Statistica Sinica Preprint No: SS-2018-0010

Title	Reliability Estimation from Left-Truncated and Right-Censored Data Using Splines
Manuscript ID	SS-2018-0010
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202018.0010
Complete List of Authors	Zhisheng Ye Weiwei Jiang and Xingqiu Zhao
Corresponding Author	Zhisheng Ye
E-mail	yez@nus.edu.sg
Notice: Accepted version subject to English editing.	

Reliability Estimation from Left-Truncated and Right-Censored Data Using Splines

May 15, 2018

Abstract

Reliability data collected from the field are often left truncated and right censored, because the data collection process usually starts much later than the installation of the first product unit, and some units are still in service at the end of the data collection. The truncation introduces sampling biases and makes analysis of the lifetime data complicated. This study develops a nonparametric likelihood-based estimation procedure for left-truncated and right-censored data using B-splines. In terms of small sample performance and large sample efficiencies, the proposed spline-based estimators for the reliability function are shown to be more efficient than the existing nonparametric estimators. We further consider nonparametric two-sample tests for left-truncated and right-censored data. The new class of tests is useful for comparing reliability of similar products. The test statistics are based on the cumulative weighted differences between the two estimated failure rates. Asymptotic distributions of proposed statistics are derived and their finite-sample properties are evaluated through Monte Carlo simulations. The performance of the proposed test statistics is compared with that of the weighted Kaplan-Meier statistics. A real-life example from high-voltage power transformers is used to illustrate this method.

Keywords: B-splines, Convergence rate, Asymptotic normality, Two-sample tests.

1 Introduction

Lifetime data collected from field operations contain important reliability information that is useful for asset management, such as preventive maintenance and remaining useful life prediction. Compared with reliability data collected from life tests, field failure data are usually subject to serious multiple censoring and truncation. In particular, the data are typically left truncated and right censored. The left truncation arises when the data collection starts later than the product launch/installation (Ye and Tang 2016). Because of the high reliability, most products will be still functioning when the data collection stops, leading to right censoring. The phenomenon is common for assets used in infrastructure facilities, such as pipes in a water supply network (Carrión et al. 2010) and power transformers used in a power grid (Hong et al. 2009). An illustration of the data generation mechanism is provided in Figure 1. The starting date is fixed for all product units. However, the installation dates (or sales dates) are generally random across the product population. The randomness in the left-truncation time results from

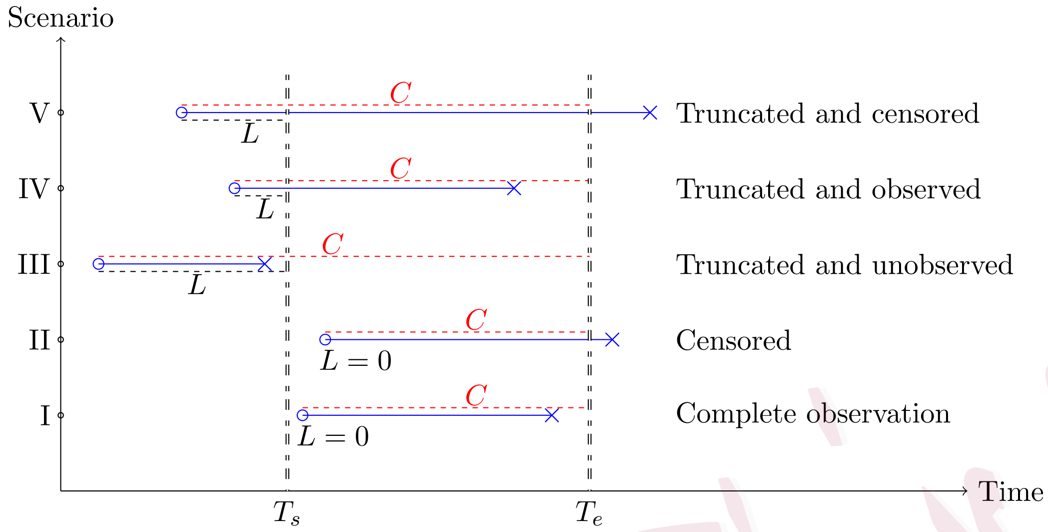


Figure 1: A schematic of the mechanism that generates left-truncated and right-censored data. The observation window in calendar date is $[T_s, T_e]$. Here, 'o' represents installation at date T_I, '$\times$' represents the failure event. The respective truncation and censoring times are $L = \max\{0, T_s - T_I\}$ and $C = \max\{0, T_e - T_I\}$.

the random installation dates. If a unit is installed before the starting date of data collection, then it is subject to left-truncation. Further, if its lifetime is longer than the truncation time, it is a left-truncated observation (possibly subject to censoring); otherwise, the unit is truncated, unobserved and the existence of the unit is unknown. The un-truncated population corresponds to those units whose installation dates are later than the starting date of data collection. The same data format is also common in survival study of clinical trials. For some further real examples, see Tsai et al. (1987), Kevin et al. (2011) and Su and Wang (2012), among others.

The vast of the literature on reliability data analysis has mainly focused on right-censored data, as life tests are an important source of reliability data. The problem of left-truncation gradually attracts research interest in recent years due to its prevalence in the increasingly important area of asset management. The recent literature on left-truncated and right-censored data features parametric models and the related inference. Hong et al. (2009) fitted a Weibull distribution to lifetime data of high-voltage power transformers from an energy company in the United States. The maximum likelihood (ML) estimates were obtained through direct maximization. An alternative method for ML estimation of the Weibull distribution is the EM algorithm developed in Balakrishnan and Mitra (2012). Parametric inferences for other distributions, such as lognormal and gamma, were developed in Balakrishnan and Mitra (2011, 2013, 2014) and Emura and Shiu (2016), among others.

A problem with the parametric inference is that the estimation results, such as the reliability function and lifetime quantiles, could be sensitive to distributional assumptions. In addition, it may be difficult, if not impossible, to check the distributional assumption in the presence of heavy truncation (Kevin et al. 2011). In view of these deficiencies, it is desirable to use nonparametric inference methods that impose less assumptions on the lifetimes. A breakthrough is found in Turnbull (1976), who proposed a nonparametric maximum likelihood estimation (NPMLE) procedure for arbitrarily censored and truncated data. He further developed a selfconsistent algorithm to compute the NPMLE, which turns out to be a special case of the EM algorithm. Frydman (1994) corrected Turnbull's algorithm to make it applicable when the data is truncated as well as interval censored. Consistency and efficiency of the NPMLE were established in a number of studies, e.g., Wang et al. (1986) and Tsai et al. (1987). The EM algorithm convergences quite slowly if the collected failure data are heavily truncated, and it can be sensitive to the initial values. With appropriate adjustment of the definition of the risk set, Tsai et al. (1987) showed that the NPMLE of the survivor function and the cumulative failure rate can be directly obtained by using the analogue of the Kaplan-Meier and Nelson-Aalen estimators.

In the study, we propose nonparametric inference for left-truncated and right-censored data using splines. A spline is a piecewise polynomial function that possesses a high degree of smoothness at the places where the polynomial pieces connect. These connection points are known as knots. Once the knots are given, it is easy to compute the splines recursively for any desired degree of the polynomial (Schumaker 2007, Chapter IV). The main advantages of the spline interpolation are its stability and calculation simplicity. When applied in nonparametric estimation, the number of parameters in the spline is usually much smaller than those in traditional nonparametric methods. This makes the estimation easier and computation less time. Therefore, spline-based nonparametric estimation has received considerable attention in recent years. In Rosenberg (1995), nonnegative B-splines, also called M-splines, are applied to estimate the hazard function of censored survival data, where the nonnegativity is guaranteed by the nonnegativity coefficients. To be specific, M-splines can be considered as a normalized version of B-splines with unit integral within the domain (Ramsay 1988). Monotonic B-splines are also widely applied in the literature (e.g., Lu et al. 2009; Xie et al. 2018), where the monotonicity is guaranteed by the nondecreasing order of coefficients. On the other hand, I-splines, whose bases are integrals from the B-splines (Ramsay 1988) are used to approximate the cumulative distribution function (CDF) in Wu and Ying (2012). I-splines naturally yield monotonicity with nonnegative coefficients, while B-splines require nondecreasing order of coefficients to ensure monotonicity. Therefore, I-splines are often used to approximate monotone functions, which may simplify the numerical computation (Wu and Ying 2012; Hong et al. 2015; Lu et al. 2007).

Motivated by the promising performance, spline basis functions are adopted for left-truncated and right-censored data. Although splines come in many different forms, they have intimate relations (Ramsay 1988; Lu et al. 2007). Using B-splines to approximate the failure rate is the same as using M-splines for the failure rate, which is further the same as using I-splines to approximate the cumulative failure rate. In our paper, we use B-splines with nonnegative constraints on the spline coefficients to approximate the failure rate, and the I-splines to approximate the cumulative failure rate. We do not pursue approximating the cumulative distribution and reliability function because the approximation induces a normalization constraint on the spline function, which complicates the maximum likelihood estimation. We show that the convergence rate of the estimated failure rate is faster than $O(n^{1/3})$. Based on the inferential results, we further develop spline-based two-sample tests for comparison of two left-truncated and right-censored datasets. The results are useful to compare reliability of similar products.

The rest of the paper is organized as follows. Section 2 formulates the spline-based likelihood estimation problem for left-truncated and right-censored data. The asymptotic properties of the spline estimators are presented in Section 3. Based on the asymptotic results, a nonparametric two-sample test is proposed to compare lifetime data from two products in Section 4. Section 5 conducts simulation studies to evaluate finite sample performance of the spline estimators. Section 6 applies the proposed spline methods to the power transformer example in Hong et al. (2009). Some technical lemmas and the proofs of the theorems are sketched in the Appendix.

2 B-Spline Approximation of the Failure Rate

Consider the lifetime T of a product unit with reliability $R(t)$, failure rate $\lambda_0(t)$ and cumulative failure rate $\Lambda_0(t)$, $t \geq 0$. The lifetime T is subject to left truncation with truncation time L, $L \geq 0$. A unit is observed only when $T > L$. The unit is further subject to right censoring with random censoring time C and $C > L$. If the observation window of the product is a fixed interval, then $C - L$ equals the length of the interval if $L > 0$. See Figure 1 for an illustration. In calendar date, we let T_I be the installation date of a random unit and $[T_s, T_e]$ be the observation interval. Then in terms of product age, the left-truncation time is $L = \max\{0, T_s - T_I\}$ and the right-censoring time is $C = \max\{0, T_e - T_I\}$. Naturally, the truncation times and censoring times are bounded because T_I cannot be earlier than the product launch date (Shen and Yan 2008; Shen 2014; Balakrishnan and Mitra 2012). Therefore, we let $[\underline{L}, \bar{L}]$ and $[\underline{C}, \bar{C}]$ be the respective supports of L and C, and assume $\bar{L}, \bar{C} < \infty$. Furthermore, suppose T and (L, C) are independent. Because of left-truncation and right-censoring, the lifetime information is only available within the interval $[\underline{L}, \bar{C}]$. As a result, the failure rate of T is identifiable in $[\underline{L}, \bar{C}]$.

