Lifetime Data Anal (2015) 21:397–418
DOI 10.1007/s10985-014-9297-5

Statistical inference based on the nonparametric maximum likelihood estimator under double-truncation

Takeshi Emura · Yoshihiko Konno · Hirofumi Michimae

Received: 19 August 2013 / Accepted: 18 June 2014 / Published online: 8 July 2014

Abstract Doubly truncated data consist of samples whose observed values fall between the right- and left- truncation limits. With such samples, the distribution function of interest is estimated using the nonparametric maximum likelihood estimator (NPMLE) that is obtained through a self-consistency algorithm. Owing to the complicated asymptotic distribution of the NPMLE, the bootstrap method has been suggested for statistical inference. This paper proposes a closed-form estimator for the asymptotic covariance function of the NPMLE, which is computationally attractive alternative to bootstrapping. Furthermore, we develop various statistical inference procedures, such as confidence interval, goodness-of-fit tests, and confidence bands to demonstrate the usefulness of the proposed covariance estimator. Simulations are performed to compare the proposed method with both the bootstrap and jackknife methods. The methods are illustrated using the childhood cancer dataset.

Keywords Asymptotic variance · Bootstrap · Confidence band · Goodness-of-fit test · Survival analysis

T. Emura (✉)
Graduate Institute of Statistics, National Central University, Zhongli, Taiwan
e-mail: emura@stat.ncu.edu.tw

Y. Konno
Department of Mathematical and Physical Sciences,
Japan Women's University, Tokyo, Japan
e-mail: konno@fc.jwu.ac.jp

H. Michimae
Department of Clinical Medicine (Biostatistics), School of Pharmacy,
Kitasato University, Tokyo, Japan
e-mail: michimaeh@pharm.kitasato-u.ac.jp

1 Introduction

Statistical methodologies for doubly truncated data have been an active research area with a variety of applications. Efron and Petrosian (1999) developed inference methods based on doubly truncated data, highlighting its importance in astronomy. In particular, due to the resolution of telescopes, the luminosity of stars may be undetected if it is either too dim or too bright, leading to double-truncation (i.e., both upper and lower truncations). Stovring and Wang (2007) considered a type of doubly truncated data to analyze the incidence and lifetime risk of diabetes that are useful statistics for public health. The childhood cancer data of North Portugal provides a similar example (Moreira and Uña-Álvarez 2010). Recently, Zhu and Wang (2012) identified a sampling bias due to double-truncation in the analysis of cancer registry data and proposed inference procedures that can properly analyze such data. In general, double-truncation is very common in fields such as astronomy, demography, and epidemiology. The methodological and theoretical developments for analyzing doubly-truncated data are attributed to Moreira and Uña-Álvarez (2010, 2012), Moreira and Keilegom (2013), Shen (2010, 2011, 2012), Emura and Konno (2012), Austin et al. (2013), and Moreira et al. (2014).

We illustrate the double-truncation occurring in the childhood cancer data discussed in Moreira and Uña-Álvarez (2010). Their data include the ages at diagnosis (T^*) of children who were diagnosed with cancer within a follow-up period between January 1, 1999 and December 31, 2003 (Fig. 1). However, they do not have any information on children who are diagnosed with cancer outside this period. Hence, the sample inclusion criterion is written as $U^* \leq T^* \leq V^*$, where U^* is the age on January 1, 1999 and $V^* = U^* + 1825$ (days) is age on December 31, 2003, leading to the double-truncation of T^* by left-truncation limit U^* and right-truncation limit V^*. Ignoring truncation causes bias in statistical inference.

Note that double-truncation is essentially different from double-censoring (i.e., both left- and right- censorings) and interval censoring. Double-truncation yields inclusion/exclusion of samples while double-censoring and interval censoring produce incomplete lifetimes of the included samples (Commenges 2002).

Efron and Petrosian (1999) first introduced the nonparametric maximum likelihood estimator (NPMLE) for $F(t) = \Pr(T^* \leq t)$ under double-truncation. Their NPMLE, denoted by $\hat{F}(t)$, takes into account the sampling bias due to double-truncation. Shen's Theorems 2 and 3 (2010) showed the uniform consistency and the asymptotic normality of the NPMLE. The asymptotic distribution is complicated, so his formula of the asymptotic variance is not explicitly written down. Moreira and Uña-Álvarez (2010) recognized the analytical difficulty in the asymptotic variance and then proposed the simple bootstrap and obvious bootstrap methods to construct the pointwise confidence interval of $F(t)$. They reported that the simple bootstrap technique is more reliable and more technically convenient than the obvious bootstrap technique. Shen (2012) circumvented the difficulty of estimating the asymptotic variance and then utilized the empirical likelihood ratio test to construct the pointwise confidence interval. Although his method may provide more accurate coverage performance than simple bootstrapping, it does not provide a variance estimator.

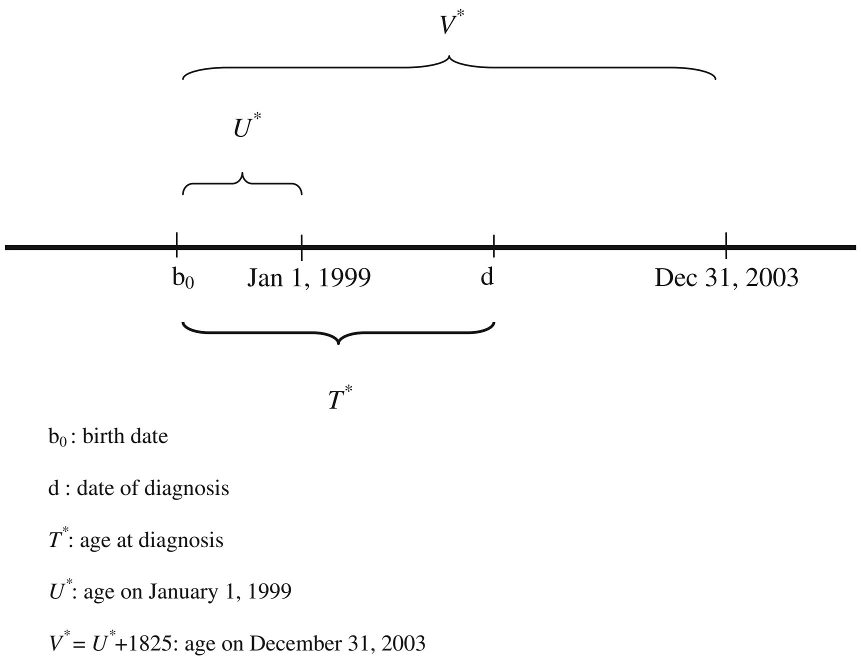

Fig. 1 The childhood cancer cases of North Portugal (Moreira and Uña-Álvarez 2010)

In this paper, we derive a closed-form estimator for $Cov\{\hat{F}(s), \hat{F}(t)\}$. For $s = t$, the estimator yields a computationally attractive alternative to the bootstrap or jackknife variance estimator. Furthermore, the estimated covariance structure is utilized to propose goodness-of-fit tests and confidence bands, both of which have not yet been developed in the literature.

The rest of the paper is organized as follows. Section 2 briefly reviews the NPMLE developed by Efron and Petrosian (1999). Section 3 presents the proposed covariance estimator. Section 4 applies the proposed estimator to develop various inference procedures, including confidence interval, goodness-of-fit tests, and confidence bands. Simulations and data analysis are given in Sects. 5 and 6, respectively. Section 7 concludes the paper.

2 The NPMLE

This paper considers doubly truncated data in which individuals can only be included in the sample if their observations fall within certain random intervals. Specifically, let T^* be a random variable of lifetime, U^* be the left-truncation limit, and V^* be the right-truncation limit. One can observe the triplet (U^*, T^*, V^*) only if $U^* \le T^* \le V^*$ holds. Therefore, the sample consists of $\{(U_j, T_j, V_j) : j = 1, \ldots, n\}$ subject to $U_j \le T_j \le V_j$. With this sampling scheme, the observations are independent and identical replications from the distribution function $\Pr(U^* \le u, T^* \le t, V^* \le v | U^* \le T^* \le V^*)$. If $\Pr(V^* = \infty) = 1$, T^* is only subject to left-truncation by

U^*; if $\Pr(U^* = 0) = 1$, T^* is only subject to right-truncation by V^*. Hence, doubly truncated data accommodates one-sided truncation. We assume throughout that T^* and (U^*, V^*) are independent as commonly imposed in the literature (Shen 2011).

Efron and Petrosian (1999) first proposed the NPMLE for $F(t) = \Pr(T^* \le t)$. Consider a discrete distribution putting probability masses $\mathbf{f} = (f_1, \ldots, f_n)^{\mathrm{T}}$ on the observed points $(T_1, \ldots, T_n)$. Let $J_{im} = \mathbf{I}\{U_i \le T_m \le V_i\}$, where $\mathbf{I}\{A\} = 1$ if A is true, and $\mathbf{I}\{A\} = 0$ if A is false. Also, let $F_i = \sum_{m=1}^{n} f_m J_{im}$ be the masses in $\mathbf{f}$ on $[U_i, V_i]$ for $i = 1, \ldots, n$. Then, it follows that $\mathbf{F} = J\mathbf{f}$, where $\mathbf{F} = (F_1, \ldots, F_n)^{\mathrm{T}}$ and J is an $n \times n$ matrix whose (i, j) component is J_{ij}.

