Ann Inst Stat Math (2011) 63:645–670
DOI 10.1007/s10463-009-0251-3

# A Wald-type variance estimation for the nonparametric distribution estimators for doubly censored data

**Tomoyuki Sugimoto**

Received: 31 March 2008 / Revised: 7 May 2009 / Published online: 31 July 2009


**Abstract** We discuss the variance estimation for the nonparametric distribution estimator for doubly censored data. We first provide another view of Kuhn–Tucker's conditions to construct the profile likelihood, and lead a Newton–Raphson algorithm as an optimization technique unlike the EM algorithm. The main proposal is an iteration-free Wald-type variance estimate based on the chain rule of differentiating conditions to construct the profile likelihood, which generalizes the variance formula in only right- or left-censored data. In this estimation procedure, we overcome some difficulties caused in directly applying Turnbull's formula to large samples and avoid a load with computationally heavy iterations, such as solving the Fredholm equations, computing the profile likelihood ratio or using the bootstrap. Also, we establish the consistency of the formulated Wald-type variance estimator. In addition, simulation studies are performed to investigate the properties of the Wald-type variance estimates in finite samples in comparison with those from the profile likelihood ratio.

**Keywords** Double censoring · Efficient Fisher information · Empirical likelihood · Integral equations · Profile likelihood · Self-consistent equations · Variance formula

## 1 Introduction

In doubly censored data, the nonparametric maximum likelihood estimator (NPMLE) cannot be written as the closed forms and the variances of the NPMLE have been estimated via some iteration methods, such as solving the Fredholm equations (Chang 1990), the bootstrap sampling (Wellner and Zhan 1997) or a self-consistent algorithm

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

under the constraint (Chen and Zhou 2003). Originally, Turnbull (1974) provided the determinant-based formula to obtain the estimated variance–covariance matrix corresponding to the inverse of a tridiagonal matrix (e.g., Klein and Moeschberger 2003, pp. 146–147). However, it is difficult to directly apply Turnbull's formula to large samples, because the values of the determinants themselves tend to blow up numerically for a larger matrix, so that the divisions of determinants in such formula become computationally unstable. On the other hand, Chang (1990) has given the asymptotic variance form of the NPMLE expressed by solving the Fredholm integral equations from the self-consistent equations. Wellner and Zhan (1997) indicated the application of the bootstrap sampling via a faster algorithm to capture the variability of the NPMLE. Chen and Zhou (2003) have proposed the self-consistent algorithm to compute the NPMLE under the constraint, so that the confidence intervals and the variances of the NPMLE are estimated based on the profile likelihood ratio. However, the three approaches cannot avoid a load with computationally heavy iterations, even if variance estimates can be numerically obtained. Still, Chang's self-consistent equation-based formula has a drawback equivalent to finding the inverse of a larger matrix for practical use. Though the bootstrap method and Chen and Zhou's method are able to obtain computationally stable variance estimates in larger samples, vast time and a great number of iterations are required to obtain the pointwise variance estimates of the NPMLE. For example, referring Wellner and Zhan (1997), assume that we use 200 resamplings and have 20 s on an average to calculate the NPMLE in a sample of $n = 1,500$, then, in the bootstrap method, the computation time needed for acquiring a variance estimate is about 67 min.

In this paper, we propose an iteration-free variance estimation for the nonparametric distribution estimator for doubly censored data. Our estimation procedure is based on a semiparametric profile likelihood and generalizes the variance formula in only right- or left-censored data. We provide another view of Kuhn–Tucker's conditions to construct the profile likelihood, by which a Newton–Raphson algorithm different from the EM algorithm is planned as an optimization technique to obtain the profile likelihood. In particular, our proposal is a Wald-type variance estimate based on the chain rule of differentiating conditions to construct the profile likelihood in order to overcome some difficulties of a Wald-type inference for semiparametric profile likelihood. In terms of gaining the inverse of the Fisher information matrix, our manner is same as Turnbull's but freed of using the form of the determinant functions. In such plug-in estimation procedure, we can gain a numerically stable solution and avoid a load with computationally heavy iterations, such as solving the Fredholm integral equations to estimate the influence functions (e.g., Chang 1990; Chen and Zhou 2003), using a self-consistent algorithm to compute the profile likelihood ratio (e.g., Chen and Zhou 2003) or applying the bootstrap method (e.g., Wellner and Zhan 1997; Zhu and Sun 2007). Unlike Turnbull's determinant-based formula, ours does not cause numerical and theoretical blowing up in large samples, which is eventually equivalent to that the denominators and numerators with determinantal forms in Turnbull's formula cancel out not to diverge, and is easier to derive an asymptotic form. In fact, to confirm the validity of the proposed estimator, we show that the Wald-type variance estimator is consistent to an asymptotic variance form, which is a direct large sample study of Turnbull's formula and provides a viewpoint different from Chang's form. Also, the Wald-type variance estimate reduces the Greenwood variance formula in right-censored data. In addition, we conduct simulation studies to investigate the performance of the Wald-type variance estimator in finite samples in comparison with that based on the profile likelihood ratio.

This paper is structured as follows: In Sects. 2.1 and 2.2, we describe the formulation and key conditions to construct the semiparametric profile likelihood. In Sect. 2.3, we provide a Newton–Raphson algorithm to obtain the profile likelihood. In Sect. 3, we propose an iteration-free Wald-type variance estimate and prove the consistency of the proposed variance estimator. In Sect. 4, we investigate the properties of the Wald-type estimates using simulation.

## 2 On a semiparametric profile likelihood

### 2.1 Notations and the NPMLE

Suppose that true event times $T_i^*$, $i = 1, \ldots, n$ and left- and right-censoring time vectors $(C_i^L, C_i^R)$, $i = 1, \ldots, n$ are independent and identically distributed as the function $F^*$ and a joint bivariate function with $C_i^L \leq C_i^R$ and independent of $T_i^*$, respectively. Let $F^L$ and $F^R$ be the marginal distribution functions of $C_i^L$ and $C_i^R$, respectively. We observe doubly censored event times $T_i$ with the censoring indicators $\Delta_i$ as

$$(T_i, \Delta_i) = \begin{cases} (T_i^*, 1) & \text{if } C_i^L < T_i^* \leq C_i^R, \\ (C_i^R, 2) & \text{if } C_i^R < T_i^*, \\ (C_i^L, 3) & \text{if } T_i^* \leq C_i^L, \end{cases}, \quad i = 1, \ldots, n. \tag{1}$$

Given data (1), we can obtain the NPMLE $\widehat{F}$ of $F^*$ via some algorithm (e.g., Turnbull 1976; Mykland and Ren 1996; Wellner and Zhan 1997). Let $T_{(1)} < T_{(2)} < \cdots < T_{(k_a)}$ and $T^j_{(1)} < T^j_{(2)} < \cdots < T^j_{(k_j)}$, $j = 1, 2, 3$ be the distinct times of $\{T_1, T_2, \ldots, T_n\}$ and of $T_i$'s with $\Delta_i = j$ for $i = 1, \ldots, n$, respectively ($k_1 + k_2 + k_3 \leq n$). The candidates of the jump points to find the NPMLE $\widehat{F}$ consist of the following three types:

(j1) all the distinct uncensored time points $(T^1_{(1)}, \ldots, T^1_{(k_1)})$,
(j2) the earliest left-censored time point $T^3_{(1)}$ and the last right-censored time point $T^2_{(k_2)}$,
(j3) some censored time points occurring in the interval $(T^1_{(j)}, T^1_{(j+1)})$ for $j = 1, \ldots, k_1 - 1$, such as the 2–3-type points defined by Mykland and Ren (1996).

The jump points of (j2) can be treated as a type of (j1). That is, when we observe $T^3_{(1)} \leq T^1_{(1)}$ and/or $T^1_{(k_1)} \leq T^2_{(k_2)}$ in the original data (1), the NPMLE $\widehat{F}$ does not change even if the left-censored observations with $(T_i, \Delta_i) = (T^3_{(1)}, 3)$ and the right-censored observations with $(T_i, \Delta_i) = (T^2_{(k_2)}, 2)$ are compulsorily reformed to $(T^3_{(1)}, 1)$ and

$(T^2_{(k_2)}, 1)$, respectively. For some discussion relevant to these, see Wellner and Zhan (1997, p. 949). Also, in the many situations of larger sample, it will be satisfied that

$$\text{there are no jump points of (j3).} \tag{2}$$

In fact, as $n$ becomes larger, the probability that the jump points of (j3) occur trends to zero (Mykland and Ren 1996).

Let $J = \{J_1, J_2, \dots, J_{k(J)}\} \in \mathcal{J}$ be a set of jump points with $J_1 < J_2 < \cdots < J_{k(J)}$ and $k_1 \le k(J) \le k_a$, where $\mathcal{J}$ is the feasible family of $J$'s. Based on the consideration for (j1–j3), let $J_{\max} = \{T_{(1)}, T_{(2)}, \dots, T_{(k_a)}\} \in \mathcal{J}$ for the largest $k(J_{\max}) = k_a$ and $J_{\min} = \{\widetilde{T}^3_{(1)}, T^1_{(1)}, T^1_{(2)}, \dots, T^1_{(k_1)}, \widetilde{T}^2_{(k_2)}\} \in \mathcal{J}$ for some smallest $k(J_{\min}) = \widetilde{k}_1$, where $\widetilde{k}_1 = k_1 + I(\widetilde{T}^3_{(1)} = T^3_{(1)}) + I(\widetilde{T}^2_{(k_2)} = T^2_{(k_2)})$ and

$$\begin{aligned}
&\widetilde{T}^3_{(1)} = T^3_{(1)} \quad \text{if} \quad T^3_{(1)} < T^1_{(1)} \quad \text{and} \quad \widetilde{T}^3_{(1)} = \emptyset \quad \text{otherwise,}\\
&\widetilde{T}^2_{(k_2)} = T^2_{(k_2)} \quad \text{if} \quad T^1_{(k_1)} < T^2_{(k_2)} \quad \text{and} \quad \widetilde{T}^2_{(k_2)} = \emptyset \quad \text{otherwise.}
\end{aligned}$$

Thus, $\mathcal{J}$ is the collection of sets that completely include $J_{\min}$ of all possible subsets generated by $J_{\max}$. Using a $J \in \mathcal{J}$, we can set $F(J) = \{F_i = F(J_i), i = 1, \dots, k(J)-1\}$ as the parametrization of a right-continuous function $F$ with left-hand limits to infer $F^*$. Then, the log empirical likelihood constructed from the data (1) is parallel to

$$l(F) = l(F_1, \dots, F_{k(J)-1}) = \log \prod_{j=1}^{k(J)} (F_j - F_{j-1})^{d_j} (1 - F_j)^{m^R_j} (F_j)^{m^L_j}, \tag{3}$$

where the constraint condition

$$\begin{aligned}
&0 = F_0 \le F_1 \le \cdots \le F_{k(J)-1} \le F_{k(J)} = 1\\
&F_{j-1} < F_j \quad \text{if} \quad d_j > 0 \quad \text{for} \quad j = 1, \dots, k(J)
\end{aligned} \tag{4}$$

is demanded. The $d_j$'s, $m^R_j$'s and $m^L_j$'s are, by reflecting the consideration for (j1)–(j3), defined as

$$\begin{aligned}
&d_1 = \sum_{i=1}^n I(\Delta_i = 3 \text{ or } 1, T_i = J_1), \quad d_{k(J)} = \sum_{i=1}^n I(\Delta_i = 2 \text{ or } 1, T_i = J_{k(J)}),\\
&d_j = \sum_{i=1}^n I(\Delta_i = 1, T_i = J_j), \quad j = 2, \dots, k(J) - 1,\\
&m^L_1 = \sum_{i=1}^n I(\Delta_i = 3, T_i \in (J_1, J_2)), \quad m^R_0 = \sum_{i=1}^n I(\Delta_i = 2, T_i < J_1),\\
&m^L_j = \sum_{i=1}^n I(\Delta_i = 3, T_i \in [J_j, J_{j+1})), \quad j = 2, \dots, k(J),\\
&m^R_j = \sum_{i=1}^n I(\Delta_i = 2, T_i \in [J_j, J_{j+1})), \quad j = 1, \dots, k(J) - 1, \quad m^R_{k(J)} = 0
\end{aligned} \tag{5}$$

depending on $J$, where $J_{k(J)+1} = \infty$. $m^L_{k(J)}$ and $m^R_0$ are not necessarily zero, but these contributions to (3) can be ignored. The form (3) of the log likelihood is general and flexible in the sense that the NPMLE can be expressed as $\widehat{F} = \text{argmax}_{\{F=F(J_{\max})\}} l(F) = \text{argmax}_{\{F=F(J), J\in\mathcal{J}\}} l(F)$ even if (2) is not satisfied. The first derivative $u_j$ of $l(F)$ w.r.t. $F_j$ for $j = 1, \ldots, k(J) - 1$ is

$$u_j = u_j(F_{j-1}, F_j, F_{j+1}) = \frac{\partial l(F)}{\partial F_j} = \frac{d_j}{F_j - F_{j-1}} - \frac{d_{j+1}}{F_{j+1} - F_j} - \frac{m^R_j}{1 - F_j} + \frac{m^L_j}{F_j}$$

and minus the second derivative $i_{j,l}$ of $l(F)$ w.r.t. $F_j$ and $F_l$ for $j, l = 1, \ldots, k(J) - 1$ is

$$\begin{aligned} i_{j,l} &= i_{j,l}(F_{j-1}, F_j, F_{j+1}) = -\frac{\partial^2 l(F)}{\partial F_j \partial F_l} \\ &= \begin{cases} \dfrac{d_j}{(F_j - F_{j-1})^2} + \dfrac{d_{j+1}}{(F_{j+1} - F_j)^2} + \dfrac{m^R_j}{(1 - F_j)^2} + \dfrac{m^L_j}{F_j^2} & \text{if} \quad j = l \\ -\dfrac{d_j}{(F_j - F_{j-1})^2} & \text{if} \quad j = l + 1 \\ -\dfrac{d_{j+1}}{(F_{j+1} - F_j)^2} & \text{if} \quad j = l - 1 \\ 0 & \text{if} \quad |j - l| \geq 2. \end{cases} \end{aligned}$$

Let $\widehat{J}$ be the set $J$ of only jump points of the NPMLE $\widehat{F}$. The following lemma shows how the condition to find the NPMLE is characterized using $u_j$'s.

