Electronic Journal of Statistics
Vol. 16 (2022) 3099–3134
ISSN: 1935-7524
https://doi.org/10.1214/21-EJS1970

Penalized nonparametric likelihood-based inference for current status data model*

Meiling Hao

School of Statistics, University of International Business and Economics
e-mail: meilinghao@uibe.edu.cn

Yuanyuan Lin†

Department of Statistics, The Chinese University of Hong Kong
e-mail: ylin@sta.cuhk.edu.hk

Kin-yat Liu and Xingqiu Zhao†

Department of Applied Mathematics, The Hong Kong Polytechnic University
e-mail: kent.liu.kin.yat@polyu.edu.hk; xingqiu.zhao@polyu.edu.hk

Abstract: Deriving the limiting distribution of a nonparametric estimate is rather challenging but of fundamental importance to statistical inference. For the current status data, we study a penalized nonparametric likelihood-based estimator for an unknown cumulative hazard function, and establish the pointwise asymptotic normality of the resulting nonparametric estimate. We also propose the penalized likelihood ratio tests for local and global hypotheses, derive their limiting distributions, and study the optimality of the global test. Simulation studies show that the proposed method works well compared to the classical likelihood ratio test.

Keywords and phrases: Current status data, functional Bahadur representation, likelihood ratio test, nonparametric inference, penalized likelihood.

Received August 2020.

Contents

*Meiling Hao's research is partly supported by the National Natural Science Foundation of China (No. 11901087) and the Program for Young Excellent Talents, UIBE (No. 19YQ15). Yuanyuan Lin's research is partially supported by the Hong Kong Research Grants Council (Grant No. 14306219 and 14306620) and Direct Grants for Research, The Chinese University of Hong Kong. Xingqiu Zhao's research is partly supported by the Research Grant Council of Hong Kong (15303319, 15306521) and The Hong Kong Polytechnic University.

†Co-corresponding authors.

1. Introduction

When analyzing survival data, interval censoring arises frequently in medical and public health studies. In particular, interval-censored data occur when the exact failure time cannot be observed, instead, it is known to lie within an interval or not. Among these data, due to the constraints, costs, features of events of interest and many other difficulties, an event of interest is observed once only but the failure time is just known before the examination time or not. This kind of data is called case I interval-censored data or current status data ([15, 25]).

The nonparametric maximum likelihood, as a widely used method in nonparametric inference, has been developed for estimation of an unknown cumulative distribution function with current status data. In particular, [1] and [9] derived the nonparametric maximum likelihood estimator (NPMLE) with current status data, while [15] and [17] established the limit distribution of the NPMLE for current status data. [3] studied a likelihood ratio test to construct pointwise confidence intervals for an unknown distribution function with current status data. This novel method was investigated further in [4, 5, 2]. The NPMLE is an important method and widely used in applications, since it does not involve any tuning parameter and it converges in distribution pointwisely to an asymptotically unbiased limiting distribution. However, due to the discontinuity of NPMLE, estimators of the density and the hazard rate cannot be directly obtained through taking derivatives. For the problem, [6] provided adaptive risk bounds for a wide class of distribution function estimators under current status data based on smoothness classes such as splines, wavelets, etc. [14] proposed the maximum smoothed likelihood and the smoothed maximum likelihood estimators. Moreover, [11] presented the maximum smoothed likelihood estimator for the interval censoring model, [12] discussed these nonparametric estimators, and [13] improved the algorithm for computation and also provided an alternative way to derive the distribution of the likelihood ratio. [27] developed an I-spline based sieve maximum likelihood method for estimating the marginal and joint distribution functions with bivariate current status data.

As the hazard and cumulative hazard functions can provide different insights about an event of interest from a survival function, and the survival function can be directly obtained from the hazard and cumulative hazard functions via simple calculations, a number of studies have been developed for estimating the hazard and the cumulative hazard. While the nonparametric estimation without smoothing is not stable, some smoothing estimation approaches have been developed. For example, the kernel-based approaches were given in [10] and [14]. In order to avoid selecting the sensitive bandwidth in estimation, spline methods have also been developed for the case of interval-censored data. Specifically, [20] used M-splines to model the hazard and I-splines for the cumulative hazard, [22] suggested using the nonnegative coefficients to model the hazard, while [7] developed the penalized B-spline basis to model the hazard. [16] considered Cox's proportional hazards model with current status data, and studied the asymptotic properties of the maximum likelihood estimators of the regression parameter and the baseline cumulative hazard function. [18] proposed likelihood ratio tests and confidence intervals for the current status data with the model studied by [16]. Furthermore, [28] developed a B-spline based semiparametric maximum likelihood approach for Cox's proportional hazards model with case II interval-censored data.

Despite the significant contributions in the literature, the limiting distributions of the spline-based estimators have not been established yet. Our goal is to fill this gap and address the theoretical challenge. An additional challenge is to study the likelihood ratio test for the global hypothesis, which has not been addressed in the current status data model. In this paper, we develop a penalized nonparametric likelihood method to estimate an unknown cumulative hazard function with current status data. In particular, a functional Bahadur representation is established. Using this technical tool, we show that the proposed estimator enjoys the pointwise asymptotic normality. Furthermore, we study the penalized likelihood ratio tests and show the optimality of the global test.

The remainder of this paper is organized as follows. In Section 2, we present estimation procedures, construct the Sobolev space with a special inner product, and give some basic results. In Section 3, we derive a functional Bahadur representation (FBR) in the space and establish the asymptotic properties of the proposed estimator. In Section 4, we develop the penalized likelihood ratio tests for local and global hypotheses. In Section 5, we present simulation results for comparing the performance of the proposed penalized likelihood ratio test and the classical likelihood ratio test [3]. Some concluding remarks are made in Section 6. All technical proofs are deferred to the Appendix.

2. Methodology

Denote U as an examination or observation time and T as a failure time with an unknown distribution function F. Then under the scenario of current status data, the observation consists of $X = (\Delta, U)$, where $\Delta = I(T \leq U)$. In this paper, we assume that the examination time is independent of the failure time. Such an assumption is less stringent than that of [6] and common for the simplest version of the current status model in survival analysis; see [26]. Let $X_i = (\Delta_i, U_i), i = 1, 2, \ldots, n$ be i.i.d copies of $X = (\Delta, U)$. Under the assumption that U is independent of T, the log-likelihood of F is

$$l_n(F) = \frac{1}{n}\sum_{i=1}^{n}\left[\Delta_i \log\{F(U_i)\} + (1-\Delta_i)\log\{1-F(U_i)\}\right]$$

by omitting the term independent of F.

Assume that there exists a small positive constant ξ such that $P(U \geq \xi) = 1$. Let Λ_0 be the true cumulative hazard function of the failure time. Assume that $\Lambda_0(t)$ is increasing and bounded away from 0 and infinity on $\mathbb{I} = [\xi, \tau]$, where τ is the end of the study period. This assumption is less stringent than (F.1) in [14], since we do not need to assume the distribution of the survival time has bounded support. Besides, the introduction of ξ is to make the definition of the inner product in (1) meaningful. In fact, we can relax this assumption to $\xi = 0$. In applications, we can choose ξ as the minimum observation time and τ as the maximum follow-up time. Moreover, it is assumed that Λ_0 belongs to the Sobolev space $\mathcal{H}^m$, where

$$\mathcal{H}^m = \{g : \mathbb{I} \mapsto \mathbb{R} | g^{(j)} \quad \text{is absolutely continuous for} \\ j = 0, 1, \ldots, m-1, g^{(m)} \in L_2(\mathbb{I})\}.$$

Here $m > 1$ is assumed to be known, and $g^{(j)}$ is the j-th derivative of any function g. Without loss of generality, we assume that $\mathbb{I} = [\xi, 1+\xi]$. Then, the log-likelihood of Λ is

$$l_n(\Lambda) \quad = \quad \frac{1}{n}\sum_{i=1}^{n}\left(\Delta_i \log[1-\exp\{-\Lambda(U_i)\}] - (1-\Delta_i)\Lambda(U_i)\right).$$

Define $l(\Lambda) = El_n(\Lambda)$ and $J(g, \tilde{g}) = \int_{\mathbb{I}} g^{(m)}(t)\tilde{g}^{(m)}(t)\,dt$. To make an inference about $\Lambda_0(t)$, we propose the following penalized log-likelihood of Λ

$$\ell_{n,\lambda}(\Lambda) \quad = \quad l_n(\Lambda) - \frac{\lambda}{2}J(\Lambda, \Lambda),$$

where $J(\Lambda, \Lambda)$ is the roughness penalty and λ is the smoothing parameter. Then, the penalized likelihood estimator of Λ_0 is defined as

$$\hat{\Lambda}_{n,\lambda} = \arg\max_{\Lambda \in \mathcal{H}^m} \ell_{n,\lambda}(\Lambda).$$

We will show that $\hat{\Lambda}_{n,\lambda}$ is increasing on $\mathbb{I}$ with probability tending to 1 in the next section.

Define $\ell_\lambda(\Lambda) = E\ell_{n,\lambda}(\Lambda)$ and the inner product in the space $\mathcal{H}^m$ as

$$\langle g, h\rangle_\lambda = E_U\left[\frac{h(U)g(U)\exp\{-\Lambda_0(U)\}}{1-\exp\{-\Lambda_0(U)\}}\right] + \lambda\int_{\mathbb{I}} g^{(m)}(t)h^{(m)}(t)\,dt, \qquad (1)$$

where E_U is the expectation with respect to U and the corresponding norm is $\|g\|_\lambda^2 = \langle g, g\rangle_\lambda$. Hence, $\mathcal{H}^m$ is the reproducing kernel Hilbert space (RKHS) with the inner product $\langle \cdot, \cdot \rangle_\lambda$. Besides, there exists a positive self-adjoint operator: $W_\lambda : \mathcal{H}^m \to \mathcal{H}^m$, such that $\langle W_\lambda f, g\rangle_\lambda = \lambda J(f,g)$ for any $f, g \in \mathcal{H}^m$. Define

$$V(f,g) = E_U \left[\frac{f(U)g(U)\exp\{-\Lambda_0(U)\}}{1-\exp\{-\Lambda_0(U)\}} \right].$$

We have

$$\langle f, g\rangle_\lambda = V(f,g) + \langle W_\lambda f, g\rangle_\lambda.$$

Let $K(\cdot,\cdot)$ be the reproducing kernel of $\mathcal{H}^m$, and let h_j and γ_j be the eigenfunctions and eigenvalues of $\mathcal{H}^m$. The properties of $K(\cdot,\cdot)$, $h_j \in \mathcal{H}^m$ and γ_j can be found in Appendix A. Let $\mathcal{S}_n(\Lambda)$ and $\mathcal{S}_{n,\lambda}(\Lambda)$ be the Fréchet derivatives of $l_n(\Lambda)$ and $\ell_{n,\lambda}(\Lambda)$, respectively. Similarly, let $\mathcal{S}(\Lambda)$ and $\mathcal{S}_\lambda(\Lambda)$ be the Fréchet derivatives of $l(\Lambda)$ and $\ell_\lambda(\Lambda)$, respectively. Let D be the Fréchet derivative operator. Then, direct calculations yield

$$\begin{aligned} \langle D\mathcal{S}_\lambda(\Lambda_0) f, g\rangle_\lambda &= -E_U \left[\frac{f(U)g(U)\exp\{-\Lambda_0(U)\}}{1-\exp\{-\Lambda_0(U)\}} \right] - \langle W_\lambda f, g\rangle_\lambda \\ &= -\langle f, g\rangle_\lambda. \end{aligned}$$

The following proposition will play a key role in the FBR.

Proposition 1. *$D\mathcal{S}_\lambda(\Lambda_0) = -id$, where id is the identity operator.*

Following Proposition 1, the first term of the Taylor expansion of $\mathcal{S}_{n,\lambda}(\Lambda)$ at Λ_0 can be approximated by $-id(\Lambda - \Lambda_0)$. This will result in a sum of the independent and identically distributed random variables.

3. Functional Bahadur representation and asymptotic normality

The functional Bahadur representation (FBR) is a key technique to establish the asymptotic normality of the estimators. The following lemma shows that the estimator is consistent in the $\|\cdot\|_\infty$ with $\|g\|_\infty = \sup_{t\in\mathbb{I}} |g(t)|$ and $\|\cdot\|_1$, which denotes that $\|\cdot\|_\lambda$ with $\lambda = 1$.

Lemma 1 (Consistency)**.** *If $\lambda n^{1-2\mu} \to 0$ as $n \to \infty$ for any $0 < \mu < 1/2$, we have $\|\hat{\Lambda}^{(j)}_{n,\lambda} - \Lambda^{(j)}_0\|_\infty = o_p(1), j = 0, 1, \ldots, m-1$, and $J(\hat{\Lambda}_{n,\lambda} - \Lambda_0, \hat{\Lambda}_{n,\lambda} - \Lambda_0) = o_p(1)$.*

Remark 1. Define $\mathcal{H}^m_0 = \{g \in \mathcal{H}^m, g(t) > 0, g'(t) > 0, t \in \mathbb{I}\}$. Assume $\Lambda_0(t) \in \mathcal{H}^m_0$, $m > 1$. By Lemma 1, one can show that

$$\lim_{n\to\infty} P(\hat{\Lambda}_{n,\lambda} \in \mathcal{H}^m_0) = 1.$$

To see this, define $A_n = \{\omega : \sup_{t\in\mathbb{I}} |\hat{\Lambda}^{(j)}_{n,\lambda}(t) - \Lambda^{(j)}_0(t)| \le \epsilon, j = 0, 1\}$. Then by Lemma 1, for any $\epsilon > 0$, $\eta > 0$, there exists a N_0 such that when $n \ge N_0$, $P(A_n) > 1 - \eta$. Thus, on A_n, we have $\hat{\Lambda}_{n,\lambda}(t) \ge \Lambda_0(t) - \epsilon \ge \Lambda_0(\xi) - \epsilon$ and

$\hat{\Lambda}'_{n,\lambda}(t) \geq \Lambda'_0(t) - \epsilon \geq C_0 - \epsilon$ with $C_0 = \min_{t\in\mathbb{I}} \Lambda'_0(t) > 0$. Therefore, by taking $\epsilon = \min(\Lambda_0(\xi)/2, C_0/2)$, we have $\hat{\Lambda}^{(j)}_{n,\lambda}(t) > 0$ for any $t \in \mathbb{I}$ on A_n with $n > N_0$, $j = 0, 1$. This yields the conclusion. Numerically, to ensure monotonicity, we use the B-spline with constrained coefficients for estimating the cumulative hazard in the simulation studies.

Next, we present the exact rate of convergence and the FBR for $\hat{\Lambda}_{n,\lambda}$.

Theorem 1 (Rate of Convergence)**.** *If* $\log\{\log(n)\}/(nh^2) \to 0$, $\lambda n^{1-2\mu} \to 0$ *as* $n \to \infty$ *for any* $0 < \mu < 1/2$*, we have* $\|\hat{\Lambda}_{n,\lambda} - \Lambda_0\|_\lambda = O_p((nh)^{-1/2} + h^m)$.

We derive a new version of the FBR.

Theorem 2 (Functional Bahadur Representation)**.** *If* $\log\{\log(n)\}/(nh^2) \to 0$, $nh^2 \to \infty$ *and* $\lambda n^{1-2\mu} \to 0$ *as* $n \to \infty$ *for any* $0 < \mu < 1/2$*, we have* $\|\hat{\Lambda}_{n,\lambda} - \Lambda_0 - \mathcal{S}_{n,\lambda}(\Lambda_0)\|_\lambda = O_p(\alpha_n)$*, where*

$$\alpha_n = h^{-1/2}\{(nh)^{-1}+h^{2m}\}+h^{-(6m-1)/(4m)}n^{-1/2}[\log\{\log(n)\}]^{1/2}\{(nh)^{-1/2}+h^m\}.$$

From Theorem 2, we can find that the bias of the estimator is very close to a sum of some independently and identically distributed random variables, which is very useful to study the asymptotic normality.

Theorem 3. *(Asymptotic Normality) Assume that* $m > 3/4+\sqrt{5}/4$, $nh^{4m-1} \to 0$ *and* $nh^3 \to \infty$ *as* $n \to \infty$*. For* $\forall t_0 \in \mathbb{I}$*, define*

$$\sigma^2_{t_0} = \lim_{h\to 0} h \sum_{j=0}^{\infty} h_j^2(t_0)/(1+\lambda\gamma_j)^2.$$

Let $\Lambda^* = (id - W_\lambda)\Lambda_0$ *be the biased true parameter. Then we have*

$$\sqrt{nh}\{\hat{\Lambda}_{n,\lambda}(t_0) - \Lambda^*(t_0)\} \xrightarrow{d} N(0, \sigma^2_{t_0}).$$

The above theorem reports the point-wise asymptotic normality of the resulting estimate.