Let $\hat{\mathbf{f}} = (\hat{f}_1, \ldots, \hat{f}_n)^{\mathrm{T}}$ be a maximizer of the likelihood function

$$L_n(\mathbf{f}) = \prod_{j=1}^{n} \frac{f_j}{F_j},$$

subject to $1 = \sum_{j=1}^{n} f_j = \mathbf{1}_n^{\mathrm{T}}\mathbf{f}$, where $\mathbf{1}_n = (1, \ldots, 1)^{\mathrm{T}}$ is n-vector of ones. The derivative of the log-likelihood is

$$\frac{\partial \log L_n(\mathbf{f})}{\partial \mathbf{f}} = \mathbf{f}^{-1} - J^T\mathbf{F}^{-1}, \tag{1}$$

where $\mathbf{f}^{-1} = (f_1^{-1}, \ldots, f_n^{-1})^{\mathrm{T}}$ and $\mathbf{F}^{-1} = (F_1^{-1}, \ldots, F_n^{-1})^{\mathrm{T}}$. This equation leads to the following algorithm for obtaining $\hat{\mathbf{f}}$:
Self-consistency algorithm (Efron and Petrosian 1999)

Step 0: Set $\hat{\mathbf{f}}_{(0)} = (1/n, \ldots, 1/n)^{\mathrm{T}}$ and $\hat{\mathbf{F}}_{(0)} = J\hat{\mathbf{f}}_{(0)}$,
Step 1: Obtain $\hat{\mathbf{f}}_{(1)}$ by $\mathbf{f}_{(1)}^{-1} = J^{\mathrm{T}}\mathbf{F}_{(0)}^{-1}$ and then replace $\hat{\mathbf{f}}_{(1)}$ with $\hat{\mathbf{f}}_{(1)}/(\mathbf{1}_n^{\mathrm{T}}\hat{\mathbf{f}}_{(1)})$; set $\hat{\mathbf{F}}_{(1)} = J\,\hat{\mathbf{f}}_{(1)}$,
Step 2: Repeat Step 1 to update $\hat{\mathbf{f}}_{(\ell+1)}$ from the previous step for $\ell = 1, 2, \ldots$; stop the algorithm when $||\hat{\mathbf{f}}_{(\ell+1)} - \hat{\mathbf{f}}_{(\ell)}|| < \varepsilon$ for a small $\varepsilon > 0$ and some norm $||\cdot||$.

The NPMLE of $F(t) = \Pr(T^* \le t)$ is defined as $\hat{F}(t) = \sum_{j=1}^{n} \mathbf{I}(T_j \le t)\hat{f}_j$. Moreira and Uña-Álvarez (2010) suggested a simple bootstrap to estimate the confidence interval of $F(t)$. A convenient alternative to bootstrapping is the jackknife. The bootstrap and jackknife algorithms are given in Appendix A.

3 Asymptotic covariance estimator of the NPMLE

This section derives a new estimator for $Cov\{\hat{F}(s), \hat{F}(t)\}$ in a closed form.

3.1 Observed information matrix

The likelihood Eq. (1) used in the self-consistency algorithm is derived by treating $\mathbf{f} = (f_1, \ldots, f_n)^{\mathrm{T}}$ as n unknown parameters. The constraint $\sum_{j=1}^{n} f_j = 1$ is then incorporated into the algorithm through standardization $\hat{\mathbf{f}}_{(1)}/(\,\mathbf{1}_n^{\mathrm{T}}\hat{\mathbf{f}}_{(1)}\,)$ at Step 1. Alternatively, we modify the likelihood Eq. (1) by directly incorporating the constraint $\sum_{j=1}^{n} f_j = 1$ and regarding $\mathbf{f}_{(-n)} = (f_1, \ldots, f_{n-1})^{\mathrm{T}}$ as $(n-1)$ unknown parameters. Here, we set $f_n = 1 - \mathbf{1}_{n-1}^{\mathrm{T}}\mathbf{f}_{(-n)}$. This treatment is crucial for deriving the proposed variance estimator. Without loss of generality, we assume that $\hat{\mathbf{f}} = (\hat{f}_1, \ldots, \hat{f}_n)^{\mathrm{T}}$ represents masses at the ordered values of $T_{(1)} < \cdots < T_{(n)}$. Especially, $\hat{f}_n$ is the mass corresponding to the largest observation $T_{(n)} = \max_j T_j$.

Using $\partial F_i/\partial f_j = J_{ij} - J_{in}$ for $j = 1, \ldots, n-1$, the score function becomes

$$\frac{\partial \log L_n(\mathbf{f})}{\partial f_j} = \frac{1}{f_j} - \left[\sum_{i=1}^{n} \frac{J_{ij}}{F_i}\right]_{f_n = 1 - \mathbf{1}_{n-1}^{\mathrm{T}}\mathbf{f}_{(-n)}} - \left[\frac{1}{f_n} - \sum_{i=1}^{n} \frac{J_{in}}{F_i}\right]_{f_n = 1 - \mathbf{1}_{n-1}^{\mathrm{T}}\mathbf{f}_{(-n)}}.$$

This is written in the matrix form as

$$\frac{\partial \log L_n(\mathbf{f})}{\partial \mathbf{f}_{(-n)}} = D\,[\mathbf{f}^{-1} - J^{\mathrm{T}}\mathbf{F}^{-1}]_{f_n = 1 - \mathbf{1}_{n-1}^{\mathrm{T}}\mathbf{f}_{(-n)}},$$

where $D = [I_{n-1} \vdots - \mathbf{1}_{n-1}]$, $\mathbf{f}^{-1} = (\,1/f_1, \ldots, 1/f_n\,)^{\mathrm{T}}$, and $\mathbf{F}^{-1} = (\,1/F_1, \ldots, 1/F_n\,)^{\mathrm{T}}$. Also, for $j, j' \in \{\,1, \ldots, n-1\,\}$,

$$-\frac{\partial^2 \log L_n(\mathbf{f})}{\partial f_{j'}\partial f_j} = \frac{\mathbf{I}(\,j = j'\,)}{f_j^2} + \left[\frac{1}{f_n^2}\right]_{f_n = 1 - \mathbf{1}_{n-1}^{\mathrm{T}}\mathbf{f}} - \left[\sum_{i=1}^{n} \frac{(J_{ij} - J_{in})(J_{ij'} - J_{in})}{F_i^2}\right]_{f_n = 1 - \mathbf{1}_{n-1}^{\mathrm{T}}\mathbf{f}}.$$

Hence, the observed information matrix is

$$i_n(\mathbf{f}) = -\frac{\partial^2 \log L_n(\mathbf{f})}{\partial \mathbf{f}_{(-n)}\partial \mathbf{f}_{(-n)}^{\mathrm{T}}} = D\left\{\mathrm{diag}\left(\frac{1}{\mathbf{f}^2}\right) - J^{\mathrm{T}}\mathrm{diag}\left(\frac{1}{\mathbf{F}^2}\right)J\right\}\bigg|_{f_n = 1 - \mathbf{1}_{n-1}^{\mathrm{T}}\mathbf{f}} D^{\mathrm{T}}, \quad (2)$$

where diag($\mathbf{a}$) is a diagonal matrix with the diagonal elements $\mathbf{a}$.

3.2 The asymptotic covariance estimator

We derive the asymptotic covariance structure of $\sqrt{n}(\,\hat{F}(t) - F(t)\,)$ and its plug-in estimator. Let $\sigma_F(\,\cdot\,)[h] : [0, \infty) \to [0, \infty)$ be defined as

$$\sigma_F(x)[h] = E\left[\mathbf{I}(U^* \le x \le V^*)\left\{\frac{h(x)}{\int \mathbf{I}(U^* \le s \le V^*)dF(s)} - \frac{\int \mathbf{I}(U^* \le s \le V^*)h(s)dF(s)}{\{\int \mathbf{I}(U^* \le s \le V^*)dF(s)\,\}^2}\right\}\right],$$

where $h : [0, \infty) \to R$ is a bounded function. Note that $\sigma_F(\cdot)[h]$ is the Fisher information (Murphy 1995, p. 189) with h being the index of infinite-dimensional parameters. Appendix B1 shows that $\sqrt{n}(\hat{F}(t) - F(t))$ converges weakly to a Gaussian process $G_F(t)$ with $E[G_F(t)] = 0$ and

$$E[G_F(s)G_F(t)] = \int w_s(x)\sigma_F^{-1}(w_t)(x)dF(x),$$

where $w_s(x) \equiv \mathbf{I}(x \le s)$ and $\sigma_F^{-1}(w_t)$ solves $\sigma_F(x)[h] = w_t(x)$ for h.