**Lemma 1** *Let $\widehat{F}(J) = \{\widehat{F}(J_i), i = 1, \ldots, k(J) - 1\}$ be $F(J)$ which satisfies $u_j = 0$, $j = 1, \ldots, k(J) - 1$ under the condition (4), if such solution exists. Then, there exists $\overline{J}$ which has the maximum $k(J)$ of all the $J$'s corresponding to $\widehat{F}(J)$, and the NPMLE $\widehat{F}$ is uniquely $\widehat{F}(\overline{J}) = \widehat{F}(\widehat{J})$.*

This lemma is an assertion eventually equivalent to Theorem 1 of Mykland and Ren (1996), Theorem 3.1 of Wellner and Zhan (1997), and so on. However, using a manner different from them which generalizes the proof method of Turnbull (1974, Lemma A2), we provide a proof of Lemma 1 including another view of Kuhn–Tucker's conditions. A significant point is that the strict concavity of $l(F)$ is easily proved by the tridiagonal property of the observed full Fisher information matrix.

*Proof of Lemma 1* First, for $F = F(J)$ over every $J \in \mathcal{J}$, we show that $l(F)$ is two times continuously differentiable and $l(F)$ is strictly concave under the condition (4). The former is easily shown, since all the $i_{j,l}$'s are continuous and bounded under (4). The latter is completed by showing that the matrix $\mathcal{I}(F)$ is positive definite, where

$\mathcal{I}(F)$ is minus the Hessian matrix of $l(F)$ with the $(j,l)$th element $i_{j,l}$, that is,

$$\mathcal{I}(F) = \begin{pmatrix} i_{1,1} & i_{1,2} & 0 & 0 & \cdots & 0 \\ i_{2,1} & i_{2,2} & i_{2,3} & 0 & \ddots & \vdots \\ 0 & i_{3,2} & i_{3,3} & i_{3,4} & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & i_{k(J)-2,k(J)-2} & i_{k(J)-2,k(J)-1} \\ 0 & 0 & \cdots & 0 & i_{k(J)-1,k(J)-2} & i_{k(J)-1,k(J)-1} \end{pmatrix}.$$

The positive definiteness of $\mathcal{I}(F)$ is equivalent to the fact that the determinant of $\mathcal{I}(F)$ is always positive, since $\mathcal{I}(F)$ is a real symmetric matrix. Further, because $\mathcal{I}(F)$ is a tridiagonal matrix, the determinant of $\mathcal{I}(F)$ is $D_{k(J)-1}$ obtained by following the recursive formula

$$D_0 = 1, \quad D_1 = i_{1,1}, \quad D_j = i_{j,j} D_{j-1} - (i_{j-1,j})^2 D_{j-2}, \quad j = 2, \ldots, k(J) - 1. \tag{6}$$

In Appendix A.1, we provide that every $D_l$ is always positive. This indicates that all the eigenvalues of $\mathcal{I}(F)$ are positive, so that $\mathcal{I}(F)$ is positive definite over every $J \in \mathcal{J}$.

Because of $J_{\max} \supseteq$ every $J \in \mathcal{J}$, the NPMLE $\widehat{F}$ is unique by the strict concavity of $l(F)$ for $F = F(J_{\max})$. Hence, $\widehat{J}$ is also unique. Then, we have $u_j = 0$, $j = 1, \ldots, k(J) - 1$ for $F = \widehat{F}$ and $J = \widehat{J}$, since the solution $\widehat{F}(\widehat{J})$ of $F(\widehat{J})$ exists over all the inner points of (4) by the strict concavity of $l(F)|_{F=F(J)}$ for every $J \in \mathcal{J}$. Also, $\overline{J}$ supplements $\widehat{J}$ with no jump points such that both $u_j|_{F=\widehat{F}} = 0$ and $u_{j+1}|_{F=\widehat{F}} = 0$ are satisfied for some $j$ (i.e., a censoring point) with $d_{j+1} = 0$ in spite of $\widehat{F}(\widehat{J}_{j+1}) = \widehat{F}(\widehat{J}_j)$, if such points exit. Therefore, we have $\widehat{F} = \widehat{F}(\overline{J}) = \widehat{F}(\widehat{J})$. □

The relation between the self-consistent estimators (SCEs) and the condition

$$u_j|_{F=\widehat{F}(J)} = 0, \quad j = 1, \ldots, k(J) - 1 \ \text{ for } \ J \in \mathcal{J} \tag{7}$$

is not more direct than that (Wellner and Zhan 1997) based on the EM algorithm and the Lagrange multipliers. We show how the condition (7) is connected with the SCEs in the following lemma, which includes important matrix expressions reused in the proof of Theorem 1 and slightly extends Lemma A1 of Turnbull (1974).

**Lemma 2** *Let* $u(F) = (u_1, \ldots, u_{k(J)-1})^{\mathrm{T}}$ *for some* $J \in \mathcal{J}$. *A linear transformation of* (7)*, that is,*

$$n^{-1} \mathcal{P}\mathcal{H} u(F)|_{F=\widehat{F}(J)} = 0$$

*is identical to the self-consistent equations, where* $\mathcal{P}$ *and* $\mathcal{H}$ *are* $(k(J)-1)\times(k(J)-1)$ *triangular matrices such that*

$$\mathcal{P}=\mathcal{P}(F)=\begin{pmatrix} p_1 & 0 & 0 & \cdots & 0 \\ \vdots & p_2 & 0 & \cdots & 0 \\ p_1 & \vdots & p_3 & \ddots & \vdots \\ p_1 & p_2 & \vdots & \ddots & 0 \\ p_1 & p_2 & p_3 & \cdots & p_{k(J)-1} \end{pmatrix} \text{ and } \mathcal{H}=\begin{pmatrix} 1 & 1 & \cdots & 1 & 1 \\ 0 & 1 & \cdots & 1 & 1 \\ \vdots & 0 & \ddots & \vdots & \vdots \\ 0 & \ddots & \ddots & 1 & 1 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

*and* $p_j = F_j - F_{j-1}$, $j = 1, \ldots, k(J)$.

*Proof of Lemma 2* Let $(\widehat{F}_1(J), \ldots, \widehat{F}_{k(J)-1}(J))$ be $\widehat{F}(J)$ defined in Lemma 1. The $j$th element of $\mathcal{H}u(F)|_{F=\widehat{F}(J)} = 0$ is

$$u_j^p = d_j/\widehat{p}_j + \sum_{l=j}^{k(J)} m_l^L/\widehat{F}_l(J) + \sum_{l=0}^{j-1} m_l^R/(1-\widehat{F}_l(J)) - \gamma_{k(J)} = 0,$$

$j = 1, \ldots, k(J)-1$, where $\gamma_{k(J)} = d_{k(J)}/\widehat{p}_{k(J)} + \sum_{l=0}^{k(J)-1} m_l^R/(1-\widehat{F}_l(J)) + m_{k(J)}^L$ and $\widehat{p}_j = \widehat{F}_j(J) - \widehat{F}_{j-1}(J)$. Then, because we also have $u_{k(J)}^p = 0$, we can lead $\gamma_{k(J)} = n$ by solving the equation $\sum_{j=1}^{k(J)} \widehat{p}_j u_j^p = 0$ on $\gamma_{k(J)}$. Therefore, it is shown that the $i$th element of $n^{-1}\mathcal{P}\mathcal{H}u(F)'|_{F=\widehat{F}(J)} = 0$, that is, $n^{-1}\sum_{j=1}^{i} \widehat{p}_j u_j^p = 0$ becomes

$$\frac{m_0^R}{n} + \sum_{j=1}^{i} \frac{\left(d_j + m_j^R + m_j^L\right)}{n} + \sum_{l=i+1}^{k(J)} \frac{\widehat{F}_i(J)}{\widehat{F}_l(J)}\frac{m_l^L}{n} - \sum_{l=0}^{i} \frac{1-\widehat{F}_i(J)}{1-\widehat{F}_l(J)}\frac{m_l^R}{n} - \widehat{F}_i(J) = 0. \tag{8}$$

The collection of (8) for $i = 1, \ldots, k(J)-1$ is identical to the self-consistent equations. □

## 2.2 Formulation: profile likelihood

Here, we will be interested in an inference on $F^*(\tau)$ only at a fixed time $\tau$. That is, let $F_{m(J)}$ be a single parameter of interest for $m(J)$ with $J_{m(J)} \le \tau < J_{m(J)+1}$ for $J \in \mathcal{J}$ and let $F_{m(J)}^{\#} = (F_1, \ldots, F_{m(J)-1}, F_{m(J)+1}, \ldots, F_{k(J)-1})$ be nuisance parameter vector excluded only $F_{m(J)}$ of interest from all the parameters $(F_1, \ldots, F_{k(J)-1})$ $(m(J) = 1, \ldots, k(J)-1)$. Then, we define

$$\widehat{F}^{\#}(F_{m(J)}) = (\widehat{F}^f(F_{m(J)}), \widehat{F}^b(F_{m(J)})) = \mathrm{argmax}_{F_{m(J)}^{\#}}\, l(F)$$

as the solution obtained by maximizing $l(F)$ over $F_{m(J)}^{\#}$ under a fixed $F_{m(J)}$, where $\widehat{F}^f(F_{m(J)}) = (\widehat{F}_1(F_{m(J)}), \ldots, \widehat{F}_{m(J)-1}(F_{m(J)}))$ and $\widehat{F}^b(F_{m(J)}) = (\widehat{F}_{m(J)+1}(F_{m(J)}),$

$\ldots, \widehat{F}_{k(J)-1}(F_{m(J)}))$ decompose $\widehat{F}^{\#}(F_{m(J)})$ into two parts before and after $F_{m(J)}$. Using $\widehat{F}^{\#}$ and

$$pl(F_{m(J)}) = l(\widehat{F}^{f}(F_{m(J)}), F_{m(J)}, \widehat{F}^{b}(F_{m(J)})), \tag{9}$$

the log profile likelihood of $F(\tau)$ is defined as

$$pl(F(\tau)) = \max_{J \in \mathcal{J}} pl(F_{m(J)}),$$

which is of course the same quantity as $pl(F_{m(J_{\max})})$. We establish a criterion to find $pl(F(\tau))$ using $u_j$'s.

**Lemma 3** *Let $\widehat{F}^{\#}_{\tau}(J)$ be $\widehat{F}^{\#}(F_{m(J)})$ such that*

$$u_j|_{F=(\widehat{F}^{f}(F_{m(J)}),F_{m(J)},\widehat{F}^{b}(F_{m(J)}))}=0, \quad j=1,\ldots,m(J)-1,\ m(J)+1,\ldots,k(J)-1,$$

*satisfying the condition (4), if such solution exists. Then, there uniquely exists $\overline{J}^{\#}$ which maximizes both of $m(J)$ and $k(J)-m(J)$ of all the $J$'s corresponding to $\widehat{F}^{\#}_{\tau}(J)$, and the log profile likelihood of $F(\tau)$ is $pl(F(\tau)) = pl(F_{m(\overline{J}^{\#})})$.*

*Proof of Lemma 3* This can be proved in the same manner as that of Lemma 1. Let $\mathcal{I}(F^{\#}_{m(J)})$ be minus the Hessian matrix of $l(F|F_{m(J)})$, which is obtained by excluding both the $m(J)$th row and column from $\mathcal{I}(F)$, where $l(F|F_{m(J)})$ denotes the function to $l(F_1,\ldots,F_{k(J)-1})$ from $F^{\#}_{m(J)}$ under a fixed $F_{m(J)}$. Because $\mathcal{I}(F^{\#}_{m(J)})$ is a real symmetric tridiagonal matrix, $\mathcal{I}(F^{\#}_{m(J)})$ is positive definite (see Appendix A.1). Hence, $l(F|F_{m(J)})$ is a strictly concave function of $F^{\#}_{m(J)}$ under the condition (4) over every $J \in \mathcal{J}$. Therefore, by the same argument as Lemma 1, a solution $\widehat{F}^{\#}_{\tau}(\widehat{J}^{\#})$ of $\widehat{F}^{\#}_{\tau}(J)$'s exists over all the inner points of (4), so that $\overline{J}^{\#}$ is uniquely obtained by adding no jump point $J_j$'s of $\widehat{F}_j(F_{m(J)})$ but with $u_j = 0$ for such solution $\widehat{F}^{\#}_{\tau}(\widehat{J}^{\#})$. □

### 2.3 A computation of profile likelihood

Here, we provide an algorithm to compute $pl(F_{m(J)})$ of (9). Chen and Zhou (2003) presented a self-consistent algorithm to compute the NPMLE under the constraint. Their algorithm is stable but has relatively larger iterations and times until convergence as well as problems of a nonparametric EM-algorithm indicated by Wellner and Zhan (1997).

Our computation is a direct Newton–Raphson procedure using quite elegant concave properties of $l(F|F_{m(J)})$ implicated from Lemma 3. Further, we find $\widehat{F}^{f}(F_{m(J)})$ and $\widehat{F}^{b}(F_{m(J)})$ can be separately computed each other. For simplicity, suppose that the jump points of (j3) do not occur in $\widehat{F}^{\#}(F_{m(J)})$. That is, we typically consider $J = J_{\min}$ only. Also, for simplicity on only notations, we write $m = m(J)$ and $k = k(J)$ by dropping $J$.

