) = \partial pl_n(\theta_t)/\partial\theta_t$ and $\mathcal{I}_n(\theta_t) = -\partial^2 pl_n(\theta_t)/\partial\theta_t^2$ be the first and minus the second derivatives of the log profile likelihood, respectively. To express $U_n(\theta_t)$ and $\mathcal{I}_n(\theta_t)$ definitely, let $\widehat{\boldsymbol{F}}(\theta_t) = (\widehat{F}_{[1]}(\theta_t), \widehat{F}_{[2]}(\theta_t), \ldots, \widehat{F}_{[n-1]}(\theta_t), \widehat{F}_{[n]}(\theta_t))$ denote the restricted NPMLE of F^* under a given value θ_t of $F(t)$ (i.e., the solution for the maximization in (4) such that $pl_n(\theta_t) = l_n(\boldsymbol{F})|_{\boldsymbol{F}=\widehat{\boldsymbol{F}}(\theta_t)}$), which satisfies

$$\begin{cases} 0 = \widehat{F}_0(\theta_t) < \widehat{F}_1(\theta_t) < \cdots < \widehat{F}_{m_t-1}(\theta_t) < \widehat{F}_{m_t}(\theta_t) = \theta_t \\ \quad \theta_t < \widehat{F}_{m_t+1}(\theta_t) < \cdots < \widehat{F}_{[n]-1}(\theta_t) < \widehat{F}_{[n]}(\theta_t) = 1 \end{cases},$$

where m_t is the interval number including t as $J_{m_t} \le t < J_{m_t+1}$. To avoid confusion, we simply call $\widehat{\boldsymbol{F}}(\theta_t)$ the profile estimator (under the constraint $F(t) = \theta_t$) in this paper. Also, let $\widehat{\boldsymbol{J}}_{\theta_t}$ be $\boldsymbol{J}$ accompanied by $\widehat{\boldsymbol{F}}(\theta_t)$, because J_j's, m_t and $[n]$ led by $\widehat{\boldsymbol{F}}(\theta_t)$ are determined depending on θ_t. Note that the condition to obtain the profile estimator $\widehat{\boldsymbol{F}}(\theta_t)$ is

$$u_j(\widehat{\boldsymbol{F}}(\theta_t)) = 0, \quad j = 1, \ldots, m_t - 1, m_t + 1, \ldots, [n] - 1 \ \text{ for } \widehat{\boldsymbol{J}}_{\theta_t}, \tag{5}$$

excluding the case of $j = m_t$ in (3). Based on these notations and the rule of derivatives, we obtain the following result for expressions of $U_n(\theta_t)$ and $\mathcal{I}_n(\theta_t)$:

Proposition 1 *In doubly censored data* (1)*, the score and Fisher functions of the log profile likelihood $pl_n(\theta_t)$ are expressed as*

$$\begin{aligned} U_n(\theta_t) &= \frac{\alpha_{m_t}}{\theta_t - \widehat{F}_{m_t-1}(\theta_t)} - \frac{\alpha_{m_t+1}}{\widehat{F}_{m_t+1}(\theta_t) - \theta_t} - \frac{\beta_{m_t}}{1-\theta_t} + \frac{\gamma_{m_t}}{\theta_t}, \\ \mathcal{I}_n(\theta_t) &= \sum_{j=1}^{[n]} \frac{\alpha_j}{(\widehat{F}_j(\theta_t) - \widehat{F}_{j-1}(\theta_t))^2} \left\{ \frac{\partial \widehat{F}_j(\theta_t)}{\partial\theta_t} - \frac{\partial \widehat{F}_{j-1}(\theta_t)}{\partial\theta_t} \right\}^2 \\ &\quad + \sum_{j=1}^{[n]-1} \frac{\beta_j}{(1-\widehat{F}_j(\theta_t))^2} \left\{ \frac{\partial \widehat{F}_j(\theta_t)}{\partial\theta_t} \right\}^2 + \sum_{j=1}^{[n]-1} \frac{\gamma_j}{\widehat{F}_j(\theta_t)^2} \left\{ \frac{\partial \widehat{F}_j(\theta_t)}{\partial\theta_t} \right\}^2. \end{aligned} \tag{6}$$

See Appendix A.1 for the details of Proposition 1. The derivative $\partial\widehat{F}_j(\theta_t)/\partial\theta_t$ included in (6) is usually obtained by solving linear equations derived by differentiating both sides of the equations (5) w.r.t. θ_t. However, in doubly censored data, we can compute $\partial\widehat{F}_j(\theta_t)/\partial\theta_t$ without such an inverse matrix expression (for further details, see Sugimoto 2011). To explain this, let $i_{j,l}(F)$ be minus the second derivative of $l_n(F)$ w.r.t. F_j and F_l. If $j \leq m_t - 1$, we progressively have

$$\frac{\partial\widehat{F}_j(\theta_t)}{\partial\widehat{F}_{j+1}(\theta_t)} = -i_{j,j+1}(\widehat{\boldsymbol{F}}(\theta_t)) \Big/ \left\{ i_{j,j}(\widehat{\boldsymbol{F}}(\theta_t)) + i_{j-1,j}(\widehat{\boldsymbol{F}}(\theta_t)) \frac{\partial\widehat{F}_{j-1}(\theta_t)}{\partial\widehat{F}_j(\theta_t)} \right\} \tag{7}$$

by starting from $\partial\widehat{F}_0(\theta_t)/\partial\widehat{F}_1(\theta_t) = 0$. While, if $m_t + 1 \leq j$, we regressively obtain

$$\frac{\partial\widehat{F}_j(\theta_t)}{\partial\widehat{F}_{j-1}(\theta_t)} = -i_{j,j-1}(\widehat{\boldsymbol{F}}(\theta_t)) \Big/ \left\{ i_{j,j}(\widehat{\boldsymbol{F}}(\theta_t)) + i_{j+1,j}(\widehat{\boldsymbol{F}}(\theta_t)) \frac{\partial\widehat{F}_{j+1}(\theta_t)}{\partial\widehat{F}_j(\theta_t)} \right\} \tag{8}$$

from $\partial\widehat{F}_{[n]}(\theta_t)/\partial\widehat{F}_{[n]-1}(\theta_t) = 0$. By the chain rule of differentiations, $\partial\widehat{F}_j(\theta_t)/\partial\theta_t$ is computed as

$$\frac{\partial\widehat{F}_j(\theta_t)}{\partial\theta_t} = \begin{cases} \dfrac{\partial\widehat{F}_j(\theta_t)}{\partial\widehat{F}_{j+1}(\theta_t)} \dfrac{\partial\widehat{F}_{j+1}(\theta_t)}{\partial\widehat{F}_{j+2}(\theta_t)} \cdots \dfrac{\partial\widehat{F}_{m_t-1}(\theta_t)}{\partial\theta_t} & \text{if} \quad j \leq m_t - 1 \\ \dfrac{\partial\widehat{F}_j(\theta_t)}{\partial\widehat{F}_{j-1}(\theta_t)} \dfrac{\partial\widehat{F}_{j-1}(\theta_t)}{\partial\widehat{F}_{j-2}(\theta_t)} \cdots \dfrac{\partial\widehat{F}_{m_t+1}(\theta_t)}{\partial\theta_t} & \text{if} \quad j \geq m_t + 1 \end{cases}. \tag{9}$$

The quantity $1/\mathcal{I}_n(\theta_t)$ evaluated at $\theta_t = \widehat{F}(t)$ is an appropriate variance formula of the NPMLE $\widehat{F}(t)$ for doubly censored data, which always gives the same result as Turnbull's determinant-based formula (Sugimoto 2011). That is, $1/\mathcal{I}_n(\widehat{F}(t))$ is the same as the m_tth diagonal elements of the inverse of the full Fisher matrix of $l_n(\widehat{F})$ (composed of all $i_{j,l}(\widehat{F})$), similar to classical profile likelihood theory. In addition, $1/\mathcal{I}_n(\widehat{F}(t))$ yields a limit identical to Chang (1990, Theorem 4.2) asymptotic variance formula under some regular conditions (e.g., as seen in Murphy and van der Vaart 1997, Theorem 2.1). Further, in right-censored data, $1/\mathcal{I}_n(\widehat{F}(t))$ reduces to Greenwood's variance formula (Sugimoto 2011, Lemma 5). Therefore, the form of $\mathcal{I}_n(\widehat{F}(t))$ provides us with a hint of how to consider martingale properties included in the difference $\widehat{F} - F^*$ or provides evidence that a martingale approach is valid.

3 Counting processes and their martingales

In the section below, we formulate martingale properties of the counting processes included in doubly censored data. Hereafter, assume the following condition:

Condition 1 *F^*, F^L and F^R are continuous functions and the supports of $F^*(t)$, $F^L(t)$ and $F^R(t)$ are included in* $[0, 1]$.

To derive forward and backward martingale properties simply, the distributions are supported by the unit interval under the continuity. One goal of this section is to prepare the properties needed to investigate the asymptotic distribution of the NPMLE in Sect. 4.

3.1 Forward properties

Here we will consider forward counting processes and their martingales. Let $\Lambda^{\mathrm{f}(1)}(t)$, $\Lambda^{\mathrm{f}(2)}(t)$ and $\Lambda^{\mathrm{f}(3)}(t)$ be the true cumulative hazard functions for T^*, C_i^R and C_i^L, respectively. For example, if F^* is differentiable, $\Lambda^{\mathrm{f}(1)}(t)$ is $\int_{(0,t]} \lambda^{\mathrm{f}(1)}(s)\vec{d}s$ with $\lambda^{\mathrm{f}(1)}(t) = \lim_{dt\downarrow 0} \Pr(t \le T_i^* < t + dt | t \le T_i^*)/dt$, where we denote $\vec{d}g(t) = g(t) - g(t_-)$ for a function g and a time t_- just prior to t. We define the forward counting processes and at-risk processes as

$$N_i^{\mathrm{f}(\jmath)}(t) = \mathbb{1}(T_i \le t, \Delta_i = \jmath) \text{ for } \jmath = 1, 2, 3 \text{ and } Y_i^{\mathrm{f}}(t) = \mathbb{1}(t \le T_i),$$

$i = 1, \dots, n$. Let $\mathscr{F}_t^{\mathrm{f}}$ be a forward filtration formulated by

$$\mathscr{F}_t^{\mathrm{f}} = \sigma\{N_i^{\mathrm{f}(1)}(s), N_i^{\mathrm{f}(2)}(s), N_i^{\mathrm{f}(3)}(s), i = 1, \dots, n : 0 \le s \le t\},$$

where $\sigma\{B\}$ is the smallest σ-algebra generated by B. In Lemma 1, we formulate forward martingale properties of the counting processes $N_i^{\mathrm{f}(1)}$, $N_i^{\mathrm{f}(2)}$ and $N_i^{\mathrm{f}(3)}$ to detect the uncensored, right- and left-censored observations, respectively.

Lemma 1 *Suppose that Condition 1 is satisfied. Let*

$$A_i^{\mathrm{f}(\jmath)}(t) = \int_{(0,t]} Y_i^{\mathrm{f}}(s)\omega^{\mathrm{f}(\jmath)}(s)\vec{d}\Lambda^{\mathrm{f}(\jmath)}(s), \quad \jmath = 1, 2, 3,$$

where $\omega^{\mathrm{f}(1)}(t) = (S^R(t) - S^L(t))S^*(t)/Q^{\mathrm{f}}(t)$, $\omega^{\mathrm{f}(2)}(t) = S^*(t)S^R(t)/Q^{\mathrm{f}}(t)$, $\omega^{\mathrm{f}(3)}(t) = (1 - S^*(t))\ S^L(t)/Q^{\mathrm{f}}(t)$ *and* $Q^{\mathrm{f}}(t) = \Pr(T_i > t) = S^R(t)S^*(t) + S^L(t)(1 - S^*(t))$*. Then, in the double-censoring model (1), the processes*

$$M_i^{\mathrm{f}(\jmath)}(t) = N_i^{\mathrm{f}(\jmath)}(t) - A_i^{\mathrm{f}(\jmath)}(t), \quad \jmath = 1, 2, 3, \quad i = 1, \dots, n$$

are square-integrable $\mathscr{F}_t^{\mathrm{f}}$*-martingales. Let* $\langle M_i^{\mathrm{f}(\imath)}, M_j^{\mathrm{f}(\jmath)}\rangle^{\mathrm{f}}$, $i, j = 1, \dots, n$, $\imath$, $\jmath = 1, 2, 3$ *be* $\mathscr{F}_t^{\mathrm{f}}$*-predictable covariation processes of* $M_i^{\mathrm{f}(\imath)}$ *and* $M_j^{\mathrm{f}(\jmath)}$*; then*

$$\langle M_i^{\mathrm{f}(\imath)}, M_j^{\mathrm{f}(\jmath)}\rangle^{\mathrm{f}}(t) = \begin{cases} A_i^{\mathrm{f}(\jmath)}(t), & \text{if } i = j \text{ and } \imath = \jmath, \\ 0, & \text{otherwise}, \end{cases} \quad (0 \le t).$$

Lemma 1 is proved in Appendix A.2. By Lemma 1, for locally bounded $\mathscr{F}_t^{\mathrm{f}}$-predicable processes $H_1^{\mathrm{f}}(t)$ and $H_2^{\mathrm{f}}(t)$,

$$\left\langle \int_{(0,t]} H_1^{\mathrm{f}}(x)\,\vec{d}M_i^{\mathrm{f}(\imath)}(x),\ \int_{(0,s]} H_2^{\mathrm{f}}(y)\,\vec{d}M_j^{\mathrm{f}(\jmath)}(y) \right\rangle^{\mathrm{f}} = \int_{(0,t\wedge s]} H_1^{\mathrm{f}}(x)H_2^{\mathrm{f}}(x)\,\vec{d}\left\langle M_i^{\mathrm{f}(\imath)}, M_j^{\mathrm{f}(\jmath)}\right\rangle^{\mathrm{f}}(x)$$

is easily shown, where $t \wedge s = \min(t, s)$.

3.2 Backward properties

Here we discuss backward martingale properties opposite to those in Sect. 3.1. Let $\Lambda^{\mathrm{b}(1)}(t)$, $\Lambda^{\mathrm{b}(2)}(t)$ and $\Lambda^{\mathrm{b}(3)}(t)$ be the true cumulative reversed-hazard functions of T^*, C_i^R and C_i^L, respectively. Similar to $\Lambda^{\mathrm{f}(1)}$, when F^* is differentiable, $\Lambda^{\mathrm{b}(1)}(t)$ is $\int_{[t,1)} \lambda^{\mathrm{b}(1)}(s)\,\overleftarrow{d}s$ with $\lambda^{\mathrm{b}(1)}(t) = \lim_{dt\downarrow 0} \Pr(t - dt < T_i^* \le t | T_i^* \le t)/dt$, where we denote $\overleftarrow{d}g(t) = g(t) - g(t_+)$ for a function g and a time t_+ just after t. Define the backward counting processes and at-risk processes as

$$N_i^{\mathrm{b}(\jmath)}(t) = \mathbb{1}(t \le T_i, \Delta_i = \jmath) \ \text{ for } \ \jmath = 1, 2, 3 \ \text{ and } \ Y_i^{\mathrm{b}}(t) = \mathbb{1}(T_i \le t),$$

$i = 1, \ldots, n$ and let $\mathscr{F}_t^{\mathrm{b}}$ be a backward filtration such that

$$\mathscr{F}_t^{\mathrm{b}} = \sigma\left\{ N_i^{\mathrm{b}(1)}(s), N_i^{\mathrm{b}(2)}(s), N_i^{\mathrm{b}(3)}(s), \quad i = 1, \ldots, n : t \le s \le 1 \right\}.$$

As a reversed version of Lemma 1, we formulate backward martingale properties of the counting processes $N_i^{\mathrm{b}(\jmath)}$, $\jmath = 1, 2, 3$. Note that $N_i^{\mathrm{b}(\jmath)}$, $\jmath = 1, 2, 3$ are left-continuous with right-hand limits.