When $T \geq L$, the unit enters into our observation, and the observed lifetime is denoted as $Y = \min(T, C)$, $Y \geq L$. Let $\delta = I(T \leq C)$ be the censoring indicator. That is, $\delta = 1$ if the lifetime is observed, and 0 if censored. The observation from the unit is thus $X = (L, Y, \delta)$. Let $X_i = (L_i, Y_i, \delta_i)$, $i = 1, \cdots, n$ be n i.i.d. copies of X, and $\mathbf{D} = \{X_1, X_2, \cdots, X_n\}$. We are interested in estimating the failure rate $\lambda(t)$ using $\mathbf{D}$. It suffices to consider the conditional log-likelihood (Wang 1987)

$$\mathcal{L}(\lambda|\mathbf{D}) = \sum_{i=1}^{n} \left\{ \delta_i \ln \lambda(Y_i) - \int_{L_i}^{Y_i} \lambda(s) ds \right\}.$$

In order to implement the spline approximation, a finite closed interval $[a, b]$ is identified first. The guideline for choosing a and b is that, they should include all the observed L_i's and Y_i's. In an application, we can let $a = \min\{L_i, i = 1, \cdots, n\}$, and $b = \max\{Y_i, i = 1, \cdots, n\}$. Given $[a, b]$, let $\mathcal{T} = \{t_j\}_1^{m_n+2l}$, with

$$a = t_1 = \cdots = t_l < t_{l+1} < \cdots < t_{l+m_n} < t_{l+m_n+1} = \cdots = t_{m_n+2l} = b,$$

be a sequence of knots that partition $[a, b]$ into $m_n + 1$ subintervals $J_j = [t_{l+j}, t_{l+j+1})$, $j = 0, 1, \cdots, m_n$. To ensure the large-sample property as discussed in the next section, the number m_n of inner knots is usually chosen as $O(n^{\nu})$ for some $\nu \in (0, 1/2)$. A common choice is $m_n = \lceil n^{1/3} \rceil$, e.g., Lu et al. (2007, 2009), and Hua and Zhang (2012). With fixed m_n, the inner knots $\{t_j\}_{l+1}^{l+m_n}$ can be either equally spaced (Lu et al. 2007), or placed at the corresponding quantiles of the distinct observation times $\{Y_i\}_1^n$ (Hua and Zhang 2012), or even the Chebyshev points. According to our simulation experience, as well as the simulation experiments reported in the literature, e.g., Zhao et al. (2013), the estimation results are insensitive to the selection of m_n and the placement of knots. For ease of implementation, we would recommend $m_n = \lceil n^{1/3} \rceil$ and equally-spaced inner knots.

With the knot sequence, $q_n = m_n + l$ spline bases, denoted as B_k, $1 \le k \le q_n$, can be constructed using a recursive formula (Schumaker 2007, Chapter IV). The class of polynomial splines of order l with the knot sequence $\mathcal{T}$ is the linear space spanned by these bases (Schumaker 2007, Theorem 4.18). To satisfy the nonnegativity constraint of the failure rate approximation, we single out a subclass of $\psi_{l,\mathcal{T}}$ as

$$\psi_{l,\mathcal{T}} = \left\{ \sum_{k=1}^{q_n} \alpha_k B_k : \alpha_k \ge 0 \right\}.$$

According to theorem 5.9 of Schumaker (2007), $\psi_{l,\mathcal{T}}$ is a class of nonnegative polynomial splines on $[a, b]$. The nonnegativity of the B-splines is guaranteed by the nonnegative coefficients. For each $h(\cdot) \in \psi_{l,\mathcal{T}}$, h is a polynomial of order l in the interval J_j for $0 \le j \le m_n$, and h is $l-2$ times continuously differentiable on $[a, b]$. Define $I_k(t) = \int_a^t B_k(s)ds$. Using the spline approximation, the log-likelihood function can be written as

$$\mathcal{L}(\boldsymbol{\alpha}|\mathbf{D}) = \sum_{i=1}^{n} \left\{ \delta_i \ln \left[\sum_{k=1}^{q_n} \alpha_k B_k(y_i) \right] + \sum_{k=1}^{q_n} \alpha_k I_k(L_i) - \sum_{k=1}^{q_n} \alpha_k I_k(y_i) \right\}. \tag{1}$$

Let $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \cdots, \hat{\alpha}_{q_n})$ be the spline coefficients that maximize (1) subject to the nonnegativity constraints $\alpha_k \ge 0$, $k = 1, \cdots, q_n$. The spline log-likelihood function (1) is concave with respect to the unknown coefficients. Therefore, the spline estimation problem is equivalent to a nonlinear convex programming subject to linear inequality constraints. The optimization can be easily solved by most software packages for scientific/statistical computation. Based on $\hat{\boldsymbol{\alpha}}$, the spline-based likelihood estimator for the failure rate is $\hat{\lambda}_n(t) = \sum_{k=1}^{q_n} \hat{\alpha}_k B_k(t)$.

3 Statistical Properties

In this section, we study the statistical properties of the spline-based likelihood estimator $\hat{\boldsymbol{\alpha}}$ with the L_2-metric d given by

$$d(\lambda_1, \lambda_2) = \|\lambda_1 - \lambda_2\|_2 = \left\{ \int |\lambda_1(t) - \lambda_2(t)|^2 dF^*(t) \right\}^{1/2},$$

where $F^*(t) = P(L \le T \le C, T \le t)$ and λ_1, λ_2 are non-negative functions. To ensure the asymptotic convergence, we first require $m_n = O(n^\nu)$, for some $\nu \in (0, 1/2)$ (Stone 1994). Below, we list the technical assumptions for the theoretical results of the proposed spline-based NPMLE.

- Condition 1: The maximum spacing of the knots satisfies

$$\Delta = \max_{l+1 \le j \le m_n + l + 1} |t_j - t_{j-1}| = O(n^{-\nu}).$$

  Moreover, there exists a constant $M > 0$ such that $\Delta/\delta \le M$ uniformly in n, where $\delta = \min_{l+1 \le j \le m_n + l + 1} |t_j - t_{j-1}|$.

- Condition 2: The interval $[a, b]$ satisfies $P(\{Y \in [a, b]\}) = 1$.

- Condition 3: there exists a constant $C_0 > 0$ such that $\lambda_0(t) \ge C_0$ for $t \in [a, b]$. In addition, the true failure rate λ_0 is differentiable up to order r and all the dirivatives are uniformly bounded by a constant M in $[a, b]$, where $r \ge 1$.

Remark 1. Condition 1 is a weak restriction on the knot sequence, which is satisfied when equally-spaced knots are used. This condition is also adopted by Stone (1994). Condition 2 requires that $[\underline{L}, \bar{C}] \subset [a, b]$. Condition 3 is needed in the proof of the asymptotic normality in Theorem 3. It usually holds in practice. Product lifetime is a nonnegative and continuous random variable. Continuous parametric lifetime distributions, such as Weibull, lognormal and inverse Gaussian distributions have been widely used to model the lifetime data. See Balakrishnan and Mitra (2011, 2012, 2013), among others. All the parametric distributions have smooth hazard rate functions. As an extension from parametric to nonparametric estimation, the smoothness assumption in Condition 3 is natural and reasonable. This assumption is also used in Wang (2005) and Zhao and Zhang (2017).

Theorem 1 (Consistency) *Suppose that Conditions 1-3 hold. Then the estimated failure rate $\hat{\lambda}_n$ converges to the true failure rate λ_0 in probability, that is,*

$$\|\hat{\lambda}_n - \lambda_0\|_2 \to_p 0, \quad \text{as } n \to \infty.$$

Theorem 2 (Rate of convergence) *Suppose that Conditions 1-3 hold. If ν is chosen to be $1/(2r+1)$, then*

$$n^{\frac{r}{2r+1}} \|\hat{\lambda}_n - \lambda_0\|_2 = O_p(1).$$

Remark 2. Theorem 2 shows that the spline likelihood estimators have a convergence rate slower than $n^{-1/2}$ but faster than $n^{-1/3}$.

To discuss the asymptotic distributions of functions of $\hat{\lambda}_n$, define

$$\mathcal{H}_r = \left\{ h(\cdot) : |h^{(r-1)}(s) - h^{(r-1)}(t)| \le c_0 |s - t| \text{ for all } a \le s, t \le b \right\},$$

where $h^{(r-1)}$ is the $(r-1)$th derivative of h, and $c_0 > 0$ is a constant. Let $\mathcal{U}_\lambda$ denote a neighborhood of the failure rate λ_0. We also define a sequence of maps G_n, mapping $\mathcal{U}_\lambda$ in the parameter space for λ into $\mathcal{L}^\infty(\mathcal{H}_r)$ as

$$G_n(\lambda)[h] = n^{-1} \sum_{i=1}^{n} \left\{ \delta_i \frac{h(Y_i)}{\lambda(Y_i)} - \int_{L_i}^{Y_i} h(t) dt \right\} = \mathbb{P}_n \phi(\lambda; X)[h].$$

The limit map $G : \mathcal{U}_\lambda \mapsto \mathcal{L}^\infty(\mathcal{H}_r)$ is

$$G(\lambda)[h] = P\phi(\lambda; X)[h] = P\left\{\delta\frac{h(Y)}{\lambda(Y)} - \int_L^Y h(t)dt\right\},$$

where $X = (L, Y, \delta)$, $\mathbb{P}_n$ and P denote the empirical measure and probability measure with $\mathbb{P}_n g = n^{-1}\sum_{i=1}^n g(X_i)$ and $Pg = \int g dP$, respectively.

Theorem 3 (Asymptotic normality) *Suppose Conditions 1-3 hold. Then for $h \in \mathcal{H}_r$,*

$$\sqrt{n}\int_a^b \frac{h(t)}{\lambda_0^2(t)}\{\hat{\lambda}_n(t) - \lambda_0(t)\}dF^*(t) = \sqrt{n}(G_n - G)(\lambda_0)[h] + o_p(1). \tag{2}$$

Remark 3. Theorem 3 does not require $\hat{\lambda}_n$ be $\sqrt{n}$-consistent. Since we assume λ_0 is differentiable, it is easy to see that F^* is differentiable with derivative denoted as $f^*(t)$. Consider the situation $\underline{L} = 0$ and $f^*(t) > 0$ for all $t \in [0, \bar{C}]$. For any fixed time $\tau \in [0, \bar{C}]$, choose $h(t) = I_{(0,\tau]}(t)\lambda_0^2(t)/f^*(t)$ to see that the estimated cumulative hazard $\hat{\Lambda}_n(\tau)$ is $\sqrt{n}$ consistent for $\Lambda_0(\tau)$. Further, a routine evaluation of the right-hand side of (2) shows that the asymptotic variance of $\hat{\Lambda}_n(\tau)$ is the same as that for the NPMLE of Λ_0 given in Wang et al. (1986). This means that the proposed method leads to efficient estimation of the cumulative failure rate. Moreover, the asymptotic normality can be used to construct new tests for the problem of multi-sample nonparametric comparison of reliabilities of left-truncated and right-censored data, as shown in the next section.

4 Nonparametric Tests

Due to technological advances and the availability of multiple suppliers, the fleet of assets used in the field usually consists of different brands or different generations of the same brand (Ye et al. 2013). The difference naturally stratifies the field failure data into several categories. The transformer failure data analyzed in Hong et al. (2009) is a typical example of this kind. It is important to know if there is any difference between categories in terms of product reliability. A knowledge of the difference can be used to select a more reliable product. If there is no difference in the reliability, the field data can be combined to achieve a more accurate estimation of the product lifetime distribution. In the literature of left-truncated and right-censored data, some extensions of nonparametric tests have been developed for two-sample comparison, such as the Wilcoxon test, the weighted Kaplan-Meier (WKM) statistics (Shen 2007) and the weighted log-rank statistics (Shen 2014). The above tests are based on estimates of the failure rates, the cumulative failure rates or the survival functions. Similarly, we use the spline-based smooth estimatior of the failure rate developed above and further propose a flexible class of nonparametric test statistics based on integrated weighted differences between the two estimated failure rates. The performance will be compared with the weighted Kaplan-Meier statistics (Shen 2007) in Section 5.