*On the forward nuisance part:* Finding $\widehat{F}^f(F_m)$ is to solve the $(m-1)$ equations $u_j = 0,\ j = 1, \ldots, m-1$. However, note that, a solution $F_{j-1}$ of the equation $u_j(F_{j-1}, F_j, F_{j+1}) = 0$ is provided as

$$\overline{F}^f_{j-1}(F_j, F_{j+1}) = \left[1 - \frac{d_j(1-F_j)(F_{j+1}-F_j)}{d_{j+1}(1-F_j)F_j + (F_{j+1}-F_j)\left(m^R_j F_j - m^L_j(1-F_j)\right)}\right] F_j, \quad (10)$$

given $F_j$ and $F_{j+1}$. If we utilize such relations at $m \geq 3$, by only putting an $\widetilde{F}_{m-1}$ under a fixed $F_m$ of interest, $\widetilde{F}_{m-2} = \overline{F}^f_{m-2}(\widetilde{F}_{m-1}, F_m)$ for $u_{m-1}(\widetilde{F}_{m-2}, \widetilde{F}_{m-1}, F_m) = 0$, $\widetilde{F}_{m-3} = \overline{F}^f_{m-3}(\widetilde{F}_{m-2}, \widetilde{F}_{m-1})$ for $u_{m-2}(\widetilde{F}_{m-3}, \widetilde{F}_{m-2}, \widetilde{F}_{m-1}) = 0, \ldots, \widetilde{F}_1 = \overline{F}^f_1(\widetilde{F}_2, \widetilde{F}_3)$ for $u_2(\widetilde{F}_1, \widetilde{F}_2, \widetilde{F}_3) = 0$ are automatically obtained one after another. Then, the equation not used yet of $u_j = 0, j = 1, \ldots, m-1$ is only $u_1(0, \widetilde{F}_1, \widetilde{F}_2) = 0$, which is a criterion to be $\widehat{F}^f(F_m)$ and know whether $\widetilde{F}_{m-1}$ put in the first is appropriate. Therefore, when $m \geq 3$, we can construct a Newton–Raphson algorithm to search $\widehat{F}^f(F_m)$ as follows.

ALGORITHM 1 (A computation to find $\widehat{F}^f(F_m)$ when $m \geq 3$)

F1. Let $l = 1$. Set an initial value $F_{m-1} = \widetilde{F}^{(1)}_{m-1}$ such that $\{F_m, \widetilde{F}^{(1)}_{m-1}, \widetilde{F}^{(1)}_{m-2}, \ldots, \widetilde{F}^{(1)}_1\}$ computed in F2 satisfies (4).

F2. Using (10) for the equations $u_j(\widetilde{F}^{(l)}_{j-1}, \widetilde{F}^{(l)}_j, \widetilde{F}^{(l)}_{j+1}) = 0$, $j = m-1, \ldots, 2$ and $\widetilde{F}^{(l)}_m = F_m$,

$$\widetilde{F}^{(l)}_{m-2} = \overline{F}^f_{m-2}(\widetilde{F}^{(l)}_{m-1}, F_m),\ \widetilde{F}^{(l)}_{m-3} = \overline{F}^f_{m-3}(\widetilde{F}^{(l)}_{m-2}, \widetilde{F}^{(l)}_{m-1}), \ldots$$
$$\ldots, \widetilde{F}^{(l)}_1 = \overline{F}^f_1(\widetilde{F}^{(l)}_2, \widetilde{F}^{(l)}_3)$$

are computed, so that $\widetilde{u}_1 = u_1(0, \widetilde{F}^{(l)}_1, \widetilde{F}^{(l)}_2)$ is computed. If $\{F_m, \widetilde{F}^{(l)}_{m-1}, \widetilde{F}^{(l)}_{m-2}, \ldots, \widetilde{F}^{(l)}_1\}$ satisfies (4), go to F3 as $h = 1$. Otherwise, go to F3 after letting $h \leftarrow 2h$ and $l \leftarrow l-1$.

F3. Stop if $\widetilde{u}_1 < \epsilon$ and $h = 1$, so that we have $\widehat{F}^f(F_m) = (\widetilde{F}^{(l)}_1, \ldots, \widetilde{F}^{(l)}_{m-1})$. Otherwise, $\widetilde{F}^{(l+1)}_{m-1}$ is updated as

$$\widetilde{F}^{(l+1)}_{m-1} = \widetilde{F}^{(l)}_{m-1} + h^{-1} u_1(0, \widetilde{F}^{(l)}_1, \widetilde{F}^{(l)}_2)/\bar{i}_1(\widetilde{F}^{(l)}_{m-1}),$$

then go to F2 after letting $l \leftarrow l+1$, where $\bar{i}_1(\widetilde{F}^{(l)}_{m-1}) = du_1(0, \widetilde{F}^{(l)}_1, \widetilde{F}^{(l)}_2)/d\widetilde{F}^{(l)}_{m-1}$.

By resetting $h \leftarrow 2h$ in F2 for too larger update to $\widetilde{F}^{(l+1)}_{m-1}$ from $\widetilde{F}^{(l)}_{m-1}$, a milder update is performed. The first derivative $\bar{i}_1(\widetilde{F}^{(l)}_{m-1})$ of $u_1(0, \widetilde{F}^{(l)}_1, \widetilde{F}^{(l)}_2)$ w.r.t. $\widetilde{F}^{(l)}_{m-1}$ is written as

$$\bar{\imath}_1\big(\widetilde{F}^{(l)}_{m-1}\big) = \frac{\partial u_1\big(0, \widetilde{F}^{(l)}_1, \widetilde{F}^{(l)}_2\big)}{\partial \widetilde{F}^{(l)}_1} \frac{d\widetilde{F}^{(l)}_1}{d\widetilde{F}^{(l)}_{m-1}} + \frac{\partial u_1\big(0, \widetilde{F}^{(l)}_1, \widetilde{F}^{(l)}_2\big)}{\partial \widetilde{F}^{(l)}_2} \frac{d\widetilde{F}^{(l)}_2}{d\widetilde{F}^{(l)}_{m-1}},$$

where $d\widetilde{F}^{(l)}_1/d\widetilde{F}^{(l)}_{m-1}$ and $d\widetilde{F}^{(l)}_2/d\widetilde{F}^{(l)}_{m-1}$ are obtained by the chain rule of differentiating (10), starting from

$$d\widetilde{F}^{(l)}_{m-2}/d\widetilde{F}^{(l)}_{m-1} = d\overline{F}^{f}_{m-2}(\widetilde{F}^{(l)}_{m-1}, F_m)/d\widetilde{F}^{(l)}_{m-1}$$

and following the computing order of $j = m-3, \ldots, 2, 1$ provided as

$$\frac{d\widetilde{F}^{(l)}_j}{d\widetilde{F}^{(l)}_{m-1}} = \frac{\partial \overline{F}^{f}_j\big(\widetilde{F}^{(l)}_{j+1}, \widetilde{F}^{(l)}_{j+2}\big)}{\partial \widetilde{F}^{(l)}_{j+1}} \frac{d\widetilde{F}^{(l)}_{j+1}}{d\widetilde{F}^{(l)}_{m-1}} + \frac{\partial \overline{F}^{f}_j\big(\widetilde{F}^{(l)}_{j+1}, \widetilde{F}^{(l)}_{j+2}\big)}{\partial \widetilde{F}^{(l)}_{j+2}} \frac{d\widetilde{F}^{(l)}_{j+2}}{d\widetilde{F}^{(l)}_{m-1}}.$$

When $m = 2$, rather than using Algorithm 1, one direct solution $\widetilde{F}_1$ of the quadratic equation

$$\begin{aligned}&(d_1 + d_2 + m^R_1 + m^L_1)\widetilde{F}^2_1\\ &\quad - \Big\{d_1 + d_2 + m^L_1 + (d_1 + m^R_1 + m^L_1)F_2\Big\}\, \widetilde{F}_1 + (d_1 + m^L_1)F_2 = 0 \end{aligned} \tag{11}$$

derived by $u_1(0, \widetilde{F}_1, F_2) = 0$ is $\widehat{F}^f(F_2) = \widetilde{F}_1$, where another solution of (11) does not satisfy (4) $(0 \le \widetilde{F}_1 \le F_2)$.

*On the backward nuisance part:* Finding $\widehat{F}^b(F_m)$ is performed similarly as in Algorithm 1 for $\widehat{F}^f(F_m)$. To solve the $(k-m-1)$ equations $u_j = 0$, $j = m+1, \ldots, k-1$, note that a solution $F_{j+1}$ of the equation $u_j(F_{j-1}, F_j, F_{j+1}) = 0$ is provided as

$$\begin{aligned}&\overline{F}^{b}_{j+1}(F_j, F_{j-1})\\ &\quad = \left[1 + \frac{d_{j+1}(1-F_j)(F_{j+1}-F_j)}{d_j(1-F_j)F_j + (F_j - F_{j-1})\Big(m^L_j(1-F_j) - m^R_j F_j\Big)}\right] F_j,\end{aligned} \tag{12}$$

given $F_j$ and $F_{j-1}$. When $m \le k-3$, by setting an $\widetilde{F}_{m+1}$ under a fixed $F_m$, $\widetilde{F}_{m+2} = \overline{F}^{b}_{m+2}(\widetilde{F}_{m+1}, F_m)$ for $u_{m+1} = 0$, $\widetilde{F}_{m+3} = \overline{F}_{m+3}(\widetilde{F}_{m+2}, \widetilde{F}_{m+1})$ for $u_{m+2} = 0, \ldots,$ $\widetilde{F}_{k-1} = \overline{F}_{k-1}(\widetilde{F}_{k-2}, \widetilde{F}_{k-3})$ for $u_2 = 0$ are automatically obtained. $u_{k-1}(\widetilde{F}_{k-2}, \widetilde{F}_{k-1}, 1) = 0$ is a criterion to be $\widehat{F}^b(F_m)$ and know whether $\widetilde{F}_{m+1}$ set in the first is appropriate. Algorithm 2 is a Newton–Raphson method to search $\widehat{F}^b(F_m)$.

ALGORITHM 2 (A computation to find $\widehat{F}^b(F_m)$ when $m \le k-3$)

B1. Let $l = 1$. Set an initial value $F_{m+1} = \widetilde{F}^{(1)}_{m+1}$ such that $\{\widetilde{F}^{(1)}_{m+1}, \widetilde{F}^{(1)}_{m+2}, \ldots, \widetilde{F}^{(1)}_{k-1}\}$ computed in B2 satisfies (4).

B2. Using (12) for the equations $u_j(\widetilde{F}^{(l)}_{j-1}, \widetilde{F}^{(l)}_j, \widetilde{F}^{(l)}_{j+1}) = 0$, $j = m+1, \ldots, k-2$ and $\widetilde{F}^{(l)}_m = F_m$,

$$\widetilde{F}^{(l)}_{m+2} = \overline{F}^b_{m+2}\big(\widetilde{F}^{(l)}_{m+1}, F_m\big)\widetilde{F}^{(l)}_{m+3} = \overline{F}^b_{m+3}\big(\widetilde{F}^{(l)}_{m+2}, \widetilde{F}^{(l)}_{m+1}\big), \ldots$$
$$\ldots, \widetilde{F}^{(l)}_{k-1} = \overline{F}^b_{k-1}\big(\widetilde{F}^{(l)}_{k-2}, \widetilde{F}^{(l)}_{k-3}\big)$$

are computed, so that $\widetilde{u}_{k-1} = u_{k-1}(\widetilde{F}^{(l)}_{k-2}, \widetilde{F}^{(l)}_{k-1}, 1)$ is computed. If $\{F_m, \widetilde{F}^{(l)}_{m+1}, \widetilde{F}^{(l)}_{m+2}, \ldots, \widetilde{F}^{(l)}_{k-1}\}$ satisfies (4), go to B3 as $h = 1$. Otherwise, go to B3 after letting $h \leftarrow 2h$ and $l \leftarrow l-1$.

B3. Stop if $\widetilde{u}_{h-1} < \epsilon$ and $h = 1$, so that we have $\widehat{F}^b(F_m) = (\widetilde{F}^{(l)}_{m+1}, \ldots, \widetilde{F}^{(l)}_{h-1})$. Otherwise, $\widetilde{F}^{(l+1)}_{m+1}$ is updated as

$$\widetilde{F}^{(l+1)}_{m+1} = \widetilde{F}^{(l)}_{m+1} + h^{-1}u_{k-1}\big(\widetilde{F}^{(l)}_{k-2}, \widetilde{F}^{(l)}_{k-1}, 1\big)\Big/\bar{i}_{k-1}\big(\widetilde{F}^{(l)}_{m+1}\big),$$

then go to B2 after letting $l \leftarrow l+1$, where $\bar{i}_{k-1}\big(\widetilde{F}^{(l)}_{m+1}\big) = du_{k-1}(\widetilde{F}^{(l)}_{k-2}, \widetilde{F}^{(l)}_{k-1}, 1)/d\widetilde{F}^{(l)}_{m+1}$.