Lemma 2 *Suppose that Condition 1 is satisfied. Let*

$$A_i^{\mathrm{b}(\jmath)}(t) = \int_{[t,1)} Y_i^{\mathrm{b}}(s)\omega^{\mathrm{b}(\jmath)}(s)\,\overleftarrow{d}\Lambda^{\mathrm{b}(\jmath)}(s), \quad \jmath = 1, 2, 3,$$

where $\omega^{\mathrm{b}(1)}(t) = (F^R(t) - F^L(t))F^*(t)\big/Q^{\mathrm{b}}(t)$, $\omega^{\mathrm{b}(2)}(t) = (1 - F^*(t))F^R(t)\big/Q^{\mathrm{b}}(t)$, $\omega^{\mathrm{b}(3)}(t) = F^*(t)F^L(t)\big/Q^{\mathrm{b}}(t)$ *and* $Q^{\mathrm{b}}(t) = \Pr(T_i \le t) = F^L(t)F^*(t) + F^R(t)(1 - F^*(t))$. *Then, in the double-censoring model* (1)*, the processes*

$$M_i^{\mathrm{b}(\jmath)}(t) = N_i^{\mathrm{b}(\jmath)}(t) - A_i^{\mathrm{b}(\jmath)}(t), \quad \jmath = 1, 2, 3, \quad i = 1, \ldots, n$$

are square-integrable $\mathscr{F}_t^{\mathrm{b}}$*-martingales. Let* $\langle M_i^{\mathrm{b}(\imath)}, M_j^{\mathrm{b}(\jmath)}\rangle^{\mathrm{b}}$, $i, j = 1, \ldots, n$, $\imath$, $\jmath = 1, 2, 3$ *be* $\mathscr{F}_t^{\mathrm{b}}$*-predictable covariation processes of* $M_i^{\mathrm{b}(\imath)}$ *and* $M_j^{\mathrm{b}(\jmath)}$*; then*

$$\left\langle M_i^{\mathrm{b}(\imath)}, M_j^{\mathrm{b}(\jmath)}\right\rangle^{\mathrm{b}}(t) = \begin{cases} A_i^{\mathrm{b}(\jmath)}(t), & \text{if } \ i = j \ \text{ and } \ \imath = \jmath, \\ 0, & \text{otherwise,} \end{cases} \qquad (t \le 1).$$

The proof of Lemma 2 is a simple reversed version of Lemma 1. By Lemma 2, for locally bounded $\mathscr{F}_t^{\mathrm{b}}$-predicable processes $H_1^{\mathrm{b}}(t)$ and $H_2^{\mathrm{b}}(t)$, we have

$$\Big\langle \int_{[t,1)} H_1^{\mathrm{b}}(x) \overleftarrow{d} M_i^{\mathrm{b}(\imath)}(x), \int_{[s,1)} H_2^{\mathrm{b}}(y) \overleftarrow{d} M_j^{\mathrm{b}(\jmath)}(y) \Big\rangle^{\mathrm{b}} = \int_{[t\vee s,1)} H_1^{\mathrm{b}}(x) H_2^{\mathrm{b}}(x) \overleftarrow{d} \big\langle M_i^{\mathrm{b}(\imath)}, M_j^{\mathrm{b}(\jmath)} \big\rangle^{\mathrm{b}}(x),$$

where $t \vee s = \max(t, s)$.

The specific cases of Lemmas 1 and 2 correspond to fundamental martingale properties usually used in right- and left-censored-only data, respectively. In doubly censored data, as a natural request, it is necessary to consider both forward and backward aspects. However, note that even in right-censored data there is a backward aspect as a special case of Lemma 2 (although it may not be particularly useful if used separately). See Patilea and Rolin (2006) for a study related to this characterization using backward counting processes.

3.3 Correlation structure between forward and backward martingales

To obtain the covariance form with the NPMLE, we need to clarify the correlation structure between forward and backward martingales. For this purpose, consider the product $M_i^{\mathrm{f}(\imath)}(t) M_j^{\mathrm{b}(\jmath)}(s)$. If $i \neq j$, the product holds a martingale structure based on either the filtration $\mathscr{F}_t^{\mathrm{f}}$ or $\mathscr{F}_s^{\mathrm{b}}$ on all $s, t \in [0, 1]$, because of the independence between $M_i^{\mathrm{f}(\imath)}(t)$ and $M_j^{\mathrm{b}(\jmath)}(s)$, $\imath, \jmath = 1, 2, 3$. If $i = j$ and $\imath = \jmath$, for the same purpose, consider

$$M_i^{\mathrm{f}(\jmath)}(t) M_i^{\mathrm{b}(\jmath)}(r(t)) = \int_{(0,t]} M_i^{\mathrm{b}(\jmath)}(r(x_-)) \overrightarrow{d} M_i^{\mathrm{f}(\jmath)}(x) + \int_{[r(t),1)} M_i^{\mathrm{f}(\jmath)}(\overline{r}(x_+)) \overleftarrow{d} M_i^{\mathrm{b}(\jmath)}(x) + \mathbb{1}(r(t) \leq t) \int_{[r(t),t]} \overrightarrow{d} N_i^{\mathrm{f}(\jmath)}(x), \tag{10}$$

where $r(\cdot)$ is a strictly monotone decreasing function on $[0, 1]$ with $r(0) = 1$ and $r(1) = 0$, such as $r(t) = 1 - t$, which reverses the forward time-direction, and $\overline{r}(t)$ is the inverse function of $r(t)$. The derivation of (10) is based on integration by parts for the Stieltjes integration (see Fleming and Harrington 1991, pp. 74–75) and

$$\textstyle\sum_{0<s\leq t} \overrightarrow{d} M_i^{\mathrm{f}(\jmath)}(s) \overleftarrow{d} M_i^{\mathrm{b}(\jmath)}(r(s)) = \sum_{0<s\leq t} \overrightarrow{d} N_i^{\mathrm{f}(\jmath)}(s) \overleftarrow{d} N_i^{\mathrm{b}(\jmath)}(r(s)).$$

Put $s = r(t)$ in (10). The two martingales $M_i^{\mathrm{f}(\jmath)}(t)$ and $M_i^{\mathrm{b}(\jmath)}(s)$ are not always uncorrelated. If $s \leq t$, a correlation occurs certainly in the third term of (10). However, considering the first and second terms of (10) together, whether $s \leq t$ or not, the structure of (10) may appear to be complicated as long as we persist with either $\mathscr{F}_t^{\mathrm{f}}$ or $\mathscr{F}_s^{\mathrm{b}}$ alone. We will avoid the insistence along single time lines and begin by considering a superposition of the forward and backward filtrations.

Let $\mathscr{F}_t^{\natural} = \mathscr{F}_t^{\mathrm{f}} \vee \mathscr{F}_{r(t)}^{\mathrm{b}}$ be the smallest σ-algebra containing all events of $\mathscr{F}_t^{\mathrm{f}}$ and $\mathscr{F}_{r(t)}^{\mathrm{b}}$, and let $N_i^{\natural(\jmath)}(t) = N_i^{\mathrm{f}(\jmath)}(t) + N_i^{\mathrm{b}(\jmath)}(r(t))$, $\jmath = 1, 2, 3$. Using the inverse $\overline{r}(t)$ of $r(t)$, $N_i^{\natural(\jmath)}(\overline{r}(t)) = N_i^{\mathrm{f}(\jmath)}(\overline{r}(t)) + N_i^{\mathrm{b}(\jmath)}(t)$. We formulate the $\mathscr{F}_t^{\natural}$-martingale property of $N_i^{\natural(\jmath)}$ as follows:

Lemma 3 *Suppose that Condition 1 is satisfied. Let* $A_i^{\natural(\jmath)}(t) = A_i^{\natural\mathrm{f}(\jmath)}(t) + A_i^{\natural\mathrm{b}(\jmath)}(r(t))$ *and*

$$\begin{cases} A_i^{\natural\mathrm{f}(\jmath)}(t) = \int_{(0,t]} Y_i^{\mathrm{f}}(x) Y_i^{\mathrm{b}}(r(x)) \upsilon^{\mathrm{f}(\jmath)}(x) \overrightarrow{d} \Lambda^{\mathrm{f}(\jmath)}(x) + \delta_i^{\mathrm{f}(\jmath)}(t) \\ A_i^{\natural\mathrm{b}(\jmath)}(t) = \int_{[t,1)} Y_i^{\mathrm{f}}(\overline{r}(x)) Y_i^{\mathrm{b}}(x) \upsilon^{\mathrm{b}(\jmath)}(x) \overleftarrow{d} \Lambda^{\mathrm{b}(\jmath)}(x) + \delta_i^{\mathrm{b}(\jmath)}(t) \end{cases}$$

for $\jmath = 1, 2, 3$ *and* $i = 1, \ldots, n$*, where* $\upsilon^{\mathrm{f}(\jmath)}$ *and* $\upsilon^{\mathrm{b}(\jmath)}$ *are provided as*

$$\begin{aligned} \upsilon^{\mathrm{f}(1)}(t) &= \frac{(S^L(t) - S^R(t))S^*(t)}{Q^{\mathrm{f}}(t) - Q^{\mathrm{f}}(r(t)_+)}, & \upsilon^{\mathrm{b}(1)}(t) &= \frac{(F^R(t) - F^L(t))F^*(t)}{Q^{\mathrm{b}}(r(t)) - Q^{\mathrm{b}}(t_-)}, \\ \upsilon^{\mathrm{f}(2)}(t) &= \frac{S^*(t)S^R(t)}{Q^{\mathrm{f}}(t) - Q^{\mathrm{f}}(r(t)_+)}, & \upsilon^{\mathrm{b}(2)}(t) &= \frac{(1 - F^*(t))F^R(t)}{Q^{\mathrm{b}}(r(t)) - Q^{\mathrm{b}}(t_-)}, \\ \upsilon^{\mathrm{f}(3)}(t) &= \frac{(1 - S^*(t))S^L(t)}{Q^{\mathrm{f}}(t) - Q^{\mathrm{f}}(r(t)_+)}, & \upsilon^{\mathrm{b}(3)}(t) &= \frac{F^*(t)F^L(t)}{Q^{\mathrm{b}}(r(t)) - Q^{\mathrm{b}}(t_-)} \end{aligned}$$

on $t < r(t)$ *(and set as zeros on* $t \geq r(t)$*), and* $\delta_i^{\mathrm{f}(\jmath)}(t)$ *and* $\delta_i^{\mathrm{b}(\jmath)}(r(t))$ *are* $\mathscr{F}_t^{\natural}$*-predictable processes such that*

$$\begin{aligned} &\delta_i^{\mathrm{f}(\jmath)}(t) = \mathbb{1}(r(t) \leq t) \int_{[r(t),t]} \overrightarrow{d} N_i^{\mathrm{f}(\jmath)}(x) \text{ and} \\ &\delta_i^{\mathrm{b}(\jmath)}(r(t)) = \mathbb{1}(r(t) \leq t) \int_{[t,r(t)]} \overleftarrow{d} N_i^{\mathrm{b}(\jmath)}(x). \end{aligned}$$

Then, for $\jmath = 1, 2, 3$*,* $i = 1, \ldots, n$*, in the double-censoring model (1), the processes*

$$M_i^{\natural(\jmath)}(t) = N_i^{\natural(\jmath)}(t) - A_i^{\natural(\jmath)}(t) \quad \text{and} \quad \begin{cases} M_i^{\natural\mathrm{f}(\jmath)}(t) = N_i^{\mathrm{f}(\jmath)}(t) - A_i^{\natural\mathrm{f}(\jmath)}(t) \\ M_i^{\natural\mathrm{b}(\jmath)}(r(t)) = N_i^{\mathrm{b}(\jmath)}(r(t)) - A_i^{\natural\mathrm{b}(\jmath)}(r(t)) \end{cases}$$

are square-integrable $\mathscr{F}_t^{\natural}$*-martingales* $(0 \leq t)$*, where* $M_i^{\natural(\jmath)}(t) = M_i^{\natural\mathrm{f}(\jmath)}(t) + M_i^{\natural\mathrm{b}(\jmath)}(r(t))$*. Further, the* $\mathscr{F}_t^{\natural}$*-predictable covariation processes* $\langle M_i^{\natural(\imath)}, M_j^{\natural(\jmath)}\rangle^{\natural}$ *satisfy* $\langle M_i^{\natural(\imath)}, M_j^{\natural(\jmath)}\rangle^{\natural}(t) = 0$ *whenever at least one of three conditions* $i \neq j$*,* $\imath \neq \jmath$ *or* $t < r(t)$ *is satisfied.*

See Appendix A.2 for the proof of Lemma 3. Lemma 3 includes a useful tool to show that the expectations of the first and second terms of (10) are zeros. It will brought by adding the identity $M_i^{\xi(\jmath)}(t) + A_i^{\xi(\jmath)}(t) = M_i^{\natural\xi(\jmath)}(t) + A_i^{\natural\xi(\jmath)}(t)$ $(= N_i^{\xi(\jmath)}(t))$, $\xi = \mathrm{f}, \mathrm{b}$. Then, the result for the product of forward and backward martingales is obtained as follows:

Lemma 4 *Suppose that Condition 1 is satisfied. Then, in the double-censoring model (1),* $\mathrm{E}[M_i^{\natural\mathrm{f}(\imath)}(t) M_j^{\natural\mathrm{b}(\jmath)}(s)] = 0$ *and*

$$\mathrm{E}[M_i^{\mathrm{f}(\imath)}(t) M_j^{\mathrm{b}(\jmath)}(s)] = \mathbb{1}(s \leq t,\ i = j,\ \imath = \jmath) \mathrm{E}\Big[\int_{[s,t]} \overrightarrow{d} N_i^{\mathrm{f}(\jmath)}(x) \Big]$$

for $i, j = 1, \ldots, n$*,* $\imath, \jmath = 1, 2, 3$.

The proof of Lemma 4 is also provided in Appendix A.2. Lemma 4 is easily extended to a multivariate version with the martingale transformation. We define $\overline{N}^{\xi(\jmath)}(t) = \sum_{i=1}^{n} N_i^{\xi(\jmath)}(t)$,

$$\overline{A}^{\xi(\jmath)}(t) = \sum_{i=1}^{n} A_i^{\xi(\jmath)}(t) \text{ and } \overline{M}^{\xi(\jmath)}(t) = \sum_{i=1}^{n} M_i^{\xi(\jmath)}(t)$$

for $\xi = \mathrm{f}, \mathrm{b}$, $\jmath = 1, 2, 3$. Let $G^{\mathrm{f}(\jmath)}(t)$ and $G^{\mathrm{b}(\jmath)}(t)$ be, respectively, bounded $\mathscr{F}_t^{\mathrm{f}}$- and $\mathscr{F}_t^{\mathrm{b}}$-predicable processes, which are also $\mathscr{F}_t^{\natural}$-predictable because

$$G^{\mathrm{f}(\jmath)}(t) = \mathrm{E}[G^{\mathrm{f}(\jmath)}(t) \mid \mathscr{F}_{t_-}^{\mathrm{f}}] = \mathrm{E}[\mathrm{E}[G^{\mathrm{f}(\jmath)}(t) \mid \mathscr{F}_{r(t_-)}^{\mathrm{b}}] \mid \mathscr{F}_{t_-}^{\mathrm{f}}] = \mathrm{E}[G^{\mathrm{f}(\jmath)}(t) \mid \mathscr{F}_{t_-}^{\natural}].$$

Further, let $\overline{M}_G^{\mathrm{f}(\jmath)}(t) = \int_{(0,t]} G^{\mathrm{f}(\jmath)}(s) \overrightarrow{d}\overline{M}^{\mathrm{f}(\jmath)}(s)$ and $\overline{M}_G^{\mathrm{b}(\jmath)}(t) = \int_{[t,1)} G^{\mathrm{b}(\jmath)}(s) \overleftarrow{d}\overline{M}^{\mathrm{b}(\jmath)}(s)$, $\jmath = 1, 2, 3$. Then, by reasons similar to those for Lemma 4, these products satisfy

$$\begin{aligned}
\mathrm{E}[\overline{M}_G^{\mathrm{f}(\imath)}(t)\overline{M}_G^{\mathrm{b}(\jmath)}(s)] &= \mathbb{1}(s \le t,\ \imath = \jmath)\mathrm{E}\left[\int_{[s,t]} G^{\mathrm{f}(\jmath)}(x)G^{\mathrm{b}(\jmath)}(x)\overrightarrow{d}\overline{N}^{\mathrm{f}(\jmath)}(x)\right] \\
&= \mathbb{1}(s \le t,\ \imath = \jmath)\mathrm{E}\left[\int_{[s,t]} G^{\mathrm{f}(\jmath)}(x)G^{\mathrm{b}(\jmath)}(x)\overrightarrow{d}\overline{A}^{\mathrm{f}(\jmath)}(x)\right].
\end{aligned} \tag{11}$$

This result (11) will be used in Sect. 4.2.1.