Consider two homogeneous groups. In group k, $k = 1, 2$, the ith observed lifetime is $X_i^{(k)} = (L_i^{(k)}, Y_i^{(k)}, \delta_i^{(k)})$. The observed data of group k are $\mathbf{D}_k = \left\{X_i^{(k)}, i = 1, 2, \cdots, n_k\right\}$. Let $n = n_1 + n_2$. Assume that the failure rate and the cumulative failure rate functions of units from group k are λ_k and Λ_k, respectively. The goal is to test $H_0 : \lambda_1 = \lambda_2 = \lambda_0$ where λ_0 denotes the unknown common failure rate function when H_0 is true. The test statistics proposed here capitalize on the spline-based estimator developed in Section 3. Let $\hat{\lambda}_n^{(k)}(t)$ and $\hat{\lambda}_n(t)$ be the B-spline maximum likelihood estimators of $\lambda_k(t)$ and $\lambda_0(t)$ based on $\mathbf{D}_k$ and the pooled data $\mathbf{D} = \mathbf{D}_1 \cup \mathbf{D}_2$, respectively. Motivated by a method that is commonly used in survival analysis (e.g., Pepe and Fleming 1989; Balakrishnan and Zhao 2009), we propose the following test statistic

$$U_n = \sqrt{n} \int_a^b W_n(t)\{\hat{\lambda}_n^{(1)}(t) - \hat{\lambda}_n^{(2)}(t)\} dF_n^*(t), \tag{3}$$

where W_n is a bounded weight process (Zhao and Zhang 2017; Balakrishnan and Zhao 2009; Andersen et al. 1993, Chapter V), and $F_n^*(t) = \frac{\sum_{i=1}^n \delta_i I(Y_i \le t)}{\sum_{i=1}^n \delta_i}$. The presence of the weight process $W_n(t)$ makes the above statistic flexible. A simple and natural choice for the weight is $W_n^{(1)}(t) = 1$. Another natural choice is $W_n^{(2)}(t) = Z_n(t) = \frac{1}{n}\sum_{i=1}^n I(L_i < t \le Y_i)$, in which case weights are proportional to the number of subjects under observation. In addition, one may choose the weight process as

$$W_n^{(3)}(t) = \frac{Z_{n_1}(t) Z_{n_2}(t)}{Z_n(t)}, \quad W_n^{(4)}(t) = 1 - Z_n(t),$$

where $Z_{n_k}(t)$ is defined as $Z_n(t)$ with the summation being only over subjects in sample k. Some weight processes similar to $W_n^{(3)}$ and $W_n^{(4)}$ have been used for recurrent event data (e.g., Andersen et al. 1993, Chapter V). Now, we state the asymptotic distribution of U_n.

Theorem 4 *Suppose $\lambda_1 = \lambda_2 = \lambda_0$, and Conditions 1-3 hold for λ_0 and the spline estimators $\hat{\lambda}_n^{(1)}, \hat{\lambda}_n^{(2)}, \hat{\lambda}_n$. Further suppose W_n are bounded weight processes and that there exists a bounded function $W(t)$ such that $W \in \mathcal{H}_r$, and*

$$\left[\int_a^b \{W_n(t) - W(t)\}^2 dt\right]^{1/2} = o_p\left(n^{-\frac{1}{2(1+2r)}}\right).$$

Also suppose that $n_1/n \to p$ as $n \to \infty$ with $0 < p < 1$. Then, U_n has an asymptotic normal distribution $N(0, \sigma_w^2)$, where

$$\sigma_w^2 = \frac{1}{p(1-p)} E\{\phi^2(\lambda_0; X)[h_w]\}$$

that can be consistently estimated by

$$\hat{\sigma}_w^2 = \frac{n}{n_1 n_2} \sum_{i=1}^n \{\phi^2(\hat{\lambda}_n; X_i)[\hat{h}_w]\}$$

with $h_w(t) = W(t)\{\lambda_0(t)\}^2$ and $\hat{h}_w(t) = W_n(t)\{\hat{\lambda}_n(t)\}^2$.

Remark 4. For the asymptotic normality of the proposed test statistics, we do not need the bounded Lipschitz condition for the selection of the weight processes, which is required by Balakrishnan and Zhao (2009).

5 Simulation Studies

To verify the performance of the proposed spline-based estimators under finite samples, a Monte Carlo simulation is conducted. In the simulation study, we choose cubic B-splines with order $l = 4$, which are popular in the literature (Lu et al. 2009; Hong et al. 2015; Xie et al. 2018). In addition, m_n is set as $\lceil n^{1/3} \rceil$. The other simulation settings follow the work of Balakrishnan and Mitra (2012).

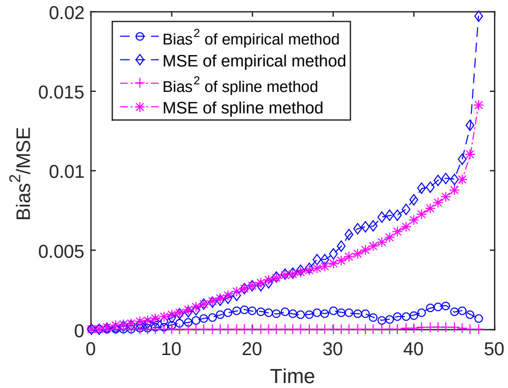
In detail, the starting date T_s of data collection is fixed as 1980 and the end date T_e is 2008. Let n be the size of observed units and p be the proportion of truncated observations, i.e., $100p\%$ of the observed units are installed before 1980. Let $T_{I,i}$ be the installation time of unit i, $i = 1, \cdots, n$, which are assigned as follows. The earliest installation date $\underline{T}_I$ is 1960. For the period $1960 - 1979$, a proportion of 0.15 is attached to each of the first five years, and the remaining proportion is distributed equally over the rest of the years of this period. For the period $1980 - 1995$, a proportion of 0.1 was attached to each of the first six years, and a proportion of 0.04 is attached to each of the rest of the years of this period. Accordingly, the left-truncation time of unit i is $L_i = \max\{0, T_s - T_{I,i}\}$ and the right-censoring time of unit i is $C_i = \max\{0, T_e - T_{I,i}\}$, $i = 1, \cdots, n$. See more details in Balakrishnan and Mitra (2012).

Four distributions are considered for the product lifetime T, i.e., Weibull, Lognormal, a mixture of two Weibull distributions, and a mixture of Lognormal and Gamma distributions (Balakrishnan and Mitra 2012, 2011, 2013). The generated data are fitted using the proposed spline method and Turnbull's NPMLE (Tsai et al. 1987). Here, we consider two fixed proportions of truncated observations, i.e., $p = 40\%$ and $p = 80\%$, and two sample sizes, i.e., $n = 100$ and 200. Based on 50,000 Monte Carlo replications, the squared biases and the Mean Squared Errors (MSEs) of the reliability estimators using the two methods are computed. The results are presented in Figures 2 - 5. From the plots, we can see that the squared biases and the mean square errors of spline-based reliability estimators are smaller than the Turnbull's NPMLE in both proportions of truncated observations. Furthermore, comparisons between Figures 2 and 3 (or Figures 4 and 5) show that when the sample size n doubles, the mean square errors of spline-based reliability estimators drop substantially, which supports the asymptotic consistency of these estimators (Theroem 1).

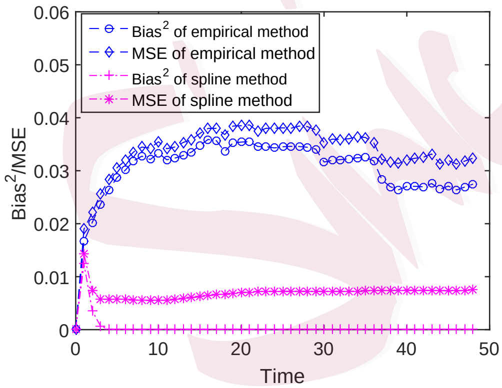
Next, we examine the finite-sample properties of the proposed two-sample test statistic U_n. Assume the lifetimes of units in the two groups follow the Weibull distributions with different values of the scale parameter α and shape parameter β. To guarantee truncation and censoring, we generate data by following the simulation setting of Balakrishnan and Mitra (2012) for each group. The null hypothesis $H_0 : \lambda_1 = \lambda_2 = \lambda_0$ is equivalent to $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$. If the null is true, then $T_n = U_n/\hat{\sigma}_w$ is approximately standard normal, where U_n in (3) can be expressed as

$$U_n = \frac{\sqrt{n}}{\sum_{i=1}^{n} \delta_i} \sum_{i=1}^{n} \delta_i W_n(Y_i) \left\{ \hat{\lambda}_1(Y_i) - \hat{\lambda}_2(Y_i) \right\},$$

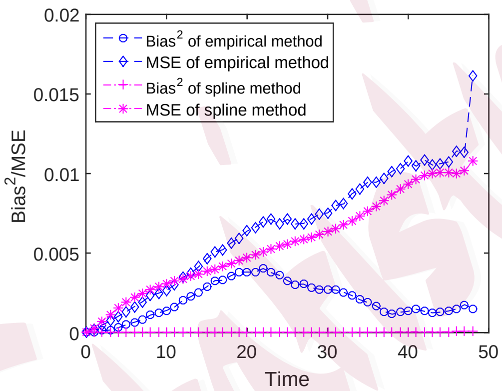
and $\hat{\sigma}_w$ is given in Theorem 4. Let T_H denote the weighted Kaplan-Meier (WKM) statistics developed by Shen (2007). Here we focus on evaluating the performance of T_n and comparing them to those of T_H. We consider two scenarios as follows:



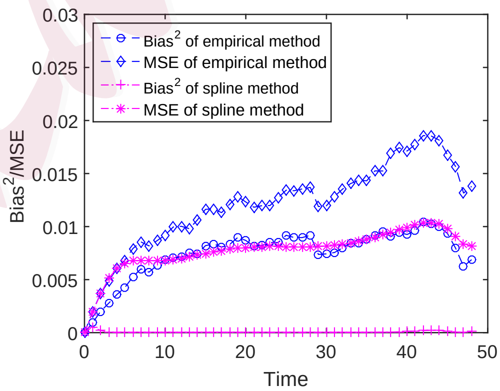
(a) $T \sim \text{Weibull}(80, 1.5)$.



(b) $T \sim \ln N(3, 5)$.