Consider the empirical estimator of $\sigma_F(x)[h]$ as

$$\hat{\sigma}_F(x)[h] = \frac{1}{n}\sum_{i=1}^{n}\mathbf{I}(U_i \le x \le V_i)\left\{\frac{1}{\hat{F}_i}h(x) - \frac{1}{\hat{F}_i^2}\sum_{k=1}^{n}J_{ik}h_k\hat{f}_k\right\},$$

where $h_k = h(T_k)$. Then, the plug-in covariance estimator is

$$\hat{E}[G_F(s)G_F(t)] = \int w_s(x)\hat{\sigma}_F^{-1}(w_t)(x)d\hat{F}(x) = \sum_{j=1}^{n}w_s(T_j)\hat{\sigma}_F^{-1}(w_t)(T_j)\hat{f}_j.$$

After some matrix calculations given in Appendix B2, one can verify

$$\sum_{j=1}^{n}w_s(T_j)\hat{\sigma}_F^{-1}(w_t)(T_j)\hat{f}_j = \mathbf{W}_s^{\mathrm{T}}\left\{\frac{i_n(\hat{\mathbf{f}})}{n}\right\}^{-1}\mathbf{W}_t,$$

where $\mathbf{W}_t = (\mathbf{I}(T_{(1)} \le t) - \mathbf{I}(T_{(n)} \le t), \ldots, \mathbf{I}(T_{(n-1)} \le t) - \mathbf{I}(T_{(n)} \le t))^{\mathrm{T}}$ and $i_n(\mathbf{f})$ is given in Eq. (2). Therefore, we obtain a plug-in covariance estimator

$$\hat{C}ov\{\hat{F}(s), \hat{F}(t)\} = \mathbf{W}_s^{\mathrm{T}}\left[D\left\{\mathrm{diag}\left(\frac{1}{\hat{\mathbf{f}}^2}\right) - J^{\mathrm{T}}\mathrm{diag}\left(\frac{1}{\hat{\mathbf{F}}^2}\right)J\right\}D^{\mathrm{T}}\right]^{-1}\mathbf{W}_t, \quad (3)$$

and a variance estimator

$$\hat{V}_{\mathrm{Info}}\{\hat{F}(t)\} = \mathbf{W}_t^{\mathrm{T}}\left[D\left\{\mathrm{diag}\left(\frac{1}{\hat{\mathbf{f}}^2}\right) - J^{\mathrm{T}}\mathrm{diag}\left(\frac{1}{\hat{\mathbf{F}}^2}\right)J\right\}D^{\mathrm{T}}\right]^{-1}\mathbf{W}_t. \quad (4)$$

Remark: Murphy (1995), Zeng and Lin (2006), Chen (2010), and Emura and Wang (2012) use similar techniques to derive variance estimators. However, none of them results in an explicit form like Eqs. (3) and (4).

4 Inference based on the asymptotic covariance estimator

This section examines various inference procedures based on the proposed covariance estimator.

4.1 Pointwise confidence interval

Applying the variance estimator $\hat{V}_{\text{Info}}\{\hat{F}(t)\}$ in Eq. (4) and the asymptotic normality, we propose a pointwise confidence interval. Log-transformation and arcsine-square root transformation are known to improve the coverage performance over the linear confidence interval (Klein and Moeschberger 2003, pp. 104–108). Here, we apply the log-transformed interval based on $\log \hat{F}(t) - \log F(t) \sim N(\, 0,\ \hat{V}_{\text{Info}}\{\, \hat{F}(t)\, \}/\hat{F}(t)^2)$. Hence, the $(1-\alpha)100\%$ confidence interval for $F(t)$ is

$$(\ \hat{F}(t)\exp[\ -z_{\alpha/2}\hat{V}_{\text{Info}}^{1/2}\{\hat{F}(t)\}/\hat{F}(t)\ ],\quad \hat{F}(t)\exp[\ z_{\alpha/2}\hat{V}_{\text{Info}}^{1/2}\{\hat{F}(t)\}/\hat{F}(t)\ ]\ ),$$

where $z_{\alpha/2}$ is the $(1-\alpha/2)100\%$ point of the standard normal distribution.

4.2 Goodness-of-fit test

We consider a goodness-of-fit test for

$$H_0:\quad F = F_0 \text{ vs. } H_1:\quad F \neq F_0,$$

where F_0 is a known continuous distribution function. Applying the continuous mapping theorem to the results of Sect. 3.2, we have

$$\sqrt{n}\sup_t |\ \hat{F}(t) - F(t)| \xrightarrow{d} \sup_t |G_F(t)|.$$

The asymptotic distribution can be easily simulated after estimating the covariance structure of $G_{F_0}(t)$ with Eq. (3). Ideally, the asymptotic distribution is approximated by that of $\max_j |G_{F_0}(t_j)|$ for fixed fine grids $t_j\,:\, j = 1,\ \ldots,\ N$ with large N. Here, we suggest a practically convenient choice of $t_j = T_{(j)}$, $j = 1,\ \ldots,\ n-1$, which leads to a simple algorithm and achieves good finite sample performance. The algorithm is stated as follows:

Kolmogorov–Smirnov test for $H_0:\quad F = F_0$ vs. $H_1:\quad F \neq F_0$;

Step 1: Calculate $K = \sup_t |\ \hat{F}(t) - F_0(t)|$ and $i_n(\hat{\mathbf{f}})$.

Step 2: Generate $\mathbf{G}^{(b)} = (G_1^{(b)}, \ldots, G_{n-1}^{(b)}) \sim N(\mathbf{0}_{n-1}, Hi_n(\hat{\mathbf{f}})^{-1}H^{\mathrm{T}})$ for $b = 1, \ldots, B$, and compute $K^{(b)} = \max_{i=1,..,n-1}|G_i^{(b)}|$, where $H = (\ \mathbf{W}_{T_{(1)}}, \ldots, \mathbf{W}_{T_{(n-1)}})^{\mathrm{T}}$.

Step 3: Reject $H_0:\ F = F_0$ with level α if $\sum_{b=1}^{B} \mathbf{I}(\ K^{(b)} > K\ )/B < \alpha$.

Similarly, we can test $H_0:\quad F = F_0$ using the Cramér–von Mises statistic

$$C = n\int_0^\infty \{\ \hat{F}(t) - F_0(t)\ \}^2 dF_n(t) = \sum_{j=1}^{n} \{\ \hat{F}(T_j) - F_0(T_j)\ \}^2,$$

where $F_n(t) = \sum_{j=1}^{n} \mathbf{I}(T_j \leq t)/n$ is the empirical distribution function.

Cramér–von Mises test for $H_0: \quad F = F_0$ vs. $H_1: \quad F \neq F_0$;

Step 1: Calculate $C = \sum_{j=1}^{n} \{ \hat{F}(T_j) - F_0(T_j) \}^2$ and $i_n(\hat{\mathbf{f}})$.
Step 2: Generate $\mathbf{G}^{(b)} = (G_1^{(b)}, \ \ldots, \ G_{n-1}^{(b)}) \sim N(\mathbf{0}_{n-1}, \ H i_n(\hat{\mathbf{f}})^{-1} H^{\mathrm{T}})$ for $b = 1, \ \ldots, \ B$, and then compute $C^{(b)} = (\mathbf{G}^{(b)})^{\mathrm{T}} \mathbf{G}^{(b)}$.
Step 3: Reject $H_0: \quad F = F_0$ with level α if $\sum_{b=1}^{B} \mathbf{I}(C^{(b)} > C)/B < \alpha$.

4.3 Confidence band

The confidence band covers the true function $F(t)$ at all t for a specified confidence level $(1-\alpha)$. We follow the construction of two most well-known confidence bands for the survival function under right-censoring, namely, the equal precision (EP) band and Hall–Wellner (HW) band (Nair 1984; Klein and Moeschberger 2003, Sect. 4.4).

Let $\psi(u)$ be a nonnegative continuous function. Applying the continuous mapping theorem to the results of Sect. 3.2, we have

$$\sqrt{n} \sup_t | \ \psi\{ \ F(t) \ \}\{ \ \hat{F}(t) - F(t) \ \}| \xrightarrow{d} \sup_t |\psi\{ \ F(t) \ \} G_F(t)|.$$

Then, the confidence bands are obtained by solving

$$1 - \alpha = \Pr \left\{ \sup_t | \ \psi\{ \ F(t) \ \}\{ \ \hat{F}(t) - F(t) \ \}| \ \leq c_{1-\alpha}(\psi) \right\},$$

where $c_{1-\alpha}(\psi)$ is the $(1-\alpha)100\%$ point of $\sup_t |\psi\{ \ F(t) \ \} G_F(t)/\sqrt{n}|$.

The EP band corresponds to $\psi(u) = \{ \ u(\ 1-u \ ) \ \}^{-1/2}$. In practice, it is desirable to make $\psi(u)$ bounded. Following Nair (1984), we alternatively use $\psi_{EP}(u) = \{ \ (\ u \vee p_1 \ )(\ 1 - u \wedge p_2 \ ) \ \}^{-1/2}$, $0 < p_1 < p_2 < 1$, to yield the EP band

$$\hat{F}(t) \pm c_{1-\alpha}(\psi_{EP}) \sqrt{\{ \ \hat{F}(t) \vee p_1 \ \}\{ \ 1 - \hat{F}(t) \wedge p_2 \ \}}.$$

We set p_1 =0.1 or 0.2 and p_2 =0.8 or 0.9 as suggested by Nair (1984).

The HW band corresponds to $\psi_{HW}(u) \equiv 1$, which is the version of Kolmogorov–Smirnov band for uncensored data. The HW band is

$$\hat{F}(t) \pm c_{1-\alpha}(\psi_{HW}),$$

where $c_{1-\alpha}(\psi_{HW})$ is obtained as the $(1-\alpha)100\%$ point for $\{ \ K^{(b)}; \ b = 1, \ \ldots, \ B \ \}$ in Step 2 of the Kolmogorov–Smirnov test.

The bootstrap is useful to validate the coverage performance of the confidence bands above. First, the bootstrap NPMLEs, denoted as $\{ \ \hat{F}_b^*, \ b = 1, \ \ldots, \ B \ \}$, are computed (see Step 1 of Appendix A). Then, approximately $(1-\alpha)100\%$ of the bootstrap NPMLEs should fall inside the band. This validation scheme will be illustrated with real data analysis.

5 Simulations

Extensive simulations have been conducted to investigate the performances of the proposed methods and to compare them with the bootstrap and jackknife methods.