4 Weak convergence based on a martingale approach

This section is organized in two parts (Sects. 4.1 and 4.2). In Sect. 4.1, we discuss a linearization of the NPMLE and conduct useful decompositions of the profile score function to apply the martingale technique developed in Sect. 3 to the NPMLE, without any discussions of asymptotic approximation. In Sect. 4.2, as our main result, we discuss the asymptotic distribution of the NPMLE based on the linearization developed in Sect. 4.1 and martingale properties provided in Sect. 3.

4.1 Linearization of the NPMLE

We here formulate a linearization convenient for investigating statistical properties of the NPMLE. For this purpose, it is essential to clarify a structure immanent in the profile score function, which is summarized in Propositions 2 and 3. Based on Proposition 4, the martingale properties on the profile score function are summarized in Lemma 5. This yields a main result (15) for the linearization of the NPMLE. Proposition 3 is obtained by reconstructing the intensity components of Lemma 5 in terms of Proposition 2, which is used in the proof of Theorem 1 (in particular, Lemma 7).

Introducing the notation $\theta_t^* = F^*(t)$ into the argument of the profile estimator to avoid double use of parentheses, similar to $\theta_t = F(t)$, the first-order Taylor expansion of $U_n(\theta_t^*)$ around the NPMLE $\theta_t = \widehat{F}(t)$ is written as

$$U_n(\theta_t^*) = \mathcal{I}_n(\widetilde{\theta}_t)(\widehat{F}(t) - F^*(t)), \tag{12}$$

where $\widetilde{\theta}_t$ is on the line segment between $\widehat{F}(t)$ and $F^*(t)$. This expansion is a foundation of our discussion. To derive the martingale properties of $\widehat{F}(t)$, we will investigate several structures which $U_n(\theta_t^*)$ and $\mathcal{I}_n(\widetilde{\theta}_t)$ possess. Thereafter, in our discussion based on (12), the parametrization of F can be limited to the case of $\boldsymbol{J} = \widehat{\boldsymbol{J}}_{\theta_t^*}$.

Denote $\widehat{\boldsymbol{J}}_{\theta_t^*} = (\widehat{J}_1, \ldots, \widehat{J}_{[n]})$ and let $\overrightarrow{\boldsymbol{J}}_j$ and $\overleftarrow{\boldsymbol{J}}_j$ be the intervals $[\widehat{J}_j, \widehat{J}_{j+1})$ and $(\widehat{J}_{j-1}, \widehat{J}_j]$, respectively. Let $F^{d*}(\cdot)$ and $\boldsymbol{F}^* = (F_1^*, \ldots, F_{[n]}^*)$ be a discretized step function of $F^*(\cdot)$ and its vector expression, respectively, that is, F^{d*} satisfies $F^{d*}(s) = F_j^*$ for $s \in \overrightarrow{\boldsymbol{J}}_j$, $j = 1, \ldots, [n]$, where $F_j^* = F^*(\widehat{J}_j)$, $j = 1, \ldots, [n]$. Note that $F^{d*}(\cdot)$ is often preferred to $F^*(\cdot)$ in investigating the structure of $U_n(\theta_t^*)$ in (12), since $F^{d*}(s) = \theta_t^*$ always holds on $s \in [\widehat{J}_{m_t}, \widehat{J}_{m_t+1})$. See Appendix A.3.1 for another definition and viewpoint of $F^{d*}(\cdot)$ and $\boldsymbol{F}^*$ needed in Sect. 4.2.2. We set $F_0^* = 0$ and $F_{[n]+1}^* = 1$ for some discussions including $\boldsymbol{F}^*$.

On expressions of the score function. To discuss the structure of $U_n(\theta_t^*)$ in (12), from Proposition 1, recall that

$$U_n(\theta_t^*) = u_{m_t}(\widehat{\boldsymbol{F}}(\theta_t^*)) = \frac{\alpha_{m_t}}{\theta_t^* - \widehat{F}_{m_t-1}(\theta_t^*)} - \frac{\alpha_{m_t+1}}{\widehat{F}_{m_t+1}(\theta_t^*) - \theta_t^*} - \frac{\beta_{m_t}}{1-\theta_t^*} + \frac{\gamma_{m_t}}{\theta_t^*}.$$

However, this expression is not convenient for obtaining a martingale structure of $U_n(\theta_t^*)$, and we should find some useful alternative expression. A key to obtain such a finding is to supplement a gap between the form of $\mathcal{I}_n(\theta_t)$ (see Proposition 1) and a conjecture $\mathrm{E}[U_n(\theta_t^*)^2] \approx \mathrm{E}[\mathcal{I}_n(\theta_t^*)]$ conveyed from the information identity. This work is first realized in the following result:

Proposition 2 *In doubly censored data* (1), $U_n(\theta_t^*)$ *is expressed as*

$$U_n(\theta_t^*) = \textstyle\sum_{j=1}^{[n]-1} u_j(\boldsymbol{F}^*)\widetilde{F}_{\theta_t^*}(\widehat{J}_j), \tag{13}$$

where $u_j(\boldsymbol{F}^*)$ *means* $u_j(F_{j-1}^*, F_j^*, F_{j+1}^*)$, *and the definition of* $\widetilde{F}_{\theta_t^*}(\cdot)$ *is provided in* (24) *of Appendix A.4. The counting process expression of* (13) *is*

$$U_n(\theta_t^*) = \textstyle\sum_{\jmath=1}^{3} \mathsf{s}_\jmath \int_0^1 \widetilde{H}_t^{(\jmath)}(s)\, \vec{d}\,\overline{N}^{\mathrm{f}(\jmath)}(s), \tag{14}$$

where $\mathsf{s}_1 = 1$, $\mathsf{s}_2 = -1$, $\mathsf{s}_3 = 1$, $\widetilde{H}_t^{(\jmath)}(s)$, $\jmath = 1, 2, 3$ *are*

$$\widetilde{H}_t^{(1)}(s) = \frac{\vec{d}\,\widetilde{F}_{\theta_t^*}(s)}{\vec{d}\,F^{d*}(s)}, \quad \widetilde{H}_t^{(2)}(s) = \frac{\widetilde{F}_{\theta_t^*}(s)}{S^{d*}(s)} \quad \text{and} \quad \widetilde{H}_t^{(3)}(s) = \frac{\widetilde{F}_{\theta_t^*}(s)}{F^{d*}(s)}$$

and $S^{d*}(\cdot) = 1 - F^{d*}(\cdot)$.

See Appendix A.4 for the proof of Proposition 2.

Remark on Proposition 2 We observe $\vec{d}\widetilde{F}_{\theta_t^*}(s) = 0$ if $\vec{d}F^{d*}(s) = 0$, because of $\vec{d}F^{d*}(s) = F^{d*}(s) - F^{d*}(s_-)$ and $\vec{d}\widetilde{F}_{\theta_t^*}(s) = \widetilde{F}_{\theta_t^*}(s) - \widetilde{F}_{\theta_t^*}(s_-)$. So, although there may be some manners to define $\widetilde{H}_t^{(1)}(s)$ on all $s \in [0, 1]$, we adopt the rule that $\widetilde{H}_t^{(1)}(s) = \widetilde{H}_t^{(1)}(\widehat{J}_{m_s})$ if $\vec{d}F^{d*}(s) = 0$.

The form of $\widetilde{F}_{\theta_t^*}(\cdot)$ defined in (24) may seem to be complicated. However, the readers can go on to read the following contents even without fully understanding the structure of $\widetilde{F}_{\theta_t^*}(\cdot)$. That is, a knowledge that $\widetilde{F}_{\theta_t^*}(s)$ has a structure similar to $\widehat{F}_{\theta_t^*}(s; \theta_t^*)$ will be sufficient to read hereafter, where $\widehat{F}_{\theta_t}(s; \theta_t)$ is the function expression for the derivatives $\partial\widehat{F}_1(\theta_t)/\partial\theta_t, \ldots, \partial\widehat{F}_{[n]-1}(\theta_t)/\partial\theta_t$ of the profile estimator $\widehat{\boldsymbol{F}}(\theta_t)$ discussed in Sect. 2.2. See Appendix 5 for detailed definition of $\widehat{F}_{\theta_t}(s; \theta_t)$.

From (14), we have a fundamental expression of $U_n(\theta_t^*)$ for a martingale approach as follows:

Lemma 5 *In doubly censored data* (1), $U_n(\theta_t^*)$ *is decomposed into the processes such that*

$$U_n(\theta_t^*) = U_n^M(t; \widetilde{H}_t) + U_n^A(t; \widetilde{H}_t) \quad \text{and} \left\{ \begin{array}{l} U_n^M(t; \widetilde{H}_t) = U_n^{M\mathrm{f}}(t; \widetilde{H}_t) + U_n^{M\mathrm{b}}(t; \widetilde{H}_t) \\ U_n^A(t; \widetilde{H}_t) = U_n^{A\mathrm{f}}(t; \widetilde{H}_t) + U_n^{A\mathrm{b}}(t; \widetilde{H}_t), \end{array} \right.$$

where the four components are

$$\begin{aligned} U_n^{M\mathrm{f}}(t; \widetilde{H}_t) &= \textstyle\sum_{J=1}^3 \mathsf{s}_J \int_{(0,t]} \widetilde{H}_t^{(J)}(s)\, \vec{d}\,\overline{M}^{\mathrm{f}(J)}(s), \\ U_n^{M\mathrm{b}}(t; \widetilde{H}_t) &= \textstyle\sum_{J=1}^3 \mathsf{s}_J \int_{(t,1)} \widetilde{H}_t^{(J)}(s)\, \overleftarrow{d}\,\overline{M}^{\mathrm{b}(J)}(s), \\ U_n^{A\mathrm{f}}(t; \widetilde{H}_t) &= \textstyle\sum_{J=1}^3 \mathsf{s}_J \int_{(0,t]} \widetilde{H}_t^{(J)}(s)\, \vec{d}\,\overline{A}^{\mathrm{f}(J)}(s) \\ \text{and}\ \ U_n^{A\mathrm{b}}(t; \widetilde{H}_t) &= \textstyle\sum_{J=1}^3 \mathsf{s}_J \int_{(t,1)} \widetilde{H}_t^{(J)}(s)\, \overleftarrow{d}\,\overline{A}^{\mathrm{b}(J)}(s). \end{aligned}$$

Lemma 5 is shown easily by (14) and the Doob–Meyer decomposition based on Lemmas 1 and 2. By Lemma 5, (12) can be re-expressed as

$$n^{-1/2}U_n^M(t; \widetilde{H}_t) + n^{-1/2}U_n^A(t; \widetilde{H}_t) = n^{-1}\mathcal{I}_n(\widetilde{\theta}_t)\sqrt{n}(\widehat{F}(t) - F^*(t)), \qquad (15)$$

which provides a useful viewpoint to study properties of the NPMLE.

The asymptotic properties of the NPMLE and $n^{-1/2}U_n^M(t; \widetilde{H}_t)$ will be discussed using (15) and Lemma 5. However, the expression of Lemma 5 is slightly difficult in order to show that $n^{-1/2}U_n^A(t; \widetilde{H}_t)$ converges in probability to zero. We therefore prepare another expression of $U_n^A(t; \widetilde{H}_t)$ obtained from Proposition 2 following an idea from Lemma 5.

Proposition 3 *In doubly censored data* (1), $U_n^A(t; \widetilde{H}_t)$ *is expressed as*

$$U_n^A(t; \widetilde{H}_t) = \sum_{j=1}^{m_t-1} u_j^{A\mathrm{f}}(\boldsymbol{F}^*)\widetilde{F}_{\theta_t^*}(\widehat{J}_j) + u_{m_t}^{A\mathrm{fb}}(\boldsymbol{F}^*) + \sum_{j=m_t+1}^{[n]-1} u_j^{A\mathrm{b}}(\boldsymbol{F}^*)\widetilde{F}_{\theta_t^*}(\widehat{J}_j),$$

where u_j^{Af} *and* $u_{m_t}^{\mathrm{Afb}}$ *are written as*

$$
\begin{aligned}
u_j^{\mathrm{Af}}(\boldsymbol{F}^*) &= \textstyle\int_{\overleftarrow{\boldsymbol{J}}_j} \overrightarrow{d}\overline{A}^{\mathrm{f}(1)}(s)/p_j^* - \int_{\overleftarrow{\boldsymbol{J}}_{j+1}} \overrightarrow{d}\overline{A}^{\mathrm{f}(1)}(s)/p_{j+1}^* \\
&\quad - \textstyle\int_{\overrightarrow{\boldsymbol{J}}_j} \overrightarrow{d}\overline{A}^{\mathrm{f}(2)}(s)/S_j^* + \int_{\overrightarrow{\boldsymbol{J}}_j} \overrightarrow{d}\overline{A}^{\mathrm{f}(3)}(s)/F_j^*, \\
u_{m_t}^{\mathrm{Afb}}(\boldsymbol{F}^*) &= \textstyle\int_{\overleftarrow{\boldsymbol{J}}_{m_t}} \overrightarrow{d}\overline{A}^{\mathrm{f}(1)}(s)/p_j^* - \int_{\overleftarrow{\boldsymbol{J}}_{m_t+1}} \overleftarrow{d}\overline{A}^{\mathrm{b}(1)}(s)/p_{j+1}^* \\
&\quad - \textstyle\int_{\overrightarrow{\boldsymbol{J}}_{m_t}} \overrightarrow{d}\overline{A}^{\mathrm{f}(2)}(s)/S_j^* + \int_{\overrightarrow{\boldsymbol{J}}_{m_t}} \overrightarrow{d}\overline{A}^{\mathrm{f}(3)}(s)/F_j^*,
\end{aligned}
$$

$p_j^* = F_j^* - F_{j-1}^*$, $S_j^* = 1 - F_j^*$, *and* u_j^{Ab} *is* u_j^{Af} *in which the notations* f *and* $\overrightarrow{d}$ *are replaced by* b *and* $\overleftarrow{d}$, *respectively.*

Proposition 3 is proved in Appendix A.4.

4.2 Distribution convergence results

First, we prepare the standard conditions required for study of the asymptotic distribution of the NPMLE as follows:

Condition 2 F^*, F^L *and* F^R *satisfy Condition* 1 *with*

$$\inf\{t : 0 < F^*(t)\} = 0,\ \sup\{t : F^*(t) < 1\} = 1,\ F^L(1) = 1 \text{ and } F^R(0) = 0.$$

$F^L(t) - F^R(t)$ *is positive on* $t \in (0, 1)$.

This condition covers that of Chang and Yang (1987) and is equivalent to that of Murphy and van der Vaart (1997, Theorem 2.1), except for the form of the support [0, 1] specified explicitly for simplicity. Condition 2 is standard in the theoretical study for doubly censored data, while there are some works under weaker or more practical conditions, such as Gu and Zhang (1993) and Yu and Li (2001). Condition 2 is assumed throughout this section and Appendix A.5.