The reason for resetting $h \leftarrow 2h$ in B2 is the same as for Algorithm 1. The first derivative $\bar{i}_{k-1}(\widetilde{F}^{(l)}_{m+1})$ of $u_{k-1}(\widetilde{F}^{(l)}_{k-2}, \widetilde{F}^{(l)}_{k-1}, 1)$ w.r.t. $\widetilde{F}^{(l)}_{m+1}$ is written as

$$\bar{i}_{k-1}(\widetilde{F}^{(l)}_{m+1}) = \frac{\partial u_{k-1}\big(\widetilde{F}^{(l)}_{k-2}, \widetilde{F}^{(l)}_{k-1}, 1\big)}{\partial \widetilde{F}^{(l)}_{k-2}} \frac{d\widetilde{F}^{(l)}_{k-2}}{d\widetilde{F}^{(l)}_{m+1}} + \frac{\partial u_{k-1}\big(\widetilde{F}^{(l)}_{k-2}, \widetilde{F}^{(l)}_{k-1}, 1\big)}{\partial \widetilde{F}^{(l)}_{k-1}} \frac{d\widetilde{F}^{(l)}_{k-1}}{d\widetilde{F}^{(l)}_{m+1}},$$

where $d\widetilde{F}^{(l)}_{k-2}/d\widetilde{F}^{(l)}_{m+1}$ and $d\widetilde{F}^{(l)}_{k-1}/d\widetilde{F}^{(l)}_{m+1}$ are obtained by the chain rule of differentiating (12), starting from

$$d\widetilde{F}^{(l)}_{m+2}/d\widetilde{F}^{(l)}_{m+1} = d\overline{F}^b_{m+2}(\widetilde{F}^{(l)}_{m+1}, F_m)/d\widetilde{F}^{(l)}_{m+1}$$

and following the computing order of $j = m+3, \ldots, k-2, k-1$ provided as

$$\frac{d\widetilde{F}^{(l)}_j}{d\widetilde{F}^{(l)}_{m+1}} = \frac{\partial \overline{F}^b_j\big(\widetilde{F}^{(l)}_{j+2}, \widetilde{F}^{(l)}_{j+1}\big)}{\partial \widetilde{F}^{(l)}_{j+1}} \frac{d\widetilde{F}^{(l)}_{j+1}}{d\widetilde{F}^{(l)}_{m+1}} + \frac{\partial \overline{F}^b_j\big(\widetilde{F}^{(l)}_{j+2}, \widetilde{F}^{(l)}_{j+1}\big)}{\partial \widetilde{F}^{(l)}_{j+2}} \frac{d\widetilde{F}^{(l)}_{j+2}}{d\widetilde{F}^{(l)}_{m+1}}.$$

When $m = k-2$, one direct solution $\widetilde{F}_{k-1}$ of the quadratic equation

$$(d_k + d_{k-1} + m^R_{k-1} + m^L_{k-1})\widetilde{F}^2_{k-1}$$
$$- \Big\{d_{k-1} + m^L_{k-1} + (d_k + m^R_{k-1} + m^L_{k-1})F_{k-2}\Big\}\widetilde{F}_{k-1} + m^L_{k-1}F_{k-2} = 0 \qquad (13)$$

from $u_{k-1}(F_{k-2}, \widetilde{F}_{k-1}, 1) = 0$ is $\widehat{F}^b(F_{k-2}) = \widetilde{F}_{k-1}$, where another of (13) does not satisfy (4) $(\widetilde{F}_{k-2} \le F_{k-1} \le 1)$.

If the jump points of (j3) occur, we can cope with such a situation by a slight correction of the algorithm described here. For example, if we need to consider the possibility of $F_{j-1} < F_j < F_{j+1}$ when $d_j = 0$ in (10), we modify Algorithm 1 by setting not only $F_{m-1}$ and $\widetilde{u}_1 = 0$ but also $F_{j-1}$ and $u_j(F_{j-1}, \widetilde{F}_j^{(l)}, \widetilde{F}_{j+1}^{(l)}) = 0$ as initial values and the criterion of convergence, respectively. Also, the correction of Algorithm 2 with the jump points of (j3) is similar to that of Algorithm 1.

## 3 A Wald-type variance formula based on derivations

In empirical likelihood with higher dimensional nuisance parameter, a Wald-type inference, for example, test, the estimated variance and associated confidence regions of parameter of interest may be computationally hard for the drawback in finding the inverse of the observed value of the full Fisher information matrix (e.g., Murphy and van der Vaart 1997). In fact, a direct application of Turnbull's formula to large samples brings blowing up numerically in the values of the determinants themselves. However, a Wald-type inference does not need to use the explicit expression of inverse matrix formula is motivated by a Newton–Raphson algorithm as in Sect. 2.3 and several recent theoretical works on semiparametric profile likelihood. Our idea is based on the chain rule of differentiating the equations required to construct the profile likelihood.

For practice, assume that the NPMLE $\widehat{F} = (\widehat{F}_1, \ldots, \widehat{F}_{k(\widehat{J})-1})$ has already been found. By Lemma 1, the condition to obtain the NPMLE $\widehat{F}$ is equivalent to keeping the $k(\widehat{J}) - 1$ equations

$$u_j(\widehat{F}_{j-1}, \widehat{F}_j, \widehat{F}_{j+1}) = 0, \quad j = 1, \ldots, k(\widehat{J}) - 1. \tag{14}$$

So, for simplicity only on notations, we write $m = m(\widehat{J})$ and $k = k(\widehat{J})$ by dropping $\widehat{J}$, jump points of the NPMLE. We are interested in estimating the pointwise variance of each $\widehat{F}_m$ $(m = 1, \ldots, k-1)$. Define $\mathcal{I}_p(F_m) = -\partial^2 pl(F_m)/\partial F_m^2, m = 1, \ldots, k-1$. A Taylor expansion of $2pl(F_m)$ at the true parameter $F_m^* = F^*(\widehat{J}_m)$ around $\widehat{F}_m$ provides

$$2pl(\widehat{F}_m) - 2pl(F_m^*) = n^{-1}\mathcal{I}_p(\widetilde{F}_m)\left\{\sqrt{n}(\widehat{F}_m - F_m^*)\right\}^2,$$

since the first derivative term is zero when $F_m = \widehat{F}_m$, where $\widetilde{F}_m$ is on the line segment between $\widehat{F}_m$ and $F_m^*$. Gu and Zhang (1993) or Chang (1990) proved that $\sqrt{n}(\widehat{F}(\cdot) - F^*(\cdot))$ converges weakly to a Gaussian process under some regular conditions. Also, Murphy and van der Vaart (1997, Theorem 2.1) showed that the log profile likelihood ratio $2pl(\widehat{F}(\tau)) - 2pl(F^*(\tau))$ is asymptotically chi-squared distributed with one degree of freedom under similar regular conditions. From these facts, we can predict that $1/n^{-1}\mathcal{I}_p(\widehat{F}_m)$ is available as a variance estimate of $\sqrt{n}(\widehat{F}_m - F_m^*)$. We now derive a useful expression of the Wald-type variance estimator based on differentiating the profile likelihood. Also, we provide its asymptotic form in Theorems 1 and 2 via lines different from Chang (1990).

Minus the second derivative $\mathcal{I}_p(F_m)$ of $pl(F_m)$ is, in a general form, written as

$$\mathcal{I}_p(F_m) = -\frac{\partial}{\partial F_m}\sum_{j=1}^{k-1}\frac{\partial l(\widehat{F}^f(F_m), F_m, \widehat{F}^b(F_m))}{\partial \widehat{F}_j(F_m)}\frac{\partial \widehat{F}_j(F_m)}{\partial F_m}$$
$$=\sum_{j=1}^{k-1}\left\{\sum_{l=1}^{k-1} i_{j,l}(\widehat{F}_{j-1}(F_m), \widehat{F}_j(F_m), \widehat{F}_{j+1}(F_m))\frac{\partial \widehat{F}_l(F_m)}{\partial F_m}\right\}\frac{\partial \widehat{F}_j(F_m)}{\partial F_m} \qquad (15)$$

since we have $u_j(\widehat{F}_{j-1}(F_m), \widehat{F}_j(F_m), \widehat{F}_{j+1}(F_m)) = 0$ for $j = 1, \ldots, m-1, m+1, \ldots, k-1$, where $F_m$ is written as $\widehat{F}_m(F_m)$ on notation. Similarly to the algorithm of Sect. 2.3, we find that the derivatives $\partial\widehat{F}_l(F_m)/\partial F_m$ in (15) can be computed separately in cases of $j \le m-1$ and $j \ge m+1$. Especially, when $F_m = \widehat{F}_m$, all the $\widehat{F}_j(\widehat{F}_m)$'s are

$$\widehat{F}_j = \widehat{F}_j(\widehat{F}_m), \quad j = 1, \ldots, k-1, \ \ m = 1, \ldots, k-1. \qquad (16)$$

By Lemma 3 and Sect. 2.3, note that the condition to obtain $\widehat{F}^f(F_m)$ is

$$u_j(\widehat{F}_{j-1}(F_m), \widehat{F}_j(F_m), \widehat{F}_{j+1}(F_m)) = 0, \ \text{ for } \ j = 1, \ldots, m-1 \qquad (17)$$

rather than (14). For simplicity, we write $\partial\widehat{F}_j/\partial\widehat{F}_{j+1} = \partial\widehat{F}_j(F_m)/\partial\widehat{F}_{j+1}(F_m)$. First, we have

$$\partial\widehat{F}_1/\partial\widehat{F}_2 = -i_{1,2}\left(0, \widehat{F}_1(F_m), \widehat{F}_2(F_m)\right)/i_{1,1}(0, \widehat{F}_1(F_m), \widehat{F}_2(F_m))$$

by differentiating $u_1(0, \widehat{F}_1(F_m), \widehat{F}_2(F_m)) = 0$ of (17). Next, given $\partial\widehat{F}_{j-1}/\partial\widehat{F}_j$ $(j \ge 2)$, we have

$$\frac{\partial\widehat{F}_j}{\partial\widehat{F}_{j+1}} = \frac{-i_{j,j+1}(\widehat{F}_{j-1}(F_m), \widehat{F}_j(F_m), \widehat{F}_{j+1}(F_m))}{i_{j,j}(\widehat{F}_{j-1}(F_m), \widehat{F}_j(F_m), \widehat{F}_{j+1}(F_m)) - \dfrac{d_j}{(\widehat{F}_j(F_m) - \widehat{F}_{j-1}(F_m))^2}\dfrac{\partial\widehat{F}_{j-1}}{\partial\widehat{F}_j}} \qquad (18)$$

by differentiating $u_j(\widehat{F}_{j-1}(F_m), \widehat{F}_j(F_m), \widehat{F}_{j+1}(F_m)) = 0$ of (17). Therefore, since all $\partial\widehat{F}_j/\partial\widehat{F}_{j+1}$, $j \le m-1$ are solved via (18) using all of (17), so that we can obtain

$$\frac{\partial\widehat{F}_j(F_m)}{\partial F_m} = \frac{\partial\widehat{F}_j}{\partial\widehat{F}_{j+1}}\frac{\partial\widehat{F}_{j+1}}{\partial\widehat{F}_{j+2}}\cdots\frac{\partial\widehat{F}_{m-1}}{\partial F_m} \quad \text{for } j \le m-1 \qquad (19)$$

by the chain rule of differentiation. In case of $j \ge m+1$, $\widehat{F}^b(F_m)$ holds the condition

$$u_j\left(\widehat{F}_{j-1}(F_m), \widehat{F}_j(F_m), \widehat{F}_{j+1}(F_m)\right) = 0, \quad j = m+1, \ldots, k-1. \qquad (20)$$

So, as well as the case of $j \leq m-1$, given $\partial \widehat{F}_{j+1}/\partial \widehat{F}_j$ $(j \leq k-1)$, we have

$$\frac{\partial \widehat{F}_j}{\partial \widehat{F}_{j-1}} = \frac{-i_{j,j-1}\left(\widehat{F}_{j-1}(F_m), \widehat{F}_j(F_m), \widehat{F}_{j+1}(F_m)\right)}{i_{j,j}\left(\widehat{F}_{j-1}(F_m), \widehat{F}_j(F_m), \widehat{F}_{j+1}(F_m)\right) - \dfrac{d_{j+1}}{\left(\widehat{F}_{j+1}(F_m) - \widehat{F}_j(F_m)\right)^2}\dfrac{\partial \widehat{F}_{j+1}}{\partial \widehat{F}_j}}, \tag{21}$$

where $\partial \widehat{F}_k/\partial \widehat{F}_{k-1} = 0$ for $\widehat{F}_k = 1$. Therefore, since all $\partial \widehat{F}_j/\partial \widehat{F}_{j-1}$, $j \geq m+1$ are solved via (21) using (20), so that we can obtain

$$\frac{\partial \widehat{F}_j(F_m)}{\partial F_m} = \frac{\partial \widehat{F}_j}{\partial \widehat{F}_{j-1}} \frac{\partial \widehat{F}_{j-1}}{\partial \widehat{F}_{j-2}} \cdots \frac{\partial \widehat{F}_{m+1}}{\partial F_m} \quad \text{for} \quad j \geq m+1 \tag{22}$$

by the chain rule of differentiation. We provide the following result relevant to the existence of these derivatives:

**Lemma 4** *Suppose that* (2) *is held. Then, under a fixed* $F_m$ *and the constraint condition* (4), $\partial \widehat{F}_j(F_m)/\partial F_m$, $j = 1, \ldots, k-1$ *uniquely exist satisfying* $0 < \partial \widehat{F}_j(F_m)/\partial F_m < 1$.

*Proof of Lemma 4* By assumption of (2), all the $d_j$'s $(j = 1, \ldots, k-1)$ are not zeros. First, we consider the case of $l \leq m-1$. Clearly, there exists $0 < \partial \widehat{F}_1/\partial \widehat{F}_2 < 1$. Since the denominator of (18) exists if $\widehat{F}_j(F_m) - \widehat{F}_{j-1}(F_m) > 0$ at $d_j > 0$ and can be written as

$$\frac{d_j(1 - \partial \widehat{F}_{j-1}/\partial \widehat{F}_j)}{(\widehat{F}_j(F_m) - \widehat{F}_{j-1}(F_m))^2} - i_{j,j+1}(\widehat{F}_{j-1}(F_m), \widehat{F}_j(F_m), \widehat{F}_{j+1}(F_m)) + m_j^R/(1 - \widehat{F}_j(F_m))^2 + m_j^L/\widehat{F}_j(F_m)^2,$$

it is inductively shown that there exist $\partial \widehat{F}_j(F_m)/\partial F_m$, $j = 1, \ldots, m-1$ satisfying $0 < \partial \widehat{F}_{j-1}/\partial \widehat{F}_j < 1$. Also, similar results are shown in the case of $j \geq m+1$. □

Even if (2) is not held in Lemma 4, $\partial \widehat{F}_j(F_m)/\partial F_m = 0$ occurs at $d_j = 0$ but $0 \leq \partial \widehat{F}_j(F_m)/\partial F_m \leq 1$, $j = 1, \ldots, k-1$ are satisfied.