We adopt the same design used in Moreira and Uña-Álvarez (2010). They consider models $U^* \sim \text{Unif}(0,\ a)$, $T^* \sim \text{Unif}(0,\ 1)$, and $V^* \sim \text{Unif}(b,\ 1)$, where $(a,\ b) = (0.25,\ 0.75)$, $(0.5,\ 0.5)$ or $(0.67,\ 0.33)$. The corresponding sample inclusion probabilities are $\Pr(U^* \leq T^* \leq V^*) = (1 - a + b)/2$ =0.75, 0.5 and 0.33, respectively. They also consider a model $U^* \sim \text{Unif}(-5,\ 15)$, $T^* \sim \text{Unif}(0,\ 15)$, and $V^* = U^* + c$, where c =5. This model is important since it yields a situation similar to the childhood cancer example.

Based on simulated samples, we compute the relevant quantities (NPMLE, confidence interval, goodness-of-fit statistic, and confidence band) for 500 repetitions. We choose B = 1000 for the number of resamplings.

5.1 Performance of the covariance estimator

For r (= 1, ..., 500)-th repetition, we compute the NPMLE $\hat{F}(s)_{(r)}$, $\hat{F}(t)_{(r)}$ and the covariance estimator $\hat{C}ov\{\ \hat{F}(s),\ \ \hat{F}(t)\ \}_{(r)}$ in Eq. (3). We compare the average of the estimated covariance

$$\frac{1}{500}\sum_{r=1}^{500}\hat{C}ov\left\{\hat{F}(s),\ \hat{F}(t)\right\}_{(r)}$$

with the sample covariance

$$\frac{1}{500}\sum_{r=1}^{500}\left\{\hat{F}(s)_{(r)} - \bar{\hat{F}}(s)\right\}\left\{\hat{F}(t)_{(r)} - \bar{\hat{F}}(t)\right\}.$$

where $\bar{\hat{F}}(s) = \sum_{r=1}^{500}\hat{F}(s)_{(r)}/500$. As shown in Table 1, the differences between the estimated covariance and the sample covariance are very small for all configurations. The sample covariance between $\hat{F}(s)_{(r)}$ and $\hat{F}(t)_{(r)}$ increases as the distance $|t - s|$ decreases, which is a similar behavior to that of the empirical distribution function from un-truncated data.

5.2 Comparison with the bootstrap and jackknife methods

We compare the performance of the proposed variance estimator ($\hat{V}_{\text{Info}}\{\hat{F}(t)\}$), the bootstrap estimator ($\hat{V}_{\text{Boot}}\{\hat{F}(t)\}$) and the jackknife estimator ($\hat{V}_{\text{Jack}}\{\hat{F}(t)\}$) for fixed t. We compute the average of the estimated standard deviation (SD)

$$\frac{1}{500}\sum_{r=1}^{500}\sqrt{\hat{V}\{\hat{F}(t)\}_{(r)}}$$

Table 1 Simulation results of the proposed covariance estimator based on 500 replications

	$(s,\ t)$	n = 100		n = 250	
		Sample covariance	Estimated covariance	Sample covariance	Estimated covariance
a= 0.25, b = 0.75	(0.2, 0.5)	0.00565	0.00595	0.00195	0.00180
	(0.4, 0.5)	0.00655	0.00668	0.00234	0.00215
	(0.2, 0.8)	0.00279	0.00366	0.00112	0.00110
a= 0.5, b = 0.5	(0.2, 0.5)	0.00641	0.00704	0.00266	0.00258
	(0.4, 0.5)	0.00833	0.00877	0.00335	0.00336
	(0.2, 0.8)	0.00378	0.00450	0.00184	0.00169
a= 0.067, b = 0.33	(0.2, 0.5)	0.00366	0.00374	0.00157	0.00141
	(0.4, 0.5)	0.00677	0.00645	0.00283	0.00245
	(0.2, 0.8)	0.00356	0.00350	0.00159	0.00130
c = 5	(3.0, 7.5)	0.01166	0.01291	0.00525	0.00511
	(6.0, 7.5)	0.02027	0.02068	0.00860	0.00845
	(3.0, 12.0)	0.00558	0.00654	0.00269	0.00260

Data are generated from $U^* \sim \text{Unif}(0,\ a)$, $T^* \sim \text{Unif}(0,\ 1)$, and $V^* \sim \text{Unif}(b,\ 1)$ in the first three cases, and from $U^* \sim \text{Unif}(-5,\ 15)$, $T^* \sim \text{Unif}(0,\ 15)$, and $V^* = U^* + c$ in the last case

Sample covariance = $\frac{1}{500}\sum_{r=1}^{500}\{\hat{F}(s)_{(r)} - \bar{\hat{F}}(s)_{(r)}\}\{\hat{F}(t)_{(r)} - \bar{\hat{F}}(t)_{(r)}\}$

Estimated covariance = $\frac{1}{500}\sum_{r=1}^{500}\hat{C}ov\{\ \hat{F}(s),\ \hat{F}(t)\ \}_{(r)}$

where $\hat{V}\{\hat{F}(t)\}_{(r)}$ is a variance estimator for the rth repetition, and compare it with $\text{SD}\{\hat{F}(t)\}$, the sample standard deviation (SD) for $\hat{F}(t)_{(r)}$, $r = 1, \ \ldots, \ 500$. The performance of the three methods are measured with the mean squared error

$$\text{MSE} = \frac{1}{500}\sum_{r=1}^{500}\left(\sqrt{\hat{V}\{\hat{F}(t)\}_{(r)}} - \text{SD}\{\hat{F}(t)\}\right)^2.$$

We also compare the performance of the three methods in terms of the coverage performance of the 95% confidence interval.

Tables 2 and 3 show the results under the models $U^* \sim \text{Unif}(0,\ a)$, $T^* \sim \text{Unif}(0,\ 1)$, and $V^* \sim \text{Unif}(b,\ 1)$, where $(a,\ b) = (0.25,\ 0.75)$ and $(0.5,\ 0.5)$, respectively. All the three variance estimators correctly capture the estimates of $\text{SD}\{\hat{F}(t)\}$. Among the three estimators, the jackknife has the smallest bias. In terms of MSE, the bootstrap is the best for small samples, while the proposed method tends to be the best for large samples. For instance, the bootstrap is the best for $n = 100$, while the proposed method is the best for n =200, 250 and 300 (Table 2). The jackknife has the largest MSE in most configurations.

All the three methods generally produce the nominal 95% coverage performance at $t = 0.5$ ($F(t) = 0.5$). However, at the tail $t = 0.2$ ($F(t) = 0.2$), the bootstrap method often results in serious under-coverage. The magnitude of the under-coverage of the bootstrap is similar to that reported in the simulation results of Moreira and Uña-Álvarez (2010). Both the proposed and the jackknife methods alleviate the under-

Table 2 Simulation results under $U^* \sim \text{Unif}(0,\ a)$, $T^* \sim \text{Unif}(0,\ 1)$, and $V^* \sim \text{Unif}(b,\ 1)$ with a= 0.25 and b = 0.75 based on 500 replications

		n = 100	n = 150	n = 200	n = 250	n = 300
$F(t) = 0.5$						
SD		0.083	0.064	0.053	0.050	0.046
ESD	Proposed	0.070	0.057	0.050	0.045	0.042
	Bootstrap	0.070	0.059	0.051	0.046	0.043
	Jackknife	0.075	0.061	0.053	0.047	0.044
MSE	Proposed	0.00219	0.00086	0.00033	0.00028	0.00026
	Bootstrap	0.00104	0.00070	0.00048	0.00038	0.00035
	Jackknife	0.00296	0.00185	0.00094	0.00075	0.00073
95%Cov	Proposed	0.930	0.942	0.950	0.946	0.938
	Bootstrap	0.920	0.938	0.950	0.942	0.948
	Jackknife	0.930	0.950	0.948	0.946	0.940
$F(t) = 0.2$						
SD		0.090	0.065	0.057	0.052	0.048
ESD	Proposed	0.069	0.056	0.049	0.045	0.041
	Bootstrap	0.069	0.058	0.051	0.046	0.042
	Jackknife	0.074	0.061	0.053	0.047	0.043
MSE	Proposed	0.00394	0.00091	0.00073	0.00052	0.00044
	Bootstrap	0.00213	0.00113	0.00094	0.00067	0.00055
	Jackknife	0.00522	0.00248	0.00189	0.00115	0.00103
95%Cov	Proposed	0.938	0.942	0.946	0.932	0.942
	Bootstrap	0.898	0.910	0.928	0.908	0.924
	Jackknife	0.940	0.948	0.952	0.938	0.950

ESD = $\frac{1}{500}\sum_{r=1}^{500}\sqrt{\hat{V}\{\hat{F}(t)\}_{(r)}}$

MSE = $\frac{1}{500}\sum_{r=1}^{500}\left(\sqrt{\hat{V}\{\hat{F}(t)\}_{(r)}} - \text{SD}\{\hat{F}(t)\}\right)^2$

95%Cov = Empirical coverage probability of the 95% confidence interval

coverage at the tail. Interestingly, the jackknife is quite competitive with the proposed method in terms of coverage performance despite the poor performance of the MSE.

Table 4 shows the results under the model $U^* \sim \text{Unif}(-5,\ 15)$, $T^* \sim \text{Unif}(0,\ 15)$, and $V^* = U^* + 5$. All the three variance estimators are nearly unbiased and their MSEs are very similar. Although the bootstrap seems to provide the best result in terms of the MSE, the three methods are quite competitive. In terms of coverage probability, the bootstrap tends to be the best.