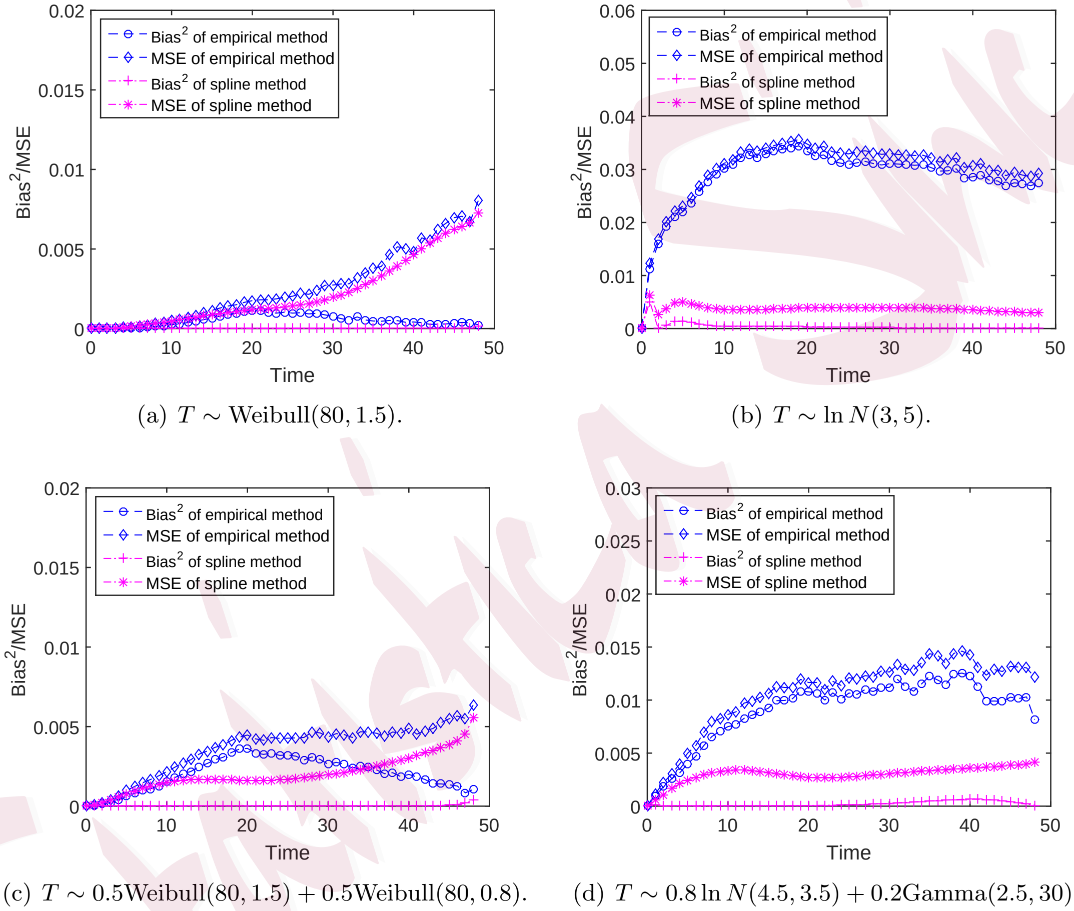


(c) $T \sim 0.5\text{Weibull}(80, 1.5) + 0.5\text{Weibull}(80, 0.8)$.



(d) $T \sim 0.8 \ln N(4.5, 3.5) + 0.2\text{Gamma}(2.5, 30)$.

Figure 2: Comparisons of the spline estimator and the NPMLE (Turnbull 1976) for estimating the reliability function when $n = 100$ and $p = 40\%$.



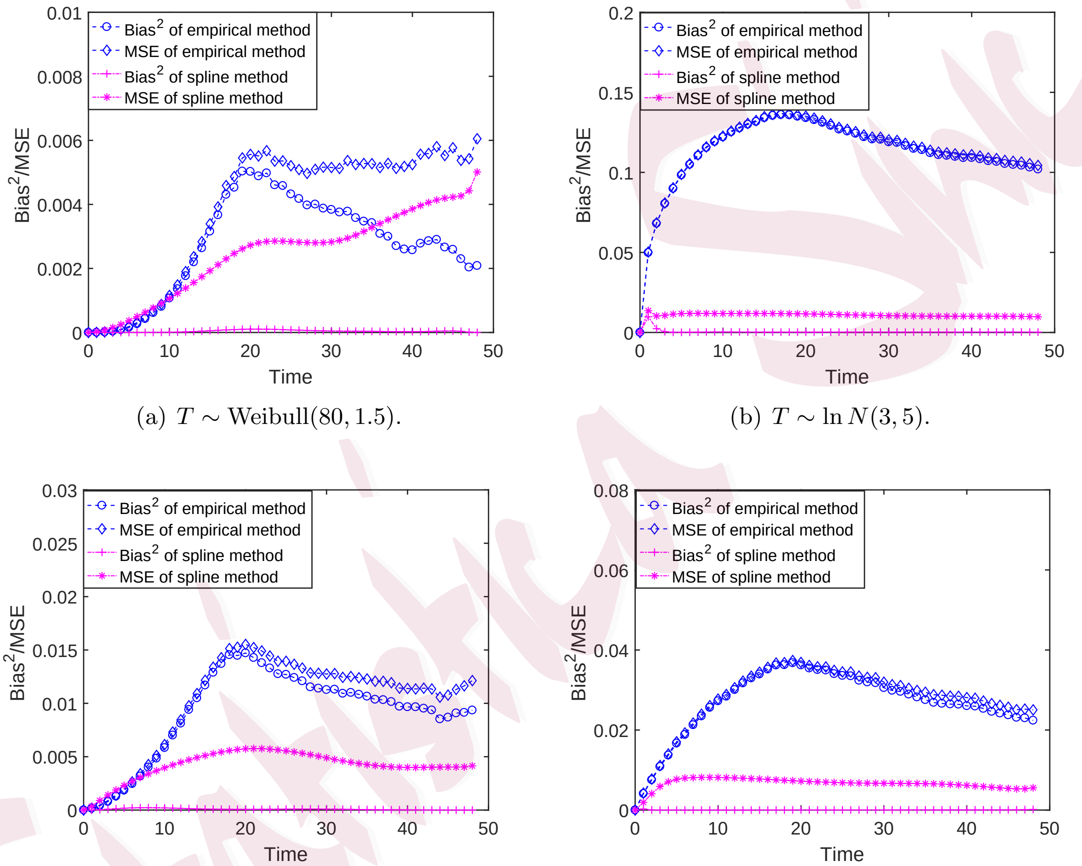
(a) $T \sim \text{Weibull}(80, 1.5)$.

(b) $T \sim \ln N(3, 5)$.

(c) $T \sim 0.5\text{Weibull}(80, 1.5) + 0.5\text{Weibull}(80, 0.8)$.

(d) $T \sim 0.8 \ln N(4.5, 3.5) + 0.2\text{Gamma}(2.5, 30)$.

Figure 3: Comparisons of the spline method and the NPMLE (Turnbull 1976) for estimating the reliability function when $n = 200$ and $p = 40\%$.



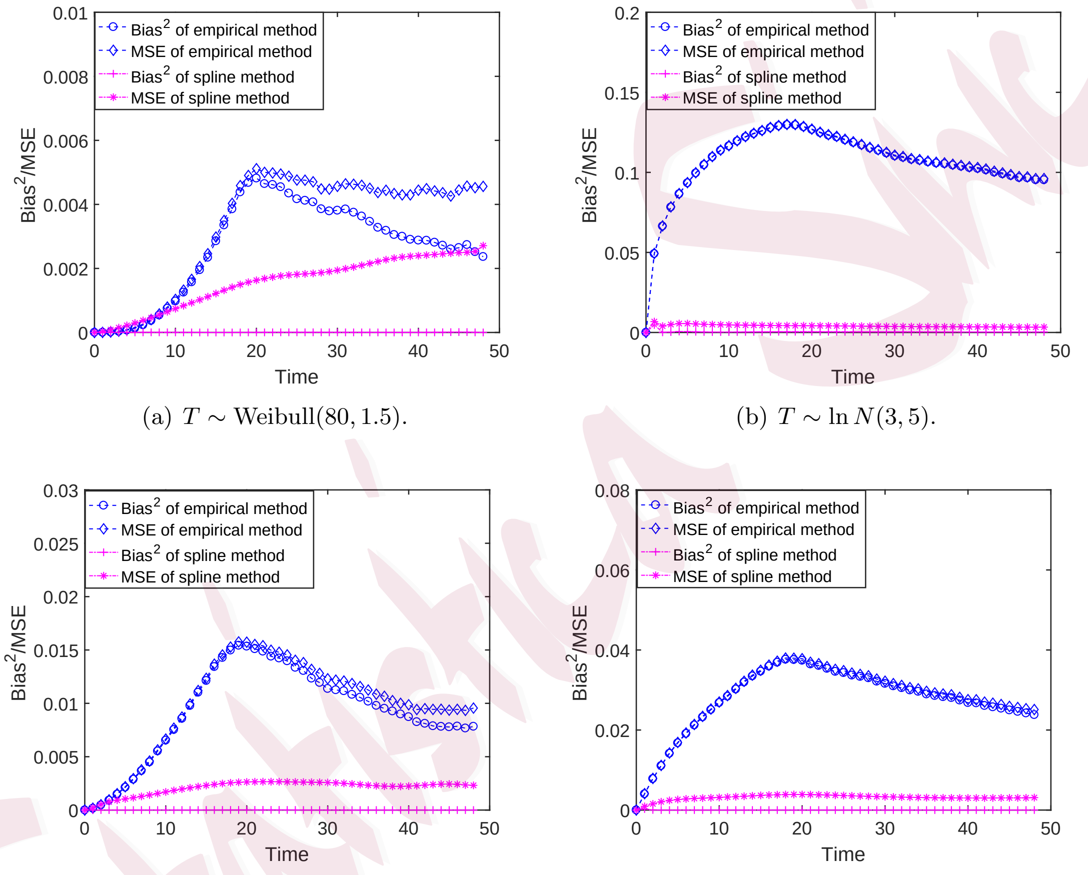
(a) $T \sim \text{Weibull}(80, 1.5)$.

(b) $T \sim \ln N(3, 5)$.

(c) $T \sim 0.5\text{Weibull}(80, 1.5) + 0.5\text{Weibull}(80, 0.8)$.

(d) $T \sim 0.8 \ln N(4.5, 3.5) + 0.2\text{Gamma}(2.5, 30)$.

Figure 4: Comparisons of the spline method and the NPMLE (Turnbull 1976) for estimating the reliability function when $n = 100$ and $p = 80\%$.



(a) $T \sim \text{Weibull}(80, 1.5)$.

(b) $T \sim \ln N(3, 5)$.

(c) $T \sim 0.5\text{Weibull}(80, 1.5) + 0.5\text{Weibull}(80, 0.8)$.

(d) $T \sim 0.8 \ln N(4.5, 3.5) + 0.2\text{Gamma}(2.5, 30)$.

Figure 5: Comparisons of the spline method and the NPMLE (Turnbull 1976) for estimating the reliability function when $n = 200$ and $p = 80\%$.

Table 1: Estimated sizes and powers of $T_n = U_n/\hat{\sigma}_w$ and T_H with Weibull Distribution (α, β), where the shape parameters $\beta_1 = \beta_2 = 1.5$, and the scale parameters $\alpha_1 = 30, \alpha_2 = 30, 40, 80$. Here, $W_n^{(k)}$ are the weight processes, $k = 1, 2, 3, 4$.

	T_n					T_H			
α_1/α_2	$W_n^{(1)}$	$W_n^{(2)}$	$W_n^{(3)}$	$W_n^{(4)}$	α_1/α_2	$W_n^{(1)}$	$W_n^{(2)}$	$W_n^{(3)}$	$W_n^{(4)}$
					$n_1 = n_2 = 100$				
30/30	0.058	0.052	0.052	0.056	30/30	0.065	0.062	0.062	0.064
30/40	0.697	0.583	0.585	0.650	30/40	0.705	0.657	0.657	0.683
30/80	1.000	0.981	0.981	0.997	30/80	1.000	1.000	1.000	1.000
					$n_1 = n_2 = 200$				
30/30	0.052	0.049	0.049	0.053	30/30	0.060	0.058	0.058	0.061
30/40	0.769	0.718	0.720	0.750	30/40	0.827	0.654	0.670	0.716
30/80	1.000	1.000	1.000	1.000	30/80	1.000	1.000	1.000	1.000

- Case 1. Two groups with the same shape parameter and different scale parameters.
- Case 2. Two groups with the same scale parameter and different shape parameters.