Further, using the relationship

$$i_{j,j}(F_{j-1}, F_j, F_{j+1}) = -i_{j,j-1}(F_{j-1}, F_j, F_{j+1}) - i_{j,j+1}(F_{j-1}, F_j, F_{j+1}) + m_j^R/(1 - F_j)^2 + m_j^L/F_j^2, \tag{23}$$

$\partial \widehat{F}_0/\partial F_m = 0$ and $\partial \widehat{F}_k/\partial F_m = 0$ for $m = 1, \ldots, k-1$, we can rewrite (15) as

$$\mathcal{I}_p(F_m) = \sum_{j=1}^{k} \frac{d_j}{\{\widehat{F}_j(F_m) - \widehat{F}_{j-1}(F_m)\}^2} \left( \frac{\partial \widehat{F}_j(F_m)}{\partial F_m} - \frac{\partial \widehat{F}_{j-1}(F_m)}{\partial F_m} \right)^2 + \sum_{j=1}^{k-1} \frac{m_j^R}{\{1-\widehat{F}_j(F_m)\}^2} \left( \frac{\partial \widehat{F}_j(F_m)}{\partial F_m} \right)^2 + \sum_{j=1}^{k-1} \frac{m_j^L}{\widehat{F}_j(F_m)^2} \left( \frac{\partial \widehat{F}_j(F_m)}{\partial F_m} \right)^2 \quad (24)$$

with better insight than Turnbull's determinant-based formula.

We now provide the asymptotic consistency of $\mathcal{I}_p(F_m)$ at $F_m = \widehat{F}_m$. Let $t_-$ and $t_+$ be times just after $t$ and just prior to $t$, respectively.

CONDITION 1 $F^*$, $F^L$ and $F^R$ are continuous functions with $T_e = \inf\{t : F^*(t) = 1\} < \infty$, $\sup\{t : F^*(t) = 0\} = 0$, $F^L(T_e) = 1$ and $F^R(0) = 0$. Further, $F^L(t_-) - F^R(t_-)$ is positive on $t \in (0, T_e)$.

This condition is almost the same as that of Murphy and van der Vaart (1997, Theorem 2.1) or Chang (1990), but we impose on the boundedness of $T_e$ to prove the following theorem simply based on the linear Fredholm integral equations.

**Theorem 1** *Suppose that Condition 1 is satisfied, and that $t_0$ and $t_e$ hold $0 < t_0 < t_e < T_e$. Then, for $m$ and $\tau$ satisfying $\widehat{J}_m \le \tau < \widehat{J}_{m+1}$, as $n \to \infty$, $n^{-1}\mathcal{I}_p(\widehat{F}_m)$ is consistent to the limit*

$$\mathscr{I}_p(\widehat{F}(\tau)) = \int_0^{T_e} dH_1(t) \left\{ \frac{1}{\widehat{F}^*(t) - \widehat{F}^*(t_-)} \frac{\partial(\widehat{F}^*(t) - \widehat{F}^*(t_-))}{\partial \widehat{F}^*(\tau)} \right\}^2 + \int_0^{T_e-} \frac{dH_2(t)}{\{1 - F^*(t)\}^2} \left( \frac{\partial \widehat{F}^*(t)}{\partial \widehat{F}^*(\tau)} \right)^2 + \int_{0+}^{T_e} \frac{dH_3(t)}{F^*(t)^2} \left( \frac{\partial \widehat{F}^*(t)}{\partial \widehat{F}^*(\tau)} \right)^2 \quad (25)$$

*uniformly on $\tau \in [t_0, t_e]$, where $H_l(t)$, $l = 1, 2, 3$ are $H_1(t) = \int_0^t (F^L(s) - F^R(s))dF^*(s)$, $H_2(t) = \int_0^t (1 - F^*(s))dF^R(s)$, $H_3(t) = \int_0^t F^*(s)dF^L(s)$, and*

$$\{\widehat{F}^*(t) - \widehat{F}^*(t_-)\}^{-1} \partial(\widehat{F}^*(t) - \widehat{F}^*(t_-))/\partial \widehat{F}^*(\tau) \quad \text{and} \quad \partial \widehat{F}^*(t)/\partial \widehat{F}^*(\tau)$$

*are deterministic limits corresponding to*

$$\{\widehat{F}(t) - \widehat{F}(t_-)\}^{-1} \partial(\widehat{F}(t) - \widehat{F}(t_-))/\partial \widehat{F}(\tau) \quad \text{and} \quad \partial \widehat{F}(t)/\partial \widehat{F}(\tau),$$

*respectively, $\widehat{F}^*(s) = F^*(s)$ in the sense of the function value. The limits of derivatives are expressed as $\partial \widehat{F}^*(t)/\partial \widehat{F}^*(\tau) = \prod_{t \le s < \tau} \partial \widehat{F}^*(s)/\partial \widehat{F}^*(s_+)$ satisfying the non-linear Volterra equations (30) if $t < \tau$ and $\partial \widehat{F}^*(t)/\partial \widehat{F}^*(\tau) = \prod_{\tau < s \le t} \partial \widehat{F}^*(s)/\partial \widehat{F}^*(s_-)$ satisfying (31) if $t > \tau$.*

*Proof of Theorem 1* For simplicity, we write $\partial\widehat{F}_j/\partial\widehat{F}_m = \partial\widehat{F}_j(F_m)/\partial F_m|_{F_m=\widehat{F}_m}$ and $\widehat{p}_j = \widehat{F}_j - \widehat{F}_{j-1}$. The main task to prove this theorem is that for every $j$ and $m$, as $n \to \infty$

$$\partial\widehat{F}_j/\partial\widehat{F}_m \text{ and } (\widehat{p}_j)^{-1}\partial\widehat{p}_j/\partial\widehat{F}_m = (\widehat{p}_j)^{-1}(\partial\widehat{F}_j/\partial\widehat{F}_m - \partial\widehat{F}_{j-1}/\partial\widehat{F}_m) \quad \text{at } d_j > 0$$

converge almost surely to their limits, respectively. Let $\widehat{x}_j = \partial\widehat{F}_j/\partial\widehat{F}_m$, $\widehat{x}^f = (\widehat{x}_1, \ldots, \widehat{x}_{m-1})^{\mathrm{T}}$, $\widehat{y}_j = (\widehat{p}_j)^{-1}\partial\widehat{p}_j/\partial\widehat{F}_m$ and $\widehat{y}^f = (\widehat{y}_1, \ldots, \widehat{y}_{m-1})^{\mathrm{T}}$. Let $\widehat{\mathcal{I}}_\tau^f$ be a submatrix of $\mathcal{I}(\widehat{F})$ formed by rows and columns $1, \ldots, m-1$. We find that $\widehat{x}^f$ solved recursively by (18) and (19) is equivalent to the solution of linear equations

$$n^{-1}\widehat{\mathcal{I}}_\tau^f\widehat{x}^f = \widehat{a}_\tau^f,$$

where $\widehat{a}_\tau^f = (0, \ldots, 0, -\widehat{i}_{m-1,m}/n)^{\mathrm{T}}$ and $\widehat{i}_{i,j} = i_{i,j}|_{F=\widehat{F}}$. These equations are derived by differentiating the both sides of (17) w.r.t. $F_m$. Also, $\widehat{x}^f$ is same as $\mathcal{H}_{m-1}^{\mathrm{T}}\widehat{\mathcal{Q}}_{m-1}\widehat{y}^f$, where $\mathcal{H}_{m-1}$ is a submatrix of $\mathcal{H}$ (of Lemma 2) formed by rows and columns $1, \ldots, m-1$ and $\widehat{\mathcal{Q}}_{m-1}$ is an $(m-1)\times(m-1)$ diagonal matrix $\mathrm{diag}(\widehat{p}_1, \ldots, \widehat{p}_{m-1})$.

First, we show that $\widehat{y}_j$'s at $d_j > 0$ converge almost surely to their limits as $n \to \infty$. Especially, since $\widehat{p}_j$, $j = 1, \ldots, k-1$ are positive on all the jump points $\{\widehat{J}_1, \ldots, \widehat{J}_{k-1}\}$ of the NPMLE, all the $\widehat{y}_j$'s are well-defined not only at $d_j > 0$ but also at all $j$'s. By some facts mentioned above, $\widehat{y}^f$ is the unique solution of $n^{-1}\widehat{\mathcal{I}}_\tau^f\mathcal{H}_{m-1}^{\mathrm{T}}\widehat{\mathcal{Q}}_{m-1}\widehat{y}^f = \widehat{a}_\tau^f$. In fact, the inverse of $n^{-1}\widehat{\mathcal{I}}_\tau^f\mathcal{H}_{m-1}^{\mathrm{T}}\widehat{\mathcal{Q}}_{m-1}$ exists, because $\widehat{\mathcal{I}}_\tau^f$ is invertible from the discussion of Appendix A.1. Here, multiplying this matrix system by an invertible $\widehat{\mathcal{P}}_{m-1}\widehat{\mathcal{R}}_{m-1}$, we consider the equations

$$n^{-1}\widehat{\mathcal{P}}_{m-1}\widehat{\mathcal{R}}_{m-1}\widehat{\mathcal{I}}_\tau^f\mathcal{H}_{m-1}^{\mathrm{T}}\widehat{\mathcal{Q}}_{m-1}\widehat{y}^f = \widehat{\mathcal{P}}_{m-1}\widehat{\mathcal{R}}_{m-1}\widehat{a}_\tau^f, \tag{26}$$

where $\widehat{\mathcal{P}}_{m-1}$ is a submatrix of $\mathcal{P}(\widehat{F})$ (of Lemma 2) formed by rows and columns $1, \ldots, m-1$ and $\widehat{\mathcal{R}}_{m-1}$ is an $(m-1)\times(m-1)$ matrix such that

$$\widehat{\mathcal{R}}_{m-1} = \begin{pmatrix} 1 & 1 & \cdots & 1 & \widehat{\nu}_{m-1} \\ 0 & 1 & \cdots & 1 & \widehat{\nu}_{m-1} \\ \vdots & 0 & \ddots & \vdots & \vdots \\ 0 & \ddots & \ddots & 1 & \vdots \\ 0 & 0 & \cdots & 0 & \widehat{\nu}_{m-1} \end{pmatrix}, \quad \widehat{\nu}_{m-1} = \frac{\widehat{p}_m + \widehat{\lambda}_{m-1}}{1+\widehat{\lambda}_{m-1}}, \quad \widehat{\lambda}_{m-1} = \frac{\widehat{\beta}_{m-1}}{\widehat{\alpha}_m}$$

with $\widehat{\alpha}_m = d_m/\widehat{p}_m^2$ and $\widehat{\beta}_{m-1} = m_{m-1}^R/(1-\widehat{F}_{m-1})^2 + m_{m-1}^L/\widehat{F}_{m-1}^2$. The equations (26) are also equivalent to the Fredholm integral equation

$$\int_0^{\widehat{J}_{m-}} \widehat{K}_\tau^f(t,s)\widehat{y}^f(s)ds = \widehat{A}_\tau^f(t)$$

in the sense of Stieltjes's integration, where $\widehat{y}^f(s)$ is the $j$th element of $\widehat{y}^f$ satisfying $\widehat{J}_j \le s < \widehat{J}_{j+1}$, $\widehat{A}^f_\tau(t)$ and $\widehat{K}^f_\tau(t,s)ds$ are written as

$$\widehat{A}^f_\tau(t) = \widehat{F}(t)\left\{d\widehat{H}_1(\widehat{J}_m)/\Delta\widehat{F}(\widehat{J}_m) + \int_{\widehat{J}_{m-1}}^{\widehat{J}_{m-}} \widehat{\beta}(u)du\right\}/(1+\widehat{\lambda}_{m-1}) \quad \text{and}$$

$$\widehat{K}^f_\tau(t,s)ds = \left\{I(t\ge s)\frac{d\widehat{H}_1(s)}{\Delta\widehat{F}(s)} + \widehat{F}(t)\frac{d\widehat{H}_1(\widehat{J}_m)}{\Delta\widehat{F}(\widehat{J}_m)} + \int_s^{\widehat{J}_{m-}} \widehat{F}(t\wedge u)\widehat{\beta}(u)du\right\}d\widehat{F}(s)$$
$$-I(s=\widehat{J}_{m-})\widehat{F}(t)\left\{\frac{d\widehat{H}_1(\widehat{J}_{m-1})}{\Delta\widehat{F}(\widehat{J}_{m-1})}\right\}(1-\Delta\widehat{F}(\widehat{J}_m))/(1+\widehat{\lambda}_{m-1})$$

with $\widehat{\beta}(u)du = d\widehat{H}_2(u)/(1-\widehat{F}(u))^2 + d\widehat{H}_3(u)/\widehat{F}(u)^2$, $\Delta\widehat{F}(t) = \widehat{F}(t) - \widehat{F}(t_-)$, $\widehat{H}_1(t) = \sum_{\{i:\widehat{J}_i\le t\}} d_i/n$, $\widehat{H}_2(t) = \sum_{\{i:\widehat{J}_i\le t\}} m_i^R/n$, $\widehat{H}_3(t) = \sum_{\{i:\widehat{J}_i\le t\}} m_i^L/n$ and $t\wedge s = \min(t,s)$. Further, note that the NPMLE or the SCE $\widehat{F}$ satisfies

$$\frac{d\widehat{H}_1(t)}{\Delta\widehat{F}(t)} = (\widehat{F}^L(t_-) - \widehat{F}^R(t_-))\delta_1(t),$$
$$\widehat{F}^R(t) = \int_0^t d\widehat{H}_2(s)/(1-\widehat{F}(s)),\ 1-\widehat{F}^L(t) = \int_t^{T_e} d\widehat{H}_3(s)/\widehat{F}(s) \tag{27}$$

and $\delta_1(t) = I(d\widehat{H}_1(t) > 0)$.