We now discuss how the limit distribution of the left side of (15) converges weakly to a distribution using a martingale approach. We show that $n^{-1/2}U_n^M(t; \widetilde{H}_t)$ converges weakly to a Gaussian process in Sect. 4.2.1, and $n^{-1/2}U_n^A(t; \widetilde{H}_t)$ converges in probability to zero in Sect. 4.2.2. Finally, the asymptotic distribution of the NPMLE can be derived as Theorem 1. For the limit forms included in Theorem 1, let us prepare the notations

$$
\begin{aligned}
&\mathscr{I}_{(t,s)}(a,b) = \int_{[a,b]} H_t^{*(1)}(x) H_s^{*(1)}(x) \overrightarrow{d}A^{*(1)}(x) \\
&+ \int_{[a,b)} H_t^{*(2)}(x) H_s^{*(2)}(x) \overrightarrow{d}A^{*(2)}(x) + \int_{(a,b]} H_t^{*(3)}(x) H_s^{*(3)}(x) \overrightarrow{d}A^{*(3)}(x),
\end{aligned}
$$

$\vec{d}A^{*(\jmath)}(s) = Q^{\mathrm{f}}(s)w^{\mathrm{f}(\jmath)}(s)\vec{d}\Lambda^{\mathrm{f}(\jmath)}(s),\ \jmath = 1, 2, 3,$

$$H_t^{*(1)}(s) = \frac{\vec{d}F_{\theta_t^*}^*(s)}{\vec{d}F^*(s)},\quad H_t^{*(2)}(s) = \frac{F_{\theta_t^*}^*(s)}{S^*(s)},\quad H_t^{*(3)}(s) = \frac{F_{\theta_t^*}^*(s)}{F^*(s)},$$

and $F_{\theta_t^*}^*(\cdot)$ of the true derivative function is a limit of $\widehat{F}_{\theta_t^*}(\cdot;\theta_t^*)$. See Appendix A.5 for further details of $F_{\theta_t^*}^*(\cdot)$.

Theorem 1 *Suppose that Condition 2 is satisfied and that* τ_0 *and* τ_1 *are arbitrary values such that* $0 < \tau_0 \le \tau_1 < 1$. *Then, as* $n \to \infty$, $\sqrt{n}(\widehat{F}(t) - F^*(t))$ *converges weakly to the Gaussian process* $\{\mathbb{G}_t^{\mathrm{f}}(t) + \mathbb{G}_t^{\mathrm{b}}(t_+)\}/\mathscr{I}_{(t,t)}(0,1)$ *on* $t \in [\tau_0, \tau_e]$ *with zero mean and covariance function of*

$$\mathrm{Cov}\left(\frac{\mathbb{G}_t^{\mathrm{f}}(t) + \mathbb{G}_t^{\mathrm{b}}(t_+)}{\mathscr{I}_{(t,t)}(0,1)}, \frac{\mathbb{G}_s^{\mathrm{f}}(s) + \mathbb{G}_s^{\mathrm{b}}(s_+)}{\mathscr{I}_{(s,s)}(0,1)}\right) = \frac{\mathscr{I}_{(t,s)}(0,1)}{\mathscr{I}_{(t,t)}(0,1)\mathscr{I}_{(s,s)}(0,1)},$$

where $\mathbb{G}_t^{\mathrm{f}}(\cdot)$ *and* $\mathbb{G}_t^{\mathrm{b}}(\cdot)$ *are Gaussian forward and backward martingale processes (indexed by a given* t*) with zero means, respectively, such that*

$$\mathrm{Cov}(\mathbb{G}_t^{\mathrm{f}}(s), \mathbb{G}_t^{\mathrm{f}}(u)) = \mathscr{I}_{(t,t)}(0, s \wedge u),\quad \mathrm{Cov}(\mathbb{G}_t^{\mathrm{b}}(s), \mathbb{G}_t^{\mathrm{b}}(u)) = \mathscr{I}_{(t,t)}(s \vee u, 1)$$
$$\text{and}\ \ \mathrm{Cov}(\mathbb{G}_t^{\mathrm{f}}(t), \mathbb{G}_s^{\mathrm{b}}(s_+)) = \mathbb{1}(s < t)\mathscr{I}_{(t,s)}(s_+, t).$$

The proof of Theorem 1 is performed following Sects. 4.2.1 and 4.2.2 and is summarized in Sect. 4.2.3. From the form of the asymptotic covariance function $\mathscr{I}_{(t,s)}(0,1)/\mathscr{I}_{(t,t)}(0,1)\mathscr{I}_{(s,s)}(0,1)$, we can know that a natural estimate (empirical version) of $\mathrm{Cov}(\widehat{F}(t), \widehat{F}(s))$ is $\mathcal{J}_n(\widehat{F}_{m_t}, \widehat{F}_{m_s})/\mathcal{I}_n(\widehat{F}_{m_t})\mathcal{I}_n(\widehat{F}_{m_s})$, which gives the same value as Turnbull's formula (see Sugimoto 2012a, Theorem 1, Lemma 1), where

$$\begin{aligned}\mathcal{J}_n(\widehat{F}_{m_t}, \widehat{F}_{m_s}) = &\sum_{j=1}^{[n]} \frac{\alpha_j}{(\widehat{F}_j - \widehat{F}_{j-1})^2}\left(\frac{\partial \widehat{F}_j}{\partial \widehat{F}_{m_t}} - \frac{\partial \widehat{F}_{j-1}}{\partial \widehat{F}_{m_t}}\right)\left(\frac{\partial \widehat{F}_j}{\partial \widehat{F}_{m_s}} - \frac{\partial \widehat{F}_{j-1}}{\partial \widehat{F}_{m_s}}\right)\\ &+ \sum_{j=1}^{[n]-1} \frac{\beta_j}{(1-\widehat{F}_j)^2}\left(\frac{\partial \widehat{F}_j}{\partial \widehat{F}_{m_t}}\right)\left(\frac{\partial \widehat{F}_j}{\partial \widehat{F}_{m_s}}\right) + \sum_{j=1}^{[n]-1} \frac{\gamma_j}{\widehat{F}_j^2}\left(\frac{\partial \widehat{F}_j}{\partial \widehat{F}_{m_t}}\right)\left(\frac{\partial \widehat{F}_j}{\partial \widehat{F}_{m_s}}\right).\end{aligned}$$

In right-censored data, $\mathcal{J}_n(\widehat{F}_{m_t}, \widehat{F}_{m_s})/\mathcal{I}_n(\widehat{F}_{m_t})\mathcal{I}_n(\widehat{F}_{m_s})$ reduces to a Greenwood-type covariance formula

$$(1-\widehat{F}_{m_t})(1-\widehat{F}_{m_s})\int_0^{t\wedge s} \frac{\vec{d}\overline{N}^{\mathrm{f}(1)}(u)}{\sum_{i=1}^n Y_i^{\mathrm{f}}(u)\{\sum_{i=1}^n Y_i^{\mathrm{f}}(u) - \vec{d}\overline{N}^{\mathrm{f}(1)}(u)\}}$$

(see Sugimoto 2012a, Corollary 1).

4.2.1 Asymptotic distribution of martingale components

Lemma 6 *Suppose the conditions of Theorem 1 are satisfied. Then,* $n^{-1/2}U_n^M(t;\widetilde{H}_t)$ *converges weakly to the Gaussian process* $\mathbb{G}_t^{\mathrm{f}}(t) + \mathbb{G}_t^{\mathrm{b}}(t_+)$ *on* $t \in [\tau_0, \tau_e]$ *with zero* *mean and covariance function such that*

$$\mathrm{Cov}\left(\mathbb{G}_t^{\mathrm{f}}(t) + \mathbb{G}_t^{\mathrm{b}}(t_+), \mathbb{G}_s^{\mathrm{f}}(s) + \mathbb{G}_s^{\mathrm{b}}(s_+)\right) = \mathscr{I}_{(t,s)}(0,1).$$

Proof of Lemma 6 Given $t \in [\tau_0, \tau_1]$, $H_t^{*(\jmath)}(s)$, $\jmath = 1, 2, 3$ are bounded uniformly on $s \in [0, 1]$. Because $\widetilde{F}_{\theta_t^*}(s)$ included in $U_n^{M\xi}(t; \widetilde{H}_t)$ is $\mathscr{F}_t^{\xi}$-predictable but not $\mathscr{F}_s^{\xi}$-predictable, $\xi = \mathrm{f}, \mathrm{b}$, we consider the processes of $U_n^{M\xi}(s; \widetilde{H}_t)$ weighted by $H_t^{*(\jmath)}/\widetilde{H}_t^{(\jmath)}$ such that

$$\begin{aligned}\overline{U}_{n,t}^{M\mathrm{f}}(s) &= \textstyle\sum_{\jmath=1}^{3} \mathsf{s}_\jmath \int_{(0,s]} H_t^{*(\jmath)}(u) \overrightarrow{d}\overline{M}^{\mathrm{f}(\jmath)}(u),\\ \overline{U}_{n,t}^{M\mathrm{b}}(s) &= \textstyle\sum_{\jmath=1}^{3} \mathsf{s}_\jmath \int_{(s,1)} H_t^{*(\jmath)}(u) \overleftarrow{d}\overline{M}^{\mathrm{b}(\jmath)}(u).\end{aligned}$$

Since $H_t^{*(\jmath)}(u)$ is $\mathscr{F}_u^{\xi}$-predictable, $\overline{U}_{n,t}^{M\xi}(s)$'s are $\mathscr{F}_s^{\xi}$-martingale processes by Lemmas 1 and 2. Referring to Sects. 3.1 and 3.2 for $\mathscr{F}_s^{\xi}$-predictable variation processes, we obtain

$$\begin{aligned}\left\langle n^{-1/2}\overline{U}_{n,t}^{M\mathrm{f}}, n^{-1/2}\overline{U}_{n,t}^{M\mathrm{f}}\right\rangle^{\mathrm{f}}(s) &= \textstyle\sum_{\jmath=1}^{3} \mathsf{s}_\jmath \int_{(0,s]} H_t^{*(\jmath)}(u)^2 n^{-1} \overrightarrow{d}\overline{A}^{\mathrm{f}(\jmath)}(u)\\ &\to_p \textstyle\sum_{\jmath=1}^{3} \mathsf{s}_\jmath \int_{(0,s]} H_t^{*(\jmath)}(u)^2 \overrightarrow{d}A^{*(\jmath)}(u) \ \text{ as } \ n \to \infty\end{aligned}$$

and a similar result for $\langle n^{-1/2}\overline{U}_{n,t}^{M\mathrm{b}}, n^{-1/2}\overline{U}_{n,t}^{M\mathrm{b}}\rangle^{\mathrm{b}}$, where $\to_p$ denotes the convergence in probability. In addition, $n^{-1/2}\overline{U}_{n,t}^{M\xi}, \xi = \mathrm{f}, \mathrm{b}$ satisfy the Lindeberg condition because $H_t^{*(\jmath)}(s)$ is bounded uniformly on $s \in [0, 1]$ under $t \in [\tau_0, \tau_1]$. By Rebolledo's central limit theorem (Andersen and Gill 1982), as $n \to \infty$, we show that

$$n^{-1/2}\overline{U}_{n,t}^{M\mathrm{f}}(s) \to_D \mathbb{G}_t^{\mathrm{f}}(s) \quad \text{and} \quad n^{-1/2}\overline{U}_{n,t}^{M\mathrm{b}}(s) \to_D \mathbb{G}_t^{\mathrm{b}}(s_+),$$

where $\to_D$ denotes the convergence in distribution. Hence, applying Slutsky's lemma for the above results and (29), we conclude that $n^{-1/2}U_n^M(t; \widetilde{H}_t)$ and $n^{-1/2}\{\overline{U}_{n,t}^{M\mathrm{f}}(t) + \overline{U}_{n,t}^{M\mathrm{b}}(t)\}$ have the same limit distribution. That is, we have

$$n^{-1/2}\{U_n^{M\mathrm{f}}(t; \widetilde{H}_t) + U_n^{M\mathrm{b}}(t; \widetilde{H}_t)\} \to_D \mathbb{G}_t^{\mathrm{f}}(t) + \mathbb{G}_t^{\mathrm{b}}(t_+) \ \text{ on } \ t \in [\tau_0, \tau_e] \tag{16}$$

for the uncorrelated relation between the two quantities. This limit distribution $\mathbb{G}_t^{\mathrm{f}}(t) + \mathbb{G}_t^{\mathrm{b}}(t_+)$ is a Gaussian process again due to the sum of Gaussian. Therefore, the weak convergence result in this lemma is shown.

Next, we discuss the result on covariance. The covariance function of the limit distribution in (16) consists of

$$\begin{aligned}&\mathrm{Cov}(\mathbb{G}_t^{\mathrm{f}}(t), \mathbb{G}_s^{\mathrm{f}}(s)) + \mathrm{Cov}(\mathbb{G}_t^{\mathrm{f}}(t), \mathbb{G}_s^{\mathrm{b}}(s_+))\\ &\qquad + \mathrm{Cov}(\mathbb{G}_t^{\mathrm{b}}(t_+), \mathbb{G}_s^{\mathrm{f}}(s)) + \mathrm{Cov}(\mathbb{G}_t^{\mathrm{b}}(t_+), \mathbb{G}_s^{\mathrm{b}}(s_+)).\end{aligned}$$

By properties of martingale processes, we have $\mathrm{Cov}(\mathbb{G}_t^{\mathrm{f}}(t), \mathbb{G}_s^{\mathrm{f}}(s)) = \mathscr{I}_{(t,s)}(0, s \wedge t)$ and $\mathrm{Cov}(\mathbb{G}_t^{\mathrm{b}}(t_+), \mathbb{G}_s^{\mathrm{b}}(s_+)) = \mathscr{I}_{(t,s)}(s_+ \vee t_+, 1)$. To investigate $\mathrm{Cov}(\mathbb{G}_t^{\mathrm{f}}(t), \mathbb{G}_s^{\mathrm{b}}(s_+))$, we use (11). Then, since we can write $\overline{U}_{n,t}^{M\mathrm{f}}(t) = \sum_{J=1}^{3} \overline{M}_{G_t}^{\mathrm{f}(J)}(t)$ and $\overline{U}_{n,s}^{M\mathrm{b}}(s) = \sum_{J=1}^{3} \overline{M}_{G_s}^{\mathrm{b}(J)}(s_+)$ putting $G_t^{\mathrm{f}(J)}(x) = H_t^{*(J)}(x)$ and $G_s^{\mathrm{b}(J)}(x) = H_s^{*(J)}(x)$, it is obtained that

$$
\begin{aligned}
&\mathrm{E}[n^{-1/2}\overline{U}_{n,t}^{M\mathrm{f}}(t)n^{-1/2}\overline{U}_{n,s}^{M\mathrm{b}}(s)] \\
&\quad = \mathbb{1}(s < t)\sum_{J=1}^{3} \mathrm{E}\left[\int_{(s,t]} G_t^{\mathrm{f}(J)}(u)G_s^{\mathrm{b}(J)}(u)n^{-1}\vec{d}\overline{A}^{\mathrm{f}(J)}(u)\right] \\
&\quad = \mathbb{1}(s < t)\sum_{J=1}^{3} \int_{(s,t]} H_t^{*(J)}(u)H_s^{*(J)}(u)\,\vec{d}A^{*(J)}(u),
\end{aligned}
$$

which is identical to $\mathrm{Cov}(\mathbb{G}_t^{\mathrm{f}}(t), \mathbb{G}_s^{\mathrm{b}}(s_+)) = \mathbb{1}(s < t)\mathscr{I}_{(t,s)}(s_+, t)$. Therefore, the asymptotic covariance of (16) can be gathered together as

$$
\mathscr{I}_{(t,s)}(0, s \wedge t) + \mathscr{I}_{(t,s)}(s_+ \wedge t_+, s \vee t) + \mathscr{I}_{(t,s)}(s_+ \vee t_+, 1) = \mathscr{I}_{(t,s)}(0, 1).
$$

□

4.2.2 Convergence of intensity components

Lemma 7 *Suppose the conditions of Theorem 1 are satisfied. Then,* $\sup_{t\in[\tau_0,\tau_e]} |n^{-1/2}U_n^A(t; \widetilde{H}_t)|$ *converges in probability to zero.*