Corollary 1. *If* $m > 3/2$, $nh^{2m} \to 0$ *and* $nh^3 \to \infty$ *as* $n \to \infty$*, then*

$$\sqrt{nh}\{\hat{\Lambda}_{n,\lambda}(t_0) - \Lambda_0(t_0)\} \xrightarrow{d} N(0, \sigma^2_{t_0}).$$

Remark 2. The choice of tuning parameter that leads to the optimal pointwise rate of convergence is $h \asymp n^{-\frac{1}{2m+1}}$. However, this rate does not satisfy the conditions of Corollary 1; in other words, the estimator needs to be under-smoothed to ensure an unbiased limiting distribution, which yields a sub-optimal pointwise rate of convergence. This shares the same spirit as the under-smoothing procedures in the literature. Thus, we need to sacrifice the optimal rate for removing bias. In practice, we apply the widely-used CV or GCV to select the tuning parameter h, as suggested by [24].

Remark 3. Corollary 1 together with the Delta-method immediately gives the pointwise confidence interval for some real-valued smooth function of $\Lambda_0(t_0)$ at any fixed point $t_0 \in \mathbb{I}$, denoted as $\rho(\Lambda_0(t))$. Let $\dot{\rho}(\cdot)$ be the first derivative of $\rho(\cdot)$. If $\dot{\rho}(\Lambda_0(t_0)) \neq 0$, we have

$$P\left(\rho\{\Lambda_0(t_0)\} \in \left[\rho\{\hat{\Lambda}_{n,\lambda}(t_0)\} \pm z_{\alpha/2}\frac{\dot{\rho}\{\Lambda_0(t_0)\}\sigma_{t_0}}{\sqrt{nh}}\right]\right) \to 1-\alpha,$$

where z_α is the lower αth quantile of the standard normal distribution function.

4. Penalized likelihood ratio test

In this section, we present the penalized likelihood ratio tests for the local and global hypotheses.

4.1. Local likelihood ratio test

For a prespecified point (t_0, ω_0), we consider the following hypothesis:

$$H_0 : \Lambda(t_0) = \omega_0 \qquad \text{versus} \qquad H_1 : \Lambda(t_0) \neq \omega_0.$$

The constrained penalized log-likelihood is defined as

$$\begin{aligned} &L_{n,\lambda}(\Lambda) \\ =&\frac{1}{n}\sum_{i=1}^{n}[\Delta_i \log\{1-\exp(-\Lambda(U_i)-\omega_0)\} - (1-\Delta_i)\{\omega_0+\Lambda(U_i)\}] - \frac{\lambda}{2}J(\Lambda,\Lambda), \end{aligned}$$

where $\Lambda \in \mathcal{H}_0 = \{\Lambda \in \mathcal{H}^m : \Lambda(t_0) = 0\}$. To test H_0 against H_1, take the penalized local likelihood ratio test (PLLRT) statistic:

$$PLLRT_{n,\lambda} = L_{n,\lambda}(\omega_0 + \hat{\Lambda}^0_{n,\lambda}) - L_{n,\lambda}(\hat{\Lambda}_{n,\lambda}),$$

where $\hat{\Lambda}^0_{n,\lambda} = \arg\max_{\Lambda \in \mathcal{H}_0} L_{n,\lambda}(\Lambda)$.

Endowed with the norm $\|\cdot\|_\lambda$, $\mathcal{H}_0$ is a closed subset in $\mathcal{H}^m$, and thus a Hilbert space. The following proposition states that $\mathcal{H}_0$ also inherits the reproducing kernel and penalty operator from the space $\mathcal{H}^m$. The proof is trivial and thus omitted.

Proposition 2. *$(\mathcal{H}_0, \langle\cdot,\cdot\rangle_\lambda)$ is a reproducing kernel Hilbert space.*

(a) The bivariate function

$$K^*(t_1,t_2) = K(t_1,t_2) - K(t_0,t_1)K(t_0,t_2)/K(t_0,t_0)$$

is a reproducing kernel for $(\mathcal{H}_0, \langle\cdot,\cdot\rangle_\lambda)$. That is, for any $t' \in I$ and $g \in \mathcal{H}_0$, we have $K^*_{t'} \equiv K^*(t',\cdot) \in \mathcal{H}_0$ and $\langle K^*_{t'}, g\rangle_\lambda = g(t')$. Besides, $\|K^*\|_\lambda \leq \sqrt{2}c_m h^{-1/2}$.

(b) The operator W_λ^* defined by $W_\lambda^* g = W_\lambda g - (W_\lambda g)(t_0)K_{t_0}/K(t_0,t_0)$ is bounded linear from $\mathcal{H}_0$ to $\mathcal{H}_0$ and satisfies $\langle W_\lambda^* g, \tilde{g}\rangle_\lambda = \lambda J(g,\tilde{g})$.

From Proposition 2, we obtain the restricted FBR for $\hat{\Lambda}^0_{n,\lambda}$, which will be used to derive the null limiting distribution. Let the first-order Fréchet derivative of $L_{n,\lambda}$ and L_n be $\mathcal{S}^0_{n,\lambda}$ and $\mathcal{S}^0_n$ respectively. Define $\mathcal{S}^0(\Lambda) = E\{\mathcal{S}^0_n(\Lambda)\}$ and $\mathcal{S}^0_\lambda(\Lambda) = \mathcal{S}^0(\Lambda) - W_\lambda^*\Lambda$. Taking the derivative of $\mathcal{S}^0_\lambda(\Lambda)$, we have

$$D\mathcal{S}^0_\lambda(\Lambda)g_1g_2 = -E\left[\frac{\Delta_i \exp\{-\Lambda(U_i) - \omega_0\} g_1(U_i)g_2(U_i)}{[1-\exp\{-\Lambda(U_i)-\omega_0\}]^2}\right] - \langle W_\lambda^* g_2, g_1\rangle_\lambda.$$

Define $\Lambda^0_0(t) = \Lambda_0(t) - \omega_0$. Thus, we have

$$\begin{aligned}
\langle D\mathcal{S}^0_\lambda(\Lambda^0_0)f, g\rangle_\lambda &= \langle D\{\mathcal{S}^0(\Lambda^0_0)\}f, g\rangle_\lambda - \langle W_\lambda^* f, g\rangle_\lambda \\
&= -E_U\left[\frac{\exp\{-\Lambda_0(U_i)\}g_1(U_i)g_2(U_i)}{1-\exp\{-\Lambda_0(U_i)\}}\right] - \langle W_\lambda^* f, g\rangle_\lambda \\
&= -\langle f, g\rangle_\lambda.
\end{aligned}$$

Following from the equation, we have the next proposition.

Proposition 3. *$D\mathcal{S}^0_\lambda(\Lambda^0_0) = -id$, where id is the identity operator.*

Proposition 4 (Rate of Convergence)**.** *Under H_0, if $(\log\log(n))/(nh^2) \to 0$, $\lambda n^{1-2\mu} \to 0$ as $n \to \infty$ for any $0 < \mu < 1/2$, we have $\|\hat{\Lambda}^0_{n,\lambda} - \Lambda^0_0\|_\lambda = O_p((nh)^{-1/2} + h^m)$.*

The proof of Proposition 4 is similar to Theorem 1 and thus omitted.

Theorem 4 (Restricted FBR)**.** *Suppose that $(\log\log(n))/(nh^2) \to 0$, $\lambda n^{1-2\mu} \to 0$ as $n \to \infty$ for any $0 < \mu < 1/2$. Under H_0, we have $\|\hat{\Lambda}^0_{n,\lambda} - \Lambda^0_0 - \mathcal{S}^0_{n,\lambda}(\Lambda^0_0)\|_\lambda = O_p(\alpha_n)$, where α_n is given in Theorem 2.*

The proof of Theorem 4 is similar to that of Theorem 2 and thus omitted. Our main result follows immediately from the Restricted FBR.

Theorem 5. *(Penalized Local Likelihood Ratio Test). Assume $m > (5+\sqrt{21})/4$, $nh^{2m} \to 0$ and $nh^4 \to \infty$ as $n \to \infty$. Also assume that for $\forall t_0 \in I$, $\sigma_{t_0} \neq 0$, and*

$$c_{t_0} = \lim_{h\to 0} V(K_{t_0}, K_{t_0})/\|K_{t_0}\|^2_\lambda \in (0,1].$$

Under H_0, as $n \to \infty$, we have

(i) $\|\hat{\Lambda}_{n,\lambda} - \hat{\Lambda}^0_{n,\lambda} - \omega_0\|_\lambda = O_p(n^{-1/2})$;

(ii)$-2nPLLRT_{n,\lambda} = n\|\hat{\Lambda}_{n,\lambda} - \hat{\Lambda}^0_{n,\lambda} - \omega_0\|^2_\lambda + o_p(1)$;

(iii)$-2nPLLRT_{n,\lambda} \xrightarrow{d} c_{t_0}\chi^2_1$.

Note that the convergence rate stated in Theorem 5 is reasonable since the restriction is local.

4.2. Global likelihood ratio test

Consider the following global hypothesis:

$$H_0^{global} : \Lambda = \Lambda_0 \qquad \text{versus} \qquad H_1 : \Lambda \neq \Lambda_0.$$

The penalized global likelihood ratio test (PGLRT) statistic is defined as

$$PGLRT_{n,\lambda} = \ell_{n,\lambda}(\Lambda_0) - \ell_{n,\lambda}(\hat{\Lambda}_{n,\lambda}).$$

Theorem 6 (Penalized Global Likelihood Ratio Test)**.** *Suppose $m > (3+\sqrt{5})/4$, $nh^{2m+1} = O(1)$, $nh^3 \to \infty$ as $n \to \infty$. Define $\sigma_\lambda^2 = \sum_{j=0}^\infty h/(1+\lambda\gamma_j)$, $\rho_\lambda^2 = \sum_{j=0}^\infty h/(1+\lambda\gamma_j)^2$, $\gamma_\lambda = \sigma_\lambda^2/\rho_\lambda^2$, $\nu_\lambda = h^{-1}\sigma_\lambda^4/\rho_\lambda^2$. Under H_0^{global}, we have*

$$(2\nu_\lambda)^{-1/2}(-2n\gamma_\lambda PGLRT_{n,\lambda} - n\gamma_\lambda\|W_\lambda\Lambda_0(t)\|_\lambda^2 - \nu_\lambda) \xrightarrow{d} N(0,1).$$

A direct examination reveals that $h \asymp n^{-d}$, where $1/(2m+1) \leq d < 1/3$ satisfies the conditions required in Theorem 6. As one can show that $n\|W_\lambda\Lambda_0\|^2 = o(h^{-1}) = o(\nu_\lambda)$, then $-2n\gamma_\lambda PGLRT_{n,\lambda}$ is asymptotically distributed as $N(\nu_\lambda, 2\nu_\lambda)$. Since $N(\nu_\lambda, 2\nu_\lambda)$ is asymptotically distributed as $\chi^2_{\nu_\lambda}$, then

$$-2n\gamma_\lambda PGLRT_{n,\lambda} \sim \chi^2_{\nu_\lambda}.$$

This shows that the Wilks phenomenon holds for the PGLRT. Since people pay much more attention to the shape of a function, the PGLRT plays a key role in practice. For example, the proposed likelihood ratio approach can also be applied to a parametric setup, e.g.,

$$H_0 : \Lambda \in \mathcal{P}_d$$

where $\mathcal{P}_d = \left\{\Lambda(t) : \Lambda(t) = \sum_{j=0}^d t^j b_j\right\}$. It follows from similar arguments as in Remark 5.4 of [24] that, the asymptotic null distribution for testing such a hypothesis is $\chi^2_{\nu_\lambda}$, which is the same as that in Theorem 6. This result fills a gap compared to the local test results in [3].

Furthermore, we show that the PGLRT achieves the optimal minimax rate of testing presented in [19] based on the uniform version of the FBR. Write $H_1 : \Lambda = \Lambda_{n_0}$, where $\Lambda_{n_0} = \Lambda_0 + \Lambda_n$, where $\Lambda_0 \in \mathcal{H}_0^m$ and Λ_n belongs to the alternative set $\mathcal{A} = \{\Lambda_n \in \mathcal{H}_0^m, \exp(\Lambda_n(t)) \leq \zeta, J(\Lambda_n, \Lambda_n) \leq \zeta\}$ for some constant $\zeta > 0$.

Theorem 7. *Let $m > (3+\sqrt{5})/4$, $h \asymp n^{-d}$, where $1/(2m+1) \leq d < 1/3$. Suppose that uniformly over $\Lambda_n \in \mathcal{A}$, $\|\hat{\Lambda}_{n,\lambda} - \Lambda_{n_0}\|_\lambda = O_p((nh)^{-1/2} + h^m)$ holds under $H_{1n} : \Lambda = \Lambda_{n_0}$. Then for any $\epsilon \in (0,1)$, there exist positive constants C and N such that*

$$\inf_{n \geq N} \inf_{\Lambda_n \in \mathcal{A}, \|\Lambda_n\|_\lambda \geq C\eta_n} P(\text{reject} \quad H_0^{global} | H_{1n} \quad \text{is} \quad \text{true}) \geq 1 - \epsilon,$$

where $\eta_n \geq \sqrt{h^{2m} + (nh^{1/2})^{-1}}$. The minimal lower bound of η_n, $n^{-2m/(4m+1)}$, can be achieved when $h = h^* = n^{-2/(4m+1)}$.

Theorem 7 shows that when $h = h^* = n^{-2/(4m+1)}$, the PGLRT can detect any local alternatives with separation rates no faster than $n^{-2m/(4m+1)}$, which turns out to be the minimax rate of testing in the sense of [19].

5. Simulation studies

To assess the performance of the penalized likelihood estimator and likelihood ratio test, we conducted a simulation study with a focus on the comparison of the proposed penalized method and the classical likelihood method [3, 4, 5, 13]. For this purpose, we consider two examples. We use 5-fold cross validation to choose the tuning parameter. Besides, we consider $m = 2$ and 3, and sample sizes $n = 600$ and 800. In addition, we calculate c_{t_0} on the basis of the eigenfunctions and eigenvalues defined in equation (1) with the cumulative hazard function replaced by the estimate of Λ_0. For each setting, we report simulation results based on 1000 repetitions using the proposed local likelihood ratio test (PLLRT) and the likelihood ratio test (LRT) proposed by [3] and [13].

Example 5.1. The failure time follows the Exponential distribution with density function

$$f(t) = \frac{1}{\mu} \exp\left(-\frac{t}{\mu}\right), t \in [0, \infty),$$

where $\mu = 2.5$, while the examination time follows the uniform distribution from 0.5 to 1.5. Specifically, we test $H_0 : \Lambda(0.82) = \Lambda_0(0.82)$ against $H_1 : \Lambda(0.82) \neq \Lambda_0(0.82)$ where Λ_0 is the cumulative hazard function of the failure time defined above. To access the power of the test, we generated failure times with density function

$$f(t) = \frac{1}{\mu + c} \exp\left(-\frac{t}{\mu + c}\right), t \in [0, \infty),$$

where $\mu = 2.5, c = 0.0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0$, and the distribution of the examination time remains unchanged. Simulation results for this example are shown in Table 1. The powers of both methods are comparable. It can be seen that the estimated sizes of the proposed test are closer to the target level 5% than that of LRT.