In Case 1, the two failure rates do not overlap. While the true failure rates intersect in Case 2. For each case, we consider two sample sizes, $n_1 = n_2 = 100$ and 200, respectively. As with Section 4, we choose the four weight processes

$$W_n^{(1)}(t) = 1, \quad W_n^{(2)}(t) = Z_n(t) = \frac{1}{n}\sum_{i=1}^{n} I(L_i < t \le Y_i),$$

$$W_n^{(3)}(t) = \frac{Z_{n_1}(t)Z_{n_2}(t)}{Z_n(t)}, \quad W_n^{(4)}(t) = 1 - Z_n(t),$$

where $Z_{n_k}(t)$ is defined as $Z_n(t)$ with the summation being only over subjects in group k. All the results reported here are based on 50000 Monte Carlo replications. Tables 1 and 2 present the estimated sizes and powers of the proposed test statistics T_n and the weighted Kaplan-Meier (WKM) statistics T_H (Shen 2007) at significance level $\alpha = 0.05$ for different cases and the four weight processes. As expected, the powers of all test statistics increase with the sample size. Under H_0, when the proportion of truncated observations is serve (40%), the proposed test T_n performs better than T_H. For Case 1, Table 1 shows good power properties of the proposed test T_n for the four weight processes. The proposed test with weight $W_n^{(1)}(t)$ has the best power performance. On the other hand, the powers heavily rely on the choices of the weight processes in Case 2, as can be seen from Table 2. The simulation results suggest that the proposed test T_n with $W_n^{(4)}(t)$ has the best power performance. The different performance of the four test statistics is due to the intersection between the two failure rate functions. The difference of the failure rate functions changes sign at the intersection point. If the weight W_n is approximately the same at the two sides of the intersection point, the value of U_n will be small, leading to poor powers of the test. The weight $W_n^{(4)}(t)$ puts unequal weight on the two sides, and thus it has the best performance. From the simulation, we would suggest using $W_n^{(4)}(t)$ for the test.

Table 2: Estimated sizes and powers of $T_n = U_n/\hat{\sigma}_w$ and T_H with Weibull Distribution (α, β), where the scale parameters $\alpha_1 = \alpha_2 = 30$, and the shape parameters $\beta_1 = 1.5, \beta_2 = 1.5, 1.2, 0.8, 0.5$. Here, $W_n^{(k)}$ are the weight processes, $k = 1, 2, 3, 4$.

	T_n					T_H			
β_1/β_2	$W_n^{(1)}$	$W_n^{(2)}$	$W_n^{(3)}$	$W_n^{(4)}$	β_1/β_2	$W_n^{(1)}$	$W_n^{(2)}$	$W_n^{(3)}$	$W_n^{(4)}$
				$n_1 = n_2 = 100$					
1.5/1.5	0.058	0.052	0.052	0.056	1.5/1.5	0.065	0.062	0.062	0.064
1.5/1.2	0.264	0.073	0.078	0.401	1.5/1.2	0.093	0.082	0.085	0.191
1.5/0.8	0.643	0.189	0.192	0.981	1.5/0.8	0.202	0.119	0.122	0.411
1.5/0.5	0.499	0.325	0.377	0.998	1.5/0.5	0.407	0.421	0.473	0.813
				$n_1 = n_2 = 200$					
1.5/1.5	0.052	0.049	0.049	0.053	1.5/1.5	0.060	0.058	0.058	0.061
1.5/1.2	0.399	0.080	0.085	0.627	1.5/1.2	0.265	0.111	0.092	0.380
1.5/0.8	0.870	0.201	0.214	0.999	1.5/0.8	0.394	0.157	0.168	0.553
1.5/0.5	0.668	0.477	0.565	1.000	1.5/0.5	0.637	0.558	0.563	0.895

6 A Real Example: Power Transformer Failure Data

The power transformer is one of the most important components in a power grid. Unexpected failures of transformers cause power shortage and lead to large economic losses. Therefore, it is important to know the failure behaviors of a transformer in the field. Such information can be extracted from field failure data of the transformers. Due to the long lifetime of a transformer and the late development of data recording systems, transformer lifetime data are left truncated and right censored. Figure 6 displays the data set "MC_Old65", which is recorded in operating time (Hong et al. 2009). The data set "MC_Old65" consists of 80 transformers and the installation dates of these transformers are recorded. The starting year of observation T_s is 1980 and the end of data collection date T_e is 2008. The earliest installation date $\underline{T}_I$ is 1950 and 69 transformers are installed before 1980. As a result, their lifetime observations are left-truncation observed. The proportion of truncated observations of tranformers is 86%. In the data set, 65 transformers continue to function after 2008 and the proportion of censored observed transformers is 81%.

We use the proposed spline method with $m_n = 5$ equally spaced inner knots, Turnbull's NPMLE and the Weibull distribution (Hong et al. 2009; Balakrishnan and Mitra 2012) to fit the data. Figure 7 presents the estimated reliability functions based on the three methods. We also tried $m_n = 4$ and 6 (no shown), and the estimated reliability function is almost the same as that with $m_n = 5$ here. Generally, the spline estimate and the empirical estimate agree quite well. The empirical estimate becomes constant when t is greater than the largest failure time, which is 42.1 in this example. By contrast, the spline method estimates the reliability function up to the largest observation time, which corresponds to a censoring time of 58. The wider range shows a greater flexibility of the spline method. Moreover, it is clear that spline-based estimator is more smooth. The comparisons of spline-based estimator and the Weibull estimator show that spline-based method can be used to assess the goodness-of-fit of a parametric model. To quality the uncertainty in the spline estimates, the random weighted bootstrap procedure (Hong et al. 2009) with 50,000 resamples is used to construct pointwise 95% confidence band

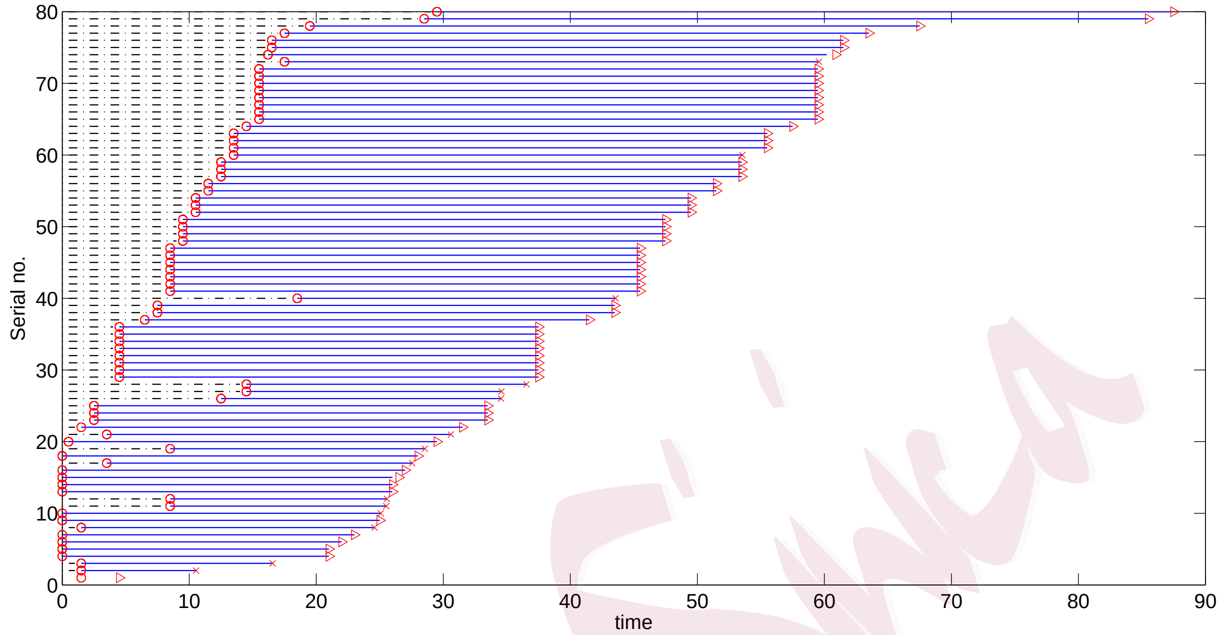


Figure 6: Service-time event plot of a subset of the transformer lifetime data: "o" represents the install time, "×" represents the failure time, "▷" represents the censored time.

of the reliability function, as shown in Figure 8.

Hong et al. (2009) also collected failure times of transformers which are from the same manufacturer as "MC_Old65" but different generations. We choose the dataset "MC_Old55" as the second group. The difference between the two groups is the type of insulation. The problem of interest here is to compare the two groups and check whether data from the two groups can be merged. The test statistics (3) developed in Section 4 are used for the comparison. We obtain $T_n = 7.643$, 10.392, 6.028 and 4.653 with $W_n(t) = W_n^{(k)}(t)$, $k = 1, 2, 3, 4$, defined in Section 5. All the values correspond to p-values $\ll 0.0001$. The proposed tests suggest that the two groups are significantly different. Therefore, the effect of insulation type can not be ignored, and the two data sets "MC_Old65" and "MC_Old55" can not be combined.

Acknowledgement

We are grateful to the Editor, an associate editor, and two referees for their insightful comments that have lead to a substantial improvement of an earlier version of the paper. Jiang and Ye's research is supported by the Natural Science Foundation of China (71601138) and Singapore AcRF Tier 1 funding (R-266-000-095-112). Zhao's research is supported in part by the National Natural Science Foundation of China (11771366) and The Hong Kong Polytechnic University.

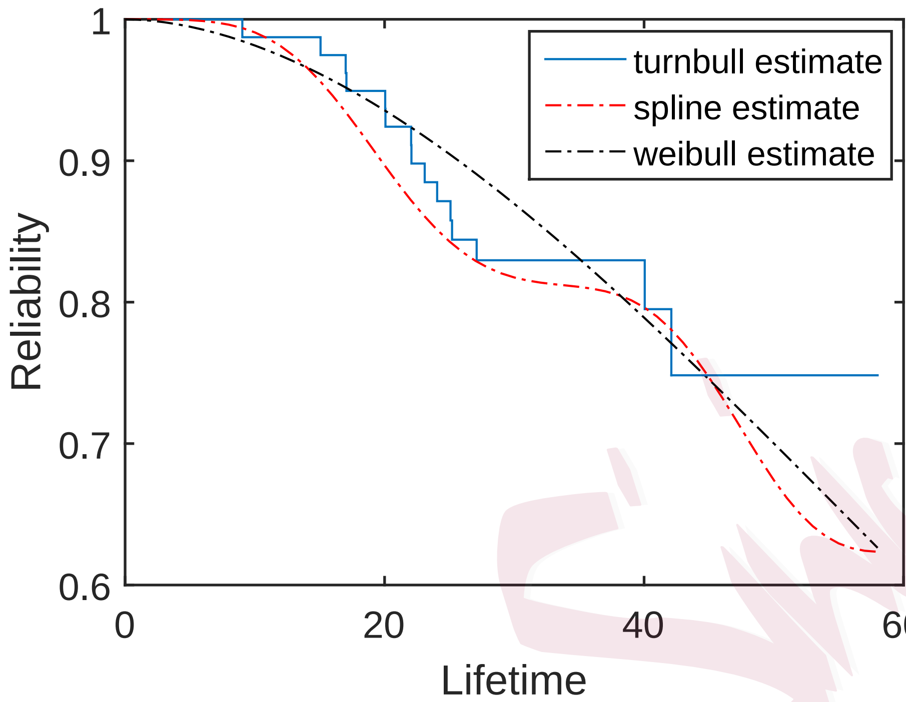


Figure 7: Comparisons of different estimators for the reliability function based on "MC_Old65" data (Hong et al. 2009): The stair line is the Trunbull estimate, the dash line is for Weibull and the chain line is the spline estimate.