Proof of Lemma 7 Let $U_\star^A(t; \widetilde{H}_t) = \sum_{j=1}^{[n]-1} u_j^*(\boldsymbol{F}^*)\widetilde{F}_{\theta_t^*}(\widehat{J}_j)$, where the definition of $u_j^*(\boldsymbol{F}^*)$ is given in Appendix A.3.1. Note the equality $U_\star^A(t; \widetilde{H}_t) = 0$ holds since $u_j^*(\boldsymbol{F}^*) = 0$ for all j's from (19). To consider $U_n^A(t; \widetilde{H}_t) - U_\star^A(t; \widetilde{H}_t)$, we use the expression of $U_n^A(t; \widetilde{H}_t)$ in Proposition 3. The differences $n^{-1}(u_j^{A\xi}(\boldsymbol{F}^*) - u_j^*(\boldsymbol{F}^*))$, $\xi = \mathrm{f}, \mathrm{fb}, \mathrm{b}$ are written as

$$
\begin{aligned}
&\int_{\overleftarrow{\boldsymbol{J}}_j} M_Q^{\mathrm{f}}(s)\,\vec{d}A^{\mathrm{f}(1)}(s)/p_j^* - \int_{\overleftarrow{\boldsymbol{J}}_{j+1}} M_Q^{\mathrm{f}}(s)\,\vec{d}A^{\mathrm{f}(1)}(s)/p_{j+1}^* \\
&\quad - \int_{\vec{\boldsymbol{J}}_j} M_Q^{\mathrm{f}}(s)\,\vec{d}A^{\mathrm{f}(2)}(s)/S_j^* + \int_{\vec{\boldsymbol{J}}_j} M_Q^{\mathrm{f}}(s)\,\vec{d}A^{\mathrm{f}(3)}(s)/F_j^*
\end{aligned}
$$

in the case of $u_j^{A\mathrm{f}}(\boldsymbol{F}^*)$, where $M_Q^{\xi}(s) = (\widehat{Q}^{\xi}(s) - Q^{\xi}(s))/Q^{\xi}(s)$, $\xi = \mathrm{f,b}$ and $\widehat{Q}^{\mathrm{f}}(s) = n^{-1}\overline{Y}^{\mathrm{f}}(s)$ and $\widehat{Q}^{\mathrm{b}}(s) = n^{-1}\overline{Y}^{\mathrm{b}}(s)$ are the empirical estimates of $Q^{\mathrm{f}}(s)$ and $Q^{\mathrm{b}}(s)$. Then, by applying the mean-value theorem in each of $u_j^{A\xi}(\boldsymbol{F}^*) - u_j^*(\boldsymbol{F}^*)$, $\xi = \mathrm{f}, \mathrm{fb}, \mathrm{b}$, there are some $M_{Q\,j}^{\xi(1)} \in (\inf_{\overleftarrow{\boldsymbol{J}}_j} M_Q^{\xi}(s), \sup_{\overleftarrow{\boldsymbol{J}}_j} M_Q^{\xi}(s))$ and $M_{Q\,j}^{\xi(2)}, M_{Q\,j}^{\xi(3)} \in (\inf_{\vec{\boldsymbol{J}}_j} M_Q^{\xi}(s), \sup_{\vec{\boldsymbol{J}}_j} M_Q^{\xi}(s))$, so that we have

$$u_j^{A\xi}(\boldsymbol{F}^*) = nM_{Qj}^{\xi(1)}\Big[n^{-1}\alpha_j^*/p_j^* - \{M_{Q\,j+1}^{\xi(1)}/M_{Qj}^{\xi(1)}\}n^{-1}\alpha_{j+1}^*/p_{j+1}^*$$
$$-\{M_{Qj}^{\xi(2)}/M_{Qj}^{\xi(1)}\}n^{-1}\beta_j^*/S_j^* - \{M_{Qj}^{\xi(3)}/M_{Qj}^{\xi(1)}\}n^{-1}\gamma_j^*/F_j^*\Big],\ \xi = \mathrm{f}, \mathrm{b}$$

and the similar expression about $u_{m_t}^{A\mathrm{fb}}(\boldsymbol{F}^*)$, where the notations of α_j^*, β_j^* and γ_j^* are given in Appendix A.3.1. By the Glivenko–Cantelli theorem and the continuity of the distributions, $M_Q^{\xi}(s_{j+1}^{\xi(1)})/M_Q^{\xi}(s_j^{\xi(1)})$ and $M_Q^{\xi}(s_j^{\xi(2)})/M_Q^{\xi}(s_j^{\xi(1)})$ converge almost surely to ones uniformly. Similarly, by adding the relation $\widehat{Q}^{\mathrm{b}}(t) - Q^{\mathrm{b}}(t) = -\{\widehat{Q}^{\mathrm{f}}(t_+) - Q^{\mathrm{f}}(t_+)\}$ further, $M_Q^{\mathrm{b}}(s_{m_t+1}^{\mathrm{b}(1)})/M_Q^{\mathrm{f}}(s_{m_t}^{\mathrm{f}(1)})$ converges almost surely to one uniformly. Also, we have

$$n^{-1}\alpha_j^*/p_j^* = \int_{\overleftarrow{\boldsymbol{J}}_j}(F^L(t) - F^R(t))\overrightarrow{d}F^*(t)/p_j^* \approx F^L(\widehat{J}_j) - F^R(\widehat{J}_j),$$
$$n^{-1}\beta_j^*/S_j^* = \int_{\overrightarrow{\boldsymbol{J}}_j} S^*(t)\overrightarrow{d}F^R(t)/S_j^* \approx F^R(\widehat{J}_{j+1}) - F^R(\widehat{J}_j),$$
$$n^{-1}\gamma_j^*/F_j^* = \int_{\overrightarrow{\boldsymbol{J}}_j} F^*(t)\overrightarrow{d}F^L(t)/F_j^* \approx F^L(\widehat{J}_{j+1}) - F^L(\widehat{J}_j)$$

approximately, which become exact with probability 1 for a sufficiently large n using Proposition 4. Thus, we find $u_j^{A\xi}(\boldsymbol{F}^*)/nM_{Qj}^{\xi(1)} \le o_P(1)$ uniformly on all j's for $\xi = \mathrm{f}, \mathrm{b}$. Because $\widetilde{F}_{\theta_t^*}(\widehat{J}_j) \to_P F_{\theta_t^*}^*(\widehat{J}_j) \le 1$ uniformly for every j from (29), as $n \to \infty$, we can conclude

$$\sup_t |n^{-1/2}U_n^A(t; \widetilde{H}_t)| \le \sup_t n^{-1/2}\left\{\sum_{j=1}^{m_t-1}|M_{Qj}^{\mathrm{f}(1)}| + \sum_{j=m_t}^{[n]-1}|M_{Qj}^{\mathrm{b}(1)}|\right\} o_P(1)$$
$$\le n^{-1}\left[\sum_{j=1}^{[n]-1}\left\{\sqrt{n}|M_{Qj}^{\mathrm{f}(1)}| + \sqrt{n}|M_{Qj}^{\mathrm{b}(1)}|\right\}\right] o_P(1) \to_P 0.$$

In fact, for either $\xi = \mathrm{f}$ or b, because $M_Q^{\xi}(\cdot)$ is a martingale process such that

$$M_Q^{\mathrm{f}}(t) = -\int_{(0,t]} n^{-1}\overrightarrow{d}\overline{M}^{\mathrm{f}(0)}(s)/Q^{\mathrm{f}}(s) \ \text{ or } \ M_Q^{\mathrm{b}}(t) = -\int_{[t,1)} n^{-1}\overleftarrow{d}\overline{M}^{\mathrm{b}(0)}(s)/Q^{\mathrm{b}}(s),$$

with $\overline{M}^{\xi(0)}(s) = \sum_{i=1}^n \sum_{J=1}^3 M_i^{\xi(J)}(s)$ using the Duhamel equation, $\sqrt{n}M_Q^{\xi}(s)$ converges weakly to a Gaussian martingale process by the martingale central limit theorem or the Donsker theorem. The limits of $\sqrt{n}M_Q^{\mathrm{f}}(s)$ for $s \in [0, t]$ and $\sqrt{n}M_Q^{\mathrm{b}}(s)$ for $s \in [t, 1]$ are Gaussian processes with some finite variances on $t \in [\tau_0, \tau_1]$, so that we can show that $n^{-1}\sum_j \sqrt{n}|M_{Qj}^{\xi(1)}|$ converges in probability to a bounded quantity, as $n \to \infty$. □

4.2.3 Summary and further considerations

Proof of Theorem 1 This is completed by applying the results of Lemmas 6 and 7 and (30) obtained under $t \in [\tau_0, \tau_1]$ to (15). □

The result of Theorem 1 is extended to [0, 1] if

$$\lim_{\tau_1 \to 1}(1 - F^R(\tau_{1-}))/\vec{d}F^*(\tau_1) = \infty \quad \text{and} \quad \lim_{\tau_0 \to 0} F^L(\tau_0)/\vec{d}F^*(\tau_{0+}) = \infty \tag{17}$$

are described definitely in Condition 2. In fact, $\mathscr{I}_{(t,t)}(0, 1)$ holds the uniform continuity on $t \in [\tau_0, \tau_1]$ and we observe $1/\mathscr{I}_{(t,t)}(0, 1) \to 0$ as $t = \tau_0 \to 0_+$ or $t = \tau_e \to 1_-$ under (17) by reasoning similar to Sugimoto (2011, Theorem 2), so that this means that $\sqrt{n}(\widehat{F}(t) - F^*(t))$ converges weakly to zero near the extreme points of t. Although (17) may resemble $\inf\{t : 0 < F^L(t)\} < \inf\{t : 0 < F^*(t)\}$ and $\sup\{t : F^*(t) < 1\} < \sup\{t : F^R(t) < 1\}$, this expression is avoided for the support [0, 1] of Condition 1. A finer alternative expression will be obtained by analogy of the behaviour of the transformed Kaplan–Meier process (Gill 1983).

Based on Theorem 1 we can construct the methods for overall tests and simultaneous confidence intervals at several times. In addition, the covariance function of the NPMLE can be easily estimated by $\mathcal{J}_n(\widehat{F}_{m_t}, \widehat{F}_{m_s})/\mathcal{I}_n(\widehat{F}_{m_t})\mathcal{I}_n(\widehat{F}_{m_s})$. It is useful as a substitution for that based on the profile likelihood ratio (Chen and Zhou 2003) in a larger sample, because this method is computationally fast. In particular, Sugimoto (2012a, Theorem 1, Lemma 1) developed an inverse formula of general tridiagonal matrix based on this discussion, which provides one expansion of this study in mathematical sciences.

5 Discussion

In doubly censored data, we formulated the martingale characterizations for the forward and backward counting processes and their correlation structure. Unlike right- or left-censoring only, the intensity processes explicitly include the true distribution functions (F^*, F^L, F^R). In doubly or interval censored data, generally it is difficult to provide the asymptotic distribution of the NPMLE or the SCEs without the infinite matrix or operator expression if we only incorporate martingale properties into the self-consistent equations. In this paper, using an expansion of the score function of the semiparametric profile likelihood, we derived a structure of the NPMLE focusing on doubly censored data. The score function was then expressed as an approximate form of the efficient score using the derivatives between the profile estimators. Based on such an expression of the score function, we showed that the NPMLE possesses a structure of a superposition of the forward and backward martingales with a bias in intensities which converges in probability to zero. Thus, while avoiding the infinite matrix or operator expression, we demonstrated that the asymptotic distribution of the NPMLE is a Gaussian process via martingale properties. This is summarized as Theorem 1, which forms the basis for overall tests and simultaneous confidence intervals at several times. The correlation structure between forward and backward martingales was used to obtain the asymptotic covariance function of the NPMLE. In future work, we are interested in determining how the martingale approach presented here can be extended, for example, under more practical conditions than Condition 2, to some regression models or to interval-censored data.

Appendix A

A.1 On derivatives of profile likelihood

Proof of Proposition 1 If the differential rule only is used, $U_n(\theta_t)$ and $\mathcal{I}_n(\theta_t)$ may be

$$U_n(\theta_t) = \sum_{j=1}^{[n]} u_j(\widehat{\boldsymbol{F}}(\theta_t)) \frac{\partial \widehat{F}_j(\theta_t)}{\partial \theta_t} \quad \text{and}$$

$$\mathcal{I}_n(\theta_t) = \sum_{j=1}^{[n]} \sum_{l=1}^{[n]} i_{j,l}(\widehat{\boldsymbol{F}}(\theta_t)) \frac{\partial \widehat{F}_j(\theta_t)}{\partial \theta_t} \frac{\partial \widehat{F}_l(\theta_t)}{\partial \theta_t} - \sum_{j=1}^{[n]} u_j(\widehat{\boldsymbol{F}}(\theta_t)) \frac{\partial^2 \widehat{F}_j(\theta_t)}{\partial \theta_t^2},$$

where the derivative notation is suitable for d rather than ∂, but the latter is used to avoid a confusion with the infinitesimal on time t, such as $dF(t)$. These forms can be reduced further following the condition (5): that is, the score function satisfies $U_n(\theta_t) = u_{m_t}(\widehat{\boldsymbol{F}}(\theta_t))$. Also, the Fisher function $\mathcal{I}_n(\theta_t)$ is obtained as (6) using (5), $\partial \widehat{F}_{m_t}(\theta_t)/\partial \theta_t = 1$ and $\partial^2 \widehat{F}_{m_t}(\theta_t)/\partial \theta_t^2 = 0$ because of $\widehat{F}_{m_t}(\theta_t) = \theta_t$. $\square$

A.2 Proofs for fundamental martingale properties

Proof of Lemma 1 Unlike right-censored-only data, the information on $Y_i^{\mathrm{f}}(t)$ does not lead a martingale property so directly for $N_i^{\mathrm{f}(1)}(t)$. We therefore set an ideal at-risk process, which we cannot always observe completely, $Y_i^{\mathrm{f}*}(t) = \mathbb{1}(C_i^L < t \leq \min(T_i^*, C_i^R))$, so that we can see

$$\begin{aligned} \mathrm{E}[\vec{d}N_i^{\mathrm{f}(1)}(t) | Y_i^{\mathrm{f}*}(t) = 1] & \\ &= \Pr(t \leq T_i^* < t + dt, C_i^L < T_i^* \leq C_i^R | C_i^L < t \leq \min(T_i^*, C_i^R)) \\ &= \vec{d}\Lambda^{\mathrm{f}(1)}(t). \end{aligned}$$

To obtain the $\mathscr{F}_t^{\mathrm{f}}$-intensity of $\vec{d}N_i^{\mathrm{f}(1)}(t)$, using $Y_i^{\mathrm{f}*}(t)$, we can lead the relationship

$$\begin{aligned} \mathrm{E}[\vec{d}N_i^{\mathrm{f}(1)}(t) | \mathscr{F}_{t-}^{\mathrm{f}}] &= Y_i^{\mathrm{f}}(t) \Pr(\vec{d}N_i^{\mathrm{f}(1)}(t) = 1, Y_i^{\mathrm{f}*}(t) = 1 | \mathscr{F}_{t-}^{\mathrm{f}}) \\ &= Y_i^{\mathrm{f}}(t) \Pr(Y_i^{\mathrm{f}*}(t) = 1 | \mathscr{F}_{t-}^{\mathrm{f}}) \mathrm{E}[\vec{d}N_i^{\mathrm{f}(1)}(t) | \mathscr{F}_{t-}^{\mathrm{f}}, Y_i^{\mathrm{f}*}(t) = 1]. \end{aligned}$$

Given $Y_i^{\mathrm{f}}(t) = 1$, the information on the ith individual extracted from $\mathscr{F}_{t-}^{\mathrm{f}}$ is equivalent to $\{Y_i^{\mathrm{f}}(t) = 1\}$. Therefore, given $Y_i^{\mathrm{f}}(t) = 1$, we can easily obtain $\mathrm{E}[\vec{d}N_i^{\mathrm{f}(1)}(t) | \mathscr{F}_{t-}^{\mathrm{f}}, Y_i^{\mathrm{f}*}(t) = 1] = \vec{d}\Lambda^{\mathrm{f}(1)}(t)$ because of $\{Y_i^{\mathrm{f}*}(t) = 1\} \subseteq \{Y_i^{\mathrm{f}}(t) = 1\}$.