Although we found no single best method across all criteria, the advantage of the proposed method over other methods appears for larger samples (n = 250 and 300). The MSE of the proposed method is smallest in majority of cases. Unlike the bootstrap that may exhibit serious under-coverage at the tails, the proposed method can alleviate the problem for large sample sizes. As for the computational cost among the three methods, the proposed method is the lowest since it merely performs the matrix algebra in Eq. (4). On the other extreme, the bootstrap requires performing the

Table 3 Simulation results under $U^* \sim \text{Unif}(0,\ a)$, $T^* \sim \text{Unif}(0,\ 1)$, and $V^* \sim \text{Unif}(b,\ 1)$ with a= 0.5 and b = 0.5 based on 500 replications

		n = 100	n = 150	n = 200	n = 250	n = 300
$F(t) = 0.5$						
SD		0.093	0.078	0.068	0.059	0.055
ESD	Proposed	0.084	0.070	0.060	0.054	0.049
	Bootstrap	0.086	0.071	0.061	0.055	0.050
	Jackknife	0.092	0.076	0.063	0.057	0.051
MSE	Proposed	0.00237	0.00227	0.00141	0.00069	0.00059
	Bootstrap	0.00125	0.00104	0.00061	0.00054	0.00040
	Jackknife	0.00385	0.00326	0.00161	0.00150	0.00087
95%Cov	Proposed	0.932	0.934	0.950	0.952	0.958
	Bootstrap	0.934	0.944	0.954	0.952	0.958
	Jackknife	0.934	0.944	0.950	0.964	0.962
$F(t) = 0.2$						
SD		0.093	0.074	0.065	0.059	0.052
ESD	Proposed	0.080	0.066	0.058	0.052	0.048
	Bootstrap	0.084	0.069	0.060	0.054	0.049
	Jackknife	0.090	0.073	0.061	0.055	0.050
MSE	Proposed	0.00270	0.00132	0.00077	0.00055	0.00035
	Bootstrap	0.00237	0.00152	0.00087	0.00071	0.00046
	Jackknife	0.00551	0.00366	0.00138	0.00119	0.00051
95%Cov	Proposed	0.932	0.938	0.946	0.928	0.926
	Bootstrap	0.908	0.902	0.934	0.924	0.910
	Jackknife	0.944	0.950	0.952	0.942	0.938

ESD = $\frac{1}{500}\sum_{r=1}^{500}\sqrt{\hat{V}\{\hat{F}(t)\}_{(r)}}$
MSE = $\frac{1}{500}\sum_{r=1}^{500}\left(\sqrt{\hat{V}\{\hat{F}(t)\}_{(r)}} - \text{SD}\{\hat{F}(t)\}\right)^2$
95%Cov = Empirical coverage probability of the 95% confidence interval

self-consistency algorithms over $B = 1{,}000$ resamplings. Hence, the proposed method would be useful when the sample size is large.

5.3 Performance of the goodness-of-fit test

First, we examine the type I error of the goodness-of-fit tests introduced in Sect. 4.2. For each run, we record the rejection/acceptance status of the goodness-of-fit tests at the $100\alpha\%$ level, and calculate the rejection rates among 500 repetitions. We also compare the null means of the tests (denoted by $E[C]$ and $E[K]$) with the resampling means (denoted by $E[C^{(b)}]$ and $E[K^{(b)}]$).

As shown in Table 5, the rejection rates (type I error rates) are in good agreement with the selected nominal sizes ($\alpha = 0.01, 0.05,$ and 0.10). In addition, the sample means $E[C]$ and $E[K]$ are close to the resampling means $E[C^{(b)}]$ and $E[K^{(b)}]$, respectively.

Table 4 Simulation results under $U^* \sim \text{Unif}(-5,\ 15)$, $T^* \sim \text{Unif}(0,\ 15)$, and $V^* = U^* + 5$ based on 500 replications

		$n = 100$	$n = 150$	$n = 200$	$n = 250$	$n = 300$
$F(t) = 0.5$						
SD		0.146	0.121	0.103	0.096	0.086
ESD	Proposed	0.146	0.120	0.105	0.094	0.086
	Bootstrap	0.147	0.120	0.105	0.094	0.086
	Jackknife	0.156	0.125	0.108	0.096	0.088
MSE	Proposed	0.00064	0.00023	0.00013	0.000065	0.000046
	Bootstrap	0.00048	0.00020	0.00010	0.000069	0.000046
	Jackknife	0.00089	0.00026	0.00014	0.000067	0.000047
95%Cov	Proposed	0.904	0.932	0.946	0.928	0.928
	Bootstrap	0.940	0.944	0.950	0.950	0.930
	Jackknife	0.910	0.936	0.946	0.940	0.936
$F(t) = 0.2$						
SD		0.101	0.084	0.071	0.064	0.057
ESD	Proposed	0.100	0.081	0.070	0.062	0.057
	Bootstrap	0.106	0.083	0.072	0.063	0.058
	Jackknife	0.107	0.084	0.072	0.064	0.058
MSE	Proposed	0.00109	0.00048	0.00027	0.00018	0.00012
	Bootstrap	0.00097	0.00044	0.00024	0.00017	0.00012
	Jackknife	0.00133	0.00053	0.00028	0.00018	0.00013
95%Cov	Proposed	0.938	0.944	0.946	0.942	0.948
	Bootstrap	0.958	0.946	0.954	0.944	0.948
	Jackknife	0.956	0.954	0.952	0.942	0.954

ESD = $\frac{1}{500}\sum_{r=1}^{500}\sqrt{\hat{V}\{\hat{F}(t)\}_{(r)}}$
MSE = $\frac{1}{500}\sum_{r=1}^{500}\left(\sqrt{\hat{V}\{\hat{F}(t)\}_{(r)}} - \text{SD}\{\hat{F}(t)\}\right)^2$
95%Cov = Empirical coverage probability of the 95 % confidence interval

However, under $(a,\ b) = (0.67, 0.33)$, the Cramér–von Mises test leads to somewhat higher rejection rates than the nominal sizes. Overall, the Kolmogorov–Smirnov test produces a slightly conservative result.

Next, we examine the power under alternative hypotheses. We focus on the case of $(a,\ b) = (0.5,\ 0.5)$ under the null $F_0(t) = t\mathbf{I}(0 < t < 1)$ and alternatives

(1) $F_1(t) = t^{1/\gamma}\mathbf{I}(0 < t < 1)$, γ=1/1.8, 1/1.6, …, 1, …, 1.6, 1.8.
(2) $F_2(t) = t\mathbf{I}(0 < t < \gamma)/\gamma$, γ=1, 0.975, 0.95, …, 0.75, 0.725.

As shown in Fig. 2, the power increases as γ departs from the null model of $\gamma = 1$. The curves for $\alpha = 0.05$ (right panels) are consistently higher than those for $\alpha = 0.01$ (left panels). It is found that the Cramér–von Mises test exhibits higher power than the Kolmogorov–Smirnov test under the alternative model (1). This conclusion, however, should not be overemphasized as the type I error rates of the Cramér–von Mises test

Table 5 Simulation results for the proposed goodness-of-fit tests under the null hypothesis based on 500 replications

		Cramér–von Mises test (C)			Kolmogorov–Smirnov test (K)		
		n			n		
		100	150	250	100	150	250
$(a,\ b) =$ (0.25, 0.75)	Reject rate at $\alpha = 0.10$	0.096	0.097	0.100	0.075	0.075	0.095
	Reject rate at $\alpha = 0.05$	0.045	0.047	0.056	0.031	0.037	0.043
	Reject rate at $\alpha = 0.01$	0.006	0.011	0.018	0.006	0.005	0.015
	$E[C]$ or $E[K]$	0.608	0.623	0.667	0.118	0.099	0.081
	$E[C^{(b)}]$ or $E[K^{(b)}]$	0.625	0.614	0.605	0.115	0.096	0.077
$(a,\ b) =$ (0.5, 0.5)	Reject rate at $\alpha = 0.10$	0.120	0.105	0.088	0.089	0.081	0.078
	Reject rate at $\alpha = 0.05$	0.063	0.055	0.045	0.037	0.033	0.030
	Reject rate at $\alpha = 0.01$	0.015	0.014	0.008	0.006	0.007	0.005
	$E[C]$ or $E[K]$	1.078	1.206	1.306	0.143	0.121	0.098
	$E[C^{(b)}]$ or $E[K^{(b)}]$	1.167	1.281	1.331	0.137	0.116	0.093
$(a,\ b) =$ (0.67, 0.33)	Reject rate at $\alpha = 0.10$	0.140	0.150	0.120	0.090	0.065	0.105
	Reject rate at $\alpha = 0.05$	0.095	0.085	0.060	0.040	0.035	0.040
	Reject rate at $\alpha = 0.01$	0.015	0.020	0.030	0.010	0.005	0.010
	$E[C]$ or $E[K]$	1.006	1.109	0.989	0.142	0.118	0.091
	$E[C^{(b)}]$ or $E[K^{(b)}]$	1.001	0.984	0.820	0.135	0.113	0.087
$c = 5$	Reject rate at $\alpha = 0.10$	0.108	0.119	0.119	0.096	0.106	0.110
	Reject rate at $\alpha = 0.05$	0.063	0.057	0.059	0.052	0.055	0.055
	Reject rate at $\alpha = 0.01$	0.015	0.014	0.014	0.010	0.011	0.013
	$E[C]$ or $E[K]$	0.417	0.411	0.412	0.104	0.084	0.067
	$E[C^{(b)}]$ or $E[K^{(b)}]$	0.403	0.403	0.401	0.103	0.085	0.066

The average of the Cramér–von Mises statistics is denoted by $E[C]$. The average of its resampling version is denoted by $E[C^{(b)}]$. $E[K]$ and $E[K^{(b)}]$ are defined similarly for the Kolmogorov–Smirnov statistics

are slightly higher than those of the Kolmogorov–Smirnov test. The results for other $(a,\ b)$ are similar.