Let $\int_0^{\tau-} K^f_\tau(t,s)y^f(s)ds = A^f_\tau(t)$ be a limit form corresponding to the above Fredholm integral equation or (26), where $y^f(s) = \{\Delta\widehat{F}^*(s)\}^{-1}\partial\Delta\widehat{F}^*(s)/\partial\widehat{F}^*(\tau)$, $\Delta\widehat{F}^*(s) = \widehat{F}^*(s) - \widehat{F}^*(s_-)$, $A^f_\tau(t) = F^*(t)(F^L(\tau_-) - F^R(\tau_-))$, $K^f_\tau(t,s)ds$ corresponds to $\widehat{K}^f_\tau(t,s)ds$ in which $\widehat{H}_l$, $l = 1,2,3$, $\widehat{F}$, $\widehat{J}_{m-1}$ and $\widehat{J}_m$ are substituted by their limits $H_l$, $l=1,2,3$, $F^*$, $\tau_-$ and $\tau$, as $\lambda_{m-1}\to 0$. The kernel satisfies that

$$\sup_{t,s<\tau\le t_e} |K^f_\tau(t,s)| < \infty, \tag{28}$$

which is shown from, for example, $\int_0^{\tau-} dH_2(u)/(1-F^*(u))^2 < F^R(\tau_-)/(1-F^*(t_{e-})) < \infty$, etc. Then, there exists the resolvent kernel corresponding to $\int_0^{\tau-} K^f_\tau(t,s)y^f(s)ds$. In fact, the determinant of finite difference approximations of this integral equation is proved to be always positive using a transformation to the matrix system such as (26) and the discussion similar to Appendix A.1, then the limit of such determinant converges to some deterministic value (the Fredholm determinant) by the Fredholm theory with the bounded kernel, such as (28).

If Condition 1 is satisfied, then, we have

$$\sup_{t\in[0,T_e]}|\widehat{F}(t) - F^*(t)| \to_{a.s.} 0,\quad \sup_{t\in[0,T_e]}|\widehat{H}_l(t) - H_l(t)| \to_{a.s.} 0,\quad l=1,2,3,$$
$$\sup_{t\in[0,T_e]}|\widehat{F}^L(t) - F^L(t)| \to_{a.s.} 0 \quad\text{and}\quad \sup_{t\in[0,T_e]}|\widehat{F}^R(t) - F^R(t)| \to_{a.s.} 0, \tag{29}$$

where $\rightarrow_{a.s.}$ denotes the convergence almost surely (see Chang and Yang 1987; Gu and Zhang 1993 or Murphy and van der Vaart 1997, Lemma A.3). By (28), we have $\sup_{t<\tau\le t_e}\left|\int_0^{\tau-} K_\tau^f(t,s)ds\right| < \infty$ and $\sup_{t<\tau\le t_e}|y^f(t)| < \infty$. So, using Condition 1, (27), (28) and (29), as $n \rightarrow \infty$, we can show

$$\sup_{t<\tau\le t_e}\left|\int_0^{\widehat{J}_m-}\widehat{K}_\tau^f(t,s)ds - \int_0^{\tau-}K_\tau^f(t,s)ds\right| \underset{a.s.}{\rightarrow} 0 \quad\text{and}\quad \sup_{t<\tau\le t_e}\left|\widehat{A}_\tau^f(t) - A_\tau^f(t)\right| \underset{a.s.}{\rightarrow} 0.$$

Also, the former result leads that the integrations of the resolvent kernel corresponding to $\int_0^{\widehat{J}_m-}\widehat{K}^f(t,s)\widehat{y}^f(s)ds$ converge almost surely to those of $\int_0^{\tau-}K^f(t,s)y^f(s)ds$ uniformly on $\tau \le t_e$ (e.g., by a reasoning similar to Sugimoto and Hamasaki 2006, Proposition A.4). Hence, based on these results on convergences, $\sup_{t<\tau\le t_e}|\widehat{y}^f(t) - y^f(t)| \rightarrow_{a.s.} 0$ can be shown. Also, we can show that $\widehat{y}^b = (\widehat{y}_{m+1},\ldots,\widehat{y}_{k-1})^{\mathrm{T}}$ converges almost surely to its limit by an argument parallel to $\widehat{y}^f$ and $y^f$ as stated above.

Second, we show that $\widehat{x}_j$'s converge almost surely to their limits as $n \rightarrow \infty$. Let $\widehat{x}^f(s)$ be the $j$th element of $\widehat{x}^f$ satisfying $\widehat{J}_j \le s < \widehat{J}_{j+1}$ and let $x^f(s) = \partial\widehat{F}^*(s)/\partial\widehat{F}^*(\tau)$ for $s < \tau$, then we have

$$\widehat{x}^f(t) = \int_0^t \widehat{y}^f(s)d\widehat{F}(s) \quad\text{and}\quad x^f(t) = \int_0^t y^f(s)dF^*(s),$$

where $\widehat{F}^*(s)$ is a notation used to define the differentiation, but $\widehat{F}^*(s) = F^*(s)$ is satisfied in the sense of the function value. Hence, using $\sup_{t<\tau\le t_e}|\widehat{y}^f(t) - y^f(t)| \rightarrow_{a.s.} 0$ and (29), $\sup_{t<\tau\le t_e}|\widehat{x}^f(t) - x^f(t)| \rightarrow_{a.s.} 0$ can be shown. Similarly, it can be proved that $\widehat{x}^b = (\widehat{x}_{m+1},\ldots,\widehat{x}_{k-1})^{\mathrm{T}}$ converges almost surely to its limit.

We have $\mathscr{I}_p(\widehat{F}(\tau)) < \infty$ on $\tau \in [t_0, t_e]$ by the boundednesses of $|y^f(t)|$, $|x^f(t)/F^*(t)|$, $|x^f(t)/(1-F^*(t))|$, etc., where for example, $|x^f(t)/F^*(t)| \le \int_0^t \sup|y^f(s)|d F^*(s)/F^*(t) = \sup|y^f(s)|$. Using (29), $\sup_{t,\tau}|\widehat{x}^f(t) - x^f(t)| \rightarrow_{a.s.} 0$, $\sup_{t,\tau}|\widehat{y}^f(t) - y^f(t)| \rightarrow_{a.s.} 0$, etc., $n^{-1}\mathcal{I}_p(\widehat{F}_m)$ converges almost surely to $\mathscr{I}_p(\widehat{F}(\tau))$ uniformly on $\tau \in [t_0, t_e]$. □

We establish that the Wald-type variance estimator is consistent to the asymptotic variance formula in the following theorem:

**Theorem 2** *Redefine $\mathscr{I}_p(\widehat{F}(\tau))$ as (25) on $\tau \in (0, T_e)$ and $\infty$ at $\tau = 0$ and $T_e$. Suppose that Condition 1 is satisfied. Then, for $m$ and $\tau$ satisfying $\widehat{J}_m \le \tau < \widehat{J}_{m+1}$, as $n \rightarrow \infty$, a variance estimator $1/n^{-1}\mathcal{I}_p(\widehat{F}_m)$ of $\sqrt{n}(\widehat{F}(\tau) - F^*(\tau))$ is consistent to $1/\mathscr{I}_p(\widehat{F}(\tau))$ uniformly on $\tau \in [0, T_e]$.*

*Proof of Theorem 2* By Theorem 1 and the continuous mapping property, as $n \rightarrow \infty$, we have $\sup_{\tau\in[t_0,t_e]}|1/n^{-1}\mathcal{I}_p(\widehat{F}_m) - 1/\mathscr{I}_p(\widehat{F}(\tau))| \rightarrow_{a.s} 0$. Clearly we can define $\partial\widehat{F}_j/\partial\widehat{F}_m = \infty$ at $m = 0$ and $k$, that is, if $\tau < \widehat{J}_1$ or $\tau \ge \widehat{J}_k$, so that we have $1/n^{-1}\mathcal{I}_p(\widehat{F}_m) = 0$.

For $t < \tau$, $x^f(t)$ is written as a product integral $x^f(t) = \prod_{t \le s < \tau} \partial \widehat{F}^*(s)/\partial \widehat{F}^*(s_+)$ with the relationship based on a nonlinear forward Volterra equation

$$\frac{\partial \widehat{F}^*(s)}{\partial \widehat{F}^*(s_+)} = \frac{dH_1(s_+)/\Delta F^*(s_+)^2}{\frac{dH_1(s_+)}{\Delta F^*(s_+)^2} + \left(1 - \frac{\partial \widehat{F}^*(s_-)}{\partial \widehat{F}^*(s)}\right)\frac{dH_1(s)}{\Delta F^*(s)^2} + \frac{dH_2(s)}{(1-F^*(s))^2} + \frac{dH_3(s)}{F^*(s)^2}}. \tag{30}$$

Also, for $\tau < t$, $x^b(t)$ is written as $\prod_{\tau < s \le t} \partial \widehat{F}^*(s)/\partial \widehat{F}^*(s_-)$ with the relationship based on a nonlinear backward Volterra equation

$$\frac{\partial \widehat{F}^*(s)}{\partial \widehat{F}^*(s_-)} = \frac{dH_1(s)/\Delta F^*(s)^2}{\frac{dH_1(s)}{\Delta F^*(s)^2} + \left(1 - \frac{\partial \widehat{F}^*(s_+)}{\partial \widehat{F}^*(s)}\right)\frac{dH_1(s_+)}{\Delta F^*(s_+)^2} + \frac{dH_2(s)}{(1-F^*(s))^2} + \frac{dH_3(s)}{F^*(s)^2}}. \tag{31}$$

Hence, we can show that $\mathscr{I}_p(\widehat{F}(\tau))$ is a continuous function on $\tau \in [t_0, t_e]$ by adding Condition 1 further. As $\tau \to T_{e-}$ or $\tau \to 0_+$, we observe $\mathscr{I}_p(\widehat{F}(\tau)) \to \infty$. In fact, for example as $\tau \to T_{e-}$ at $t = T_e$, the first integrand of (25) is

$$\begin{aligned}\lim_{\tau \to T_{e-},\, t=T_e} \left[\{\Delta \widehat{F}^*(t)\}^{-1} \frac{\partial \Delta \widehat{F}^*(t)}{\partial \widehat{F}^*(\tau)}\right]^2 dH_1(t) &= (F^L(T_{e-}) - F^R(T_{e-}))/dF^*(T_e) \\ &= \infty\end{aligned}$$

by $\partial \widehat{F}^*(T_e)/\partial \widehat{F}^*(t) = 0$. Similarly, as $\tau \to t_+$ at $t = 0_+$, we observe the situation of $\tau \to 0_+$. Therefore, Theorem 2 can be proved by adding the results of $\tau \to T_e$ and 0 via $\epsilon$-$\delta$ arguments. □

To see another aspect that $1/\mathcal{I}_p(\widehat{F}_m)$ is a reasonable variance estimate of $\widehat{F}_m$, we can provide that $1/\mathcal{I}_p(\widehat{F}_m)$ is identical to the Greenwood variance formula in right-censored data:

**Lemma 5** *In right-censored data, $1/\mathcal{I}_p(\widehat{F}_m)$ reduces to the Greenwood variance formula*

$$\sum_{j=1}^{m} d_j/r_j(r_j - d_j), \quad where \quad r_j = \sum_{l=j}^{k}(d_l + m_l^R).$$

In this paper, we do not provide a proof of Lemma 5, since this can be proved using the mathematical induction and via some algebra calculi, but such a proof is excessively long. However, for a brief check, readers will be able to confirm Lemma 5 numerically.

## 4 Simulation

The single parameter $F(\tau)$ of interest is equivalent to $F_m$ corresponding to $m$ which satisfies $\widehat{J}_m \leq \tau < \widehat{J}_{m+1}$, where the notations of $m$ and $k$ are $m(\widehat{J})$ and $k(\widehat{J})$ as in Sect. 3. Let $\widehat{V}^{\text{Wald}}(\widehat{F}(\tau)) = 1/\mathcal{I}_p(\widehat{F}_m)$ be the Wald-type variance estimate of $\widehat{F}(\tau)$ defined in Sect. 3. We carry out simulation studies to investigate the performance of $n\widehat{V}^{\text{Wald}}(\widehat{F}(\tau))$, which is a Wald-type variance estimate of $\sqrt{n}(\widehat{F}_m - F_m^*)$. As a competitor of $\widehat{V}^{\text{Wald}}(\widehat{F}(\tau))$, we set another estimate

$$\widehat{V}^{\text{LR}}(\widehat{F}(\tau)) = \exp\left\{\widetilde{h}^{-1}\sum_{l=1}^{\widetilde{h}}\log\left(\frac{(\widehat{F}_m - F_m^{(l)})^2}{2\{pl(\widehat{F}_m) - pl(F_m^{(l)})\}}\right)\right\}$$

based on the profile likelihood ratio for given $F_m = F_m^{(l)}, l = 1, \ldots, \widetilde{h}$.