Example 5.2. We use the setting of Example 5.1 except that the failure time distribution is changed to the Pareto distribution with density function

$$f(t) = \frac{1}{\sigma}\left(1 + \frac{kt}{\sigma}\right)^{-1-\frac{1}{k}}, t \in [0, \infty),$$

where $\sigma = k = 2.5$. We test $H_0 : \Lambda(0.82) = \Lambda_0(0.82)$ against $H_1 : \Lambda(0.82) \neq \Lambda_0(0.82)$ where Λ_0 is the cumulative hazard function of the Pareto distribution defined above. To assess the power of the test, we generated failure times with density function

$$f(t) = \frac{1}{\sigma + c}\left(1 + \frac{kt}{\sigma + c}\right)^{-1-\frac{1}{k}}, t \in [0, \infty),$$

Table 1
The estimated size and power of the proposed test (PLLRT) in comparison to the NPMLE method (LRT) in Example 5.1

	PLLRT $(m=2)$		**PLLRT** $(m=3)$		**LRT**	
c	$n=600$	$n=800$	$n=600$	$n=800$	$n=600$	$n=800$
0.0	0.064	0.059	0.067	0.050	0.072	0.033
0.5	0.292	0.355	0.214	0.277	0.256	0.314
1.0	0.629	0.761	0.560	0.692	0.675	0.778
1.5	0.862	0.959	0.833	0.932	0.923	0.963
2.0	0.951	0.988	0.945	0.985	0.984	0.996
2.5	0.985	0.994	0.980	0.989	0.999	1.000
3.0	0.987	0.999	0.984	0.996	1.000	1.000

Table 2
The estimated size and power of the proposed test (PLLRT) in comparison to the NPMLE method (LRT) in Example 5.2

	PLLRT $(m=2)$		**PLLRT** $(m=3)$		**LRT**	
c	$n=600$	$n=800$	$n=600$	$n=800$	$n=600$	$n=800$
0.0	0.061	0.044	0.047	0.040	0.074	0.050
0.5	0.161	0.180	0.126	0.122	0.198	0.226
1.0	0.370	0.474	0.279	0.360	0.463	0.558
1.5	0.596	0.739	0.528	0.652	0.741	0.832
2.0	0.787	0.896	0.726	0.841	0.911	0.960
2.5	0.902	0.955	0.857	0.939	0.971	0.991
3.0	0.947	0.973	0.913	0.972	0.991	0.999

Table 3
The mean squared error (MSE) of the proposed method in comparison to the NPMLE method.

	Proposed$(m=2)$		**Proposed**$(m=3)$		**NPMLE**	
	$n=600$	$n=800$	$n=600$	$n=800$	$n=600$	$n=800$
Example 5.1	0.0107	0.0038	0.0094	0.0044	0.0056	0.0043
Example 5.2	0.0442	0.0035	0.0151	0.0033	0.0036	0.0028

where $c = 0.0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0$, and the distribution of the examination time remains unchanged. Simulation results for this example are shown in Table 2. It can be seen that the estimated sizes and powers of both methods are comparable.

For each setting, the pointwise averages, the coverage probability of the pointwise 95% confidence intervals, and the estimated standard errors of $\hat{\Lambda}_{n,\lambda}(t)$ are calculated. The simulation results of Example 5.1 are displayed in Figures 1–3, and those of Example 5.2 are showed in Figures 4–6. Table 3 summarizes the mean squared error (MSE) of the estimate in each setting. The simulation results based the NPMLE proposed by [3] and [13] are also included for comparison. For each case, the estimates given by our proposed method are closer to $\Lambda_0(t)$, particularly when t is near the boundaries, than that given by [3] and [13]. When the sample size increases, the reduction of MSE of the proposed method is faster than that of NPMLE. This result confirms that the convergence rate of the proposed method is higher than that of the NPMLE. The coverage probabilities of both methods are reasonable. Besides, the simulation results indicate

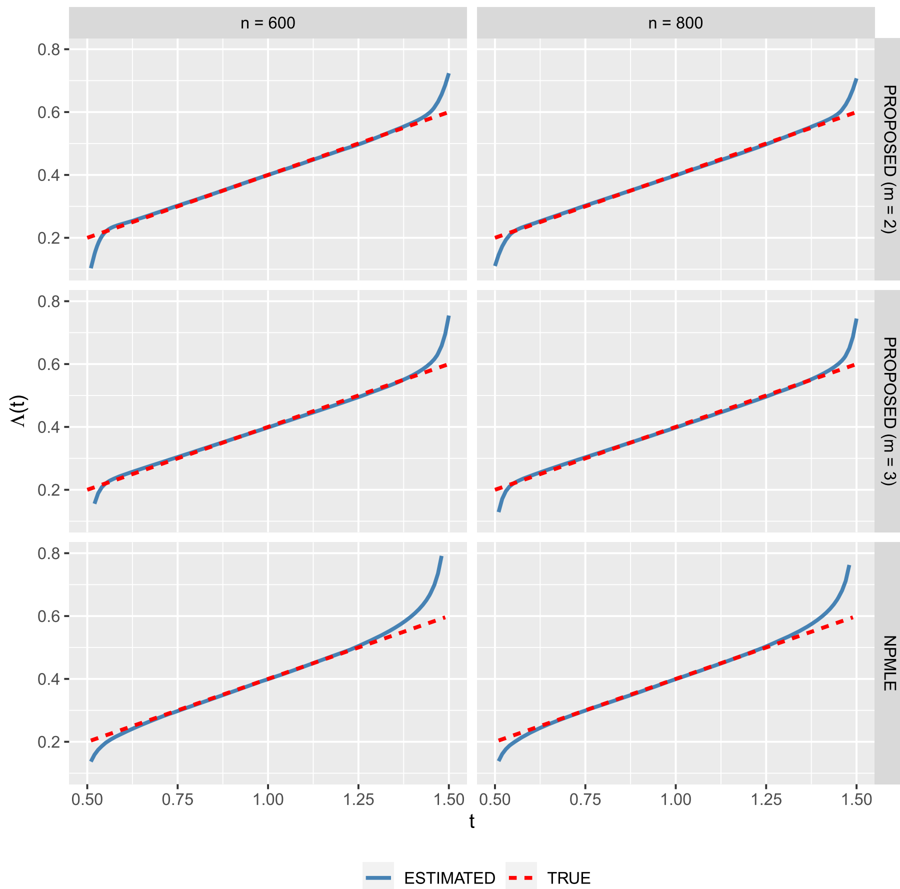

Fig 1. *Estimates of the cumulative hazard function in Example 5.1. Note: The solid blue lines represent* $\hat{\Lambda}_{n,\lambda}(t)$ *; the dashed red lines represent* $\Lambda_0(t)$*. The bottom panel, labelled NPMLE, is the results based on the estimator proposed by Banerjee and Wellner's.*

that the proposed method with $m = 2$ and $m = 3$ has similar performance.

6. Closing remarks

This article focuses on the development of nonparametric inference for the cumulative hazard function with case I interval-censored data or current status data. It is well known that the convergence rate of the NPMLE of an unknown distribution function is $n^{-1/3}$ rather than the standard rate $n^{-1/2}$ due to interval-censoring. To improve the convergence rate, we develop a penalized likelihood method using smoothing techniques. To establish the asymptotic properties of the proposed estimator, we derive a functional Bahadur representation of the estimator in the Sobolev space with a proper inner product, which plays a key

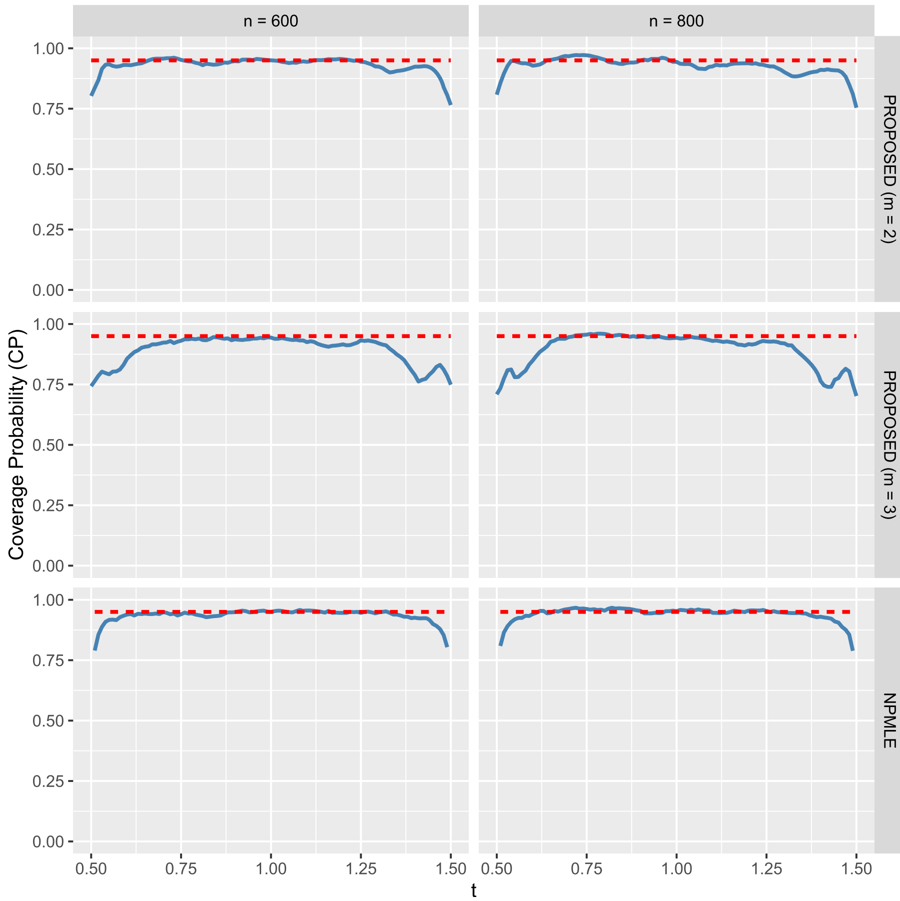

FIG 2. *The coverage probability of the pointwise* 95% *confidence intervals in Example 5.1. Note: The solid blue lines represent the coverage probability; the dashed red horizontal lines represent the target coverage probability,* 95%. *The bottom panel, labelled NPMLE, is the results based on the estimator proposed by Banerjee and Wellner's.*

role for nonparametric inference. Furthermore, we develop the penalized likelihood ratio tests for both local and global hypotheses. In particular, the proposed penalized global likelihood ratio test is able to detect any local alternatives with minimax separation rate in the sense of [19], while the classical likelihood ratio test for the global hypothesis in the current status data model has not been addressed, to the best of our knowledge. Simulation studies demonstrate that the proposed estimator outperforms the classical NPMLE and the penalized likelihood ratio test is more powerful than the classical likelihood ratio test, as expected.

Note that for case II interval-censored data, the convergence rate of the nonparametric maximum likelihood estimator of an unknown distribution function

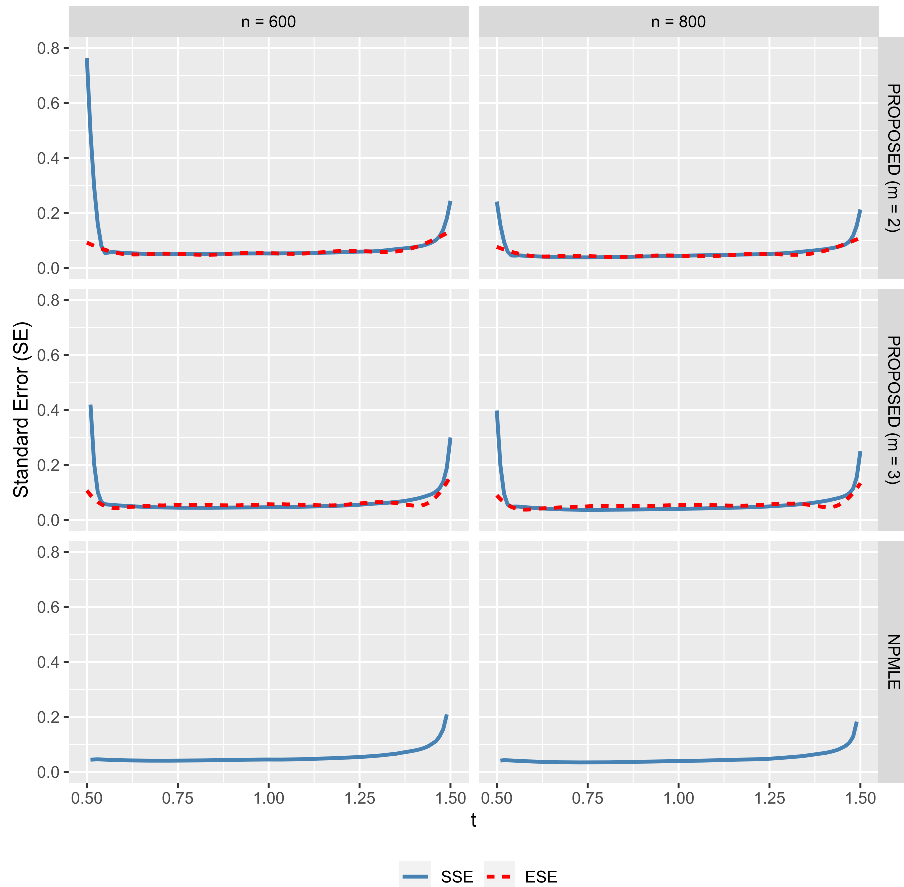

FIG 3. *The simulated standard error (SSE) and the estimated standard error (ESE) in Example 5.1. Note: The solid blue lines represent the SSE; the dashed red lines represent ESE. The bottom panel, labelled NPMLE, is the results based on the estimator proposed by Banerjee and Wellner's.*

is $n^{-1/3}$ and the limiting distribution of the estimator is still unknown. Further interesting research is to investigate the limiting distribution of penalized nonparametric maximum likelihood estimator through deriving a functional Bahadur representation of the estimator, which is very challenging in the presence of case II interval-censoring.

Appendix

The Appendix contains the proofs of the main results in the main text and the properties of the reproducing kernel and the eigensystems. In the following, we will denote different positive constants by C which may take different values in

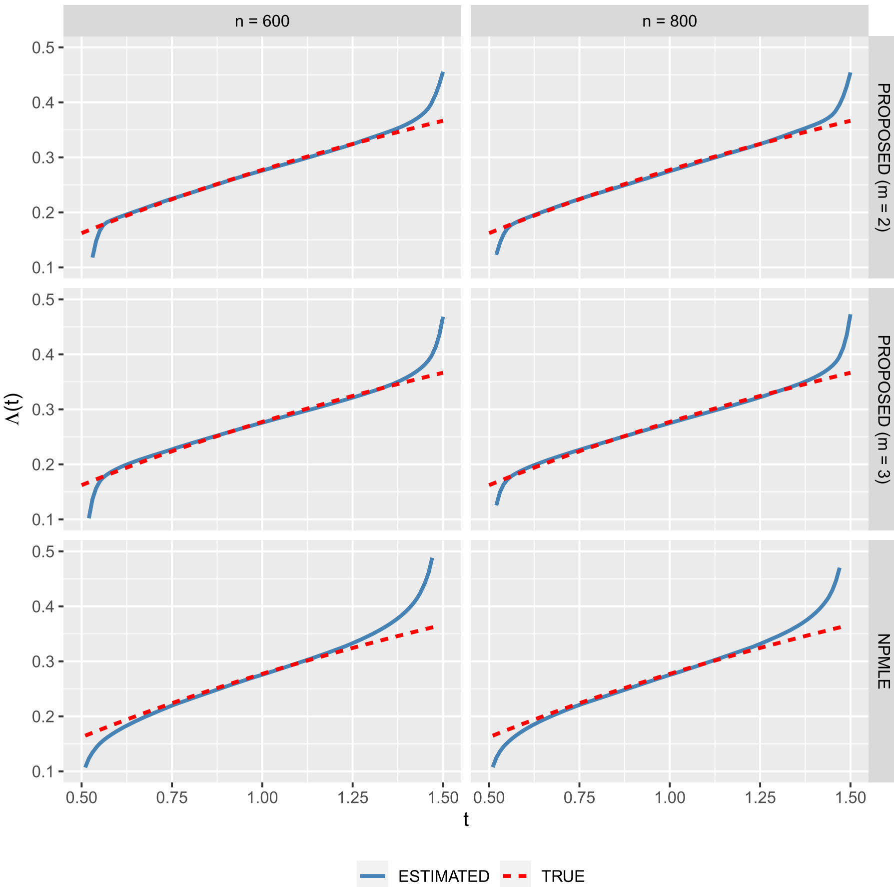

FIG 4. *Estimates of the cumulative hazard function in Example 5.2. Note: The solid blue lines represent* $\hat{\Lambda}_{n,\lambda}(t)$ *; the dashed red lines represent* $\Lambda_0(t)$*. The bottom panel, labelled NPMLE, is the results based on the estimator proposed by Banerjee and Wellner's.*

different places, while "$a \lesssim b$" means "$a \leq Cb$" and "$a \gtrsim b$" means "$a \geq Cb$".

Appendix A

In this section, we state the properties of the reproducing kernel $K(\cdot,\cdot)$, the eigensystems and how to compute the eigensystems.

First, the reproducing kernel $K(\cdot,\cdot)$ of $\mathcal{H}^m$ defined on $\mathbb{I} \times \mathbb{I}$ satisfies the following properties:

(P_1) $K_t(\cdot) = K(t,\cdot)$ and $\langle K_t, g\rangle_\lambda = g(t)$ for any g in $\mathcal{H}^m$ and any $t \in \mathbb{I}$.