Second, given $Y_i^{\mathrm{f}}(t) = 1$, we can obtain

$$\Pr(Y_i^{\mathrm{f}*}(t) = 1 | \mathscr{F}_{t-}^{\mathrm{f}}) = \frac{\Pr(C_i^L < t \le \min(T_i^*, C_i^R))}{\Pr(Y_i^{\mathrm{f}}(t) = 1)} = \frac{\Pr(C_i^L < t \le C_i^R) S^*(t_-)}{\Pr(Y_i^{\mathrm{f}}(t) = 1)},$$

where $\Pr(C_i^L < t \le C_i^R)$ is obtained as $1 - \Pr(t \le C_i^L) - \Pr(t > C_i^R) = S^R(t_-) - S^L(t_-)$. The event patterns included in $\{Y_i^{\mathrm{f}}(t) = 1\}$ consist of $\{t \le T_i^*, C_i^L < T_i^* \le C_i^R\}$, $\{t \le C_i^R, C_i^R < T_i^*\}$ or $\{t \le C_i^L, T_i^* \le C_i^L\}$. For example, the probability of the second event is $-(1 - F^*(s_-))F^R(s_-) + \int_{[t,1]} F^R(s_-) \vec{d} F^*(s)$ using the Stieltjes integration by the right continuity of $F^*(t)$ and $F^R(t)$. By summing up the three probabilities, we have

$$Q^{\mathrm{f}}(t) = \Pr(Y_i^{\mathrm{f}}(t) = 1) = S^R(t_-)S^*(t_-) + S^L(t_-)(1 - S^*(t_-)).$$

Therefore, these provide $\mathrm{E}[\vec{d} N_i^{\mathrm{f}(1)}(t) | \mathscr{F}_{t-}^{\mathrm{f}}] = Y_i^{\mathrm{f}}(t) w_0^{\mathrm{f}(1)}(t) \vec{d} \Lambda^{\mathrm{f}(1)}(t) = \vec{d} A_i^{\mathrm{f}(1)}(t)$ and then $M_i^{\mathrm{f}(1)}(t)$ is an $\mathscr{F}_t^{\mathrm{f}}$-martingale and square-integrable, because it is clear that $N_i^{\mathrm{f}(1)}$ is adapted to $\{\mathscr{F}_t^{\mathrm{f}} : t \ge 0\}$ and $\mathrm{E}|N_i^{\mathrm{f}(1)}(t)| \le 1 < \infty$ and $\mathrm{E}[A_i^{\mathrm{f}(1)}(t)] \le 1$ for all t. Results for $M_i^{\mathrm{f}(2)}(t)$ and $M_i^{\mathrm{f}(3)}(t)$ are similar to $M_i^{\mathrm{f}(1)}(t)$.

For the latter part of this lemma, note that $\vec{d} N_i^{\mathrm{f}(2)}(t)$ and $\vec{d} N_i^{\mathrm{f}(3)}(t)$ never jump at the same time t because of (1) even if C_i^L and C_i^R are not mutually independent. Hence, $\langle M_i^{\mathrm{f}(2)}, M_i^{\mathrm{f}(3)} \rangle^{\mathrm{f}}(t) = 0$ is shown, for example, using Fleming and Harrington (1991, Lemma 2.6.1). The other martingales $M_i^{\mathrm{f}(\iota)}$ and $M_j^{\mathrm{f}(\jmath)}$ are mutually independent unless $i = j$ and $\iota = \jmath$. So, it is shown as usual that $\langle M_i^{\mathrm{f}(\iota)}, M_j^{\mathrm{f}(\jmath)} \rangle^{\mathrm{f}}(t) = \mathbb{1}(i = j, \iota = \jmath) A_i^{\mathrm{f}(\jmath)}(t)$. □

Proof of Lemma 2 This lemma is the time-reversed result of Lemma 1 and can be shown by approaches similar to Lemma 1. □

Proof of Lemma 3 By the definitions $\mathscr{F}_t^{\natural} = \mathscr{F}_t^{\mathrm{f}} \vee \mathscr{F}_{r(t)}^{\mathrm{b}}$ and $N_i^{\natural(\jmath)}(t) = N_i^{\mathrm{f}(\jmath)}(t) + N_i^{\mathrm{b}(\jmath)}(r(t))$, if $\{N_i^{\mathrm{f}(\jmath)}(t) = 1\} \subset \mathscr{F}_t^{\mathrm{f}}$, $N_i^{\mathrm{b}(\jmath)}(r(t))$ is $\mathscr{F}_t^{\natural}$-predictable and $N_i^{\natural(\jmath)}(t)$ too. In fact, if we already observe $\vec{d} N_i^{\mathrm{f}(\jmath)}(x) = 1$ at a time x before a current time t $(x < t)$, we can know that the present $N_i^{\mathrm{b}(\jmath)}(r(t))$ will jump surely at the future time x $(< r(t) < r(x))$. Inversely, if $\{N_i^{\mathrm{b}(\jmath)}(r(t)) = 1\} \subset \mathscr{F}_{r(t)}^{\mathrm{b}}$, $N_i^{\mathrm{f}(\jmath)}(t)$ and $N_i^{\natural(\jmath)}(t)$ are $\mathscr{F}_t^{\natural}$-predictable. For these reasons, after $Y_i^{\mathrm{b}}(r(t)) = 0$ or $Y_i^{\mathrm{f}}(t) = 0$ occurs (under $r(t) \le t$), $\delta_i^{\mathrm{f}(\jmath)}(t)$ and $\delta_i^{\mathrm{b}(\jmath)}(r(t))$ are added in $\mathscr{F}_t^{\natural}$-predictable components of $N_i^{\mathrm{f}(\jmath)}(t)$ and $N_i^{\mathrm{b}(\jmath)}(r(t))$, respectively. Hence, $N_i^{\natural(\jmath)}(t)$ is at risk with the intensity controlled by the conditional probability when $Y_i^{\mathrm{f}}(t) Y_i^{\mathrm{b}}(r(t)) = 1$. So, we have $\mathrm{E}[\vec{d} N_i^{\natural(\jmath)}(t) | \mathscr{F}_{t-}^{\natural}] = \mathrm{E}[\vec{d} N_i^{\mathrm{f}(\jmath)}(t) | \mathscr{F}_{t-}^{\natural}] + \mathrm{E}[\overleftarrow{d} N_i^{\mathrm{b}(\jmath)}(r(t)) | \mathscr{F}_{t-}^{\natural}]$ and their components are expressed as

$$\mathrm{E}[\vec{d} N_i^{\mathrm{f}(\jmath)}(t) | \mathscr{F}_{t-}^{\natural}] = Y_i^{\mathrm{f}}(t) Y_i^{\mathrm{b}}(r(t)) \mathrm{E}[\vec{d} N_i^{\mathrm{f}(\jmath)}(t) | Y_i^{\mathrm{f}}(t) Y_i^{\mathrm{b}}(r(t)) = 1] + \vec{d} \delta_i^{\mathrm{f}(\jmath)}(t),$$
$$\mathrm{E}[\vec{d} N_i^{\mathrm{b}(\jmath)}(r(t)) | \mathscr{F}_{t-}^{\natural}] = Y_i^{\mathrm{f}}(t) Y_i^{\mathrm{b}}(r(t)) \mathrm{E}[\vec{d} N_i^{\natural(\jmath)}(r(t)) | Y_i^{\mathrm{f}}(t) Y_i^{\mathrm{b}}(r(t)) = 1] + \overleftarrow{d} \delta_i^{\mathrm{b}(\jmath)}(r(t)).$$

Here note that $N_i^{\natural(\jmath)}(t) = \int_{(0,t]} \vec{d}N_i^{\mathrm{f}(\jmath)}(x) + \int_{[r(t),1)} \overleftarrow{d}N_i^{\mathrm{b}(\jmath)}(x)$ and $\overleftarrow{d}N_i^{\mathrm{b}(\jmath)}(r(t)) = N_i^{\mathrm{b}(\jmath)}(r(t)) - N_i^{\mathrm{b}(\jmath)}(r(t_-))$ with $N_i^{\mathrm{b}(\jmath)}(r(t_-)) = N_i^{\mathrm{b}(\jmath)}(r(t)_+)$. Then, because

$$\begin{aligned} &\mathrm{E}[\vec{d}N_i^{\mathrm{f}(\jmath)}(t)|Y_i^{\mathrm{f}}(t)Y_i^{\mathrm{b}}(r(t)) = 1] = \Pr(\vec{d}N_i^{\mathrm{f}(\jmath)}(t) = 1|Y_i^{\mathrm{f}}(t)Y_i^{\mathrm{b}}(r(t)) = 1) \\ \text{and }\ &\mathrm{E}[\overleftarrow{d}N_i^{\mathrm{b}(\jmath)}(t)|Y_i^{\mathrm{f}}(t)Y_i^{\mathrm{b}}(r(t)) = 1] = \Pr(\overleftarrow{d}N_i^{\mathrm{b}(\jmath)}(r(t)) = 1|Y_i^{\mathrm{f}}(t)Y_i^{\mathrm{b}}(r(t)) = 1), \end{aligned}$$

the above conditional expectations are $v^{\mathrm{f}(\jmath)}(t)\vec{d}\Lambda^{\mathrm{f}(\jmath)}(t)$ and $v^{\mathrm{b}(\jmath)}(r(t))\vec{d}\Lambda^{\mathrm{b}(\jmath)}(r(t))$, respectively. In fact, the denominator is $\Pr(Y_i^{\mathrm{f}}(t)Y_i^{\mathrm{b}}(r(t)) = 1) = Q^{\mathrm{f}}(t) - Q^{\mathrm{f}}(r(t)_+) = Q^{\mathrm{b}}(r(t)) - Q^{\mathrm{b}}(t_-)$ because of $Y_i^{\mathrm{f}}(t)Y_i^{\mathrm{b}}(r(t)) = \mathbb{1}(t \le T_i \le r(t))$, while the numerators are obtained similarly to the derivation in Lemma 1. Therefore, $M_i^{\natural\mathrm{f}(\jmath)}(t)$, $M_i^{\natural\mathrm{b}(\jmath)}(r(t))$ and $M_i^{\natural(\jmath)}(t)$ are $\mathscr{F}_t^{\natural}$-martingales, $\jmath = 1, 2, 3$, $\xi = \mathrm{f}, \mathrm{b}$ and they are square-integrable because $\mathrm{E}[A_i^{\natural\xi(\jmath)}(t)] = \mathrm{E}[A_i^{\xi(\jmath)}(t)] \le 1$ for all t. □

Proof of Lemma 4 If $i \ne j$, we have $\mathrm{E}[M_i^{\natural\mathrm{f}(\iota)}(t)M_j^{\natural\mathrm{b}(\jmath)}(s)] = 0$ and $\mathrm{E}[M_i^{\mathrm{f}(\iota)}(t)M_j^{\mathrm{b}(\jmath)}(s)] = 0$, as already described in Sect. 3.3. Here we first show $\mathrm{E}[M_i^{\natural\mathrm{f}(\jmath)}(t)M_i^{\natural\mathrm{b}(\jmath)}(r(t))] = 0$. Note that $M_i^{\natural\mathrm{f}(\jmath)}(t)$ and $M_i^{\natural\mathrm{b}(\jmath)}(r(t))$, $\jmath = 1, 2, 3$ are right-continuous with $M_i^{\natural\mathrm{f}(\jmath)}(0) = M_i^{\natural\mathrm{b}(\jmath)}(r(0)) = 0$, and $\sum_{0<s\le t} \vec{d}M_i^{\natural\mathrm{f}(\jmath)}(s)\overleftarrow{d}M_i^{\natural\mathrm{b}(\jmath)}(r(s)) = 0$. Similarly to the derivation of (10), we have

$$\begin{aligned} M_i^{\natural\mathrm{f}(\jmath)}(t)M_i^{\natural\mathrm{b}(\jmath)}(r(t)) = &\int_{(0,t]} M_i^{\natural\mathrm{b}(\jmath)}(r(x_-))\vec{d}M_i^{\natural\mathrm{f}(\jmath)}(x) \\ &+ \int_{[r(t),1)} M_i^{\natural\mathrm{f}(\jmath)}(\bar{r}(x_+))\overleftarrow{d}M_i^{\natural\mathrm{b}(\jmath)}(x). \end{aligned}$$

The right-side of the above equation is obviously an $\mathscr{F}_t^{\natural}$-martingale by Lemma 3, so that the expectation is zero. Next, we show $\mathrm{E}[M_i^{\mathrm{f}(\jmath)}(t)M_i^{\mathrm{b}(\jmath)}(r(t))] = \mathbb{1}(r(t) \le t)\mathrm{E}[\int_{[r(t),t]} \vec{d}N_i^{\mathrm{f}(\jmath)}(x)]$. Using the relationship $\overline{M}^{\xi(\jmath)}(t) = \overline{M}^{\natural\xi(\jmath)}(t) + (\overline{A}^{\natural\xi(\jmath)}(t) - \overline{A}^{\xi(\jmath)}(t))$, the first and second terms of (10) are

$$\begin{aligned} &\int_{(0,t]} M_i^{\mathrm{b}(\jmath)}(r(x_-))\vec{d}M_i^{\natural\mathrm{f}(\jmath)}(x) + \int_{[r(t),1)} M_i^{\mathrm{f}(\jmath)}(\bar{r}(x_+))\overleftarrow{d}M_i^{\natural\mathrm{b}(\jmath)}(x) \\ &\quad + \int_{(0,t]} M_i^{\mathrm{b}(\jmath)}(r(x_-))\vec{d}(A_i^{\natural\mathrm{f}(\jmath)}(x) - A_i^{\mathrm{f}(\jmath)}(x)) \\ &\quad + \int_{[r(t),1)} M_i^{\mathrm{f}(\jmath)}(\bar{r}(x_+))\overleftarrow{d}(A_i^{\natural\mathrm{b}(\jmath)}(x) - A_i^{\mathrm{b}(\jmath)}(x)). \end{aligned} \tag{18}$$

Since $M_i^{\mathrm{f}(\jmath)}(x_-)$ and $M_i^{\mathrm{b}(\jmath)}(r(x_-))$ are $\mathscr{F}_x^{\natural}$-predictable because $\mathscr{F}_x^{\mathrm{f}}, \mathscr{F}_{r(x)}^{\mathrm{b}} \subset \mathscr{F}_x^{\natural}$, the first and second terms of (18) are $\mathscr{F}_t^{\natural}$-martingales. For the third term of (18), we observe

$$\mathrm{E}[M_i^{\mathrm{b}(J)}(r(x_-))\vec{d}A_i^{\natural\mathrm{f}(J)}(x)] = \mathrm{E}\left[M_i^{\mathrm{b}(J)}(r(x_-))\mathrm{E}[\vec{d}N_i^{\mathrm{f}(J)}(x)|\mathscr{F}_{x_-}^{\natural}]\right]$$

$$\begin{aligned}\mathrm{E}[M_i^{\mathrm{b}(J)}(r(x_-))\vec{d}A_i^{\mathrm{f}(J)}(x)] &= \mathrm{E}[M_i^{\mathrm{b}(J)}(r(x_-))\mathrm{E}[\vec{d}N_i^{\mathrm{f}(J)}(x)|\mathscr{F}_{x_-}^{\mathrm{f}}]] \\ &= \mathrm{E}\left[\mathrm{E}\left[M_i^{\mathrm{b}(J)}(r(x_-))\mathrm{E}[\vec{d}N_i^{\mathrm{f}(J)}(x)|\mathscr{F}_{x_-}^{\mathrm{f}}]|\mathscr{F}_{r(x_-)}^{\mathrm{b}}\right]\right] \\ &= \mathrm{E}\left[M_i^{\mathrm{b}(J)}(r(x_-))\mathrm{E}[\vec{d}N_i^{\mathrm{f}(J)}(x)|\mathscr{F}_{x_-}^{\natural}]\right],\end{aligned}$$

so that we have $\int_{(0,t]}\mathrm{E}[M_i^{\mathrm{b}(J)}(r(x_-))\vec{d}(A_i^{\natural\mathrm{f}(J)}(x) - A_i^{\mathrm{f}(J)}(x))] = 0$. Similar relation and result are also observed about the fourth term of (18). Thus, putting $s = r(t)$, we have

$$\mathrm{E}[M_i^{\mathrm{f}(J)}(t)M_i^{\mathrm{b}(J)}(s)] = \mathbb{1}(s \le t)\mathrm{E}\left[\int_{[s,t]}\vec{d}N_i^{\mathrm{f}(J)}(x)\right].$$

All results for this lemma are shown. □

A.3 Preliminary for the linearization of the NPMLE

A.3.1 Discrete-true function

We provide another definition and viewpoint of the step function F^{d*} and its vector expression $\boldsymbol{F}^*$ introduced in Sect. 4.1. In particular, the equation (19) shown below is used in Sect. 4.2.2.