Let $U(a, b)$ be a uniform random number on $[a, b]$. We generate doubly censored samples of size $n = 100, 400, 900, 1,600$ that the true event times $T_i^*, i = 1, \ldots, n$ are independently sampled from $U(0, 1)$ throughout, which is equivalent to $F^*(t) = 1-t$. We set three models for censoring distributions:

- In censoring model I, the left- and the right-censoring times are $C_i^L = \min(U_1^{\text{I}}, U_2^{\text{I}})$ and $C_i^R = \max(U_1^{\text{I}}, U_2^{\text{I}}), i = 1, \ldots, n$ for mutually independently sampled $U_1^{\text{I}} = U(0, 0.3)$ and $U_2^{\text{I}} = U(0.1, 1)$. Model I has the left- and right-censoring rates of about 15 and 44%.
- In censoring model II, we set $C_i^L = \min(U_1^{\text{II}}, U_2^{\text{II}})$ and $C_i^R = \max(U_1^{\text{II}}, U_2^{\text{II}})$, $i = 1, \ldots, n$ for mutually independently sampled $U_1^{\text{II}} = U(0, 0.9)$ and $U_2^{\text{II}} = U(0.1, 1)$. Both the left- and right-censoring rates in model II are about 34%.
- In censoring model III, let $C_i^L = U^{\text{III}}$ and $C_i^R = U^{\text{III}} + 0.4, i = 1, \ldots, n$ for independently sampled $U^{\text{III}} = U(0, 0.6)$. The left- and right-censoring rates in model III are both 30%.

We consider three different situations of moderate censoring via models I-III. Under these settings, we generate 1,000 simulated data in each model.

To compute $\widehat{V}^{\text{LR}}(\widehat{F}(\tau))$, let $\widetilde{h} = 5$ and each $F_m^{(l)}$ is generated as a random value around an NPMLE $\widehat{F}_m$. We use the algorithm of Sect. 2.3 and select five $F_m^{(l)}$'s which satisfy $pl(\widehat{F}_m) - pl(F_m^{(l)}) \leq 5$ to avoid extreme values of $F_m^{(l)}$ and achieve the convergence condition $\epsilon = 0.001$ described in Sect. 2.3. Here, for simplicity, the set $J$ of jump points to find $\widehat{F}^{\#}(F_m^{(l)})$ is restricted to $\widehat{J}$ only. On the other hand, we do not need a special setting to compute $\widehat{V}^{\text{Wald}}(\widehat{F}(\tau))$. That is, as soon as we find the NPMLE, we can easily obtain $\widehat{V}^{\text{Wald}}(\widehat{F}(\tau))$ by the plug-in procedure computing (24) using (18), (19), (21) and (22) for (16).

In Table 1, we show the simulated averages of $\widehat{F}(\tau)$, $n(\widehat{F}(\tau) - F^*(\tau))^2$, $n\widehat{V}^{\text{Wald}}(\widehat{F}(\tau))$ and $n\widehat{V}^{\text{LR}}(\widehat{F}(\tau))$ and simulated SDs (standard deviations) of $n\widehat{V}^{\text{Wald}}(\widehat{F}(\tau))$ and $n\widehat{V}^{\text{LR}}(\widehat{F}(\tau))$ at $\tau = 0.1, 0.2, \ldots, 0.9$ obtained from 1,000 samples of size $n$. We can see that the simulated averages of $n\widehat{V}^{\text{Wald}}(\widehat{F}(\tau))$ appropriately approximate to those of $n(\widehat{F}(\tau) - F^*(\tau))^2$ except some points of $\tau$'s nearer to both ends of 0

**Table 1** Simulated averages of $\widehat{F}(\tau)$, $n(\widehat{F}(\tau) - F^*(\tau))^2$, $n\widehat{V}^{\text{Wald}}(\widehat{F}(\tau))$ and $n\widehat{V}^{\text{LR}}(\widehat{F}(\tau))$ and simulated SDs of $n\widehat{V}^{\text{Wald}}(\widehat{F}(\tau))$ and $n\widehat{V}^{\text{LR}}(\widehat{F}(\tau))$

Outputs for censoring model I

| | | Averages | | | | SDs | |
|---|---|---|---|---|---|---|---|
| $n$ | $\tau$ | $\widehat{F}$ | $n(\widehat{F} - F^*)^2$ | $n\widehat{V}^{\text{Wald}}$ | $n\widehat{V}^{\text{LR}}$ | $n\widehat{V}^{\text{Wald}}$ | $n\widehat{V}^{\text{LR}}$ |
| 100 | 0.1 | 0.0982 | 0.3575 | 0.1606 | 0.1559 | 0.0415 | 0.0419 |
| | 0.2 | 0.2113 | 0.1879 | 0.1819 | 0.1772 | 0.0256 | 0.0261 |
| | 0.3 | 0.3094 | 0.2108 | 0.2300 | 0.2261 | 0.0206 | 0.0213 |
| | 0.4 | 0.4033 | 0.2787 | 0.2839 | 0.2813 | 0.0211 | 0.0217 |
| | 0.5 | 0.5010 | 0.3173 | 0.3326 | 0.3314 | 0.0245 | 0.0246 |
| | 0.6 | 0.6006 | 0.3541 | 0.3732 | 0.3731 | 0.0367 | 0.0364 |
| | 0.7 | 0.7017 | 0.4462 | 0.4052 | 0.4063 | 0.0646 | 0.0642 |
| | 0.8 | 0.8155 | 0.7873 | 0.4209 | 0.4228 | 0.1096 | 0.1091 |
| | 0.9 | 0.9424 | 0.8001 | 0.4160 | 0.4185 | 0.1654 | 0.1652 |
| 400 | 0.1 | 0.1040 | 0.1529 | 0.1535 | 0.1489 | 0.0225 | 0.0227 |
| | 0.2 | 0.2027 | 0.1678 | 0.1786 | 0.1739 | 0.0130 | 0.0132 |
| | 0.3 | 0.3024 | 0.2158 | 0.2287 | 0.2248 | 0.0111 | 0.0114 |
| | 0.4 | 0.4016 | 0.2639 | 0.2870 | 0.2844 | 0.0105 | 0.0108 |
| | 0.5 | 0.5012 | 0.2956 | 0.3364 | 0.3352 | 0.0123 | 0.0124 |
| | 0.6 | 0.6014 | 0.3476 | 0.3767 | 0.3768 | 0.0180 | 0.0178 |
| | 0.7 | 0.7020 | 0.4008 | 0.4074 | 0.4086 | 0.0300 | 0.0297 |
| | 0.8 | 0.8011 | 0.4400 | 0.4287 | 0.4307 | 0.0537 | 0.0535 |
| | 0.9 | 0.9083 | 0.7216 | 0.4299 | 0.4324 | 0.1169 | 0.1168 |
| 900 | 0.1 | 0.1016 | 0.1461 | 0.1522 | 0.1477 | 0.0148 | 0.0149 |
| | 0.2 | 0.2009 | 0.1748 | 0.1778 | 0.1730 | 0.0090 | 0.0092 |
| | 0.3 | 0.3009 | 0.2298 | 0.2280 | 0.2241 | 0.0076 | 0.0078 |
| | 0.4 | 0.4004 | 0.2807 | 0.2866 | 0.2841 | 0.0068 | 0.0070 |
| | 0.5 | 0.5006 | 0.3268 | 0.3367 | 0.3355 | 0.0079 | 0.0079 |
| | 0.6 | 0.6002 | 0.3820 | 0.3774 | 0.3774 | 0.0123 | 0.0122 |
| | 0.7 | 0.7002 | 0.4215 | 0.4093 | 0.4105 | 0.0207 | 0.0206 |
| | 0.8 | 0.8002 | 0.4425 | 0.4317 | 0.4338 | 0.0367 | 0.0366 |
| | 0.9 | 0.9008 | 0.5033 | 0.4435 | 0.4454 | 0.0793 | 0.0794 |
| 1,600 | 0.1 | 0.1018 | 0.1562 | 0.1533 | 0.1488 | 0.0116 | 0.0117 |
| | 0.2 | 0.2013 | 0.1783 | 0.1782 | 0.1735 | 0.0067 | 0.0069 |
| | 0.3 | 0.3013 | 0.2256 | 0.2282 | 0.2242 | 0.0056 | 0.0058 |
| | 0.4 | 0.4009 | 0.2852 | 0.2867 | 0.2841 | 0.0052 | 0.0053 |
| | 0.5 | 0.5008 | 0.3453 | 0.3364 | 0.3352 | 0.0060 | 0.0060 |
| | 0.6 | 0.6005 | 0.3885 | 0.3770 | 0.3771 | 0.0091 | 0.0090 |
| | 0.7 | 0.6997 | 0.4189 | 0.4088 | 0.4100 | 0.0154 | 0.0153 |
| | 0.8 | 0.7999 | 0.4446 | 0.4323 | 0.4344 | 0.0274 | 0.0273 |

**Table 1** continued

Outputs for censoring model I

| $n$ | $\tau$ | Averages $\widehat{F}$ | $n(\widehat{F}-F^*)^2$ | $n\widehat{V}^{\text{Wald}}$ | $n\widehat{V}^{\text{LR}}$ | SDs $n\widehat{V}^{\text{Wald}}$ | $n\widehat{V}^{\text{LR}}$ |
|---|---|---|---|---|---|---|---|
| | 0.9 | 0.9002 | 0.4678 | 0.4476 | 0.4499 | 0.0626 | 0.0625 |
| 100 | 0.1 | 0.0648 | 0.7904 | 0.3374 | 0.3263 | 0.1126 | 0.1195 |
| | 0.2 | 0.2053 | 0.5748 | 0.3182 | 0.3128 | 0.0604 | 0.0624 |
| | 0.3 | 0.3107 | 0.3173 | 0.3169 | 0.3130 | 0.0377 | 0.0404 |
| | 0.4 | 0.4084 | 0.3127 | 0.3200 | 0.3170 | 0.0259 | 0.0299 |
| | 0.5 | 0.5025 | 0.2915 | 0.3221 | 0.3206 | 0.0215 | 0.0262 |
| | 0.6 | 0.5986 | 0.3123 | 0.3207 | 0.3207 | 0.0260 | 0.0300 |
| | 0.7 | 0.7003 | 0.3349 | 0.3171 | 0.3185 | 0.0404 | 0.0426 |
| | 0.8 | 0.8111 | 0.5858 | 0.3149 | 0.3182 | 0.0674 | 0.0785 |
| | 0.9 | 0.9434 | 0.7655 | 0.3269 | 0.3293 | 0.1107 | 0.1106 |
| 400 | 0.1 | 0.0969 | 0.8177 | 0.3569 | 0.3296 | 0.0894 | 0.1133 |
| | 0.2 | 0.2050 | 0.3208 | 0.3199 | 0.3144 | 0.0330 | 0.0334 |
| | 0.3 | 0.3037 | 0.3296 | 0.3193 | 0.3153 | 0.0188 | 0.0191 |
| | 0.4 | 0.4030 | 0.3297 | 0.3222 | 0.3197 | 0.0126 | 0.0128 |
| | 0.5 | 0.5021 | 0.3103 | 0.3235 | 0.3226 | 0.0106 | 0.0105 |
| | 0.6 | 0.6017 | 0.2945 | 0.3217 | 0.3223 | 0.0127 | 0.0125 |
| | 0.7 | 0.7013 | 0.3049 | 0.3173 | 0.3192 | 0.0189 | 0.0185 |
| | 0.8 | 0.8003 | 0.3092 | 0.3165 | 0.3198 | 0.0328 | 0.0325 |
| | 0.9 | 0.9081 | 0.6702 | 0.3542 | 0.3554 | 0.0958 | 0.0952 |
| 900 | 0.1 | 0.1031 | 0.4459 | 0.3632 | 0.3560 | 0.0625 | 0.0662 |
| | 0.2 | 0.2021 | 0.3038 | 0.3179 | 0.3124 | 0.0223 | 0.0225 |
| | 0.3 | 0.3018 | 0.3240 | 0.3184 | 0.3143 | 0.0130 | 0.0132 |
| | 0.4 | 0.4011 | 0.3280 | 0.3222 | 0.3197 | 0.0085 | 0.0087 |
| | 0.5 | 0.5012 | 0.3144 | 0.3240 | 0.3231 | 0.0068 | 0.0068 |
| | 0.6 | 0.6007 | 0.3345 | 0.3223 | 0.3229 | 0.0083 | 0.0081 |
| | 0.7 | 0.7001 | 0.3309 | 0.3185 | 0.3204 | 0.0132 | 0.0130 |
| | 0.8 | 0.7998 | 0.3406 | 0.3183 | 0.3222 | 0.0230 | 0.0229 |
| | 0.9 | 0.9010 | 0.4337 | 0.3649 | 0.3625 | 0.0642 | 0.0729 |
| 1,600 | 0.1 | 0.1023 | 0.3577 | 0.3638 | 0.3577 | 0.0438 | 0.0444 |
| | 0.2 | 0.2012 | 0.2961 | 0.3181 | 0.3126 | 0.0162 | 0.0164 |
| | 0.3 | 0.3010 | 0.3259 | 0.3185 | 0.3144 | 0.0099 | 0.0101 |
| | 0.4 | 0.4009 | 0.3274 | 0.3262 | 0.3237 | 0.0065 | 0.0066 |
| | 0.5 | 0.5010 | 0.3349 | 0.3237 | 0.3228 | 0.0050 | 0.0050 |
| | 0.6 | 0.6008 | 0.3224 | 0.3223 | 0.3227 | 0.0064 | 0.0063 |
| | 0.7 | 0.7002 | 0.3271 | 0.3184 | 0.3203 | 0.0098 | 0.0097 |
| | 0.8 | 0.8004 | 0.3100 | 0.3185 | 0.3222 | 0.0170 | 0.0187 |
| | 0.9 | 0.9008 | 0.3744 | 0.3688 | 0.3553 | 0.0488 | 0.0567 |