(P_2) There exists a constant c_m which only depends on m such that $\|K_t\|_\lambda \leq c_m h^{-1/2}$ for $\forall t \in \mathbb{I}$, where $h = \lambda^{1/2m}$. Thereby, for any $g \in \mathcal{H}^m$, we have $\|g\|_\infty \leq c_m h^{-1/2}\|g\|_\lambda$.

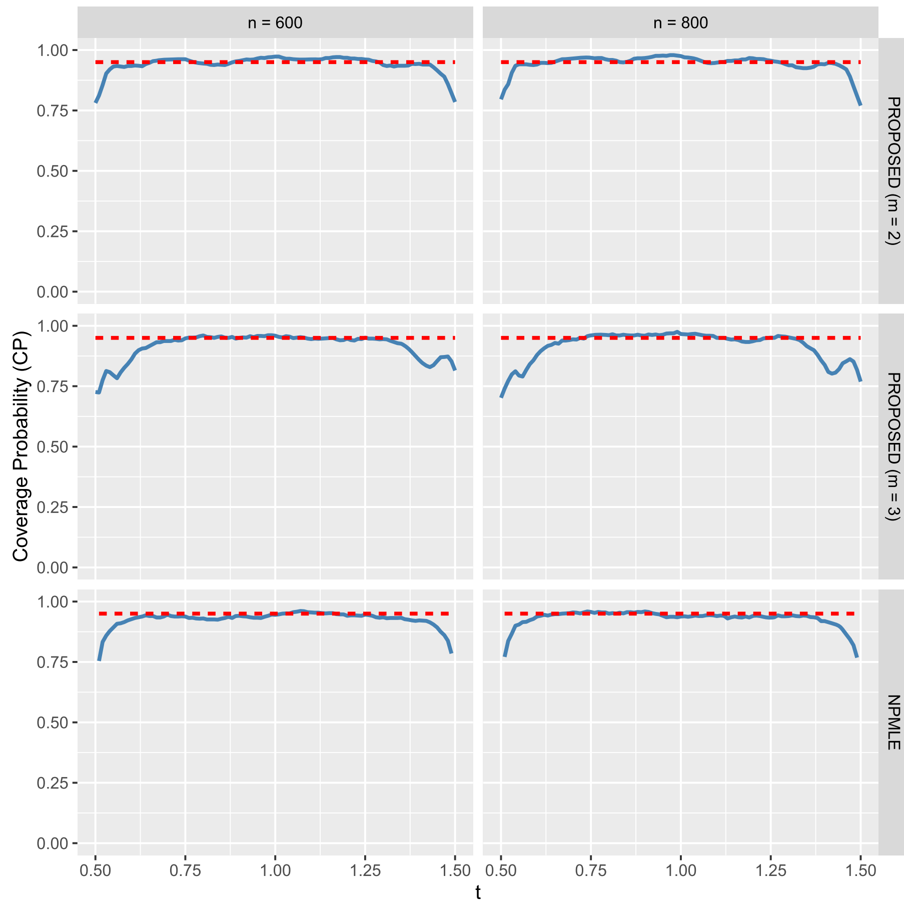

Fig 5. *The coverage probability of the pointwise* 95% *confidence intervals in Example* 5.2. *Note: The solid blue lines represent the coverage probability; the dashed red horizontal lines represent the target coverage probability,* 95%. *The bottom panel, labelled NPMLE, is the results based on the estimator proposed by Banerjee and Wellner's.*

The eigenfunctions $h_j \in \mathcal{H}^m$ and the eigenvalues γ_j satisfy the following properties:

(P_3) $\sup_{j\in N} \|h_j\|_\infty < \infty$, $\gamma_j \asymp j^{2m}$.
(P_4) $V(h_i, h_j) = \delta_{ij}$, $J(h_i, h_j) = r_j\delta_{ij}$, where δ_{ij} is a Kronecker's delta, which means that when $i = j$, $\delta_{ij} = 1$; otherwise, it's 0.
(P_5) For any $g \in \mathcal{H}^m$, we have $g = \sum_j V(g, h_j)h_j$ with a convergence in the $\|\cdot\|_\lambda$-norm.
(P_6) For any $g \in \mathcal{H}^m$ and $t \in \mathbb{I}$, we have $\|g\|_\lambda^2 = \sum_j V(g, h_j)^2(1 + \lambda\gamma_j)$, $K_t(\cdot) = \sum_j h_j(t)h_j(\cdot)/(1 + \lambda\gamma_j)$ and $W_\lambda h_j(\cdot) = (\lambda\gamma_j)/(1 + \lambda\gamma_j)h_j(\cdot)$.

It follows from [24] that the underling eigensystem $(\gamma_j, h_j(\cdot))$ can be chosen

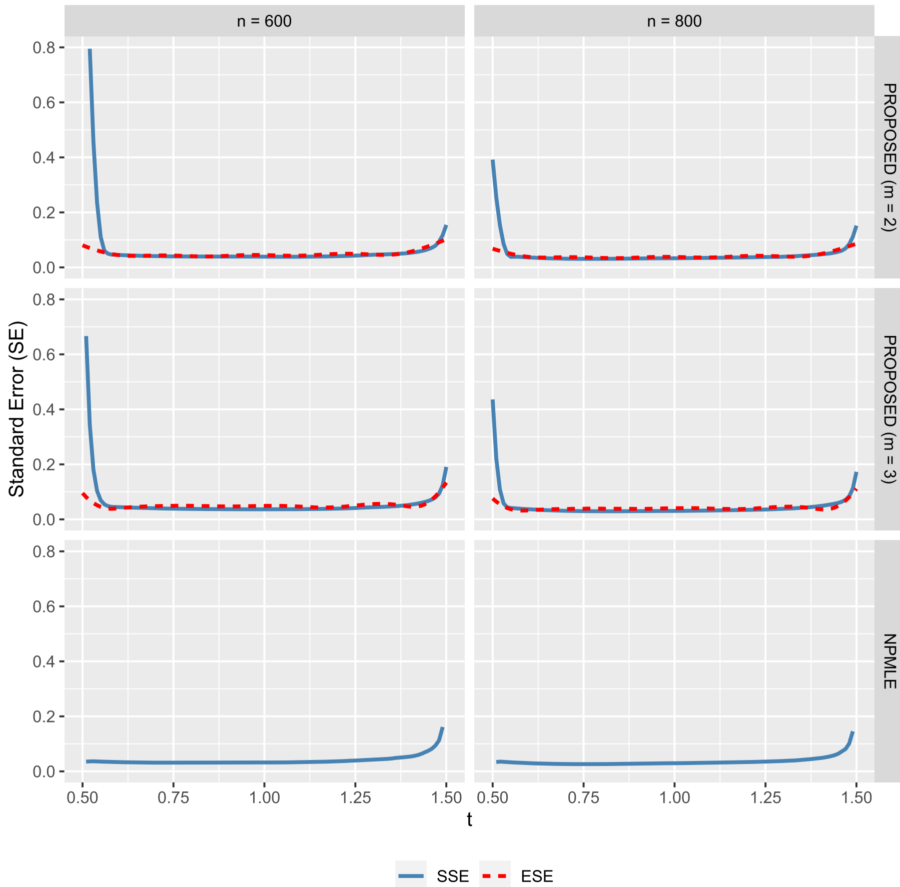

Fig 6. *The simulated standard error (SSE) and the estimated standard error (ESE) in Example 5.2. Note: The solid blue lines represent the SSE; the dashed red lines represent ESE. The bottom panel, labelled NPMLE, is the results based on the estimator proposed by Banerjee and Wellner's.*

as the (normalized) solution of the following ODE functions:

$$(-1)^m h_j^{(2m)}(t) = \gamma_j \frac{\exp\{-\Lambda_0(t)\}}{1-\exp\{-\Lambda_0(t)\}}\pi(t)h_j(t), t \in \mathbb{I},$$
$$h_j^{(k)}(\xi) = h_j^{(k)}(1+\xi) = 0, k = m, m+1, \ldots, 2m-1, \tag{2}$$

where $\pi(\cdot)$ is the density of U.

Appendix B

In order to prove Lemma 1, we need the following lemma.

Lemma B.1. *For any $g \in \mathcal{H}^m$, we have $J(g,g) \leq C_0 V(g,g)$ with C_0 being independent of g.*

Proof. Using the properties of the eigensystems (P4)-(P5), we have for any $g \in \mathcal{H}^{(m)}$, $g = \sum_j V(g,h_j)h_j$. It follows from [8] that there exists a constant C_0 such that $\{\sum_j V(g,h_j)^2\gamma_j^2\} \leq C_0 V(g,g)$ with C_0 being independent of g. Thus,

$$\begin{aligned} J(g,g) &= J\left(\sum_j V(g,h_j)h_j, g\right) = \sum_j V(g,h_j)J(h_j,g) = \sum_j V(g,h_j)^2\gamma_j \\ &\leq \{\sum_j V(g,h_j)^2\}^{1/2}\{\sum_j V(g,h_j)^2\gamma_j^2\}^{1/2} \leq V(g,g)^{1/2}C_0 V(g,g)^{1/2} \\ &\leq C_0 V(g,g). \end{aligned}$$

□

Proof of Lemma 1

Let $q_n(t) \in \mathcal{H}^m$ satisfying $\|q_n\|_\infty = O(n^{-1/2+\mu})$. Define

$$\begin{aligned} &H_n(\alpha) \\ &= \frac{1}{n}\sum_{i=1}^n \left[\Delta_i \log\{1-\exp(-\Lambda_0(U_i)-\alpha q_n(U_i))\} - (1-\Delta_i)\{\Lambda_0(U_i)+\alpha q_n(U_i)\}\right] \\ &\quad - \frac{\lambda}{2}\int_{\mathbb{I}}\{\Lambda_0^{(m)}(t)+\alpha q_n^{(m)}(t)\}^2\,dt \end{aligned}$$

Then, the derivative of $H_n(\alpha)$ is

$$\begin{aligned} H_n'(\alpha) =& \frac{1}{n}\sum_{i=1}^n \left[\Delta_i \frac{q_n(U_i)}{1-\exp\{-\Lambda_0(U_i)-\alpha q_n(U_i)\}} - q_n(U_i)\right] \\ &- \lambda\int_{\mathbb{I}}\Lambda_0^{(m)}(t)q_n^{(m)}(t)\,dt - \alpha\lambda\int_0^1 (q_n^{(m)})^2\,dt \\ =& \frac{1}{n}\sum_{i=1}^n \Bigg[\Delta_i q_n(U_i)\left\{\frac{1}{1-\exp(-\Lambda_0(U_i)-\alpha q_n(U_i))} - \frac{1}{1-\exp(-\Lambda_0(U_i))}\right\} \\ &\qquad + q_n(U_i)\left\{\Delta_i\frac{1}{1-\exp\{-\Lambda_0(U_i)\}} - 1\right\}\Bigg] \\ &- \lambda\int_{\mathbb{I}}\Lambda_0^{(m)}(t)q_n^{(m)}(t)\,dt - \alpha\lambda\int_{\mathbb{I}}(q_n^{(m)}(t))^2\,dt \\ =& -\alpha\left[\frac{1}{n}\sum_{i=1}^n \Delta_i\frac{q_n^2(U_i)\exp\{-\Lambda_0(U_i)\}}{[1-\exp\{-\Lambda_0(U_i)\}]^2}(1+o_p(1)) + \lambda\int_{\mathbb{I}} q_n^{(m)}(t)^2\,dt\right] \\ &+ \frac{1}{n}\sum_{i=1}^n q_n(U_i)\left[\Delta_i\frac{1}{1-\exp\{-\Lambda_0(U_i)\}} - 1\right] - \lambda\int_{\mathbb{I}}\Lambda_0^{(m)}(t)q_n^{(m)}(t)\,dt \\ \equiv& I_1 + I_2 + I_3. \end{aligned}$$

It follows directly from $\|q_n\|_\infty = O_p(n^{-1/2+\mu})$ that $I_1 = O_P(n^{-1+2\mu})$. Direct calculations yield that

$$E\left(\frac{I_2}{n^{-1/2+\mu}}\right)^2 = \frac{1}{n}E\left[\left\{\frac{q_n(U_i)}{n^{-1/2+\mu}}\right\}^2\left\{\Delta_i\frac{1}{1-\exp\{-\Lambda_0(U_i)\}}-1\right\}^2\right] = O_p(n^{-1}),$$

and $EI_2 = 0$.

Thus, by the central limit theorem, we can get $I_2 = O_p(n^{-1+\mu})$. As $q_n \in \mathcal{H}^m$, we have $\|q_n^{(m)}\|_\infty = O(1)$. As $\lambda n^{1-2\mu} \to 0$, $I_3 = o_p(n^{-1+2\mu})$. Hence, $H_n'(\alpha)\alpha < 0$. Since

$$H_n''(\alpha) = -\left[\frac{1}{n}\sum_{i=1}^n \Delta_i \frac{q_n^2(U_i)\exp\{-\Lambda_0(U_i)-\alpha q_n(U_i)\}}{[1-\exp\{-\Lambda_0(U_i)-\alpha q_n(U_i)\}]^2} + \lambda\int_{\mathbb{I}} q_n^{(m)}(t)^2\,dt\right],$$

then $H_n'(\alpha)$ is a nonincreasing function. So $\hat\Lambda_{n,\lambda}(t) \in [\Lambda_0(t) - |\alpha q_n(t)|, \Lambda_0(t) + |\alpha q_n(t)|]$. Then $\|\hat\Lambda_{n,\lambda} - \Lambda_0\|_\infty \le |\alpha|\|q_n\|_\infty \to 0$. Thereby, $V(\|\hat\Lambda_{n,\lambda}-\Lambda_0\|_\infty, \|\hat\Lambda_{n,\lambda}-\Lambda_0\|_\infty) = o_p(1)$. It follows from Lemma B.1 that $J(\hat\Lambda_{n,\lambda}-\Lambda_0, \hat\Lambda_{n,\lambda}-\Lambda_0) = o_p(1)$. Then, $\|\hat\Lambda_{n,\lambda} - \Lambda_0\|_{\lambda=1} = o_p(1)$.

Following [23], there exist two B-spline functions $\hat g_{m,\lambda}, \hat g_{m,0}$ such that $\|\hat\Lambda_{n,\lambda}^{(j)} - \hat g_{m,\lambda}^{(j)}\|_\infty = o(1)$, $\|\hat g_{m,0}^{(j)} - \Lambda_0^{(j)}\|_\infty = o(1)$, for $j = 0, 1, \dots, m-1$. Note that

$$\begin{aligned} &\|\hat g_{m,\lambda} - \hat g_{m,0}\|_\infty \\ \le\ &\|\hat\Lambda_{n,\lambda} - \hat g_{m,\lambda}\|_\infty + \|\hat g_{m,0} - \Lambda_0\|_\infty + \|\hat\Lambda_{n,\lambda} - \Lambda_0\|_\infty = o_p(1). \end{aligned}$$

Then it follows from Corollary 6.21 in [23] that $\|\hat g_{m,\lambda}^{(j)} - \hat g_{m,0}^{(j)}\|_\infty = o_p(1)$, $j = 1, 2, \cdots, m-1$. Thereby, we have

$$\|\hat\Lambda_{n,\lambda}^{(j)} - \Lambda_0^{(j)}\|_\infty \le \|\hat\Lambda_{n,\lambda}^{(j)} - \hat g_{m,\lambda}^{(j)}\|_\infty + \|\hat g_{m,0}^{(j)} - \hat g_{m,\lambda}^{(j)}\|_\infty + \|\Lambda_0^{(j)} - \hat g_{m,0}^{(j)}\|_\infty = o_p(1).$$

As a result, $\|\hat\Lambda_{n,\lambda}^{(j)} - \Lambda_0^{(j)}\|_\infty = o_p(1), j = 1, 2, \dots, m-1$.