Note that $\vec{d}A^{*(J)}(s)$, $J = 1, 2, 3$ defined in Sect. 4.2 are used for limit expressions, because $\vec{d}A^{*(J)}(s)$ is equivalent to the quantity such as $\overleftarrow{d}A^{*(J)}(s) = Q^{\mathrm{b}}(s)w^{\mathrm{b}(J)}(s)\overleftarrow{d}\Lambda^{\mathrm{b}(J)}(s)$ under Condition 1 (continuous model). By Lemmas 1 and 2, we can set the "pseudo-true" data as

$$\alpha_j^* = \int_{\overleftarrow{\boldsymbol{J}}_j} n\vec{d}A^{*(1)}(s), \quad \beta_j^* = \int_{\vec{\boldsymbol{J}}_j} n\vec{d}A^{*(2)}(s) \quad \text{and} \quad \gamma_j^* = \int_{\vec{\boldsymbol{J}}_j} n\vec{d}A^{*(3)}(s),$$

$j = 0, 1, \ldots, [n] + 1$ with rules that $\widehat{J}_{[n]+1} = 1$ and γ_1^* and $\beta_{[n]}^*$ are substituted by $\gamma_1^* + \gamma_0^*$ and $\beta_{[n]}^* + \beta_{[n]+1}^*$ for convenience. This results in $\sum_{j=1}^{[n]+1}\alpha_j^* + \sum_{j=0}^{[n]}\beta_j^* + \sum_{j=1}^{[n]+1}\gamma_j^* = n$. Let $l_n^*(\boldsymbol{F})$ be $l_n(\boldsymbol{F})$ in which α_j, β_j and γ_j are replaced by α_j^*, β_j^* and γ_j^* in the case of $\boldsymbol{J} = \widehat{\boldsymbol{J}}_{\theta_t^*}$. We can then define $\boldsymbol{F}^*$ and equivalently $F^{d*}(\cdot)$ as the NPMLE for such pseudo-true data, that is, $\boldsymbol{F}^* = \mathrm{argmax}_{\{F_1,\ldots,F_{[n]}\}} l_n^*(\boldsymbol{F})$. So, letting $u_j^*(\boldsymbol{F}) = \partial l_n^*(\boldsymbol{F})/\partial F_j$, then $\boldsymbol{F}^*$ satisfies

$$\begin{aligned}u_j^*(\boldsymbol{F}^*) &= \int_{\overleftarrow{\boldsymbol{J}}_j} n\vec{d}A^{*(1)}(s)/p_j^* - \int_{\overleftarrow{\boldsymbol{J}}_{j+1}} n\vec{d}A^{*(1)}(s)/p_{j+1}^* \\ &\quad - \int_{\vec{\boldsymbol{J}}_j} n\vec{d}A^{*(2)}(s)/S_j^* + \int_{\vec{\boldsymbol{J}}_j} n\vec{d}A^{*(3)}(s)/F_j^* = 0 \qquad (19)\end{aligned}$$

for $j = 1, \ldots, [n]$. Further, by a linear transformation such as Sugimoto (2011, Lemma 2), F^{d*} $(= 1 - S^{d*})$ satisfies the self-consistent equations

$$F^{d*}(t) = \sum_{j=1}^{3} A^{*(j)}(t) - \int_{(0,t]} \frac{S^{d*}(t)}{S^{d*}(s)} \vec{d} A^{*(2)}(s) + \int_{(t,1)} \frac{F^{d*}(t)}{F^{d*}(s)} \vec{d} A^{*(3)}(s)$$

for the discrete times $t = 0, \widehat{J}_1, \ldots, \widehat{J}_{[n]+1}$. The above equations extended to all t are identical to the self-consistent equations which the true function $F^*(t)$ should satisfy.

A.3.2 Function expressions for the profile estimators and their derivatives.

As introduced in Sect. 4.1, denote $\widehat{F}(s; \theta_t) = \widehat{F}_j(\theta_t)$ and $\widehat{F}_{\theta_t}(s; \theta_t) = \partial \widehat{F}_j(\theta_t)/\partial\theta_t$ if $s \in [\widehat{J}_j, \widehat{J}_{j+1})$ as function expressions of the profile estimators and their derivatives given in Sect. 2.2. Then, $\widehat{F}_{\theta_t}(s; \theta_t)$ of (9) has a structure of the product integral and is written as

$$\widehat{F}_{\theta_t}(s; \theta_t) = \begin{cases} \pi_{s \le u < t} \left(1 - \vec{d}\widehat{K}(u; \theta_t)\right) \text{ if } \ s \le t \\ \pi_{t < u \le s} \left(1 - \overleftarrow{d}\widehat{K}(u; \theta_t)\right) \text{ if } \ s > t \end{cases},$$

where

$$\vec{d}\widehat{K}(s; \theta_t) = 1 - \frac{\partial \widehat{F}(s; \theta_t)}{\partial \widehat{F}(s_+; \theta_t)} = \frac{\vec{d}\widehat{F}_{\theta_t}(s_+; \theta_t)}{\widehat{F}_{\theta_t}(s_+; \theta_t)}$$

$$\text{and} \quad \overleftarrow{d}\widehat{K}(s; \theta_t) = 1 - \frac{\partial \widehat{F}(s; \theta_t)}{\partial \widehat{F}(s_-; \theta_t)} = \frac{\overleftarrow{d}\widehat{F}_{\theta_t}(s_-; \theta_t)}{\widehat{F}_{\theta_t}(s_-; \theta_t)}.$$

Further, following (7) and (8), $d\widehat{K}(s; \theta_t)$ satisfies

$$\begin{cases} 1 - \vec{d}\widehat{K}(s; \theta_t) = \dfrac{\alpha_{j+1}/\widehat{p}_{j+1}(\theta_t)^2}{\vec{d}\widehat{K}(\widehat{J}_{j-1};\theta_t)\frac{\alpha_j}{\widehat{p}_j(\theta_t)^2} + \frac{\alpha_{j+1}}{\widehat{p}_{j+1}(\theta_t)^2} + \frac{\beta_j}{\widehat{S}_j(\theta_t)^2} + \frac{\gamma_j}{\widehat{F}_j(\theta_t)^2}} \\ 1 - \overleftarrow{d}\widehat{K}(s; \theta_t) = \dfrac{\alpha_j/\widehat{p}_j(\theta_t)^2}{\overleftarrow{d}\widehat{K}(\widehat{J}_{j+1};\theta_t)\frac{\alpha_{j+1}}{\widehat{p}_{j+1}(\theta_t)^2} + \frac{\alpha_j}{\widehat{p}_j(\theta_t)^2} + \frac{\beta_j}{\widehat{S}_j(\theta_t)^2} + \frac{\gamma_j}{\widehat{F}_j(\theta_t)^2}} \end{cases}$$

at $s = \widehat{J}_j$ and is zero if $s \ne \widehat{J}_j$, where $\vec{d}\widehat{K}(\widehat{J}_0; \theta_t) = 1$ and $\overleftarrow{d}\widehat{K}(\widehat{J}_{[n]}; \theta_t) = 1$, $\widehat{S}_j(\theta_t) = 1 - \widehat{F}_j(\theta_t)$ and $\widehat{p}_j(\theta_t) = \widehat{F}_j(\theta_t) - \widehat{F}_{j-1}(\theta_t)$. This type of expression of $\widehat{F}_{\theta_t}(s; \theta_t)$ is helpful to understand the contents as discussed in Sects. 4.1 and 4.2 and Appendix A.4.

A.4 Proofs for expressions of the score function

Proof of Proposition 2 To derive the equation (13) of Proposition 2, we consider two types of Taylor expansions. The first of the expansions is

$$\frac{\alpha_j}{F_j^* - \widehat{F}_{j-1}(F_j^*)} = \frac{\alpha_j}{F_j^* - F_{j-1}^*} - i_{j-1,j}(\widetilde{F}_{j-1}^{*(0)}, F_j^*)\left(\widehat{F}_{j-1}(F_j^*) - F_{j-1}^*\right), \quad (20)$$

$$\frac{\alpha_{j+1}}{\widehat{F}_{j+1}^*(F_j^*) - F_j^*} = \frac{\alpha_{j+1}}{F_{j+1}^* - F_j^*} + i_{j+1,j}(F_j^*, \widetilde{F}_{j+1}^{*(0)})\left(\widehat{F}_{j+1}(F_j^*) - F_{j+1}^*\right), \quad (21)$$

where $\widetilde{F}_{j-1}^{*(0)}$ and $\widetilde{F}_{j+1}^{*(0)}$ are inner points on the line segments between F_{j-1}^* and $\widehat{F}_{j-1}(F_j^*)$ and F_{j+1}^* and $\widehat{F}_{j+1}(F_j^*)$, respectively. The second of the expansions is based on the Taylor approximations of

$$u_{j-1}(\widehat{F}_{j-2}(F_j^*), \widehat{F}_{j-1}(F_j^*), F_j^*) - u_{j-1}(\widehat{F}_{j-2}(F_{j-1}^*), F_{j-1}^*, F_j^*)$$

in $\widehat{F}_{j-1}(F_j^*)$ around F_{j-1}^* and

$$u_{j+1}(F_j^*, \widehat{F}_{j+1}(F_j^*), \widehat{F}_{j+2}(F_j^*)) - u_{j+1}(F_j^*, F_{j+1}^*, \widehat{F}_{j+2}(F_{j+1}^*))$$

in $\widehat{F}_{j+1}(F_j^*)$ around F_{j+1}^*. Then, since $\widehat{F}_{j-2}(F_j^*) = \widehat{F}_{j-2}(\widehat{F}_{j-1}(F_j^*))$ and $\widehat{F}_{j+2}(F_j^*) = \widehat{F}_{j+2}(\widehat{F}_{j+1}(F_j^*))$ are satisfied, note that we can treat $\widehat{F}_{j-2}(F_j^*)$ and $\widehat{F}_{j+2}(F_j^*)$ as the functions of $\widehat{F}_{j-1}(F_j^*)$ and $\widehat{F}_{j+1}(F_j^*)$, respectively. Because $u_{j-1}(\widehat{F}_{j-2}(F_j^*), \widehat{F}_{j-1}(F_j^*), F_j^*) = 0$ and $u_{j+1}(F_j^*, \widehat{F}_{j+1}(F_j^*), \widehat{F}_{j+2}(F_j^*)) = 0$, by such Taylor approximations, we have

$$\begin{aligned} u_{j-1}(F_{j-2}^*, F_{j-1}^*, F_j^*) &+ \left\{\alpha_{j-1}/\left(F_{j-1}^* - \widehat{F}_{j-2}(F_{j-1}^*)\right) - \alpha_{j-1}/\left(F_{j-1}^* - F_{j-2}^*\right)\right\} \\ &= \widetilde{i}^{\mathrm{f}}(\widetilde{F}_{j-2}^{*(2)}, \widetilde{F}_{j-1}^{*(1)}, F_j^*)\left(\widehat{F}_{j-1}(F_j^*) - F_{j-1}^*\right), \quad (22) \\ u_{j+1}(F_j^*, F_{j+1}^*, F_{j+2}^*) &+ \left\{\alpha_{j+2}/\left(\widehat{F}_{j+2}(F_{j+1}^*) - F_{j+1}^*\right) - \alpha_{j+2}/\left(F_{j+2}^* - F_{j+1}^*\right)\right\} \\ &= \widetilde{i}^{\mathrm{b}}(F_j^*, \widetilde{F}_{j+1}^{*(1)}, \widetilde{F}_{j+2}^{*(2)})\left(\widehat{F}_{j+1}(F_j^*) - F_{j+1}^*\right), \quad (23) \end{aligned}$$

where $\widetilde{F}_{j-1}^{*(1)}$ and $\widetilde{F}_{j+1}^{*(1)}$ are another inner points on the line segments between F_{j-1}^* and $\widehat{F}_{j-1}(F_j^*)$ and F_{j+1}^* and $\widehat{F}_{j+1}(F_j^*)$, respectively, $\widetilde{F}_{j\pm 2}^{*(2)}$ mean $\widetilde{F}_{j-2}^{*(2)} = \widehat{F}_{j-2}(\widetilde{F}_{j-1}^{*(1)})$

and $\widetilde{F}^{*(2)}_{j+2} = \widehat{F}_{j+2}(\widetilde{F}^{*(1)}_{j+1})$, and

$$\widetilde{i}^{\mathrm{f}}(\widetilde{F}^{*(2)}_{j-2}, \widetilde{F}^{*(1)}_{j-1}, F^*_j) = i_{j-1,j-1}(\widetilde{F}^{*(2)}_{j-2}, \widetilde{F}^{*(1)}_{j-1}, F^*_j) + i_{j-2,j-1}(\widetilde{F}^{*(2)}_{j-2}, \widetilde{F}^{*(1)}_{j-1}) \frac{\partial \widehat{F}_{j-2}(\widetilde{F}^{*(1)}_{j-1})}{\partial \widetilde{F}^{*(1)}_{j-1}},$$

$$\widetilde{i}^{\mathrm{b}}(F^*_j, \widetilde{F}^{*(1)}_{j+1}, \widetilde{F}^{*(2)}_{j+2}) = i_{j+1,j+1}(F^*_j, \widetilde{F}^{*(1)}_{j+1}, \widetilde{F}^{*(2)}_{j+2}) + i_{j+2,j+1}(\widetilde{F}^{*(1)}_{j+1}, \widetilde{F}^{*(2)}_{j+2}) \frac{\partial \widehat{F}_{j+2}(\widetilde{F}^{*(1)}_{j+1})}{\partial \widetilde{F}^{*(1)}_{j+1}}.$$

The expression of $u_{m_t}(\widehat{\boldsymbol{F}}(\theta^*_t))$ from Proposition 1 is transformed by repeatedly using (20) and (22) to cancel $\widehat{F}_{j-1}(F^*_j) - F^*_{j-1}$ until getting $j = 1$ if $j \le m_t$, and using (21) and (23) to cancel $\widehat{F}_{j+1}(F^*_j) - F^*_{j+1}$ until $j = [n] - 1$ if $j \ge m_t$, so that we have

$$U_n(\theta^*_t) = \textstyle\sum_{j=1}^{[n]-1} u_j(F^*_{j-1}, F^*_j, F^*_{j+1}) \widetilde{F}_{\theta^*_t}(\widehat{J}_j),$$

where

$$\widetilde{F}_{\theta^*_t}(s) = \begin{cases} \prod_{j=m_s}^{m_t-1} \left\{ -i_{j,j+1}(\widetilde{F}^{*(0)}_j, F^*_{j+1}) \big/ \widetilde{i}^{\mathrm{f}}(\widetilde{F}^{*(2)}_{j-1}, \widetilde{F}^{*(1)}_j, F^*_{j+1}) \right\} & \text{if } s \le t, \\ \prod_{j=m_t+1}^{m_s} \left\{ -i_{j,j-1}(F^*_{j-1}, \widetilde{F}^{*(0)}_j) \big/ \widetilde{i}^{\mathrm{b}}(F^*_{j-1}, \widetilde{F}^{*(1)}_j, \widetilde{F}^{*(2)}_{j+1}) \right\} & \text{if } s > t. \end{cases} \tag{24}$$

The above expression provides (13) of this proposition.

Similarly to $\widehat{F}_{\theta_t}(s; \theta_t)$, $\widetilde{F}_{\theta^*_t}(s)$ satisfies a product integral structure such that $\widetilde{F}_{\theta^*_t}(s) = \boldsymbol{\pi}_{s \le u < t}(1 - \overrightarrow{d}\widetilde{K}(u; t))$ if $s \le t$ and $\widetilde{F}_{\theta^*_t}(s) = \boldsymbol{\pi}_{t < u \le s}(1 - \overleftarrow{d}\widetilde{K}(u; t))$ otherwise, where $\overrightarrow{d}\widetilde{K}(s; t)$ is

$$1 - \overrightarrow{d}\widetilde{K}(s; t) = \frac{\alpha_{j+1}/(\widetilde{p}^{*(0)}_{j+1})^2}{\left(1 - \frac{\partial \widehat{F}_{j-1}(\widetilde{F}^{*(1)}_j)}{\partial \widetilde{F}^{*(1)}_j}\right) \frac{\alpha_j}{(\widetilde{p}^{*(2)}_j)^2} + \frac{\alpha_{j+1}}{(\widetilde{p}^{*(1)}_{j+1})^2} + \frac{\beta_j}{(\widetilde{S}^{*(1)}_j)^2} + \frac{\gamma_j}{(\widetilde{F}^{*(1)}_j)^2}}$$

at $s = \widehat{J}_j$ and is zero otherwise, $\overleftarrow{d}\widetilde{K}(s; t)$ is a reversed version of $\overrightarrow{d}\widetilde{K}(s; t)$, as similar to the relation between $\overrightarrow{d}\widehat{K}(s; \theta_t)$ and $\overleftarrow{d}\widehat{K}(s; \theta_t)$, and $\widetilde{p}^{*(0)}_j = F^*_j - \widetilde{F}^{*(0)}_{j-1}$, $\widetilde{p}^{*(1)}_j = F^*_j - \widetilde{F}^{*(1)}_{j-1}$, $\widetilde{p}^{*(2)}_j = F^{*(1)}_j - \widetilde{F}^{*(2)}_{j-1}$ and $\widetilde{S}^{*(1)}_j = 1 - \widetilde{F}^{*(1)}_j$.