5.4 Performance of the confidence band

We investigate the coverage performance of the EP and HW bands introduced in Sect. 4.3. The EP band is calculated under p_1 =0.2 and p_2 =0.8. For each run, we check if the bands completely cover the true F. The coverage rates over the 500 replications are given in Table 6. Overall, the coverage rates reflect the nominal levels and are particularly accurate when $1 - \alpha = 0.99$. The EP band has slightly more accurate coverage compared to the HW band, especially at levels $1 - \alpha = 0.90$ and 0.95. This is because the HW band exhibits slight over-coverage, which parallels the conservative results of the Kolmogorov–Smirnov test.

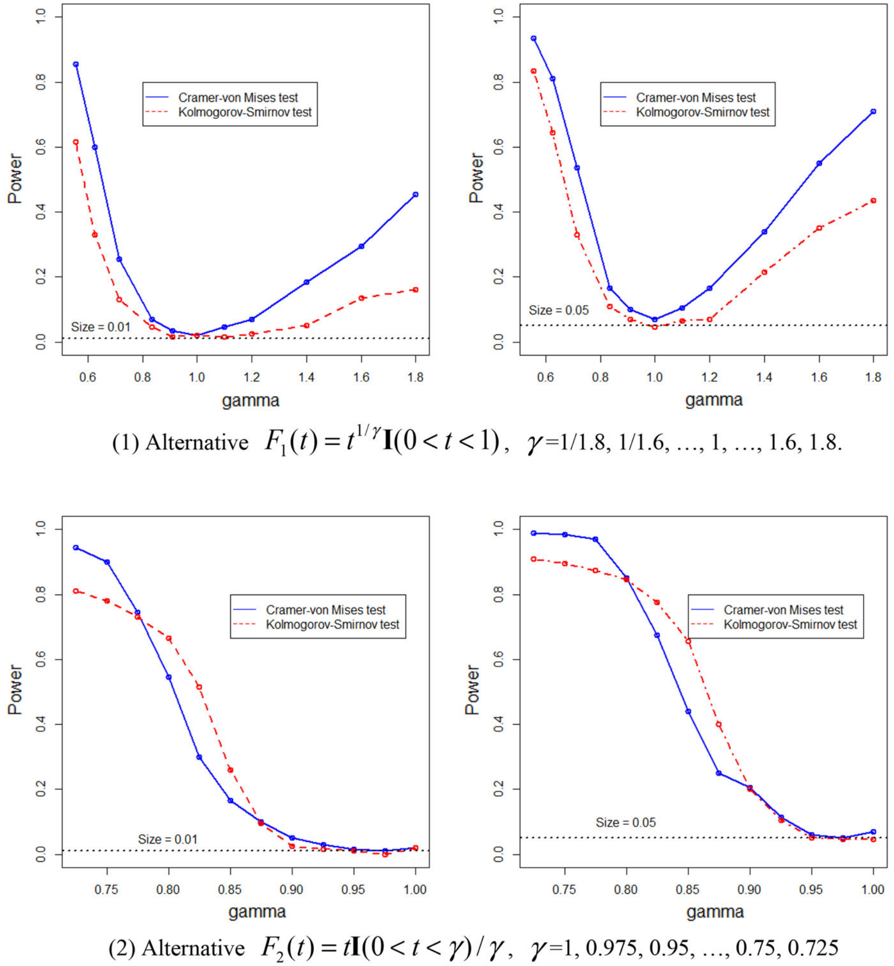

Fig. 2 The power curves for the proposed goodness-of-fit tests with sizes α =0.01 (*left panel*) and α =0.05 (*right panel*) based on n = 150. The value $\gamma = 1$ corresponds to the null, while $\gamma \neq 1$ corresponds to the alternative. (1) Alternative $F_1(t) = t^{1/\gamma}\mathbf{I}(0 < t < 1)$, γ=1/1.8, 1/1.6, …, 1, …, 1.6, 1.8. (2) Alternative $F_2(t) = t\mathbf{I}(0 < t < \gamma)/\gamma$, γ=1, 0.975, 0.95, …, 0.75, 0.725

6 Data analysis

We analyzed the childhood cancer data from Moreira and Uña-Álvarez (2010) as described in Sect. 1. The sample consists of 409 children with $\{ (U_j, T_j, V_j) : j = 1, \ldots, 409 \}$ subject to double-truncation $U_j \le T_j \le V_j$, where T_j is the age (in days) at diagnosis, U_j is the age at the start of follow-up (January 1, 1999), and $V_i = U_i + 1825$ is the age at the end of follow-up (December 31, 2003). The primary interest here is inference of the distribution function $F(t) = \Pr(T^* \le t)$, where T^* is the pre-truncated age at diagnosis. We depict the NPMLE $\hat{F}(t)$ in Fig. 3. The resulting curve is virtually identical to that reported in Moreira and Uña-Álvarez (2010). They

Table 6 Coverage rates of the proposed confidence bands at the $100(1-\alpha)\%$ level based on 500 replications

	Nominal level	$n = 100$	$n = 150$	$n = 250$
EP (equal precision) band				
$a = 0.25, b = 0.75$	$1-\alpha = 0.900$	0.904	0.924	0.902
	$1-\alpha = 0.950$	0.958	0.952	0.950
	$1-\alpha = 0.990$	0.990	0.990	0.984
$a = 0.5, b = 0.5$	$1-\alpha = 0.900$	0.908	0.910	0.918
	$1-\alpha = 0.950$	0.964	0.954	0.960
	$1-\alpha = 0.990$	0.990	0.988	0.990
$a = 0.67, b = 0.33$	$1-\alpha = 0.900$	0.915	0.905	0.910
	$1-\alpha = 0.950$	0.950	0.955	0.950
	$1-\alpha = 0.990$	0.985	0.995	0.985
$c = 5$	$1-\alpha = 0.900$	0.894	0.894	0.876
	$1-\alpha = 0.950$	0.928	0.940	0.932
	$1-\alpha = 0.990$	0.984	0.986	0.986
HW (Hall–Wellner) band				
$a = 0.25, b = 0.75$	$1-\alpha = 0.900$	0.927	0.926	0.905
	$1-\alpha = 0.950$	0.969	0.963	0.957
	$1-\alpha = 0.990$	0.994	0.995	0.985
$a = 0.5, b = 0.5$	$1-\alpha = 0.900$	0.912	0.919	0.922
	$1-\alpha = 0.950$	0.963	0.967	0.970
	$1-\alpha = 0.990$	0.994	0.993	0.995
$a = 0.67, b = 0.33$	$1-\alpha = 0.900$	0.910	0.935	0.895
	$1-\alpha = 0.950$	0.960	0.965	0.960
	$1-\alpha = 0.990$	0.990	0.995	0.990
$c = 5$	$1-\alpha = 0.900$	0.904	0.894	0.890
	$1-\alpha = 0.950$	0.948	0.945	0.945
	$1-\alpha = 0.990$	0.990	0.989	0.987

provide pointwise confidence intervals using the bootstrap. In this paper, we provide additional inference procedures using goodness-of-fit tests and confidence bands.

For goodness-of-fit tests, we set the following two hypotheses:

$$H_{01}: \quad F(t) = \frac{t}{5475}\mathbf{I}(0 < t < 5475) + \mathbf{I}(t \geq 5475)$$

and

$$H_{02}: \quad F(t) = \left(\frac{t}{5475}\right)^{3/4}\mathbf{I}(0 < t < 5475) + \mathbf{I}(t \geq 5475),$$

where $5{,}475 = 15 \times 365$ (days) is the maximum age to be defined as childhood cancer (15 years old). Here, H_{01} implies that childhood cancer occurs uniformly over all ages

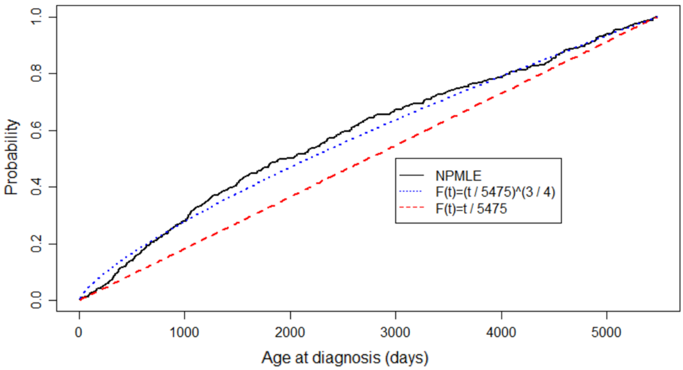

Fig. 3 The NPMLE $\hat{F}(t)$ of the distribution function of ages at diagnosis for childhood cancer (*solid line*). The hypothesized curves are H_{01} : $F(t) = (t/5475)\ \mathbf{I}(0 < t < 5475) + \mathbf{I}(t \geq 5475)$ (*dashed line*), H_{02} : $F(t) = (t/5475)^{3/4}\mathbf{I}(0 < t < 5475) + \mathbf{I}(t \geq 5475)$ (*dotted line*)