Proof of Theorem 1

Set $g = \hat\Lambda_{n,\lambda} - \Lambda_0$ and write

$$\begin{aligned} l_{n,\lambda}(g+\Lambda_0) - l_{n,\lambda}(\Lambda_0) &= \mathcal{S}_{n,\lambda}(\Lambda_0)g + \frac{1}{2}D\mathcal{S}_{n,\lambda}(\Lambda_0)gg + \frac{1}{6}D^2\mathcal{S}_{n,\lambda}(g^*)ggg \\ &\equiv I_1 + I_2 + I_3, \end{aligned}$$

where $g^* = \Lambda_0 + \alpha g$ with $\alpha \in [0,1]$. Note that

$$\begin{aligned} |6I_3| &= |D^2\mathcal{S}_{n,\lambda}(g^*)ggg| \\ &= \Big|\frac{1}{n}\sum_{i=1}^n \Delta_i \frac{[1+\exp\{-g^*(U_i)\}]\exp\{-g^*(U_i)\}g^3(U_i)}{[1-\exp\{-g^*(U_i)\}]^3}\Big| \end{aligned}$$

$$
\begin{aligned}
&\le \|g\|_\infty \Big|\frac{1}{n}\sum_{i=1}^n \Delta_i \frac{[1+\exp\{-g^*(U_i)\}]\exp\{-g^*(U_i)\}g^2(U_i)}{[1-\exp\{-g^*(U_i)\}]^3}\Big| \\
&\lesssim \|g\|_\infty \Big|\frac{1}{n}\sum_{i=1}^n \Delta_i \frac{\exp\{-\Lambda_0(U_i)\}g^2(U_i)}{[1-\exp\{-\Lambda_0(U_i)\}]^2}\Big| \\
&\le \frac{\|g\|_\infty}{n}\Big|\sum_{i=1}^n \Big[\Delta_i \frac{\exp\{-\Lambda_0(U_i)\}g^2(U_i)}{[1-\exp\{-\Lambda_0(U_i)\}]^2} - E\Big(\Delta_i \frac{\exp\{-\Lambda_0(U_i)\}g^2(U_i)}{[1-\exp\{-\Lambda_0(U_i)\}]^2}\Big)\Big]\Big| \\
&\quad +\|g\|_\infty E_U\left[\frac{\exp\{-\Lambda_0(U_i)\}g^2(U_i)}{\{1-\exp(-\Lambda_0(U_i))\}^2}\right] \\
&= \frac{\|g\|_\infty}{n}|\langle \sum_{i=1}^n [\psi(\Delta_i,U_i,g)K_{U_i} - E\{\psi(\Delta,U,g)K_U\}], g\rangle_\lambda| \\
&\quad +\|g\|_\infty E_U\Big(\frac{\exp\{-\Lambda_0(U_i)\}g^2(U_i)}{[1-\exp\{-\Lambda_0(U_i)\}]^2}\Big),
\end{aligned}
$$

where

$$\psi(\Delta_i,U_i,g) = \Delta_i \frac{\exp\{-\Lambda_0(U_i)\}g(U_i)}{[1-\exp\{-\Lambda_0(U_i)\}]^2}.$$

Define

$$\tilde{\psi}(\Delta_i,U_i,g) = \frac{[1-\exp\{-\Lambda_0(\xi)\}]^2}{\exp\{-\Lambda_0(\xi)\}} c_m^{-1}h^{1/2}\psi(\Delta_i,U_i,g), \quad i=1,2,\dots,n.$$

Then,

$$
\begin{aligned}
&|\tilde{\psi}(\Delta_i,U_i,g) - \tilde{\psi}(\Delta_i,U_i,f)| \\
&= \frac{(1-\exp(-\Lambda_0(\xi)))^2}{\exp(-\Lambda_0(\xi))} c_m^{-1}h^{1/2}\frac{\Delta_i\exp\{-\Lambda_0(U_i)\}}{[1-\exp\{-\Lambda_0(U_i)\}]^2}|f(U_i)-g(U_i)| \\
&\le c_m^{-1}h^{1/2}\|f-g\|_\infty.
\end{aligned}
$$

Lemma 1 shows that $g \in \mathcal{F} = \{g \in \mathcal{H}^m(\mathbb{I}), \|g\|_\infty \le 1, J(g,g) \le c_m^{-2}h\lambda^{-1}\}$ when n is large enough with c_m defined in (P_2). It then follows from [24] that there exits a set B_n with $\lim_{n\to} P(B_n) = 1$ such that on B_n,

$$
\begin{aligned}
&\|\sum_{i=1}^n \Big[\tilde{\psi}(\Delta_i,U_i,g)K_{U_i} - E\{\tilde{\psi}(\Delta_i,U_i,g)K_{U_i}\}\Big]\|_\lambda \\
&\le (n^{1/2}\|g\|_\infty^{1-1/(2m)}+1)\{5\log\log(n)\}^{1/2}.
\end{aligned}
$$

Thereby, on B_n, we have

$$
\begin{aligned}
&\frac{\|g\|_\infty}{n}|\langle \sum_{i=1}^n [\psi(\Delta_i,U_i,g)K_{U_i} - E\{\psi(\Delta,U,g)K_{U_i}\}], g\rangle_\lambda| \\
\lesssim\ & c_m h^{-1/2}\frac{\|g\|_\infty\|g\|_\lambda}{n}(n^{1/2}\|g\|_\infty^{1-1/(2m)}+1)\{5\log\log(n)\}^{1/2}.
\end{aligned}
$$

Since

$$\|g\|_\infty E_U\left[\frac{\exp\{-\Lambda_0(U_i)\}g^2(U_i)}{1-\exp\{-\Lambda_0(U_i)\}}\right] \quad \leq \quad \|g\|_\infty\|g\|_\lambda^2,$$

then on B_n,

$$\begin{aligned}|6I_3| &\lesssim c_m h^{-1/2}\frac{\|g\|_\infty\|g\|_\lambda}{n}(n^{1/2}\|g\|_\infty^{1-1/(2m)}+1)\{5\log\log(n)\}^{1/2}+\|g\|_\infty\|g\|_\lambda^2 \\ &\lesssim \frac{c_m^2}{n^{1/2}h}\{\log\log(n)\}^{1/2}\|g\|_\lambda^2+\|g\|_\infty\|g\|_\lambda^2. \qquad (3)\end{aligned}$$

Hence, from $(n^{1/2}h)^{-1}\{\log\log(n)\}^{1/2}=o(1)$ and $\|g\|_\infty=o_p(1)$, we have $|6I_3|=o_p(1)\|g\|_\lambda^2$.

It follows from the definition of $\mathcal{S}_{n,\lambda}(\Lambda_0)$ that $\|\mathcal{S}_{n,\lambda}(\Lambda_0)\|_\lambda=O_p((nh)^{-1/2}+\lambda^{1/2})$. Thereby,

$$|I_1|=|\mathcal{S}_{n,\lambda}(\Lambda_0)g|\leq\|\mathcal{S}_{n,\lambda}(\Lambda_0)\|_\lambda\|g\|_\lambda=O_p((nh)^{-1/2}+\lambda^{1/2})\|g\|_\lambda.$$

For I_2, we write

$$\begin{aligned}2I_2 &= D\mathcal{S}_{n,\lambda}(\Lambda_0)gg \\ &= \{D\mathcal{S}_{n,\lambda}(\Lambda_0)gg-ED\mathcal{S}_{n,\lambda}(\Lambda_0)gg\}+ED\mathcal{S}_{n,\lambda}(\Lambda_0)gg \\ &= -\|g\|_\lambda^2+\{D\mathcal{S}_{n,\lambda}(\Lambda_0)gg-ED\mathcal{S}_{n,\lambda}(\Lambda_0)gg\} \\ &= -\|g\|_\lambda^2+\frac{1}{n}\sum_{i=1}^n\left(\Delta_i\frac{g^2(U_i)\exp\{-\Lambda_0(U_i)\}}{[1-\exp\{-\Lambda_0(U_i)\}]^2}-E_U\frac{g^2(U_i)\exp\{-\Lambda_0(U_i)\}}{1-\exp\{-\Lambda_0(U_i)\}}\right).\end{aligned}$$

In view of (3), we have

$$\lim_{n\to\infty}P\left(\Big|\frac{1}{n}\sum_{i=1}^n\left[\Delta_i\frac{g^2(U_i)\exp(-\Lambda_0(U_i))}{\{1-\exp(-\Lambda_0(U_i))\}^2}-E_U\left\{\frac{g^2(U_i)\exp(-\Lambda_0(U_i))}{1-\exp(-\Lambda_0(U_i))}\right\}\right]\Big| \lesssim c_m h^{-1/2}\|g\|_\lambda\left\{\frac{1}{\sqrt{n}}\|g\|_\infty^{1-1/(2m)}+\frac{1}{n}\right\}\{5\log\log(n)\}^{1/2}\right)=1.$$

It follows from the definition of $\hat{\Lambda}_{n,\lambda}$ that $I_1+I_2+I_3\geq 0$. Thus,

$$\lim_{n\to\infty}P\Big(\|g\|_\lambda^2(1+o_p(1))\lesssim\{\|g\|_\lambda^{2-1/(2m)}n^{-1/2}c_m^{2-1/(2m)}h^{-1+1/(4m)}\{5\log\log(n)\}^{1/2} + n^{-1}h^{-1/2}\{5\log\log(n)\}^{1/2}\|g\|_\lambda+((nh)^{-1/2}+\lambda^{1/2})\|g\|_\lambda\}\Big)=1.$$

Therefore,

$$\lim_{n\to\infty}P\Big(\|g\|_\lambda\lesssim\{\|g\|_\lambda^{1-1/(2m)}n^{-1/2}c_m^{2-1/(2m)}h^{-1+1/(4m)}\{5\log\log(n)\}^{1/2} + n^{-1}h^{-1/2}\{5\log\log(n)\}^{1/2}+((nh)^{-1/2}+\lambda^{1/2})\}\Big)=1.$$

Since $(nh^{1/2})^{-1}\{5\log\log(n)\}^{1/2} = o((nh)^{-1/2})$,

$$\lim_{n\to\infty} P(\|g\|_{n,\lambda} \lesssim \{(nh)^{-1/2} + \lambda^{1/2} + \|g\|_\infty^{1-1/(2m)} n^{-1/2}\{5\log\log(n)\}^{1/2} c_m h^{-1/2}\})$$
$$= 1.$$

Moreover, as $\|g\|_\infty = o_p(1)$ and $(nh)^{-1/2}\{5\log\log(n)\}^{1/2} = o(1)$, we have

$$\|g\|_\lambda = O_p((nh)^{-1/2} + h^m).$$

Proof of Theorem 2

Denote $g = \hat{\Lambda}_{n,\lambda} - \Lambda_0$. By Theorem 1, we have $\|g\|_\lambda = O_p((nh)^{-1/2} + h^m)$. Then, there exists a constant C such that $B_n = \{\|g\|_\lambda \le r_n \equiv C((nh)^{-1/2} + h^m)\}$ has a large probability to occur. Define $\tilde{g} = d_n^{-1} g$ with $d_n = c_m r_n h^{-1/2}$. Since $h = o(1)$ and $\{\log\log(n)\}(nh^2)^{-1} \to 0$, $d_n = o(1)$. Besides, on B_n, $\|\tilde{g}\|_\infty \le 1$ and $J(\tilde{g}, \tilde{g}) = d_n^{-2}\lambda^{-1}(\lambda J(g,g)) \le d_n^{-2}\lambda^{-1}\|g\|_\lambda^2 \le c_m^{-2}\lambda^{-1}h$. Thus, on B_n, we have $\tilde{g} \in \mathcal{F}$, where $\mathcal{F} = \{g : \|g\|_\infty \le 1, J(g,g) \le c_m^{-2}h\lambda^{-1}\}$. By the Taylor expansion,

$$\begin{aligned}
&\mathcal{S}_n(\hat{\Lambda}_{n,\lambda}) - \mathcal{S}_n(\Lambda_0) - \{\mathcal{S}(\hat{\Lambda}_{n,\lambda}) - \mathcal{S}(\Lambda_0)\}\\
=&\frac{1}{n}\sum_{i=1}^n \left[-K_{U_i} + \Delta_i \frac{K_{U_i}}{1 - \exp\{-\hat{\Lambda}_{n,\lambda}(U_i)\}}\right]\\
&+ \frac{1}{n}\sum_{i=1}^n \left[K_{U_i} - \Delta_i \frac{K_{U_i}}{1 - \exp\{-\Lambda_0(U_i)\}}\right]\\
&- E\left[\Delta \frac{K_U}{1 - \exp\{-\hat{\Lambda}_{n,\lambda}(U)\}} - \Delta \frac{K_U}{1 - \exp\{-\Lambda_0(U)\}}\right]\\
=&\frac{1}{n}\sum_{i=1}^n \left[\Delta_i \frac{K_{U_i}}{1 - \exp\{-\hat{\Lambda}_{n,\lambda}(U_i)\}} - \Delta_i \frac{K_{U_i}}{1 - \exp\{-\Lambda_0(U_i)\}}\right]\\
&- E\left[\Delta \frac{K_U}{1 - \exp\{-\hat{\Lambda}_{n,\lambda}(U)\}} - \Delta \frac{K_U}{1 - \exp\{-\Lambda_0(U)\}}\right]\\
=&-\frac{1}{n}\sum_{i=1}^n \left(\Delta_i \frac{K_{U_i}\exp\{-\Lambda_0(U_i)\}g(U_i)}{[1 - \exp\{-\Lambda_0(U_i)\}]^2} - E\Delta_i \frac{K_{U_i}\exp\{-\Lambda_0(U_i)\}g(U_i)}{[1 - \exp\{-\Lambda_0(U_i)\}]^2}\right)\\
&+ \frac{1}{n}\left(\sum_{i=1}^n \Delta_i \frac{\exp\{-\Lambda_0(U_i)\}K_{U_i}g^2(U_i)[1 + \exp\{-\Lambda_0(U_i)\}]}{[1 - \exp\{-\Lambda_0(U_i)\}]^3}\right.\\
&\qquad\left. - E\Delta_i \frac{\exp\{-\Lambda_0(U_i)\}K_{U_i}g^2(U_i)[1 + \exp\{-\Lambda_0(U_i)\}]}{[1 - \exp\{-\Lambda_0(U_i)\}]^3}\right)(1 + o_p(1))\\
\equiv& I_1 + I_2.
\end{aligned}$$

Note that

$$I_1 = -\frac{1}{n}\sum_{i=1}^n \left[\Delta_i \frac{K_{U_i}\exp\{-\Lambda_0(U_i)\}g(U_i)}{[1 - \exp\{-\Lambda_0(U_i)\}]^2} - E\left\{\Delta_i \frac{K_{U_i}\exp\{-\Lambda_0(U_i)\}g(U_i)}{[1 - \exp\{-\Lambda_0(U_i)\}]^2}\right\}\right]$$

$$= -\frac{1}{n}\sum_{i=1}^{n}\left\{\phi(\Delta_i, U_i, g)K_{U_i} - E\phi(\Delta_i, U_i, g)K_{U_i}\right\},$$

where

$$\phi(\Delta_i, U_i, g) = \Delta_i \frac{\exp\{-\Lambda_0(U_i)\}g(U_i)}{[1-\exp\{-\Lambda_0(U_i)\}]^2}.$$

Define

$$\tilde{\phi}(\Delta_i, U_i, \tilde{g}) = \frac{[1-\exp\{-\Lambda_0(\xi)\}]^2}{\exp\{-\Lambda_0(\xi)\}} d_n^{-1}\phi(\Delta_i, U_i, d_n\tilde{g})c_m^{-1}h^{1/2}.$$

Then,

$$\begin{aligned} & |\tilde{\phi}(\Delta_i, U_i, \tilde{g}) - \tilde{\phi}(\Delta_i, U_i, \tilde{f})| \\ \leq\ & \frac{[1-\exp\{-\Lambda_0(\xi)\}]^2}{\exp\{-\Lambda_0(\xi)\}} d_n^{-1}c_m^{-1}h^{1/2}|\phi(\Delta_i, U_i, d_n\tilde{g}) - \phi(\Delta_i, U_i, d_n\tilde{f})| \\ =\ & \frac{[1-\exp\{-\Lambda_0(\xi)\}]^2}{\exp\{-\Lambda_0(\xi)\}} d_n^{-1}c_m^{-1}h^{1/2}\left|\frac{\Delta_i\exp\{-\Lambda_0(U_i)\}}{[1-\exp\{-\Lambda_0(U_i)\}]^2}d_n\{\tilde{g}(U_i) - \tilde{f}(U_i)\}\right| \\ \leq\ & c_m^{-1}h^{1/2}\|\tilde{f} - \tilde{g}\|_\infty. \end{aligned}$$

It follows from [24] that there exists an event $A_n \subset B_n$ such that $\lim_n P(B_n - A_n) = 0$, and on A_n,

$$\begin{aligned} &\left\|\sum_{i=1}^{n}\left\{\tilde{\phi}(\Delta_i, U_i, \tilde{g})K_{U_i} - E\tilde{\phi}(\Delta_i, U_i, \tilde{g})K_{U_i}\right\}\right\|_\lambda \\ &\leq (n^{1/2}h^{-(2m-1)/(4m)}\|\tilde{g}\|_\lambda^{1-1/(2m)} + 1)\{5\log\log(n)\}^{1/2}. \end{aligned}$$