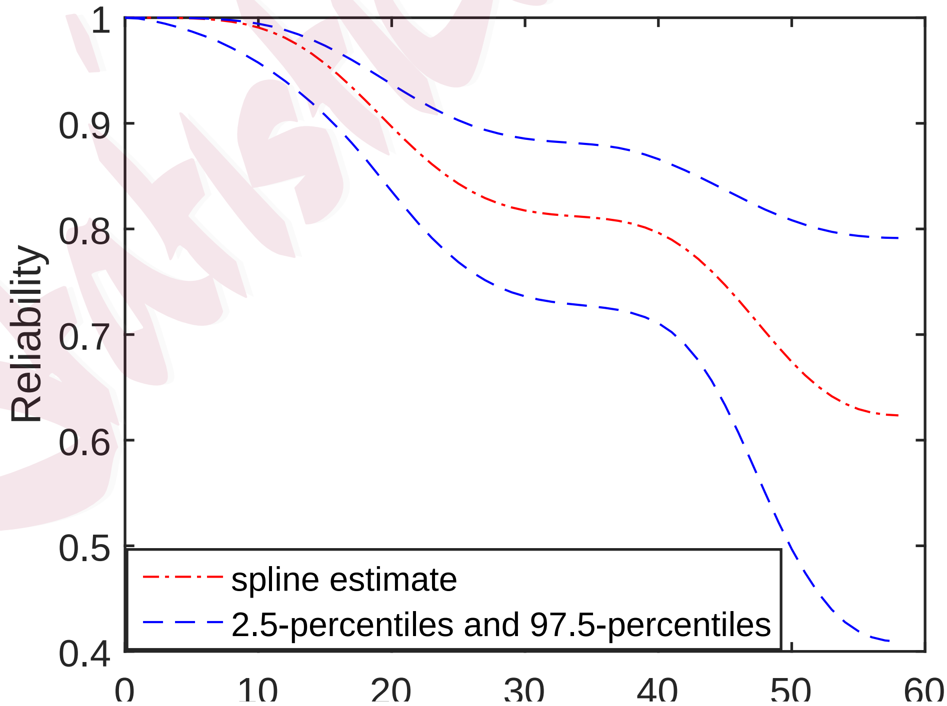


Figure 8: Spline estimates of the reliability function based on 'MC_Old65" (Hong et al. 2009), and the pointwise 95% two-sided confidence band based on 50,000 simulations.

Appendix

Proof of Theorem 1 (Consistency)

The log-likelihood function for λ is

$$\mathcal{L}(\lambda|\mathbf{D}) = \sum_{i=1}^{n} \{\delta_i \ln \lambda(Y_i) - [\Lambda(Y_i) - \Lambda(L_i)]\} .$$

With the knot sequence $\mathcal{T} = \{t_j\}_1^{m_n+2l}$ specified in Section 2, there exists a spline $\lambda_n(t) \in \psi_{l,\mathcal{T}}$ with order $l \geq r+2$ such that $\|\lambda_n(t) - \lambda_0\|_\infty = \sup_{t\in[a,b]} |\lambda_n(t) - \lambda_0(t)| = O(n^{-\nu r})$, according to Corollary 6.21 of Schumaker (2007, page 227). Choose a positive function $h_n \in \psi_{l,\mathcal{T}}$ such that $\|h_n\|_2^2 = O(n^{-2\nu r}+n^{-(1-\nu)/2})$. Therefore, for any $\alpha > 0$, $\|\lambda_n-\lambda_0+\alpha h_n\|_2^2 = O(n^{-2\nu r}+n^{-(1-\nu)/2})$ for sufficiently large n.

Denote $n\mathbb{M}_n(\lambda) = \mathcal{L}(\lambda|\mathbf{D})$ and $H_n(\alpha) = \mathbb{M}_n(\lambda_n + \alpha h_n)$. The first and second derivatives of H_n are

$$\begin{aligned} H_n'(\alpha) &= n^{-1}\sum_{i=1}^{n}\left\{\frac{\delta_i h_n(Y_i)}{\lambda_n(Y_i) + \alpha h_n(Y_i)} - \int_{L_i}^{Y_i} h_n(x)dx\right\}, \\ H_n''(\alpha) &= -n^{-1}\sum_{i=1}^{n}\frac{\delta_i h_n^2(Y_i)}{[\lambda_n(Y_i) + \alpha h_n(Y_i)]^2} < 0. \end{aligned}$$

Thus $H_n'(\alpha)$ is a non-increasing function. Therefore, to prove Theorem 1, it is sufficient to show that, for any $\alpha_0 > 0$, $H_n'(\alpha_0) < 0$ and $H_n'(-\alpha_0) > 0$ except on an event with probability converging to zero. Then $\hat{\lambda}_n$ must be between $\lambda_n - \alpha_0 h_n$ and $\lambda_n + \alpha_0 h_n$ with probability converging to one, so that $P(\|\hat{\lambda}_n - \lambda_n\|_2 \leq \alpha_0\|h_n\|_2) \to 1$ as $n \to \infty$. We first show $H_n'(\alpha_0) < 0$. Express $H_n'(\alpha_0)$ as

$$H_n'(\alpha_0) = \underbrace{(\mathbb{P}_n - P)\left\{\frac{\delta h_n(Y)}{\lambda_n(Y) + \alpha_0 h_n(Y)} - \int_L^Y h_n(x)dx\right\}}_{I_{n1}} + \underbrace{P\left\{\frac{\delta h_n(Y)}{\lambda_n(Y) + \alpha_0 h_n(Y)} - \int_L^Y h_n(x)dx\right\}}_{I_{n2}}$$

Given $\eta > 0$, define the class $\mathcal{F}_{\eta,n} = \{\lambda :\ \lambda \in \psi_{l,\mathcal{T}},\ d(\lambda, \lambda_n) \leq \eta\}$. According to Condition 1, there exists a positive integer N such that when $n > N$,

$$d(\lambda, \lambda_0) \leq d(\lambda, \lambda_n) + d(\lambda_n, \lambda_0) \leq \eta + O(N^{-\mu r}) < 2\eta,$$

where $\lambda \in \mathcal{F}_{\eta,n}, n > N$. Let $\mathcal{F}_\eta = \cup_{n\geq N}\mathcal{F}_{\eta,n}$. Corollary 6.21 of Schumaker (2007, page 227) shows that for any function $\lambda \in \mathcal{F}_\eta$, λ has uniformly bounded derivatives up to order $l-1$. Then according to Corollary 2.7.4 of van der Vaart and Wellner (1996, page 158), we can find that given ε such that $0 < \varepsilon \leq \eta$, $\mathcal{F}_\eta$ can be covered by a set of ε-brackets $\{[\underline{\lambda}_k, \bar{\lambda}_k] : k = 1, 2, \cdots, (1/\varepsilon)^{\frac{c_0}{l}}\}$, where c_0 is a constant depending on l. For any $\lambda \in \mathcal{F}_\eta$, there exists a bracket $[\underline{\lambda}_k, \bar{\lambda}_k]$, such that $\underline{\lambda}_k(t) \leq \lambda(t) \leq \bar{\lambda}_k(t)$ for all $t \in [a, b]$, where $d^2(\underline{\lambda}_k, \bar{\lambda}_k) = \int |\underline{\lambda}_k - \bar{\lambda}_k|^2 dF^*(t) \leq \varepsilon^2$, $k = 1, 2, \cdots, (1/\varepsilon)^{\frac{c_0}{l}}$. Then we have

$$d(\underline{\lambda}_k, \lambda_0) \leq d(\underline{\lambda}_k, \lambda) + d(\lambda, \lambda_n) + d(\lambda_n, \lambda_0) < \varepsilon + 2\eta,$$

where $\lambda \in \mathcal{F}_\eta, n > N$. Then by the converse of Lemma 7.1 from Wellner and Zhang (2007, Page 2140), we get $\sup_{t\in[a,b]} |\underline{\lambda}_k - \lambda_0| \leq c_1(\varepsilon + 2\eta)^{2/3}$, c_1 is constant. Since λ_0 is positive and bounded on $[a, b]$, there exists $c_2 > 0$ such that $\underline{\lambda}_k > c_2 > 0$. Similarly , there exists a positive c_3 such that $\bar{\lambda}_k > c_3 > 0$. That means $\underline{\lambda}_k$ and $\bar{\lambda}_k$ have the positive lower bounds.

Define the class sequence

$$\mathfrak{F}_{\eta,n} = \left\{ \frac{\delta}{\lambda}(\lambda - \lambda_n) - \int_L^Y (\lambda - \lambda_n)dx : \lambda \in \mathcal{F}_{\eta,n} \right\},$$

and let

$$\mathfrak{F}_\eta = \cup_{n>N} \mathfrak{F}_{\eta,n} = \left\{ \frac{\delta}{\lambda}(\lambda - \lambda_n) - \int_L^Y (\lambda - \lambda_n)dx : \lambda \in \mathcal{F}_\eta \right\}.$$

Then let

$$\underline{m}_k(X) = \delta - \frac{\delta}{\underline{\lambda}_k}\lambda_n - \int_L^Y (\bar{\lambda}_k - \lambda_n)dx$$

and

$$\bar{m}_k(X) = \delta - \frac{\delta}{\bar{\lambda}_k}\lambda_n - \int_L^Y (\underline{\lambda}_k - \lambda_n)dx.$$

Clearly, the class $\mathfrak{F}_\eta$ is covered by the set $[\underline{m}_k, \bar{m}_k]$, $k = 1, 2, \cdots, (1/\varepsilon)^{\frac{c_0}{l}}$. Therefore, to prove that the uniformly bounded class $\mathfrak{F}_\eta$ is a Donsker class, we need to show $P(\bar{m}_k - \underline{m}_k)^2 \lesssim \varepsilon^2$. Let

$$f = \bar{m}_k - \underline{m}_k = \frac{\delta\lambda_n}{\underline{\lambda}_k\bar{\lambda}_k}(\bar{\lambda}_k - \underline{\lambda}_k) + \int_L^Y (\bar{\lambda}_k - \underline{\lambda}_k)dx.$$

By the Cauchy-Schawartz inequality,

$$Pf^2 \lesssim P(\bar{\lambda}_k - \underline{\lambda}_k)^2 \lesssim d^2(\bar{\lambda}_k, \underline{\lambda}_k) \leq \varepsilon^2.$$

It is followed that

$$\sup_{f\in\mathfrak{F}_\eta} \rho_P(f) = \sup_{f\in\mathfrak{F}_\eta} \{P(f - Pf)^2\}^{1/2} \leq \sup_{f\in\mathfrak{F}_\eta} \{Pf^2\}^{1/2} \lesssim \varepsilon \to 0,$$

as $\varepsilon \to 0$. Due to the relationship between P-Donsker and asymptotic equicontinuity (Corollary 2.3.12 van der Vaart and Wellner 1996, page 115), we can show that $\mathfrak{F}_\eta$ is a Donsker class. And this is equal to the following fact : $(\mathfrak{F}_\eta, \rho_P)$ is totally bounded and $E\sqrt{n}\|\mathbb{P}_n - P\|_{\mathfrak{F}_\eta} \to 0$ for $\varepsilon \to 0$. Hence, $I_{n1} = O_p(n^{-1/2})$.