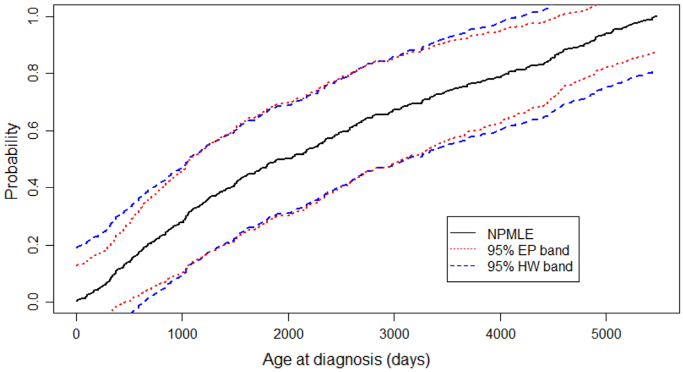

Fig. 4 The NPMLE and its 95% confidence bands. The *dotted line* is the EP (equal precision) band and the *dashed line* is the HW (Hall–Wellner) band

under 15 years, while H_{02} implies that the occurrence of childhood cancer decreases as their age increases. Figure 3 depicts the two hypothesized curves along with the NPMLE. The curve for H_{02} fits better than the curve for H_{01}. Indeed, the Cramér–von Mises test rejects H_{01} at 10 % significance level (P-value = 0.094), while does not reject H_{02} (P-value = 0.732). Similar results are found through the Kolmogorov–Smirnov test (P-value = 0.099 for H_{01} and P-value = 0.797 for H_{02}).

Figure 4 displays the 95 % EP and HW bands based on the algorithm in Sect. 4.3. The EP band is calculated under p_1 =0.1 and p_2 =0.9. The EP and HW bands are generally competitive, but the EP band is slightly narrower in the tails. This is qualitatively similar to the EP and HW bands for right-censored data. Now, we validate the coverage performance using the bootstrap as mentioned in Sect. 4.3. The EP band covers 950 out of the 1000 bootstrap NPMLEs and the HW band covers 964 out of the 1000 bootstrap NPMLEs. Hence, the coverage level is close to the nominal 95 %.

We compare the three variance estimators (proposed, bootstrap and jackknife) for selected values of t. The computation time required for the three estimators are also

Table 7 Variance estimates of the NPMLE based on the childhood cancer data

	Proposed: $\sqrt{\hat{V}_{\text{Info}}\{\hat{F}(t)\}}$	Bootstrap: $\sqrt{\hat{V}_{\text{Boot}}\{\hat{F}(t)\}}$	Jackknife: $\sqrt{\hat{V}_{\text{Jack}}\{\hat{F}(t)\}}$
Variance estimate at t = 750.0	0.0469	0.0464	0.0473
Computation time (s)	(0.25)	(342.16)	(118.37)
Variance estimate at t = 2,083.5	0.0817	0.0814	0.0815
Computation time (s)	(0.22)	(313.53)	(115.83)
Variance estimate at t = 4,251.0	0.0599	0.0665	0.0644
Computation time (s)	(0.23)	(317.48)	(117.37)

The three variance estimators are calculated at t = 750.0, t = 2,083.5, and t = 4,251.0, corresponding to the 20, 50, and 80 percentiles of observed ages at diagnosis, respectively. Required computation times for the three methods are also compared

compared. As shown in Table 7, the three estimates produce very similar results for all t. On the other hand, the computation time required for the proposed method is much shorter than those of the resampling-based methods.

7 Conclusion

This paper introduced a simple and explicit covariance estimator of the NPMLE using the observed information matrix. This technique provides various inference procedures, including pointwise confidence interval, goodness-of-fit, and confidence band.

Our simulations showed that the major advantage of the proposed variance estimator over the bootstrap and jackknife was for the larger samples (n = 250 and 300). The data analysis demonstrated the reduced computational time for the proposed method vis-à-vis the bootstrap and jackknife methods. Hence, the proposed method is most useful when the sample size is very large, which often occurs in demography and epidemiology (e.g., Stovring and Wang 2007). In such large-scale studies, the proposed method may be the best possible choice for statistical inference.

For goodness-of-fit procedures, we developed the Kolmogorov–Smirnov and Cramér–von Mises tests with the null distributions simulated by the proposed covariance structure. The simulations showed that these tests have proper type I error rates and power. Applying the tests to the childhood cancer data, we rejected the scientific assumption that childhood cancer occurs uniformly over all ages below 15 years. This conclusion could not have been derived without developing the goodness-of-fit procedures.

Acknowledgments We would like to thank the editor, the associate editor and the two reviewers for their helpful comments and corrections that greatly improved the manuscript. This work was financially supported by the National Science Council of Taiwan (NSC101-2118-M008-002-MY2) to T. Emura, and a Grant-in-Aid for a Research Fellow of the Japan Society for the Promotion of Science to H. Michimae (No. 23570036). The work of Y. Konno was partially supported by Grant-in-Aid for Scientific Research(C) (No. 25330043 and 21500283).

Appendix A: Bootstrap and jackknife algorithms

Simple bootstrap algorithm (Moreira and Uña-Álvarez 2010):

Step 1: For each $b = 1, \ldots, B$, draw bootstrap resamples $\{ (U^*_{jb}, T^*_{jb}, V^*_{jb}) : j = 1, \ldots, n \}$ from $\{ (U_j, T_j, V_j) : j = 1, \ldots, n \}$, and then compute the NPMLE $\hat{F}^*_b(t)$ from them.
Step 2: Compute the bootstrap variance estimator

$$\hat{V}_{\text{Boot}}\{\hat{F}(t)\} = \frac{1}{B-1}\sum_{b=1}^{B}\{\hat{F}^*_b(t) - \bar{F}^*(t)\}^2,$$

where $\bar{F}^*(t) = \frac{1}{B}\sum_{b=1}^{B}\hat{F}^*_b(t)$, and take the $(\alpha/2) \times 100\%$ and $(1-\alpha/2)\times 100\%$ points of $\{ \hat{F}^*_b(t) : b = 1, \ldots, B \}$ for the$(1-\alpha)\times 100\%$ confidence interval.

Jackknife algorithm:

Step 1: For each $i = 1, \ldots, n$, delete the i th sample from $\{ (U_j, T_j, V_j) : j = 1, \ldots, n \}$, and then compute the NPMLE $\hat{F}_{(-i)}(t)$ from the remaining $n-1$ samples.
Step 2: Compute the jackknife variance estimator

$$\hat{V}_{\text{Jack}}\{\hat{F}(t)\} = \frac{n-1}{n}\sum_{i=1}^{n}\{\hat{F}_{(-i)}(t) - \bar{F}_{(\cdot)}(t)\}^2,$$

where $\bar{F}_{(\cdot)}(t) = \frac{1}{n}\sum_{i=1}^{n}\hat{F}_{(-i)}(t)$, and the log-transformed $(1-\alpha)\times 100\%$ confidence interval

$$\left(\hat{F}(t)\exp[-z_{\alpha/2}\hat{V}^{1/2}_{\text{Jack}}\{\hat{F}(t)\}/\hat{F}(t)],\ \ \hat{F}(t)\exp[z_{\alpha/2}\hat{V}^{1/2}_{\text{Jack}}\{\hat{F}(t)\}/\hat{F}(t)] \right).$$

Appendix B: Asymptotic theory

Appendix B1. Weak convergence of $\sqrt{n}(\hat{F}(t) - F(t))$

Although not stated explicitly, we assume that the identifiability conditions (Shen 2010, p. 836) are satisfied. Consider the log-likelihood function

$$\ell_n(F)/n = \sum_{i=1}^{n}(\log f_j - \log F_j)/n.$$

For any $h \in Q$, where Q is the set of all uniformly bounded functions, let $H(t) = \int_0^t h(s)dF(s)$ and $\hat{H}(t) = \int_0^t h(s)d\hat{F}(s)$ where h satisfies the constraint $\hat{H}(\infty) = 1$. Suppose that $\hat{F}$ is the maximizer of $\ell_n(F)$. Then for any $h \in Q$ and $\varepsilon \geq 0$, we have $\ell_n(\hat{F} + \varepsilon\hat{H}) \leq \ell_n(\hat{F})$. Hence, the score function $\partial\ell_n(F + \varepsilon H)/\partial\varepsilon|_{\varepsilon=0}$ is equal to

$$\Psi_n(F)[h] \equiv \frac{1}{n}\sum_{i=1}^{n}\left[h(T_i) - \frac{\int \mathbf{I}(U_i \le s \le V_i)h(s)dF(s)}{\int \mathbf{I}(U_i \le s \le V_i)dF(s)}\right],$$

for any $h \in Q$. The expectation is defined as

$$\Psi(F)[h] \equiv E\left[h(T^*) - \frac{\int \mathbf{I}(U^* \le s \le V^*)h(s)dF(s)}{\int \mathbf{I}(U^* \le s \le V^*)dF(s)}\right].$$

Consider $\Psi_n(F)[h]$ as a random function defined on Q. Accordingly, consider a random map $\Theta \to l^\infty(Q)$, defined by $F \mapsto \Psi_n(F)[\cdot]$. Then, the equation $\Psi_n(F)[\cdot] = 0$ is considered the estimating function that takes its value on $l^\infty(Q)$. It follows that the NPMLE is the Z-estimator that satisfies $\Psi_n(\hat{F})[\cdot] = 0$ (van der Vaart and Wellner 1996, p. 309). In the following, we assume that certain regularity conditions for the asymptotic theory for the Z-estimator hold, which include the asymptotic approximation condition, the Fréchet differentiability of the map, and the invertibility of the derivative map.