Then, on event A_n,

$$\begin{aligned} \|I_1\|_\lambda &= \frac{1}{n}\left\|\sum_{i=1}^{n}\{\phi(\Delta_i, U_i, g) - E\phi(\Delta_i, U_i, g)\}\right\|_\lambda \\ &\lesssim \frac{1}{n}(n^{1/2}h^{-(2m-1)/(4m)}\|\tilde{g}\|_\infty^{1-1/(2m)} + 1)\{5\log\log(n)\}^{1/2}d_nc_mh^{-1/2} \\ &= (n^{-1/2}h^{-(2m-1)/(4m)}\|\tilde{g}\|_\infty^{1-1/(2m)} + n^{-1})\{5\log\log(n)\}^{1/2}d_nc_mh^{-1/2}. \end{aligned}$$

As $\|\tilde{g}\|_\infty \leq 1$, then, on A_n,

$$\begin{aligned} \|I_1\|_\lambda &\lesssim (n^{-1/2}h^{-(6m-1)/(4m)} + n^{-1}h^{-1})\{5\log\log(n)\}^{1/2}c_m^2r_n \\ &= O_p(n^{-1/2}h^{-(6m-1)/(4m)}\{5\log\log(n)\}^{1/2}\{(nh)^{-1/2} + h^m\}). \end{aligned}$$

For the main part of I_2 (ignoring $o_p(1)I_2$, still denote as I_2), we have

$$I_2 = \frac{1}{n}\sum_{i=1}^{n}\left\{\Delta_i\frac{\exp\{-\Lambda_0(U_i)\}K_{U_i}g^2(U_i)[1+\exp\{-\Lambda_0(U_i)\}]}{[1-\exp\{-\Lambda_0(U_i)\}]^3}\right.$$

$$-E\left(\Delta_i \frac{\exp\{-\Lambda_0(U_i)\}K_{U_i}g^2(U_i)[1+\exp\{-\Lambda_0(U_i)\}]}{[1-\exp\{-\Lambda_0(U_i)\}]^3}\right)\Bigg\}.$$

Write

$$\varphi(\Delta_i, U_i, g) = \Delta_i \frac{\exp\{-\Lambda_0(U_i)\}K_{U_i}g^2(U_i)[1+\exp\{-\Lambda_0(U_i)\}]}{[1-\exp\{-\Lambda_0(U_i)\}]^3},$$

$$\tilde{\varphi}(\Delta_i, U_i, \tilde{g}) = \frac{[1-\exp\{-\Lambda_0(\xi)\}]^3}{2\exp\{-\Lambda_0(\xi)\}} d_n^2 c_m h^{-1/2} \varphi(\Delta_i, U_i, d_n\tilde{g}).$$

Again, it follows from [24] and $nh^2 \to \infty$ that with n large enough,

$$\|I_2\|_\lambda = O_p(r_n h^{-1/2}\|I_1\|_\lambda) = o_p(\|I_1\|_\lambda).$$

Thereby, it is not hard to show that

$$\begin{aligned} & \|\mathcal{S}_n(\hat{\Lambda}_{n,\lambda}) - \mathcal{S}_n(\Lambda_0) - \{\mathcal{S}(\hat{\Lambda}_{n,\lambda}) - \mathcal{S}(\Lambda_0)\}\|_\lambda \\ = \ & O_p(n^{-1/2}h^{-(6m-1)/(4m)}\{\log\log(n)\}^{1/2}\{(nh)^{-1/2} + h^m\}). \end{aligned}$$

On the other hand,

$$\begin{aligned} & \mathcal{S}_n(\hat{\Lambda}_{n,\lambda}) - \mathcal{S}_n(\Lambda_0) - \{\mathcal{S}(\hat{\Lambda}_{n,\lambda}) - \mathcal{S}(\Lambda_0)\} \\ = \ & \mathcal{S}_{n,\lambda}(\hat{\Lambda}_{n,\lambda}) - \mathcal{S}_{n,\lambda}(\Lambda_0) - \{\mathcal{S}_\lambda(\hat{\Lambda}_{n,\lambda}) - \mathcal{S}_\lambda(\Lambda_0)\} \\ = \ & -\mathcal{S}_{n,\lambda}(\Lambda_0) - \{\mathcal{S}_\lambda(\hat{\Lambda}_{n,\lambda}) - \mathcal{S}_\lambda(\Lambda_0)\} \\ = \ & g - \mathcal{S}_{n,\lambda}(\Lambda_0) - \int_{\mathbb{I}}\int_{\mathbb{I}} sD^2\mathcal{S}_\lambda(\Lambda_0 + ss'g)g^2\, ds\, ds'. \end{aligned}$$

Since $\|\int_{\mathbb{I}}\int_{\mathbb{I}} sD^2\mathcal{S}_\lambda(\Lambda_0+ss'g)g^2\,ds\,ds'\|_\lambda \le \int_{\mathbb{I}}\int_{\mathbb{I}} \|D^2\mathcal{S}_\lambda(\Lambda_0+ss'g)g^2\|_\lambda\,ds\,ds'$ and

$$\|D^2\mathcal{S}_\lambda(\Lambda_0 + ss'g)g^2\|_\lambda = O_p(h^{-1/2}\{(nh)^{-1/2} + h^m\}^2),$$

then $\|g - \mathcal{S}_{n,\lambda}(\Lambda_0)\|_\lambda = O_p(\alpha_n)$, where

$$\alpha_n = h^{-1/2}\{(nh)^{-1}+h^{2m}\}+n^{-1/2}h^{-(6m-1)/(4m)}\{\log\log(n)\}^{1/2}\{(nh)^{-1/2}+h^m\}.$$

Proof of Theorem 3

Define $Rem_n = \hat{\Lambda}_{n,\lambda} - \Lambda^* - \mathcal{S}_n(\Lambda_0)$. It follows from the Functional Bahadur representation that $\|Rem_n\|_\lambda = O_p(\alpha_n)$. As $nh^3 \to \infty$, $nh^{4m-1} \to 0$ and $m > (3+\sqrt{5})/4$, we have $\alpha_n = o(n^{-1/2})$. Since $\|\mathcal{S}_n(\Lambda_0)\|_\lambda = O_p((nh)^{-1/2})$, Rem_n is negligible compared with $\mathcal{S}_n(\Lambda_0)$. Next, we intend to show the asymptotic distribution of $(nh)^{-1/2}\{\hat{\Lambda}_{n,\lambda}(t_0) - \Lambda^*(t_0)\}$. We will use the fact that for any $t \in \mathbb{I}$ and any $g \in \mathcal{H}^m$, $\langle K_t, g\rangle_\lambda = g(t)$. Hence, for any fixed $t_0 \in \mathbb{I}$, since

$$-(nh)^{1/2}\langle K_{t_0}, \mathcal{S}_n(\Lambda_0)\rangle_\lambda = (nh)^{1/2}\frac{1}{n}\sum_{i=1}^n K_{t_0}(U_i)\left[1 - \frac{\Delta_i}{1-\exp\{-\Lambda_0(U_i)\}}\right],$$

then,

$$\begin{aligned} & |(nh)^{1/2}\langle K_{t_0}, \hat{\Lambda}_{n,\lambda} - \Lambda^* - \mathcal{S}_n(\Lambda_0)\rangle_\lambda| \\ \leq \quad & \|K_{t_0}\|_\lambda \|\hat{\Lambda}_{n,\lambda} - \Lambda^* - \mathcal{S}_n(\Lambda_0)\|_\lambda (nh)^{1/2} = o_p(1). \end{aligned}$$

As $\mathrm{Var}\Big\{K_{t_0}(U_i)\big(1 - \Delta_i[1 - \exp\{-\Lambda_0(U_i)\}]^{-1}\big)\Big\} = V(K_{t_0}, K_{t_0})$ and $hV(K_{t_0}, K_{t_0}) \to \sigma^2_{t_0} < c^2_m$,

$$(nh)^{1/2}\langle K_{t_0}, \mathcal{S}_n(\Lambda_0)\rangle_\lambda \xrightarrow{d} N(0, \sigma^2_{t_0})$$

as $n \to \infty$. We finish the proof of Theorem 3.

Proof of Theorem 5

Clearly, part (i) can be obtained from part (ii) and part (iii). Here, we only need to give the proofs of part (ii) and part (iii).

Proof of Theorem 5(ii). For notational convenience, denote $\hat{\Lambda} = \hat{\Lambda}_{n,\lambda}$, $\hat{\Lambda}^0 = \hat{\Lambda}^0_{n,\lambda}$, $g = \hat{\Lambda}^0 + \omega_0 - \hat{\Lambda}$. By Theorem 4, we have

$$\|g\|_\lambda = \|\hat{\Lambda}^0 + \omega_0 - \hat{\Lambda}\|_\lambda \leq \|\hat{\Lambda}^0 + \omega_0 - \Lambda_0\|_\lambda + \|\hat{\Lambda} - \Lambda_0\|_\lambda = O_p(r_n),$$

where $r_n = (nh)^{-1/2} + h^m$. By the Taylor expansion, we have

$$\begin{aligned} PLLRT_{n,\lambda} &= L_{n,\lambda}(\omega_0 + \hat{\Lambda}^0) - L_{n,\lambda}(\hat{\Lambda}) \\ &= \mathcal{S}_{n,\lambda}(\hat{\Lambda})(\omega_0 + \hat{\Lambda}^0 - \hat{\Lambda}) + \int_{\mathbb{I}}\int_{\mathbb{I}} sD\mathcal{S}_{n,\lambda}(\hat{\Lambda} + ss'g)gg\, ds\, ds'. \end{aligned}$$

By the definition of $\mathcal{S}_{n,\lambda}(\hat{\Lambda}) = 0$, $\mathcal{S}_{n,\lambda}(\hat{\Lambda})(\omega_0 + \hat{\Lambda}^0 - \hat{\Lambda}) = 0$. Then,

$$\begin{aligned} & PLLRT_{n,\lambda} \\ =& \int_{\mathbb{I}}\int_{\mathbb{I}} sD\mathcal{S}_{n,\lambda}(\hat{\Lambda} + ss'g)gg\, ds\, ds' \\ =& \int_{\mathbb{I}}\int_{\mathbb{I}} s[D\mathcal{S}_{n,\lambda}(\hat{\Lambda} + ss'g)gg - D\mathcal{S}_{n,\lambda}(\Lambda_0)gg] ds\, ds' + \int_{\mathbb{I}}\int_{\mathbb{I}} sD\mathcal{S}_{n,\lambda}(\Lambda_0)gg\, ds\, ds' \\ =& \int_{\mathbb{I}}\int_{\mathbb{I}} s[D\mathcal{S}_{n,\lambda}(\hat{\Lambda} + ss'g)gg - D\mathcal{S}_{n,\lambda}(\Lambda_0)gg] ds\, ds' + \frac{1}{2}[D\mathcal{S}_{n,\lambda}(\Lambda_0)gg - D\mathcal{S}_\lambda(\Lambda_0)gg] \\ & + \frac{1}{2}D\mathcal{S}_\lambda(\Lambda_0)gg \\ \equiv& I_1 + I_2 + I_3. \end{aligned}$$

Define $\tilde{g} = \hat{\Lambda} + ss'g - \Lambda_0$, for any $0 \leq s, s' \leq 1$, $\|\tilde{g}\|_\lambda = \|\hat{\Lambda} - \Lambda_0 + ss'g\|_\lambda \leq \|\hat{\Lambda} - \Lambda_0\|_\lambda + \|g\|_\lambda = O_p(r_n)$. Then,

$$D\mathcal{S}_{n,\lambda}(\hat{\Lambda} + ss'g)gg = D\mathcal{S}_{n,\lambda}(\tilde{g} + \Lambda_0)gg$$

$$= -\frac{1}{n}\sum_{i=1}^{n}\Delta_i \frac{g^2(U_i)\exp\{-\tilde{g}(U_i)-\Lambda_0(U_i)\}}{[1-\exp\{-\tilde{g}(U_i)-\Lambda_0(U_i)\}]^2} - \lambda\int_{\mathbb{I}}\{g^{(m)}(t)\}^2\,dt.$$

Note that

$$\begin{aligned}
&|D\mathcal{S}_{n,\lambda}(\hat{\Lambda}+ss'g)gg - D\mathcal{S}_{n,\lambda}(\Lambda_0)gg| \\
&= \Big|\frac{1}{n}\sum_{i=1}^{n}\Delta_i \frac{g^2(U_i)\exp\{-\tilde{g}(U_i)-\Lambda_0(U_i)\}}{[1-\exp\{-\tilde{g}(U_i)-\Lambda_0(U_i)\}]^2} - \frac{1}{n}\sum_{i=1}^{n}\Delta_i \frac{g^2(U_i)\exp\{-\Lambda_0(U_i)\}}{[1-\exp\{-\Lambda_0(U_i)\}]^2}\Big| \\
&= \Big|\frac{1}{n}\sum_{i=1}^{n}\Delta_i g^2(U_i)\Big[-\Big\{\frac{1}{1-\exp\{-\Lambda_0(U_i)-\tilde{g}(U_i)\}} - \frac{1}{1-\exp\{-\Lambda_0(U_i)\}}\Big\} \\
&\qquad + \Big\{\frac{1}{[1-\exp\{-\Lambda_0(U_i)-\tilde{g}(U_i)\}]^2} - \frac{1}{[1-\exp\{-\Lambda_0(U_i)\}]^2}\Big\}\Big]\Big| \\
&\le \Big|\frac{1}{n}\sum_{i=1}^{n}\Delta_i g^2(U_i)\frac{\exp\{-\Lambda_0(U_i)\}\tilde{g}(U_i)}{[1-\exp\{-\Lambda_0(U_i)\}]^2}\Big|\{1+o_p(1)\} \\
&+ \Big|\frac{1}{n}\sum_{i=1}^{n}\Delta_i g^2(U_i)\frac{\exp\{-\Lambda_0(U_i)\}\tilde{g}(U_i)}{[1-\exp\{-\Lambda_0(U_i)\}]^3}\Big|\{1+o_p(1)\} \\
&\equiv \tilde{I}_1 + \tilde{I}_2.
\end{aligned}$$

For the main part of $\tilde{I}_1$ (ignoring the term $o_p(1)\tilde{I}_1$, still denoted as $\tilde{I}_1$), we have

$$\begin{aligned}
\tilde{I}_1 &\le \|\tilde{g}\|_\infty \Big|\frac{1}{n}\sum_{i=1}^{n}\Big[\Delta_i \frac{g^2(U_i)\exp\{-\Lambda_0(U_i)\}}{[1-\exp\{-\Lambda_0(U_i)\}]^2} - E_U\Big\{\frac{g^2(U_i)\exp\{-\Lambda_0(U_i)\}}{[1-\exp\{-\Lambda_0(U_i)\}]}\Big\}\Big]\Big| \\
&\quad + \|\tilde{g}\|_\infty E_U\Big\{\frac{g^2(U_i)\exp\{-\Lambda_0(U_i)\}}{[1-\exp\{-\Lambda_0(U_i)\}]}\Big\} \\
&\le \|\tilde{g}\|_\infty \Big|\frac{1}{n}\sum_{i=1}^{n}\Big[\Delta_i \frac{g^2(U_i)\exp\{-\Lambda_0(U_i)\}}{[1-\exp\{-\Lambda_0(U_i)\}]^2} - E_U\Big\{\frac{g^2(U_i)\exp\{-\Lambda_0(U_i)\}}{1-\exp\{-\Lambda_0(U_i)\}}\Big\}\Big]\Big| \\
&\quad + \|\tilde{g}\|_\infty \|g\|_\lambda^2 \\
&\equiv I_{11} + I_{12}.
\end{aligned}$$

From the proof of Theorem 2, we have

$$I_{11} = \|\tilde{g}\|_\infty O_p(r_n \alpha_n'),$$

where $\alpha_n' = n^{-1/2}\{(nh)^{-1/2} + h^m\}h^{-(6m-1)/(4m)}\{\log\log(n)\}^{1/2}$. Then,

$$|\tilde{I}_1| = \|\tilde{g}\|_\infty O_p(r_n\alpha_n') + \|\tilde{g}\|_\infty O_p(r_n^2).$$

Similarly, we have $|\tilde{I}_2| = \|\tilde{g}\|_\infty O_p(r_n\alpha_n') + \|\tilde{g}\|_\infty O_p(r_n^2)$. Thus,

$$|I_1| = \|\tilde{g}\|_\infty O_p(r_n\alpha_n') + \|\tilde{g}\|_\infty O_p(r_n^2).$$

Under the conditions of λ, we have $n^{-1/2}h^{-(6m-1)/(4m)}\{\log\log(n)\}^{1/2} = o(1)$. Thus $\alpha_n' = o(r_n)$ and

$$|2I_1| = \|\tilde{g}\|_\infty O_p(r_n^2) \le h^{-1/2} r_n O_p(r_n^2) = O_p(h^{-1/2} r_n^3).$$