Finally, using

$$\alpha_j = \textstyle\int_{\overleftarrow{\boldsymbol{J}}_j} \overrightarrow{d}\overline{N}^{\mathrm{f}(1)}(s), \;\; \beta_j = \int_{\overrightarrow{\boldsymbol{J}}_j} \overrightarrow{d}\overline{N}^{\mathrm{f}(2)}(s) \;\; \text{and} \;\; \gamma_j = \int_{\overrightarrow{\boldsymbol{J}}_j} \overrightarrow{d}\overline{N}^{\mathrm{f}(3)}(s),$$

we can easily see that (13) is transformed to the Eq. (14) in terms of the counting processes. □

Proof of Proposition 3 To obtain a further expression of $U_n(\theta_t^*)$ in Proposition 2, let

$$U_n^{\mathrm{f}}(\theta_t^*) = \sum_{j=1}^{m_t} u_j(\boldsymbol{F}^*)\widetilde{F}_{\theta_t^*}(\widehat{J}_j) + \alpha_{m_t+1}/p_{m_t+1}^*$$
$$\text{and} \quad U_n^{\mathrm{b}}(\theta_t^*) = \sum_{j=m_t+1}^{[n]-1} u_j(\boldsymbol{F}^*)\widetilde{F}_{\theta_t^*}(\widehat{J}_j) - \alpha_{m_t+1}/p_{m_t+1}^* \tag{25}$$

such that $U_n(\theta_t^*) = U_n^{\mathrm{f}}(\theta_t^*) + U_n^{\mathrm{b}}(\theta_t^*)$. Then, we decompose α_j, β_j and γ_j included in $u_j(\boldsymbol{F}^*)$ as $\alpha_j = \int_{\vec{\boldsymbol{J}}_j}\{\vec{d}\overline{A}^{\mathrm{f}(1)}(s) + \vec{d}\overline{M}^{\mathrm{f}(1)}(s)\}$, $\beta_j = \int_{\vec{\boldsymbol{J}}_j}\{\vec{d}\overline{A}^{\mathrm{f}(2)}(s) + \vec{d}\overline{M}^{\mathrm{f}(2)}(s)\}$ and $\gamma_j = \int_{\vec{\boldsymbol{J}}_j}\{\vec{d}\overline{A}^{\mathrm{f}(3)}(s) + \vec{d}\overline{M}^{\mathrm{f}(3)}(s)\}$ by Lemma 1 if $j \le m_t$. On the other hand, if $j \ge m_t+1$, α_j, β_j and γ_j are, respectively, decomposed as $\int_{\overleftarrow{\boldsymbol{J}}_j}\{\overleftarrow{d}\overline{A}^{\mathrm{b}(1)}(s) + \overleftarrow{d}\overline{M}^{\mathrm{b}(1)}(s)\}$ and $\int_{\vec{\boldsymbol{J}}_j}\{\overleftarrow{d}\overline{A}^{\mathrm{b}(\jmath)}(s) + \overleftarrow{d}\overline{M}^{\mathrm{b}(\jmath)}(s)\}$, $\jmath = 2, 3$ by Lemma 2. Hence, we obtain an expression of U_n as follows:

$$U_n(\theta_t^*) = \sum_{j=1}^{m_t-1}\left\{u_j^{A\mathrm{f}}(\boldsymbol{F}^*) + u_j^{M\mathrm{f}}(\boldsymbol{F}^*)\right\}\widetilde{F}_{\theta_t^*}(\widehat{J}_j) + \{u_{m_t}^{A\mathrm{fb}}(\boldsymbol{F}^*) + u_{m_t}^{M\mathrm{fb}}(\boldsymbol{F}^*)\}$$
$$+\sum_{j=m_t+1}^{[n]-1}\left\{u_j^{A\mathrm{b}}(\boldsymbol{F}^*) + u_j^{M\mathrm{b}}(\boldsymbol{F}^*)\right\}\widetilde{F}_{\theta_t^*}(\widehat{J}_j), \tag{26}$$

where $u_j^{A\mathrm{f}}$ and $u_{m_t}^{A\mathrm{fb}}$ are already provided as in Proposition 3, $u_j^{A\mathrm{b}}$ is

$$u_j^{A\mathrm{b}}(\boldsymbol{F}^*) = \int_{\overleftarrow{\boldsymbol{J}}_j}\frac{\overleftarrow{d}\overline{A}^{\mathrm{b}(1)}(s)}{p_j^*} - \int_{\overleftarrow{\boldsymbol{J}}_{j+1}}\frac{\overleftarrow{d}\overline{A}^{\mathrm{b}(1)}(s)}{p_{j+1}^*} - \int_{\vec{\boldsymbol{J}}_j}\frac{\overleftarrow{d}\overline{A}^{\mathrm{b}(2)}(s)}{S_j^*} + \int_{\vec{\boldsymbol{J}}_j}\frac{\overleftarrow{d}\overline{A}^{\mathrm{b}(3)}(s)}{F_j^*},$$

and obviously $u_j^{M\mathrm{f}}$, $u_j^{M\mathrm{fb}}$ and $u_j^{M\mathrm{b}}$ are $u_j^{A\mathrm{f}}$, $u_j^{A\mathrm{fb}}$ and $u_j^{A\mathrm{b}}$ in which the intensity components $\overline{A}^{\xi(\jmath)}$ are replaced by the corresponding martingale components $\overline{M}^{\xi(\jmath)}$ ($\jmath = 1, 2, 3$, $\xi = \mathrm{f}, \mathrm{b}$), respectively. By arranging (26), we have the expression of $U_n^A(t; \widetilde{H}_t)$ as in Proposition 3. $\square$

A.5 On consistency results

There are several ways of showing that the NPMLE $\widehat{F}$ or the SCEs (self-consistent estimators) are asymptotically consistent to F^*. Discussions in many other studies are based on the self-consistent equations (see, e.g., Tsai and Crowley 1985; Chang and Yang 1987; Gu and Zhang 1993). We briefly describe only the consistency results needed to derive the asymptotic distribution of the NPMLE below, since most of these detailed proofs are provided by Sugimoto (2012b).

If Condition 2 is satisfied, we have

$$\sup_{t\in[0,1]}|\widehat{F}(t) - F^*(t)| \to_p 0 \text{ and } \sup_{s,t\in[0,1]}|\widehat{F}(s;\theta_t^*) - F^*(s)| \to_p 0 \text{ as } n \to \infty \tag{27}$$

(see Chang and Yang 1987; Murphy and van der Vaart 1997 and Sugimoto 2012b, Theorem 1). Based on Proposition 4, it is shown that $\sup_{s\in[0,1]}|F^{d*}(s) - F^*(s)| =$

$\sup_j |F^*(\widehat{J}_j) - F^*(\widehat{J}_{j-1})|$ is bounded by $O(\log n/n)$ in probability. Using this fact, we have

$$\sup_{s\in[0,1]} \sqrt{n}|F^{d*}(s) - F^*(s)| \to_p 0 \quad \text{as} \ \ n \to \infty. \tag{28}$$

Given Condition 2 and $t \in [\tau_0, \tau_1]$ for $0 < \tau_0 \le \tau_1 < 1$, by a uniform continuity of $\widetilde{F}_{\theta_t^*}(s)$ on $\widetilde{K}(s;t)$, (27) and (28), we can show that $H_t^{*(\jmath)}(s)$, $\jmath = 1, 2, 3$ defined in Sect. 4.2 are bounded uniformly on $s \in [0, 1]$ and

$$\sup_{t\in[\tau_0,\tau_1], s\in[0,1]} |\widetilde{H}_t^{(\jmath)}(s) - H_t^{*(\jmath)}(s)| \to_p 0, \quad \jmath = 1, 2, 3 \ \ \text{as} \ \ n \to \infty \tag{29}$$

(for further details, see Sugimoto 2012b, Theorem 4, Corollary 3, Section 5), where the true derivative $F^*_{\theta_t^*}(\cdot)$ included in $H_t^{*(\jmath)}(\cdot)$ is

$$F^*_{\theta_t^*}(s) = \prod_{s\le u<t}(1 - \overrightarrow{d}K^*(u;t)) \ \text{ if } \ s < t \ \text{ and}$$
$$F^*_{\theta_t^*}(s) = \prod_{t<u\le s}(1 - \overleftarrow{d}K^*(u;t)) \ \text{ if } \ t < s$$

with

$$1 - \overrightarrow{d}K^*(u;t) = \frac{\overrightarrow{d}A^{*(1)}(u_+)/\overrightarrow{d}F^*(u_+)^2}{\overrightarrow{d}K^*(u_-)\frac{\overrightarrow{d}A^{*(1)}(u)}{\overrightarrow{d}F^*(u)^2} + \frac{\overrightarrow{d}A^{*(1)}(u_+)}{\overrightarrow{d}F^*(u_+)^2} + \int_{u_-}^{u}\frac{\overrightarrow{d}A^{*(2)}(u)}{S^*(u)^2} + \int_{u_-}^{u}\frac{\overrightarrow{d}A^{*(3)}(u)}{F^*(u)^2}},$$

$$1 - \overleftarrow{d}K^*(u;t) = \frac{\overrightarrow{d}A^{*(1)}(u)/\overrightarrow{d}F^*(u)^2}{\overleftarrow{d}K^*(u_+;t)\frac{\overrightarrow{d}A^{*(1)}(u_+)}{\overrightarrow{d}F^*(u_+)^2} + \frac{\overrightarrow{d}A^{*(1)}(u)}{\overrightarrow{d}F^*(u)^2} + \int_{u_-}^{u}\frac{\overrightarrow{d}A^{*(2)}(u)}{S^*(u)^2} + \int_{u_-}^{u}\frac{\overrightarrow{d}A^{*(3)}(u)}{F^*(u)^2}}$$

(if F^* is not strictly increasing, u_+ and u_- mean times just after and prior to $\inf\{x : F^*(x) - F^*(u) > 0\}$ and $\sup\{x : F^*(u) - F^*(x) > 0\}$, respectively). Hence, by (27), (28) and (29), we have

$$\sup_{t\in[\tau_0,\tau_1]} |1/n^{-1}\mathcal{I}_n(\widetilde{\theta}_t) - 1/\mathscr{I}_{(t,t)}(0,1)| \to_p 0 \quad \text{as} \ \ n \to \infty \tag{30}$$

(Sugimoto 2012b, Theorem 5).

A.6 Auxiliary result

Proposition 4 *Suppose that F^* is continuous function and $F^L(t) - F^R(t)$ is positive on the support of F^*. Then, $\sup_j n^{-1}\alpha_j^*$ and $\sup_j |F^*(\widehat{J}_j) - F^*(\widehat{J}_{j-1})|$ are bounded by $O(\log n/n)$ in probability.*

Proof of Proposition 4 Note that $A^{*(1)}(\widehat{J}_j) = \int_{(0,\widehat{J}_j]}(F^L(s) - F^R(s))\vec{d}F^*(s)$, $j = 1, \ldots, k$ are i.i.d. uniform random variables on $[0, A^{*(1)}(1)]$, where jump points $\widehat{J}_j$ at censoring points are excluded but do not occur with probability tending to 1 as $n \to \infty$. Because of $n^{-1}\alpha_j^* = A^{*(1)}(\widehat{J}_j) - A^{*(1)}(\widehat{J}_{j-1})$, we have $\sup_j n^{-1}\alpha_j^* \le O_p(\log n/n)$ by the result on the maximal spacing (Slud 1978), which also leads $|F^*(\widehat{J}_j) - F^*(\widehat{J}_{j-1})| \le O_p(\log n/n)$ using a relation $n^{-1}\alpha_j^* \ge (F^*(\widehat{J}_j) - F^*(\widehat{J}_{j-1}))\inf_s(F^L(s) - F^R(s))$. □

Acknowledgments The author is grateful to the Editors and two anonymous referees for their constructive comments and helpful suggestions that led to an improvement of this paper.

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.

Andersen, P. K., Borgan, Ø., Gill, R. D., Keiding, N. (1993). *Statistical models based on counting processes*. New York: Springer.

Cai, T., Cheng, S. (2004). Semiparametric regression analysis for doubly censored data. *Biometrika*, *91*, 277–290.

Chang, M. N. (1990). Weak convergence of a self-consistent estimator of the survival function with doubly censored data. *The Annals of Statistics*, *15*, 391–404.

Chang, M. N., Yang, G. L. (1987). Strong consistency of a nonparametric estimator of the survival function with doubly censored data. *The Annals of Statistics*, *15*, 1536–1547.

Chen, K., Zhou, M. (2003). Non-parametric hypothesis testing and confidence intervals with doubly censored data. *Lifetime Data Analysis*, *9*, 71–91.

Fleming, T. R., Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. New York: Wiley.

Gehan, E. A. (1965). A generalized two-sample Wilcoxon test for doubly censored data. *Biometrika*, *52*, 650–653.

Gentleman, R., Geyer, C. J. (1994). Maximum likelihood for interval censored data: consistency and computation. *Biometrika*, *81*, 618–623.

Gill, R. D. (1983). Large sample behaviour of the product-limit estimator on the whole line. *The Annals of Statistics*, *11*, 49–58.

Gu, M. G., Zhang, C.-H. (1993). Asymptotic properties of self-consistent estimators based on doubly censored data. *The Annals of Statistics*, *21*, 611–624.

Kim, J. S. (2003). Maximum likelihood estimation for the proportional hazards model with partly interval-censored data. *Journal of the Royal Statistical Society, Series B*, *65*, 489–502.

Murphy, S. A., van der Vaart, A. W. (1997). Semiparametric likelihood ratio inference. *The Annals of Statistics*, *25*, 1471–1509.

Mykland, P. A., Ren, J.-J. (1996). Algorithms for computing self-consistent and maximum likelihood estimators with doubly censored data. *The Annals of Statistics*, *24*, 1740–1764.

Oakes, D. (2000). Survival analysis. *Journal of the American Statistical Association*, *95*, 282–285.

Patilea, V., Rolin, J.-M. (2006). Product-limit estimators of the survival function with twice censored data. *The Annals of Statistics*, *34*, 925–938.

Slud, E. (1978). Entropy and maximal spacings for random partitions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, *41*, 341–352.

Sugimoto, T. (2011). Wald-type variance estimation for the nonparametric distribution estimators in doubly censored data. *Annals of the Institute of Statistical Mathematics*, *63*, 645–670.

Sugimoto, T. (2012a). On an inverse formula of a tridiagonal matrix. *Operators and Matrices*, *6*, 465–480.

Sugimoto, T. (2012b). On consistencies of the profile likelihood estimators and their derivatives of the distribution function in doubly censored data. *Communications in Statistics: Theory and Methods* (in press).

Therneau, M. T., Grambsch, P. M. (2000). *Modeling surivival data: Extending the Cox model*. New York: Springer.

Tsai, W.-Y., Crowley, J. (1985). A large sample study of generalized maximum likelihood estimators from incomplete data via self-vonsistency. *The Annals of Statistics*, *13*, 1317–1334.
Turnbull, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *Journal of the American Statistical Association*, *69*, 169–173.
Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society, Series B*, *38*, 290–295.
Wellner, J. A., Zhang, Y. (1997). A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data. *Journal of the American Statistical Association*, *92*, 945–959.
Yu, Q. Q., Li, L. X. (2001). Asymptotic properties of self-consistent estimators with doubly-censored data. *Acta Mathematica Sinica*, *17*, 581–594.