For the second term, since

$$P\left\{ \frac{\delta h_n(Y)}{\lambda_0(Y)} - \int_L^Y h_n(x)dx \right\} = 0,$$

we have, by adding and subtracting terms,

$$I_{n2} = P\left\{ \frac{\delta h_n(Y)}{\lambda_n(Y) + \alpha_0 h_n(Y)} - \frac{\delta h_n(Y)}{\lambda_0(Y)} \right\}.$$

Define $m(s) = \frac{1}{\lambda_0 + s\Delta}$, where $\Delta = \lambda_n - \lambda_0 + \alpha_0 h_n$, $0 \leq s \leq 1$. By the Taylor expansion, there exists $\theta \in (0,1)$ such that

$$m(s) = m(0) + m'(\theta)s = \frac{1}{\lambda_0} + \left(-\frac{\Delta}{(\lambda_0+\theta\Delta)^2}\right)s.$$

Therefore,

$$I_{n2} \le E\left\{-\frac{\Delta}{(\lambda_0+\theta\Delta)^2}\right\}h_n \lesssim -Eh_n^2 = O(n^{-2\nu r} + n^{-(1-\nu)/2}).$$

Since $n^{-2vr} + n^{-(1-v)/2} > n^{-1/2}$, we have

$$H_n'(\alpha_0) \le O_p(n^{-1/2}) - O(n^{-2\nu r} + n^{-(1-\nu)/2}) < 0,$$

except on an event with probability converging to zero. The same arguments show that $H_n'(-\alpha_0) > 0$ with probability converging to 1.

Proof of Theorem 2 (Rate of Convergence)

Denote $m_\lambda(X) = \delta \ln \lambda(Y) - \int_L^Y \lambda(u)du$ and define $M(\lambda) = Pm_\lambda(X)$ and $\mathbb{M}_n(\lambda) = \mathbb{P}_n m_\lambda(X)$. Then the log-likelihood function can be written as $n\mathbb{P}_n m_\lambda(X)$. Given $\eta > 0$, define the class

$$\mathcal{F}_\eta = \{\lambda | \lambda \in \psi_{l,\mathcal{T}}, d(\lambda,\lambda_0) \le \eta\}.$$

By the result of Theorem 1, $\hat\lambda_n \in \mathcal{F}_\eta$ for sufficiently large n. For $\eta > 0$ and any $\varepsilon < \eta$,

$$\log N_{[\,]}\{\varepsilon, \psi_{l,\mathcal{T}}, L_2(P)\} \le cq_n \log(\eta/\varepsilon),\ \ J_{[\,]}\{\eta, \psi_{l,\mathcal{T}}, L_2(P)\} \le c_0 q_n^{\frac{1}{2}}\eta,$$

where $q_n = m_n + l$ is the number of spline base functions, c and c_0 are constants (Shen and Wong 1994, Page 597). Therefore for each $\lambda \in \mathcal{F}_\eta$, there exists a bracket $[\underline{\lambda}_k, \bar\lambda_k]$, such that

$$\underline{\lambda}_k(t) \le \lambda(t) \le \bar\lambda_k(t)$$

for all $t \in [a,b]$, where $d^2(\underline{\lambda}_k, \bar\lambda_k) = \int |\underline{\lambda}_k - \bar\lambda_k|^2 dF^*(t) \le \varepsilon^2$, $k = 1, 2, \cdots, (\eta/\varepsilon)^{cq_n}$. Moreover, $\underline{\lambda}_k$ and $\bar\lambda_k$ are bounded on $[a,b]$ and have positive lower bounds.

Since λ_0 is the maximum of $M(\lambda)$, the first derivative is zero at λ_0 and the second derivative is negative definite. According to the Taylor expansion,

$$M(\lambda) = M(\lambda_0) + 0 + \frac{M''(\lambda_0)}{2}(\lambda-\lambda_0)^2 + o(\lambda-\lambda_0)^2.$$

Thus, for $\lambda \in \mathcal{F}_\eta$, $M(\lambda_0) - M(\lambda) \gtrsim d^2(\lambda, \lambda_0)$. Next, define the class

$$\mathcal{M}_\eta = \{m_\lambda(x) - m_{\lambda_0}(x) : \lambda \in \mathcal{F}_\eta\}.$$

Let

$$\underline{\mathfrak{m}}_k(X) = \delta \ln \underline{\lambda}_k(Y) - \int_L^Y \bar\lambda_k(x)dx - m_{\lambda_0}(x)$$

and

$$\bar{\mathfrak{m}}_k(X) = \delta \ln \bar{\lambda}_k(Y) - \int_L^Y \underline{\lambda}_k(x)dx - m_{\lambda_0}(x).$$

Clearly, the class $\mathcal{M}_\eta$ is covered by the set $[\underline{\mathfrak{m}}_k, \bar{\mathfrak{m}}_k]$, $k = 1, 2, \cdots, (\eta/\varepsilon)^{cq_n}$. To prove the uniformly bounded class $\mathcal{M}_\eta$ is a Donsker class, we need to show that (equicontinuity condition)

$$\|\bar{\mathfrak{m}}_k - \underline{\mathfrak{m}}_k\|_2^2 \lesssim \varepsilon^2.$$

Let

$$f = \bar{\mathfrak{m}}_k - \underline{\mathfrak{m}}_k = \underbrace{\delta\left(\log \bar{\lambda}_k - \underline{\lambda}_k\right)}_{I_1} + \underbrace{\int_L^Y \left(\bar{\lambda}_k - \underline{\lambda}_k\right) dx}_{I_2}.$$

Since $\sup_{t\in[a,b]} |\underline{\lambda}_k - \lambda_0| \le \varepsilon_1$ and $\sup_{t\in[a,b]} |\bar{\lambda}_k - \lambda_0| \le \varepsilon_2$ by converse theorem of Lemma 7.1 from Wellner and Zhang (2007, Page 2140), the boundedness of Y, L, δ and λ_0 yields the boundedness of I_1 and I_2 on $[a, b]$. Then according to Cauchy Schwartz inequality,

$$|f|^2 \le 2\left\{\delta^2 \left(\log \frac{\bar{\lambda}_k}{\underline{\lambda}_k}\right)^2 + \left[\int_L^Y (\bar{\lambda}_k - \underline{\lambda}_k)dx\right]^2\right\}.$$

By taylor expansion,

$$\log \frac{\bar{\lambda}_k}{\underline{\lambda}_k} = \frac{1}{\theta}\left(\bar{\lambda}_k - \underline{\lambda}_k\right),$$

where θ between $\bar{\lambda}_k$ and $\underline{\lambda}_k$. Since $\bar{\lambda}_k$ and $\underline{\lambda}_k$ are bounded functions on $[a, b]$, there exists a constant c_1 such that

$$\log \frac{\bar{\lambda}_k}{\underline{\lambda}_k} < c_1\left(\bar{\lambda}_k - \underline{\lambda}_k\right).$$

Therefore

$$P(|f|^2) \lesssim d^2(\bar{\lambda}_k, \underline{\lambda}_k) \le \varepsilon^2.$$

Then according to the Lemma 3.4.2 of van der Vaart and Wellner (1996, Page 324), we obtain

$$E_P\|n^{1/2}(\mathbb{P} - P)\|_{\mathcal{M}_\eta} \lesssim J_{[\,]}(\eta, \mathcal{M}_\eta, L_2(P))\left\{1 + \frac{J_{[\,]}(\eta, \mathcal{M}_\eta, L_2(P))}{\eta^2 n^{1/2}}\right\}. \quad (4)$$

The right-hand side of (4) yields $\phi_n(\eta) = c_2(q_n^{(1/2)}\eta + q_n/n^{1/2})$. It is easy to see that $\phi(\eta)/\eta$ is decreasing in η, and

$$r_n^2\phi(\frac{1}{r_n}) = r_n q_n^{1/2} + r_n^2 q_n/n^{1/2} \le n^{1/2}$$

yields $r_n = n^{\frac{1-\nu}{2}}$, where $0 < \nu < 1/2$. Hence, $n^{\frac{1-\nu}{2}} d(\hat{\lambda}_n, \lambda_0) = O_p(1)$ by Theorem 3.4.1 of van der Vaart and Wellner (1996, Page 322). If $\nu = \frac{r}{2r+1}$, the rate of convergence of $\hat{\lambda}_n$ is $\frac{r}{2r+1}$, which is the same as the optimal rate in nonparametric regression.

proof of Theorem 3 (Asymptotic normality)

According to Theorem 1 of Zhao and Zhang (2017, page 933), we need the following conditions to establish the asymptotic normality.

- A1. $\sqrt{n}(\mathbb{P}_n - P)(\phi(\hat{\lambda}_n; X)[h] - \phi(\lambda_0; X)[h]) = o_p(1)$.
- A2. $\sqrt{n}(G_n - G)(\lambda_0)[h]$ convergences in distribution to a tight Gaussian process on $l^\infty(\mathcal{H}_r)$.
- A3. $G(\lambda_0)[h] = 0$ and $G_n(\hat{\lambda}_n)[h] = o_p(n^{-1/2})$.
- A4. $G(\lambda)[h]$ is the Fréchet-differentiable at λ_0 with a continuous derivative, denoted by $\dot{G}_{\lambda_0}[h]$.
- A5. $G(\hat{\lambda}_n)[h] - G(\lambda_0)[h] - \dot{G}_{\lambda_0}(\hat{\lambda}_n - \lambda_0)[h] = o_p(n^{-1/2})$, where $\dot{G}_{\lambda_0}(\hat{\lambda}_n - \lambda_0)[h]$ is the directional derivative at λ_0 in the direction $(\lambda - \lambda_0)$.

Then we need to verify conditions A1-A5 above.