Further, it can be easily checked that

$$|2I_2| = |D\mathcal{S}_{n,\lambda}(\Lambda_0)gg - D\mathcal{S}_\lambda(\Lambda_0)gg| = O_p(r_n \alpha_n').$$

As $I_3 = -\|g\|_\lambda^2/2$,

$$PLLRT_{n,\lambda} = -\frac{\|g\|_\lambda^2}{2} + O_p(h^{-1/2} r_n^3 + r_n \alpha_n').$$

It follows from $nh^{2m} \to 0$ that $nh^{2m+1} \to 0$. Together with $nh^4 \to \infty$, $h^{-1/2} r_n^3 + r_n \alpha_n' = o(n^{-1})$. As a result,

$$-2nPLLRT_{n,\lambda} = n\|\hat{\Lambda}^0 + \omega_0 - \hat{\Lambda}\|_\lambda^2 + o_p(1). \qquad \square$$

Proof of Theorem 5(iii). As $-2nPLLRT_{n,\lambda} = n\|\hat{\Lambda}^0 + \omega_0 - \hat{\Lambda}\|_\lambda^2 + o_p(1)$, it suffices to derive the asymptotic property of $n\|\hat{\Lambda}^0 + \omega_0 - \hat{\Lambda}\|_\lambda^2$. Note that

$$\begin{aligned} & n^{1/2}\|\hat{\Lambda}^0 + \omega_0 - \hat{\Lambda} - \mathcal{S}_{n,\lambda}^0(\Lambda_0^0) + \mathcal{S}_{n,\lambda}(\Lambda_0)\|_\lambda \\ \le\ & n^{1/2}\|\hat{\Lambda}^0 + \omega_0 - \mathcal{S}_{n,\lambda}^0(\Lambda_0^0)\|_\lambda + n^{1/2}\|\hat{\Lambda} - \mathcal{S}_{n,\lambda}(\Lambda_0)\|_\lambda \\ =\ & O_p(n^{1/2}\alpha_n) = o_p(1). \end{aligned}$$

Hence, we only need to focus on $n^{1/2}\{\mathcal{S}_{n,\lambda}^0(\Lambda_0^0) - \mathcal{S}_{n,\lambda}(\Lambda_0)\}$. Note that

$$\begin{aligned} & \mathcal{S}_{n,\lambda}^0(\Lambda_0^0) \\ =\ & \frac{1}{n}\sum_{i=1}^n \left[\frac{\Delta_i K_{U_i}^*}{1 - \exp\{-\Lambda_0^0(U_i) - \omega_0\}} - K_{U_i}^*\right] - W_\lambda^* \Lambda_0^0 \\ =\ & \frac{1}{n}\sum_{i=1}^n \left[\frac{\Delta_i}{1 - \exp\{-\Lambda_0^0(U_i) - \omega_0\}}\left\{K_{U_i} - \frac{K_{U_i}(t_0)K_{t_0}}{K(t_0,t_0)}\right\}\right. \\ & \left. - \left\{K_{U_i} - \frac{K_{U_i}(t_0)K_{t_0}}{K(t_0,t_0)}\right\}\right] \\ & - \left\{W_\lambda \Lambda_0 - \frac{W_\lambda(\Lambda_0)(t_0)K_{t_0}}{K(t_0,t_0)}\right\}. \end{aligned}$$

Thus,

$$\begin{aligned} & \mathcal{S}_{n,\lambda}^0(\Lambda_0^0) - \mathcal{S}_{n,\lambda}(\Lambda_0) \\ =\ & \frac{K_{t_0}}{K(t_0,t_0)}\left[\frac{1}{n}\sum_{i=1}^n \left\{\frac{-\Delta_i K_{U_i}(t_0)}{1 - \exp\{-\Lambda_0(U_i)\}} + K_{U_i}(t_0)\right\} + (W_\lambda \Lambda_0)(t_0)\right]. \end{aligned}$$

Thereby,

$$n^{1/2}\|\mathcal{S}^0_{n,\lambda}(\Lambda^0_0) - \mathcal{S}_{n,\lambda}(\Lambda_0)\|_\lambda$$
$$= \left|\frac{1}{\sqrt{K(t_0,t_0)}}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n\left\{\frac{-\Delta_i K_{U_i}(t_0)}{1-\exp\{-\Lambda_0(U_i)\}} + K_{U_i}(t_0)\right\} + (W_\lambda\Lambda_0)(t_0)\right]\right|.$$

As $nh^{2m} \to 0$ as $n \to \infty$,

$$\frac{\sqrt{n}(W_\lambda\Lambda_0)(t_0)}{\|K_{t_0}\|_\lambda} \le \frac{\sqrt{nh}(W_\lambda\Lambda_0)(t_0)}{h^{1/2}\|V^{1/2}(K_{t_0},K_{t_0})\|_\lambda}$$
$$= O(1)\frac{\sqrt{nh}(W_\lambda\Lambda_0)(t_0)}{\sigma_{t_0}} = O(\sqrt{n}h^m) = o(1).$$

Consequently,

$$\frac{1}{\sqrt{K(t_0,t_0)}}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n\left\{\frac{-\Delta_i K_{U_i}(t_0)}{1-\exp(-\Lambda_0(U_i))} + K_{U_i}(t_0)\right\}\right.$$
$$\left. + (W_\lambda\Lambda_0)(t_0)\right] \xrightarrow{d} N(0, c_{t_0})$$

as $n \to \infty$, where

$$c_{t_0} = \lim_{h\to 0}\frac{V(K_{t_0},K_{t_0})}{\|K_{t_0}\|^2} \in (0,1].$$

Finally, we have shown $-2nPLLRT_{n,\lambda} \xrightarrow{d} c_{t_0}\chi^2_1$ as $n \to \infty$. □

Proof of Theorem 6

For ease of presentation, we denote $g = \Lambda_0 - \hat{\Lambda}_{n,\lambda}$ and $r_n = (nh)^{-1/2} + h^m$. By the Taylor expansion,

$$\begin{aligned} PGLRT_{n,\lambda} &= l_{n,\lambda}(\Lambda_0) - l_{n,\lambda}(\hat{\Lambda}_{n,\lambda}) \\ &= \mathcal{S}_{n,\lambda}(\hat{\Lambda}_{n,\lambda})(\Lambda_0 - \hat{\Lambda}_{n,\lambda}) + \int_{\mathbb{I}}\int_{\mathbb{I}} sD\mathcal{S}_{n,\lambda}(\hat{\Lambda}_{n,\lambda} + ss'g)\,ds\,ds' \\ &\equiv I_1 + I_2. \end{aligned}$$

According to the the definition of $\mathcal{S}_{n,\lambda}$, $|I_1| = 0$. It follows from similar lines of the proofs of Theorem 5(ii) that

$$|I_2| = -\frac{\|g\|^2_\lambda}{2} + O_p(h^{-1/2}r_n^3 + r_n\alpha'_n),$$

where $\alpha'_n = h^{-(6m-1)/(4m)}n^{-1/2}\{\log\log(n)\}^{1/2}r_n$. Thus,

$$PGLRT_{n,\lambda} = -\frac{\|g\|^2_\lambda}{2} + O_p(h^{-1/2}r_n^3 + r_n\alpha'_n).$$

Under the conditions that $m > (3+\sqrt{5})/4$, $nh^{2m+1} = O(1)$ and $nh^3 \to \infty$, we have

$$-2nPGLRT_{n,\lambda} = n\|g\|_\lambda^2 + o_p(h^{-1/2}).$$

Under the null hypothesis H_0^{global} and Theorem 2, $\|\hat{\Lambda}_{n,\lambda} - \Lambda_0 - \mathcal{S}_{n,\lambda}(\Lambda_0)\| = O_p(\alpha_n)$. And it follows from Theorem 3 that $n^{1/2}\alpha_n = o(1)$. Thus,

$$n^{1/2}\|g\|_\lambda = n^{1/2}\|\mathcal{S}_{n,\lambda}(\Lambda_0)\|_\lambda + o_p(1).$$

Next, we study the leading term $\|\mathcal{S}_{n,\lambda}(\Lambda_0)\|_\lambda$. Write

$$\begin{aligned} n\|\mathcal{S}_{n,\lambda}(\Lambda_0)\|_\lambda^2 &= n\Big\|\frac{1}{n}\sum_{i=1}^n \left[-K_{U_i} + \Delta_i \frac{K_{U_i}}{1-\exp\{-\Lambda_0(U_i)\}}\right] - W_\lambda\Lambda_0\Big\|_\lambda^2 \\ &= \frac{1}{n}\Big\|\sum_{i=1}^n \left[-K_{U_i} + \Delta_i \frac{K_{U_i}}{1-\exp\{-\Lambda_0(U_i)\}}\right]\Big\|_\lambda^2 \\ &\quad -2\langle\sum_{i=1}^n \left[-K_{U_i} + \Delta_i \frac{K_{U_i}}{1-\exp\{-\Lambda_0(U_i)\}}\right], \\ &\quad W_\lambda\Lambda_0\rangle_\lambda + n\|W_\lambda\Lambda_0\|_\lambda^2. \end{aligned}$$

We first approximate $\|W_\lambda\Lambda_0\|_\lambda$. Define

$$m_\lambda(j) = |V(\Lambda_0, h_j)|^2\gamma_j \frac{\lambda\gamma_j}{1+\lambda\gamma_j}, \qquad j = 0, 1, 2, \ldots.$$

Then, $|m_\lambda(j)|$ is a sequence of functions satisfying $|m_\lambda(j)| \le |V(\Lambda_0, h_j)|^2\gamma_j \equiv m(j)$. Since $\Lambda_0 \in \mathcal{H}^m$, we have $\sum_j |V(\Lambda_0, h_j)|^2\gamma_j = \int_N m(j)\,d\mu(j) = J(\Lambda_0, \Lambda_0) < \infty$, where $\mu(\cdot)$ is the counting measure. As $\lim_{\lambda\to 0} m_\lambda(j) = 0$,

$$\lim_{\lambda\to 0}\sum_j |V(\Lambda_0, h_j)|^2 \frac{\lambda\gamma_j^2}{1+\lambda\gamma_j} = \lim_{\lambda\to 0}\int_N m_\lambda(j)\,d\mu(j) = 0$$

by the Lebesgue dominated convergence theorem. That is,

$$\|W_\lambda\Lambda_0\|_\lambda^2 = \sum_j |V(\Lambda_0, h_j)|^2 \frac{\lambda^2\gamma_j^2}{1+\lambda\gamma_j} = o(\lambda),$$

which implies that

$$\begin{aligned} &E\Big|\langle\sum_{i=1}^n \left[-K_{U_i} + \Delta_i \frac{K_{U_i}}{1-\exp\{-\Lambda_0(U_i)\}}\right], W_\lambda\Lambda_0\rangle\Big|^2 \\ = \;& E\Big|\sum_{i=1}^n \left[\frac{\Delta_i}{1-\exp\{-\Lambda_0(U_i)\}} - 1\right] W_\lambda\Lambda_0\Big|^2 \\ = \;& nE_U\Big[\frac{\exp\{-\Lambda_0(U_i)\}}{1-\exp\{-\Lambda_0(U_i)\}}(W_\lambda\Lambda_0)^2\Big] \end{aligned}$$

$$\leq \quad n\|W_\lambda(\Lambda_0(t))\|_\lambda^2 = o(n\lambda).$$

Thus, it follows from $nh^{2m+1} = O(1)$ that

$$\begin{aligned} & \langle\sum_{i=1}^{n}\left[-K_{U_i} + \Delta_i \frac{K_{U_i}}{1-\exp\{-\Lambda_0(U_i)\}}\right], W_\lambda\Lambda_0\rangle_\lambda \\ = \quad & o_p((n\lambda)^{1/2}) = o_p(n^{1/2}h^m) = o_p(h^{-1/2}). \end{aligned}$$

Hence,

$$n\|\mathcal{S}_{n,\lambda}(\Lambda_0)\|_\lambda^2 = \frac{1}{n}\Big\|\sum_{i=1}^{n}\left[-K_{U_i} + \Delta_i \frac{K_{U_i}}{1-\exp\{-\Lambda_0(U_i)\}}\right]\Big\|_\lambda^2 + o_p(h^{-1}).$$

In what follows, we study the limiting property of

$$n^{-1}\|\sum_{i=1}^{n} -K_{U_i} + \Delta_i K_{U_i}[1-\exp\{-\Lambda_0(U_i)\}]^{-1}\|_\lambda^2.$$

Direct calculations yield that

$$\begin{aligned} & \frac{1}{n}\Big\|\sum_{i=1}^{n}\left[-K_{U_i} + \Delta_i \frac{K_{U_i}}{1-\exp\{-\Lambda_0(U_i)\}}\right]\Big\|_\lambda^2 \\ & = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\Delta_i}{1-\exp\{-\Lambda_0(U_i)\}} - 1\right]^2 K_{U_i}(U_i) + \frac{1}{n}W_n, \end{aligned}$$

where

$$W_n = \sum_{i\neq j}\left[\frac{\Delta_i}{1-\exp\{-\Lambda_0(U_i)\}} - 1\right]\left[\frac{\Delta_j}{1-\exp\{-\Lambda_0(U_j)\}} - 1\right]\langle K_{U_i}, K_{U_j}\rangle_\lambda.$$

Define

$$W_{ij} = 2\left[\frac{\Delta_i}{1-\exp\{-\Lambda_0(U_i)\}} - 1\right]\left[\frac{\Delta_j}{1-\exp\{-\Lambda_0(U_j)\}} - 1\right]\langle K_{U_i}, K_{U_j}\rangle_\lambda,$$

and $W_n = \sum_{1\leq i<j\leq n} W_{ij}$ such that W_n is clean ([21]). Next, we intend to derive the limiting distribution of W_n. Let $\sigma_n^2 = \mathrm{Var}(W_n)$. Then,

$$\begin{aligned} \sigma_n^2 &= \frac{n(n-1)}{2}E(W_{ij}^2) \\ &= 2n(n-1)E\left(\left[\frac{\Delta_i}{1-\exp\{-\Lambda_0(U_i)\}} - 1\right]\left[\frac{\Delta_j}{1-\exp\{-\Lambda_0(U_j)\}} - 1\right]K_{U_i}(U_j)\right)^2 \\ &= 2n(n-1)\sum_{l=0}^{\infty}\frac{1}{(1+\lambda\gamma_l)^2}. \end{aligned}$$

Write $G_1 = \sum_{i<j} E(W_{ij}^4)$, $G_2 = \sum_{i<j<k} E\{W_{ij}^2 W_{ik}^2\} + E\{W_{ji}^2 W_{jk}^2\} + E\{W_{ki}^2 W_{kj}^2\}$, and

$$G_3 = \sum_{i<j<k<l} E\{W_{ij}W_{ik}W_{lj}W_{lk}\} + E\{W_{ij}W_{il}W_{kj}W_{kl}\} + E\{W_{ik}W_{il}W_{jk}W_{jl}\}.$$

It follows from Proposition 3.2 of [21] that, if G_1, G_2, G_3 are all of lower order than σ_n^4, then $\sigma_n^{-1}W_n$ converges in distribution to the standard normal distribution. Note that

$$\begin{aligned} &E\{W_{ij}^4\} \\ &= 16E\left(\left[\frac{\Delta_i}{1-\exp\{-\Lambda_0(U_i)\}} - 1\right]\left[\frac{\Delta_j}{1-\exp\{-\Lambda_0(U_j)\}} - 1\right]\langle K_{U_i}, K_{U_j}\rangle_\lambda\right)^4 \\ &= O(h^{-4}). \end{aligned}$$

Then, $G_1 = O(n^2h^{-4})$. By the Cauchy-Schwarz inequality, we have

$$EW_{ij}^2 W_{ik}^2 \leq (EW_{ij}^4)^{1/2}(EW_{ik}^4)^{1/2} = O(h^{-4}).$$

Thus, $G_2 = O(n^3h^{-4})$. Straightforward calculations give that

$$E\{W_{ij}W_{ik}W_{lj}W_{lk}\} = 16\sum_{j=0}^{\infty}\frac{1}{(1+\lambda\gamma_j)^4} = O(h^{-1}).$$