Then, one can write

$$0 = n^{1/2}\Psi_n(\hat{F})[h] = n^{1/2}\Psi_n(F)[h] + n^{1/2}\dot{\Psi}_F(\hat{F} - F)[h] + o_P(1), \tag{5}$$

where $\dot{\Psi}_F(\hat{F} - F)[h]$ is the derivative of $\Psi_n(F)[h]$ at F with direction $\hat{F} - F$. It follows from the form of $\Psi(F)[\cdot]$ that

$$\dot{\Psi}_F(\hat{F} - F)[h] = \frac{d}{dt}\Psi\{\ \hat{F} + t(\hat{F} - F)\ \}[h]|_{t=0} = -\int \sigma_F(x)[h]d(\hat{F} - F)(x). \tag{6}$$

It follows from Eqs. (5) and (6) that the NPMLE satisfies the asymptotic linear expression

$$\begin{aligned}&\sqrt{n}\int \sigma_F(x)[h]d(\hat{F} - F)(x)\\ &\quad= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[h(T_i) - \frac{\int I(U_i \le s \le V_i)h(s)dF(s)}{\int I(U_i \le s \le V_i)dF(s)}\right] + o_P(1),\end{aligned} \tag{7}$$

where the right-side converges weakly to a mean zero Gaussian process with the covariance structure

$$\begin{aligned}&E\left[h(T^*) - \frac{\int \mathbf{I}(U^* \le s \le V^*)h(s)dF(s)}{\int \mathbf{I}(U^* \le s \le V^*)dF(s)}\right]\left[h'(T^*) - \frac{\int \mathbf{I}(U^* \le s \le V^*)h'(s)dF(s)}{\int \mathbf{I}(U^* \le s \le V^*)dF(s)}\right]\\ &= \int \sigma_F(x)[h]h'(x)dF(x),\end{aligned}$$

for bounded functions h and h'. The desired weak convergence of $\sqrt{n}(\ \hat{F}(t) - F(t)\ )$ is obtained by setting $h = \sigma_F^{-1}(w_t)$ in Eq. (7).

Appendix B2: Proof of $\sum_{j=1}^{n} w_s(T_j)\hat{\sigma}_F^{-1}(w_t)(T_j)\hat{f}_j = \mathbf{W}_s^{\mathrm{T}} \left\{ \frac{i_n(\mathbf{f})}{n} \right\}^{-1} \mathbf{W}_t$

It follows that

$$\hat{\sigma}_F(T_j)[h] = \frac{1}{n}\sum_{i=1}^{n} J_{ij}\left\{ \frac{h_j}{\hat{F}_i} - \frac{1}{\hat{F}_i^2}\sum_{k=1}^{n} J_{ik} h_k \hat{f}_k \right\} = \frac{1}{n}\left[\frac{h_j \hat{f}_j}{\hat{f}_j^2} - \sum_{i=1}^{n}\sum_{k=1}^{n} \frac{J_{ij} J_{ik}}{\hat{F}_i^2} h_k \hat{f}_k \right]. \tag{8}$$

Note that

$$J^{\mathrm{T}} \mathrm{diag}\left(\frac{1}{\mathbf{F}^2}\right) J = \begin{bmatrix} \sum_{i=1}^{n} \frac{J_{i1} J_{i1}}{F_i^2} & \cdots & \sum_{i=1}^{n} \frac{J_{i1} J_{in}}{F_i^2} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} \frac{J_{in} J_{i1}}{F_i^2} & \cdots & \sum_{i=1}^{n} \frac{J_{in} J_{in}}{F_i^2} \end{bmatrix}.$$

Hence, Eq. (8) with $h = h'$ and $\sigma_F(x)[h'] = w_t(x) = \mathbf{I}(x \le t)$ yield

$$\begin{bmatrix} w_t(T_1) \\ \vdots \\ w_t(T_n) \end{bmatrix} = \frac{1}{n}\left[\left\{ \mathrm{diag}\left(\frac{1}{\hat{\mathbf{f}}^2}\right) - J^{\mathrm{T}} \mathrm{diag}\left(\frac{1}{\hat{\mathbf{F}}^2}\right) J \right\}\Big|_{\hat{f}_n = 1 - \mathbf{1}_{n-1}^{\mathrm{T}} \hat{\mathbf{f}}} \right] \begin{bmatrix} h_1 \hat{f}_1 \\ \vdots \\ h_n \hat{f}_n \end{bmatrix}$$

$$= \frac{1}{n}\left[\left\{ \mathrm{diag}\left(\frac{1}{\hat{\mathbf{f}}^2}\right) - J^{\mathrm{T}} \mathrm{diag}\left(\frac{1}{\hat{\mathbf{F}}^2}\right) J \right\}\Big|_{\hat{f}_n = 1 - \mathbf{1}_{n-1}^{\mathrm{T}} \hat{\mathbf{f}}} \right] D^{\mathrm{T}} \begin{bmatrix} h_1 \hat{f}_1 \\ \vdots \\ h_{n-1} \hat{f}_{n-1} \end{bmatrix},$$

where the last equation uses the constraint $\sum_{j=1}^{n} h_j \hat{f}_j = 0$. Multiplying D for both sides, and taking the inverse of the information matrix,

$$\begin{bmatrix} \hat{\sigma}_F^{-1}(w_t)(T_1)\hat{f}_1 \\ \vdots \\ \hat{\sigma}_F^{-1}(w_t)(T_{n-1})\hat{f}_{n-1} \end{bmatrix} = \left\{ \frac{i_n(\hat{\mathbf{f}})}{n} \right\}^{-1} \begin{bmatrix} w_t(T_1) - w_t(T_n) \\ \vdots \\ w_t(T_1) - w_t(T_n) \end{bmatrix}.$$

It follows that

$$\sum_{j=1}^{n} w_s(T_j)\hat{\sigma}_F^{-1}(w_t)(T_j)\hat{f}_j = \sum_{j=1}^{n-1} \{\, w_s(T_j) - w_s(T_n) \,\}\hat{\sigma}_F^{-1}(w_t)(T_j)\hat{f}_j$$

$$= \left[\, w_s(T_1) - w_s(T_n) \cdots w_s(T_{n-1}) - w_s(T_n) \,\right] \left\{ \frac{i_n(\hat{\mathbf{f}})}{n} \right\}^{-1} \begin{bmatrix} w_t(T_1) - w_t(T_n) \\ \vdots \\ w_t(T_1) - w_t(T_n) \end{bmatrix}$$

$$= \mathbf{W}_s^{\mathrm{T}} \left\{ \frac{i_n(\hat{\mathbf{f}})}{n} \right\}^{-1} \mathbf{W}_t.$$

References

Austin D, Simon DK, Betensky RA (2013) Computationally simple estimation and improved efficiency for special cases of double truncation. Lifetime Data Anal. doi:10.1007/s10985-013-9287-z

Chen YH (2010) Semiparametric marginal regression analysis for dependent competing risks under an assumed copula. J R Stat Soc B 72:235–251

Commenges D (2002) Inference for multi-state models from interval-censored data. Stat Methods Med Res 11:167–182

Efron B, Petrosian V (1999) Nonparametric method for doubly truncated data. J Am Stat Assoc 94:824–834

Emura T, Wang W (2012) Nonparametric maximum likelihood estimation for dependent truncation data based on copulas. J Multivar Anal 110:171–188

Emura T, Konno Y (2012) Multivariate normal distribution approaches for dependently truncated data. Stat Papers 53:133–149

Klein JP, Moeschberger ML (2003) Survival analysis: techniques for censored and truncated data. Springer, New York

Moreira C, Uña-Álvarez J (2010) Bootstrapping the NPMLE for doubly truncated data. J Nonparametr Stat 22:567–583

Moreira C, Uña-Álvarez J (2012) Kernel density estimation with doubly-truncated data. Electron J Stat 6:501–521

Moreira C, Keilegom IV (2013) Bandwidth selection for kernel density estimation with doubly truncated data. Comput Stat Data Anal 61:107–123

Moreira C, Uña-Álvarez J, Meira-Machado L (2014) Nonparametric regression with doubly truncated data. Comput Stat Data Anal. doi:10.1016/j.csda.2014.03.017

Murphy SA (1995) Asymptotic theory for the frailty model. Ann Stat 23:182–198

Nair VN (1984) Confidence bands for survival functions with censored data: a comparative study. Technometrics 26:265–275

Shen PS (2010) Nonparametric analysis of doubly truncated data. Ann Inst Stat Math 62:835–853

Shen PS (2011) Testing quasi-independence for doubly truncated data. J Nonparametr Stat 23:1–9

Shen PS (2012) Empirical likelihood ratio with doubly truncated data. J Appl Stat 38:2345–2353

Stovring H, Wang MC (2007) A new approach of nonparametric estimation of incidence and lifetime risk based on birth rates and incidence events. BMC Med Res Methodol 7:53

van der Vaart AW, Wellner JA (1996) Weak convergence and empirical process. Springer-Varlag, New York

Zeng D, Lin DY (2006) Efficient estimation of semiparametric transformation models for counting processes. Biometrika 93:627–640

Zhu H, Wang MC (2012) Analyzing bivariate survival data with interval sampling and application to cancer epidemiology. Biometrika 99:345–361