**Table 1** continued

Outputs for censoring model I

| | | Averages | | | | SDs | |
|---|---|---|---|---|---|---|---|
| $n$ | $\tau$ | $\widehat{F}$ | $n(\widehat{F} - F^*)^2$ | $n\widehat{V}^{\text{Wald}}$ | $n\widehat{V}^{\text{LR}}$ | $n\widehat{V}^{\text{Wald}}$ | $n\widehat{V}^{\text{LR}}$ |
| 100 | 0.1 | 0.0841 | 0.7062 | 0.2998 | 0.2965 | 0.0898 | 0.0903 |
| | 0.2 | 0.2114 | 0.4554 | 0.2915 | 0.2887 | 0.0515 | 0.0518 |
| | 0.3 | 0.3125 | 0.2693 | 0.2833 | 0.2811 | 0.0282 | 0.0286 |
| | 0.4 | 0.4066 | 0.2635 | 0.2712 | 0.2697 | 0.0148 | 0.0152 |
| | 0.5 | 0.5023 | 0.2425 | 0.2654 | 0.2648 | 0.0086 | 0.0086 |
| | 0.6 | 0.6015 | 0.2612 | 0.2704 | 0.2709 | 0.0158 | 0.0153 |
| | 0.7 | 0.7014 | 0.2957 | 0.2806 | 0.2819 | 0.0318 | 0.0314 |
| | 0.8 | 0.8065 | 0.4216 | 0.2857 | 0.2876 | 0.0573 | 0.0570 |
| | 0.9 | 0.9335 | 0.6114 | 0.2828 | 0.2850 | 0.1128 | 0.1127 |
| 400 | 0.1 | 0.1055 | 0.4863 | 0.3023 | 0.2984 | 0.0610 | 0.0616 |
| | 0.2 | 0.2060 | 0.2871 | 0.2947 | 0.2917 | 0.0283 | 0.0284 |
| | 0.3 | 0.3040 | 0.2697 | 0.2852 | 0.2828 | 0.0151 | 0.0153 |
| | 0.4 | 0.4026 | 0.2556 | 0.2733 | 0.2717 | 0.0074 | 0.0076 |
| | 0.5 | 0.5015 | 0.2504 | 0.2670 | 0.2664 | 0.0036 | 0.0036 |
| | 0.6 | 0.6009 | 0.2581 | 0.2731 | 0.2735 | 0.0077 | 0.0074 |
| | 0.7 | 0.7005 | 0.2915 | 0.2843 | 0.2857 | 0.0156 | 0.0154 |
| | 0.8 | 0.8001 | 0.2829 | 0.2929 | 0.2948 | 0.0300 | 0.0298 |
| | 0.9 | 0.9032 | 0.4205 | 0.2928 | 0.2947 | 0.0676 | 0.0676 |
| 900 | 0.1 | 0.1040 | 0.2836 | 0.2980 | 0.2945 | 0.0414 | 0.0414 |
| | 0.2 | 0.2023 | 0.2884 | 0.2927 | 0.2897 | 0.0189 | 0.0190 |
| | 0.3 | 0.3015 | 0.2740 | 0.2847 | 0.2822 | 0.0099 | 0.0100 |
| | 0.4 | 0.4008 | 0.2730 | 0.2735 | 0.2719 | 0.0049 | 0.0050 |
| | 0.5 | 0.5007 | 0.2634 | 0.2678 | 0.2672 | 0.0024 | 0.0024 |
| | 0.6 | 0.6004 | 0.2680 | 0.2740 | 0.2744 | 0.0050 | 0.0049 |
| | 0.7 | 0.7001 | 0.2976 | 0.2854 | 0.2868 | 0.0106 | 0.0104 |
| | 0.8 | 0.7998 | 0.2786 | 0.2936 | 0.2956 | 0.0202 | 0.0200 |
| | 0.9 | 0.8999 | 0.3127 | 0.2975 | 0.2997 | 0.0453 | 0.0452 |
| 1,600 | 0.1 | 0.1035 | 0.3322 | 0.2992 | 0.2958 | 0.0322 | 0.0322 |
| | 0.2 | 0.2018 | 0.2956 | 0.2934 | 0.2903 | 0.0142 | 0.0143 |
| | 0.3 | 0.3011 | 0.2956 | 0.2954 | 0.2930 | 0.0077 | 0.0078 |
| | 0.4 | 0.4011 | 0.2861 | 0.2838 | 0.2822 | 0.0038 | 0.0039 |
| | 0.5 | 0.5010 | 0.2794 | 0.2777 | 0.2771 | 0.0018 | 0.0018 |
| | 0.6 | 0.6009 | 0.2849 | 0.2838 | 0.2842 | 0.0039 | 0.0038 |
| | 0.7 | 0.7003 | 0.2906 | 0.2902 | 0.2916 | 0.0079 | 0.0078 |
| | 0.8 | 0.8008 | 0.2800 | 0.2933 | 0.2953 | 0.0144 | 0.0143 |
| | 0.9 | 0.9011 | 0.2956 | 0.2957 | 0.2983 | 0.0315 | 0.0315 |

and 1 in cases where $n$ is smaller. As expected, it turns out that the precision of such approximations improves as $n$ becomes larger. Especially, the simulated averages and SDs of $n\widehat{V}^{\mathrm{Wald}}(\widehat{F}(\tau))$ are almost equivalent to those of $n\widehat{V}^{\mathrm{LR}}(\widehat{F}(\tau))$, although the computing speed of $\widehat{V}^{\mathrm{Wald}}(\widehat{F}(\tau))$ is much faster than that of $\widehat{V}^{\mathrm{LR}}(\widehat{F}(\tau))$. In fact, for example, when $n = 1,600$, the times to compute $\widehat{V}^{\mathrm{Wald}}(\widehat{F}(\tau))$ and $\widehat{V}^{\mathrm{LR}}(\widehat{F}(\tau))$ at $\tau =$ all the jump points of $\widehat{J}$ (not only deciles) except the computation time of the NPMLE were about 0.2 and 69 s on the average, respectively. For 1,000 samples of $n = 1,600$ in this simulation, we required about 4 min to obtain all the pointwise Wald-type variance estimates but about 19.3 hours to get all the pointwise variance estimates based on the profile likelihood ratio.

These results indicate that the proposed iteration-free Wald-type variance estimation is far superior to that based on the profile likelihood ratio in terms of computing speeds and mutually equivalent in the performance of estimators. Therefore, the proposed method is appropriate for practical use, in particular, in the situation of larger samples where it is difficult to estimate the variance of the NPMLE $\widehat{F}$ by solving the Fredholm integral equations, using the profile likelihood ratio or the bootstrap sampling.

## 5 Discussion

In this paper, we propose the iteration-free Wald-type variance estimation for the NPMLE for doubly censored data, based on the semiparametric profile likelihood. Our variance formula is same as Turnbull (1974) in terms of gaining the inverse of the Fisher information matrix, but ours is freed of using the form of the determinant functions and can compute numerically stable solutions in larger sample than Turnbull's. It is shown that the Wald-type variance estimator is consistent to an asymptotic variance formula using the Fredholm integral equations. Though our asymptotic variance formula may be seen unlike that of Chang (1990), it is identical to Chang's asymptotic formula as long as the log profile likelihood ratio is asymptotically chi-squared distributed with one degree of freedom. Simulation studies show that the proposed method appropriately approximates to the sample variances of the NPMLE, in particular, in the situation of larger samples where it is difficult to perform the methods, for example, by solving the Fredholm integral equations (e.g., Chang 1990; Chen and Zhou 2003), using the profile likelihood ratio (e.g., Chen and Zhou 2003) or the bootstrap sampling (e.g., Wellner and Zhan 1997; Zhu and Sun 2007), because of computationally hard iterations and vast computation time.

In future work, the performance of the Wald-type variance estimation may be improved using some transformation (log, log–log and so on) of distribution function, similarly to some modifications of the Greenwood variance formula in right-censored data. Also, even if we are interested in the covariance estimation between two percent points of the distribution, we will be able to easily extend the Wald-type estimation procedure based on the profile likelihood in this paper to cases with two parameters of interest. Finally, this research seems to serve as one material to study whether a martingale approach is available in doubly censored data.

# A Appendix

## A.1 On the positive definiteness of $\mathcal{I}(F), \mathcal{I}(F^{\#}_{m(J)})$ and $\mathcal{I}^f_\tau$

The content below generalizes the result and manner provided by the proof of Turnbull (1974, Lemma A2) in which the case only of $J = \mathcal{J} = \{T^1_{(1)}, T^1_{(2)}, \ldots, T^1_{(k_1)}\}$ in our notaion is considered.

As stated in the proof of Lemma 1, the determinant $D_{k(J)-1}$ of $\mathcal{I}(F)$ is calculated via $D_j$, $j = 1, \ldots, k(J) - 1$ following the rule (6). Since $\mathcal{I}(F), \mathcal{I}(F^{\#}_{m(J)})$ and $\mathcal{I}^f_\tau$ are real symmetric matrices, their positive definitenesses are obtained by showing that every $D_j$ is always positive.

Let $\alpha_j = d_j/(F_j - F_{j-1})^2$. We consider the following sequence:

$$\widetilde{D}_0 = 1, \ \widetilde{D}_1 = \alpha_1 + \alpha_2 \quad \text{and} \quad \widetilde{D}_j = (\alpha_j + \alpha_{j+1})\widetilde{D}_{j-1} - \alpha_j^2 \widetilde{D}_{j-2}, \qquad j = 2, \ldots, k(J) - 1.$$

Such $\widetilde{D}_j$ is the determinant of the $j \times j$ symmetric tridiagonal matrix $\mathcal{A}_j$ such that the $l$th diagonal and the $(l-1, l)$th elements of $\mathcal{A}_j$ are $\alpha_l + \alpha_{l+1}$ and $\alpha_l$, respectively. Then, we find that $\widetilde{D}_j$ is generally written as

$$\widetilde{D}_j = \prod_{l=1, \alpha_l \neq 0}^{j+1} \alpha_l \left( \sum_{l=1, \alpha_l \neq 0}^{j+1} 1/\alpha_l \right), \quad j = 1, \ldots, k(J) - 1.$$

If the jump point of the type (j3) occurs at $F_j$, then, because of $\alpha_j = 0$, the terms with $\alpha_j$ in the right side of the above equation are eliminated. However, since the first term $\alpha_1$ is always positive, all the $\widetilde{D}_j$'s are positive, so that all the $\mathcal{A}_j$'s are positive definite. Also, the $j$th diagonal and the $(j-1, j)$th elements of $\mathcal{I}(F)$ are written as $i_{j,j} = \alpha_j + \alpha_{j+1} + \beta_j$ and $i_{j-1,j} = -\alpha_j$, where $\beta_j = m^R_j/(1 - F_j)^2 + m^L_j/F^2_j$, $j = 1, \ldots, k(J) - 1$.

Let $\mathcal{B}_j$ be the $j \times j$ diagonal matrix of which the $l$th diagonal element is $\beta_l$. Of course, we have $\mathcal{I}(F) = \mathcal{A}_{k(J)-1} + \mathcal{B}_{k(J)-1}$. Since all the $\mathcal{A}_j$'s and $\mathcal{B}_j$'s are positive and non-negative definite, respectively, $x^{\mathrm{T}}\mathcal{A}_j x + x^{\mathrm{T}}\mathcal{B}_j x > 0$ are held for non-zero vectors $x$. This indicates that all the eigenvalues of $\mathcal{A}_j + \mathcal{B}_j$, $j = 1, \ldots, k(J) - 1$ are positive. Since the determinant of $\mathcal{A}_j + \mathcal{B}_j$ is $D_j$, all the $D_j$'s are always positive. Therefore, $\mathcal{I}(F)$ is positive definite, since all the eigenvalues of $\mathcal{I}(F)$ are positive. Also, from these considerations, not only the conclusion of the positive definiteness, we can easily lead

$$D_j \geq \widetilde{D}_j, \quad j = 1, \ldots, k(J) - 1$$

by decomposing $D_j$ repeatedly until $\widetilde{D}_j$ appears by the multi-linearity of the determinant.

By the similar argument, $\mathcal{I}(F^{\#}_{m(J)})$ defined in the proof of Lemma 1 is positive definite, since the determinants of submatrices generated by $\mathcal{I}(F^{\#}_{m(J)})$ are always positive. Further, $\mathcal{I}_\tau^f$ defined in the proof of Theorem 1 is also positive definite because $\mathcal{I}_\tau^f = \mathcal{A}_{m(J)} + \mathcal{B}_{m(J)}$.

**Acknowledgments** The author is grateful to anonymous reviewers and the Associate Editor for their constructive comments and helpful suggestions. This work was supported by Grant-in-Aid for Young Scientists (B) (No. 19700263).

## References

Chang, M. N. (1990). Weak convergence of a self-consistent estimator of the survival function with doubly censored data. *Annals of Statistics, 18*, 391–404.

Chang, M. N., Yang, G. L. (1987). Strong consistency of a nonparametric estimator of the survival function with doubly censored data. *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.

Gill, R. D., Johansen, S. (1990). A survey of product-integration with a view toward application in survival analysis. *Annals of Statistics, 18*, 1501–1555.

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

Klein, J. P., Moeschberger, M. L. (2003). *Survival analysis: techniques for censored and truncated data*, (2nd ed.). New York: Springer.

Murphy, S. A., van der Vaart, A. W. (1997). Semiparametric likelihood ratio inference. *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. *Annals of Statistics, 24*, 1740–1764.

Sugimoto, T., Hamasaki, T. (2006). Properties of estimators of baseline hazard functions in a semiparametric cure model. *Annals of the Institute of Statistical Mathematics, 58*, 647–674.

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., Zhan, 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.

Zhu, C., Sun, J. (2007). Variance estimation of a survival function with doubly censored failure time data. In J.-L. Auget, N. Balakrishnan, M. Mesbah, G. Molenberghs (Eds.), *Advances in statistical methods for the health sciences* (pp. 225–235). Boston: Birkhäuser.