Therefore, $G_3 = O(n^4h^{-1})$. Since $\sigma_n^4 = (\sigma_n^2)^2 = O(n^4h^{-2})$, $nh^3 \to \infty$ and $h = o(1)$, G_1, G_2, G_3 are of smaller order than that of σ_n^4. Hence,

$$\sigma_n^{-1}W_n \xrightarrow{d} N(0,1).$$

It then follows from $\rho_\lambda^2 = \sum_{j=0}^{\infty} h/(1+\lambda\gamma_j)^2$ that

$$\frac{1}{\sqrt{2h^{-1}}n\rho_\lambda}W_n \xrightarrow{d} N(0,1). \tag{4}$$

Next, consider

$$\frac{1}{n}\sum_{i=1}^n\left[\frac{\Delta_i}{1-\exp\{-\Lambda_0(U_i)\}} - 1\right]^2\langle K_{U_i}, K_{U_i}\rangle_\lambda.$$

It can be easily checked that

$$E\left(\left[\frac{\Delta_i}{1-\exp\{-\Lambda_0(U_i)\}} - 1\right]^2\langle K_{U_i}, K_{U_i}\rangle_\lambda\right)^2 = O(\|K_U\|_\lambda^4) = O(h^{-2}),$$

and

$$E\left(\sum_{i=1}^n\left[\frac{\Delta_i}{1-\exp\{-\Lambda_0(U_i)\}} - 1\right]^2\langle K_{U_i}, K_{U_i}\rangle_\lambda - h^{-1}\sigma_\lambda^2\right)^2$$

$$\leq \quad nE\left(\left[\frac{\Delta_i}{1-\exp\{-\Lambda_0(U_i)\}}-1\right]^2\langle K_{U_i},K_{U_i}\rangle_\lambda\right)^2 = O(nh^{-2}),$$

where $\sigma_\lambda^2 = \sum_{j=0}^{\infty} h/(1+\lambda\gamma_j)$. Thus,

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\Delta_i}{1-\exp\{-\Lambda_0(U_i)\}}-1\right]^2\langle K_{U_i},K_{U_i}\rangle_\lambda = h^{-1}\sigma_\lambda^2 + O_p((n^{1/2}h)^{-1}). \quad (5)$$

It follows from (4) and (5) that $n\|\mathcal{S}_{n,\lambda}\|_\lambda^2 = O_p(h^{-1})$. Hence,

$$n^{1/2}\|\mathcal{S}_{n,\lambda}\|_\lambda = O_p(h^{-1/2}).$$

Finally,

$$\begin{aligned} -2nPGLRT_{n,\lambda} &= \{n^{1/2}\|\mathcal{S}_{n,\lambda}\|_\lambda + o_p(1)\}^2 + o_p(h^{-1/2}) \\ &= n\|\mathcal{S}_{n,\lambda}\|_\lambda^2 + o_p(h^{-1/2}). \end{aligned} \quad (6)$$

Combining (4), (5) and (6), we have

$$(2h^{-1}\sigma_\lambda^4/\rho_\lambda^2)^{-1/2}(-2n\gamma_\lambda PGLRT_{n,\lambda} - n\gamma_\lambda\|W_\lambda\Lambda_0(t)\|_\lambda^2 - h^{-1}\sigma_\lambda^4/\rho_\lambda^2) \xrightarrow{d} N(0,1).$$

The proof of Theorem 6 is complete.

Proof of Theorem 7

First, it can be verified by straightforward calculations that $m > (3+\sqrt{5})/4$ and $h \asymp n^{-d}$, where $1/(2m+1) \leq d < 1/3$ and satisfies those conditions in Theorem 6. Then, we only need to consider $\Lambda_{n_0} = \Lambda_0 + \Lambda_n$ for $\Lambda_n \in \mathcal{A}$. Write

$$\begin{aligned} -2n\cdot PGLRT_{n,\lambda} &= -2n\{l_{n,\lambda}(\Lambda_0) - l_{n,\lambda}(\Lambda_{n_0})\} - 2n\{l_{n,\lambda}(\Lambda_{n_0}) - l_{n,\lambda}(\hat{\Lambda}_{n,\lambda})\} \\ &\equiv I_1 + I_2. \end{aligned} \quad (7)$$

Regarding I_1, we define

$$\begin{aligned} R_i &= (\Delta_i \log[-\exp\{-\Lambda_0(U_i)\}] - (1-\Delta_i)\Lambda_0(U_i)) \\ &\quad - (\Delta_i \log[1-\exp\{-\Lambda_{n_0}(U_i)\}] - (1-\Delta_i)\Lambda_{n_0}(U_i)) \\ &= \Delta_i\left(\log[1-\exp\{-\Lambda_0(U_i)\}] - \log[1-\exp\{-\Lambda_{n_0}(U_i)\}]\right) - (1-\Delta_i)\Lambda_n(U_i). \end{aligned}$$

Obviously,

$$ER_i^2 = O(\|\Lambda_n\|_\lambda^2 + \|\Lambda_n\|_\lambda^4).$$

Then,

$$E\left\{\Big|\sum_{i=1}^{n}(R_i - ER_i)\Big|^2\right\} \leq nER_i^2 = (n\|\Lambda_n\|_\lambda^2 + n\|\Lambda_n\|_\lambda^4),$$

and

$$n\,[l_{n,\lambda}(\Lambda_0) - l_{n,\lambda}(\Lambda_{n_0}) - E\{l_{n,\lambda}(\Lambda_0) - l_{n,\lambda}(\Lambda_{n_0})\}] = O_p(n^{1/2}\|\Lambda_n\|_\lambda^2 + n^{1/2}\|\Lambda_n\|_\lambda).$$

On the other hand, in view of the fact that $D\mathcal{S}_\lambda(g)\Lambda_n\Lambda_n < 0$ for any $g \in \mathcal{H}^m$, there exists constant $c' > 0$ such that

$$E\{D\mathcal{S}_{n,\lambda}(\Lambda_{n_0}^*)\Lambda_n\Lambda_n\} \le c'E\{D\mathcal{S}_{n,\lambda}(\Lambda_{n_0})\Lambda_n\Lambda_n\} = \frac{-c'\|\Lambda_n\|_\lambda^2}{2}.$$

Then,

$$\begin{aligned}
&E\{l_{n,\lambda}(\Lambda_0) - l_{n,\lambda}(\Lambda_{n_0})\}\\
&= E\{\mathcal{S}_{n,\lambda}(\Lambda_{n_0})(-\Lambda_n) + \frac{1}{2}D\mathcal{S}_{n,\lambda}(\Lambda_{n_0}^*)\Lambda_n\Lambda_n\}\\
&\le \lambda J(\Lambda_{n_0}, \Lambda_n) - \frac{c'\|\Lambda_n\|_\lambda^2}{2}\\
&\le \lambda\{J(\Lambda_n, \Lambda_n) + J(\Lambda_0, \Lambda_n)\} - \frac{c'\|\Lambda_n\|_\lambda^2}{2}\\
&\le \lambda\{J(\Lambda_n, \Lambda_n) + J(\Lambda_0, \Lambda_0)^{1/2}J(\Lambda_n, \Lambda_n)^{1/2}\} - \frac{c'\|\Lambda_n\|_\lambda^2}{2}\\
&= O(\lambda) - \frac{c'\|\Lambda_n\|_\lambda^2}{2}.
\end{aligned}$$

Combining the above results, we obtain

$$\begin{aligned}
I_1 &\ge n\|\Lambda_n\|_\lambda^2 + O_p(n\lambda + n^{1/2}\|\Lambda_n\|_\lambda + n^{1/2}\|\Lambda_n\|_\lambda^2)\\
&= n\|\Lambda_n\|_\lambda^2\{1 + O_p(\lambda\|\Lambda_n\|_\lambda^{-2} + n^{-1/2}\|\Lambda_n\|_\lambda^{-1} + n^{-1/2})\}. \qquad (8)
\end{aligned}$$

As for I_2, under H_{1n}, it follows from $\|\hat{\Lambda}_{n,\lambda} - \Lambda_{n_0}\| = O_p((nh)^{-1/2} + h^m)$ and the FBR that, for any $\epsilon \in (0,1)$, there exists a positive constant C and an integer N such that

$$\inf_{n\ge N}\inf_{\Lambda_n\in\mathcal{A}} P_{\Lambda_{n_0}}\left(\|\hat{\Lambda}_{n,\lambda} - \Lambda_{n_0} - S_{n,\lambda}(\Lambda_{n_0})\|_\lambda \le Cr_n\right) \ge 1 - \epsilon, \qquad (9)$$

where $r_n = (nh)^{-1/2} + h^m$. Similar to the proofs of Theorem 6, we can show that I_2 has the same limiting distribution as that in Theorem 6, uniformly over any $\Lambda_n \in \mathcal{A}$. In other words,

$$(2\nu_{n_0})^{-1/2}(I_2 - n\|W_\lambda\Lambda_{n_0}\|_\lambda^2 - h^{-1}\sigma_{n_0,\lambda}^2) = O_p(1), \qquad (10)$$

uniformly over any $\Lambda_n \in \mathcal{A}$, where $\nu_{n_0} = h^{-1}\sigma_{n_0,\lambda}^4/\rho_{n_0,\lambda}^2$, $\sigma_{n_0,\lambda}^2$, $\rho_{n_0,\lambda}^2$ are of the form of σ_λ^2, ρ_λ^2 with the eigenvalues and eigenvectors derived under Λ_{n_0}. Let V_{n_0} and V_0 be the V functions defined in Section 2. Then, for any $f \in \mathcal{H}^m$,

$$|V_{n_0}(f,f) - V_0(f,f)| = \zeta V_0(f,f)\|\Lambda_n\|_\infty.$$

It follows from [24] that

$$\sigma^2_{n_0,\lambda} - \sigma^2_\lambda = O(h^{-1/2}\|\Lambda_n\|_\lambda). \tag{11}$$

Combining (8), (10) and (11), we have

$$\begin{aligned}
&(2\nu_n)^{-1/2}(-2nr_\lambda PGLRT_{\lambda,n} - \nu_n)\\
=&(2\nu_n)^{-1/2}(-r_\lambda(I_1+I_2) - \nu_n)\\
=&(2\nu_n)^{-1/2}r_\lambda(I_2 - n\|W_\lambda\Lambda_{n_0}\|^2_\lambda - h^{-1}\sigma^2_{n_0,\lambda}) + (2\nu_n)^{-1/2}r_\lambda n\|W_\lambda\Lambda_{n_0}\|^2_\lambda\\
&+ (2\nu_n)^{-1/2}r_\lambda I_1 + (2\nu_n)^{-1/2}r_\lambda h^{-1}(\sigma^2_{n_0,\lambda} - \sigma^2_\lambda)\\
\geq&O_p(1) + (2\nu_n)^{-1/2}r_\lambda n\|\Lambda_n\|^2_\lambda(1 + O_p(\lambda\|\Lambda_n\|^{-2}_\lambda + n^{-1/2}\|\Lambda_n\|^{-1}_\lambda + n^{-1/2}))\\
&+ O(h^{-1}\|\Lambda_n\|_\lambda),
\end{aligned}$$

where $O_p(1)$ holds uniformly in $\mathcal{A}$, $\nu_n = h^{-1}\sigma^4_\lambda/\rho^2_\lambda$, and r_λ is defined in Theorem 6. Let $\lambda\|\Lambda_n\|^{-2}_\lambda \leq 1/C$, $n^{-1/2}\|\Lambda_n\|^{-1}_\lambda \leq 1/C$, $Ch^{-1}\|\Lambda_n\|_\lambda \leq (nh^{1/2})\|\Lambda_n\|^2_\lambda$, and $\|\Lambda_n\|^2_\lambda \geq C(nh^{1/2})^{-1}$ for some sufficiently large constant C. Then,

$$|(2\nu_n)^{-1/2}(-2nr_\lambda PGLRT_{\lambda,n} - \nu_n)| \geq c_\alpha,$$

where c_α is the critical value (based on $N(0,1)$) for rejecting H_0^{global} at significance level α. In other words,

$$\|\Lambda_n\|^2_\lambda \gtrsim (h^{2m} + (nh^{1/2})^{-1}). \tag{12}$$

Combining (9) and (12), the main results in Theorem 7 are proved.

Acknowledgments

The authors are grateful to the Editor, the Associate Editor and the anonymous reviewers for their professional review and insightful comments that lead to substantial improvements in the paper.

References

[1] Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W., Silverman, E., *et al*. (1955). An empirical distribution function for sampling with incomplete information. *Ann. Math. Statist.* **26**, 641–647. MR0073895

[2] Banerjee, M. (2007). Likelihood based inference for monotone response models. *Ann. Statist.* **35**, 931–956. MR2341693

[3] Banerjee, M. and Wellner, J. A. (2001). Likelihood ratio tests for monotone functions. *Ann. Statist.* **29**, 1699–1731. MR1891743

[4] Banerjee, M. and Wellner, J. A. (2005a). Confidence intervals for current status data. *Scand. J. Statist.* **32**, 405–424. MR2204627

[5] Banerjee, M. and Wellner, J. A. (2005b). Score statistics for current status data: comparisons with likelihood ratio and Wald statistics. *Int. J. Biostat.* **1**, Article 3. MR2232228

[6] Brunel, E. and Comte, F. (2009). Cumulative distribution function estimation under interval censoring case 1. *Electron. J. Stat.* **3**, 1–24. MR2471584

[7] Cai, T. and Betensky, R. A. (2003). Hazard regression for interval-censored data with penalized spline. *Biometrics* **59**, 570–579. MR2004262

[8] Cox, D. D. and O'Sullivan, F. (1990). Asymptotic analysis of penalized likelihood and related estimators. *Ann. Math. Statist.* **18**, 1676–1695. MR1074429

[9] Eeden, V. (1956). Maximum likelihood estimation of ordered probabilities. *Indag. Math.* **18**, 444–455. MR0083859

[10] Eubank, R. L. (1999). *Nonparametric Regression and Spline Smoothing*. CRC press. MR1680784

[11] Groeneboom, P. (2014). Maximum smoothed likelihood estimators for the interval censoring model. *Ann. Statist.* **42**, 2092–2137. MR3262478

[12] Groeneboom, P. and Jongbloed, G. (2014). *Nonparametric Estimation Under Shape Constraints*. Cambridge University Press, Cambridge. MR3445293

[13] Groeneboom, P. and Jongbloed, G. (2015). Nonparametric confidence intervals for monotone functions. *Ann. Statist.* **43**, 2019–2054. MR3375875

[14] Groeneboom, P., Jongbloed, G. and Witte, B. I. (2010). Maximum smoothed likelihood estimation and smoothed maximum likelihood estimation in the current status model. *Ann. Statist.* **38**, 352–387. MR2589325

[15] Groeneboom, P. and Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation.* Birkhäuser-Verlag, Basel. MR1180321

[16] Huang, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *Ann. Statist.* **24**, 540–568. MR1394975

[17] Huang, J. and Wellner, J. A. (1995). Asymptotic normality of the NPMLE of linear functionals for interval censored data, case 1. *Statist. Neerl.* **49**, 153–163. MR1345376

[18] Murphy, S. A. and van der Vaart, A. W. (1997). Semiparametric likelihood ratio inference. *Ann. Statist.* **25**, 1471–1509. MR1463562

[19] Ingster, Y.I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives $I-III$. *Math. Methods Statist.* **2**, 85–114; **3**, 171–189; **4**, 249–268. MR1259685

[20] Joly, P., Commenges, D., and Letenneur, L. (1998). A penalized likelihood approach for arbitrarily censored and truncated data: application to age-specific incidence of dementia. *Biometrics* **54**, 185–194.

[21] Jong, D. P. (1987). A central limit theorem for generalized quadratic forms. *Probab. Theory Related Fields* **75**, 261–277. MR0885466

[22] Rosenberg, P. S. (1995). Hazard function estimation using B-splines. *Biometrics* **51**, 874–887.

[23] Schumaker, L. L. (1981). *Spline Functions: Basic Theory*. John Wiley

and Sons, New York. MR0606200

[24] Shang, Z. and Cheng, G. (2013). Local and global asymptotic inference in smoothing spline models. *Ann. Statist.* **41**, 2608–2638. MR3161439

[25] Sun, J. (2006). *The Statistical Analysis of Interval-censored Failure Time Data*. Springer, New York. MR2287318

[26] van der Vaart, A. and van der Laan, M. J. (2006). Estimating a survival distribution with current status data and high-dimensional covariates. *Int. J. Biostat.* **2**(1), Article 9. MR2306498

[27] Wu, Y. and Zhang, Y. (2012). Partially monotone tensor spline estimation of the joint distribution function with bivariate current status data. *Ann. Statist.* **40**, 1609–1636. MR3015037

[28] Zhang, Y., Hua, L. and Huang, J. (2010). A spline-based semiparametric maximum likelihood estimation method for the Cox model with interval-censored data. *Scand. J. Statist.* **37**, 338–354. MR2682304