For A1, given $\varepsilon > 0$, define the class

$$\mathcal{G}_n(\varepsilon)[h] = \{\phi(\lambda; X)[h] - \phi(\lambda_0; X)[h] : \lambda \in \psi_{l,\mathcal{T}} \text{ such that } d(\lambda, \lambda_0) \leq \varepsilon, h \in \mathcal{H}_r\}.$$

Let

$$\underline{g}_k(X) = P\left[\frac{\delta}{\bar{\lambda}_k}h - \frac{\delta}{\lambda_0}h\right]$$

and

$$\bar{g}_k(X) = P\left[\frac{\delta}{\underline{\lambda}_k}h - \frac{\delta}{\lambda_0}h\right],$$

where $\underline{\lambda}_k$ and $\bar{\lambda}_k$ are similar defined in the proof of Theorem 2. Clearly, the class $\mathcal{G}_n(\varepsilon)[h]$ is covered by the set $[\underline{g}_k, \bar{g}_k]$, $k = 1, 2, \cdots, (\eta/\varepsilon)^{cq_n}$. Let

$$f = \bar{g}_k - \underline{g}_k = P\left[\frac{\delta}{\underline{\lambda}_k}h - \frac{\delta}{\bar{\lambda}_k}h\right] = P\left[\delta h(Y)\frac{\bar{\lambda}_k - \underline{\lambda}_k}{\bar{\lambda}_k\underline{\lambda}_k}\right].$$

By the Cauchy-Schawartz inequality,

$$Pf^2 = P\left[\delta h(Y)\frac{\bar{\lambda}_k - \underline{\lambda}_k}{\bar{\lambda}_k\underline{\lambda}_k}\right]^2 \lesssim P\left[\frac{1}{\bar{\lambda}_k\underline{\lambda}_k}\right]^2 \left(\bar{\lambda}_k - \underline{\lambda}_k\right)^2,$$

where the last inequality bolds due to $h \in \mathcal{H}_r$. Due to the result of Theorem 1, we can find $\lambda \in \psi_{l,\mathcal{T}}$ such that $d(\lambda, \lambda_0) \leq \varepsilon$. Therefore, $d(\underline{\lambda}_k, \lambda_0) \leq d(\underline{\lambda}_k, \lambda) + d(\lambda, \lambda_0) < 2\varepsilon$. Then by converse theorem of Lemma 7.1 from Wellner and Zhang (2007, Page 2140), we get $\sup_{t\in[a,b]} |\underline{\lambda}_k - \lambda_0| \leq c_1\varepsilon^{2/3}$, c_1 is constant. Since λ_0 is positive and bounded on $[a, b]$, there exists a constant $c_2 > 0$ such that $\underline{\lambda}_k > c_2 > 0$. Similarly as $\bar{\lambda}_k$. So $\bar{\lambda}_k$ and $\underline{\lambda}_k$ have the positive lower bounds. Furthermore, using the fact that $\bar{\lambda}_k$and $\underline{\lambda}_k$ have the positive lower bounds, we have

$$Pf^2 \lesssim P\left(\bar{\lambda}_k - \underline{\lambda}_k\right)^2 \lesssim d^2\left(\bar{\lambda}_k - \underline{\lambda}_k\right) \leq \varepsilon^2.$$

Then according to the Lemma 3.4.2 of van der Vaart and Wellner (1996, Page 324), we obtain

$$E_P\|n^{1/2}(\mathbb{P} - P)\|_{\mathcal{G}_n(\varepsilon)[h]} \lesssim J_{[\,]}(\varepsilon, \mathcal{G}_n(\varepsilon)[h], L_2(P))\left\{1 + \frac{J_{[\,]}(\varepsilon, \mathcal{G}_n(\varepsilon)[h], L_2(P))}{\varepsilon^2 n^{1/2}}\right\}. \quad (5)$$

Theorem 1 shows that $d(\hat{\lambda}_n, \lambda_0) \to 0$ almost surely. Hence that by converse theorem of Lemma

7.1 from Wellner and Zhang (2007, Page 2140), we have

$$\sup_{t\in[a,b]} |\hat{\lambda}_n(t) - \lambda_0(t)| \to 0 \text{ almost surely.}$$

Moreover, Theorem 2 shows that $n^{\frac{r}{2r+1}}\|\hat{\lambda}_n - \lambda_0\|_2 = O_p(1)$ with $r > 1$. Therefore we have $\phi(\hat{\lambda}_n; X)[h] - \phi(\lambda_0; X)[h] \in \mathcal{G}_n(\varepsilon_n)[h]$ with $\varepsilon_n = O(n^{-r/(1+2r)})$. Moreover, for any $\phi(\lambda; X)[h] - \phi(\lambda_0; X)[h] \in \mathcal{G}_n(\varepsilon_n)[h]$, exists $M > 0$, such that

$$P(\phi(\lambda; X)[h] - \phi(\lambda_0; X)[h])^2 \lesssim \varepsilon_n^2 \text{ and } \sup_{h\in\mathcal{H}_r} |\phi(\lambda; X)[h] - \phi(\lambda_0; X)[h]| < M.$$

Hence, we have

$$\begin{aligned}
E_P\|n^{1/2}(\mathbb{P} - P)\|_{\mathcal{G}_n(\varepsilon_n)[h]} &\lesssim J_{[\,]}(\varepsilon_n, \mathcal{G}_n(\varepsilon_n)[h], L_2(P)) \left\{1 + \frac{J_{[\,]}(\varepsilon_n, \mathcal{G}_n(\varepsilon_n)[h], L_2(P))}{{\varepsilon_n}^2 n^{1/2}}\right\} \\
&\lesssim q_n^{1/2}\varepsilon_n + q_n n^{-1/2} \\
&= O(n^{1/2(1+2r)-r/(1+2r)}) + O(n^{1/(1+2r)-1/2}) \\
&= o(1).
\end{aligned}$$

Therefore, we have

$$\sqrt{n}(\mathbb{P}_n - P)(\phi(\hat{\lambda}_n; X)[h] - \phi(\lambda_0; X)[h]) = o_p(1)$$

uniformly in h.

For A2, since $\mathcal{H}_r$ is a Donsker class and the function $\phi(\lambda_0; X)[h]$ is a bounded Lipschitz function with respect to $\mathcal{H}_r$, we have the class $\{\phi(\lambda_0; X)[h] : h \in \mathcal{H}_r\}$ is Donsker (Theorem 2.10.6 van der Vaart and Wellner 1996, Page 192). Then based on Theorem 3.10.12 (van der Vaart and Wellner 1996, Page 407), $\sqrt{n}(G_n - G)(\lambda_0)[h]$ convergences in distribution to a tight Gaussian process on $l^\infty(\mathcal{H}_r)$.

To prove the third part A3, clearly $G(\lambda_0)[h] = 0$. Note that $\hat{\lambda}_n = \sum_{j=1}^{q_n} \hat{\alpha}_j B_j(t)$ satisfies the following score function

$$n^{-1}\sum_{i=1}^n \left\{\frac{\delta_i B_j(Y_i)}{\hat{\lambda}_n(Y_i)} - \int_{L_i}^{Y_i} B_j(x)dx\right\} = 0,\ j = 1, \cdots, q_n.$$

Thus, for any $h_n = \sum_{j=1}^{q_n} \alpha_j B_j \in \varphi_{l,\mathcal{T}}$, we have

$$n^{-1}\sum_{i=1}^n \left\{\frac{\delta_i h_n(Y_i)}{\hat{\lambda}_n(Y_i)} - \int_{L_i}^{Y_i} h_n(x)dx\right\} = 0,$$

that is, $G_n(\hat{\lambda}_n)[h_n] = 0$ for any $h_n \in \varphi_{l,\mathcal{T}}$. Moreover, for any $h \in \mathcal{H}_r$, there exists $h_n \in \varphi_{l,\mathcal{T}}$

such that $\|h - h_n\|_\infty = O(n^{-r\nu})$. Therefore, we have

$$\begin{aligned}
G_n(\hat{\lambda}_n)[h] = G_n(\hat{\lambda}_n)[h-h_n] &= \left\{G_n(\hat{\lambda}_n)[h-h_n] - G_n(\lambda_0)[h-h_n]\right\} + G_n(\lambda_0)[h-h_n] - G_0(\lambda_0)[h-h_n] \\
&= n^{-1}\sum_{i=1}^n \delta_i \left\{\frac{1}{\hat{\lambda}_n(Y_i)} - \frac{1}{\lambda_0(Y_i)}\right\}[h(Y_i) - h_n(Y_i)] + (G_n - G)(\lambda_0)[h-h_n] \\
&\lesssim d(\hat{\lambda}_n, \lambda_0)\|h - h_n\|_\infty + o_p(n^{-1/2}) \\
&= o_p(n^{-1/2}),
\end{aligned}$$

where the proof of A2 leads to that $(G_n - G)(\lambda_0)[h-h_n]$ convergences in distribution to a tight Gaussian process.

For A4, by the assumption of smoothness, $G(\lambda)[h]$ is the Fréchet-differentiable at λ_0 with a continuous derivative, denoted by $\dot{G}_{\lambda_0}[h]$. Moreover, the directional derivative $\dot{G}_{\lambda_0}(\hat{\lambda}_n - \lambda_0)[h]$ at λ_0 in the direction $(\lambda - \lambda_0)$ can be defined as

$$\begin{aligned}
\dot{G}_{\lambda_0}(\hat{\lambda}_n - \lambda_0)[h] = \triangledown_{(\lambda-\lambda_0)}G(\lambda_0)[h] &= \lim_{\varepsilon\to 0}\frac{G(\lambda_0 + \varepsilon(\lambda-\lambda_0))[h] - G(\lambda_0)[h]}{\varepsilon} \\
&= -P\left[\delta h(Y)\frac{\lambda(Y) - \lambda_0(Y)}{\lambda_0^2(Y)}\right] \\
&= -\int \frac{h(t)}{\lambda_0^2(t)}(\lambda(t) - \lambda_0(t))dF^*(t),
\end{aligned}$$

where $F^*(t) = P(L \le T \le C, T \le t)$.

Then for A5, we can prove

$$\begin{aligned}
G(\hat{\lambda}_n)[h] - G(\lambda_0)[h] - \dot{G}_{\lambda_0}(\hat{\lambda}_n - \lambda_0)[h] &= P\left[\delta h(Y)\left(\frac{1}{\hat{\lambda}_n(Y)} - \frac{1}{\lambda_0(Y)}\right)\right] + P\left[\delta h(Y)\frac{\hat{\lambda}_n(Y) - \lambda_0(Y)}{\lambda_0^2(Y)}\right] \\
&= P\left[\frac{\delta h(Y)}{\hat{\lambda}_n(Y)\lambda_0^2(Y)}\left\{\hat{\lambda}_n(Y) - \lambda_0(Y)\right\}^2\right] \\
&= O_p(d^2(\hat{\lambda}_n, \lambda_0)) = O_p(n^{-2r/(1+2r)}) = o_p(n^{-1/2}).
\end{aligned}$$

Thus it follows from Theorem 1(Zhao and Zhang 2017, Page 934) that

$$\sqrt{n}\int \frac{h(t)}{\lambda_0^2(t)}(\lambda(t) - \lambda_0(t))dF^*(t) = -\sqrt{n}\dot{G}_{\lambda_0}(\hat{\lambda}_n - \lambda_0)[h] = \sqrt{n}(G_n - G)(\lambda_0)[h] + o_p(1).$$

Proof of Theorem 4

We first note that U_n can be rewritten as

$$\begin{aligned}
U_n &= \frac{\sqrt{n}}{\sum_{i=1}^n \delta_i}\sum_{i=1}^n \delta_i W_n(Y_i)\left\{\hat{\lambda}_n^{(1)}(Y_i) - \hat{\lambda}_n^{(2)}(Y_i)\right\} = \sqrt{n}\mathbb{P}_n\left[W_n^{(k)}(Y)\left\{\hat{\lambda}_n^{(1)}(Y) - \hat{\lambda}_n^{(2)}(Y)\right\}\right] \\
&= \sqrt{n}\mathbb{P}_n\left[W_n^{(k)}(Y)\left\{\hat{\lambda}_n^{(1)}(Y) - \lambda_0(Y)\right\}\right] - \sqrt{n}\mathbb{P}_n\left[W_n^{(k)}(Y)\left\{\hat{\lambda}_n^{(2)}(Y) - \lambda_0(Y)\right\}\right] \\
&= U_n^{(1)} - U_n^{(2)}.
\end{aligned}$$

Then we define $U = \left[U_n^{(1)}, U_n^{(2)}\right]$ and note that $U_n^{(l)}$ can be written as

$$U_n^{(l)} = U_{1n}^{(l)} + U_{2n}^{(l)} + \sqrt{\frac{n}{n_l}} U_{3n}^{(l)},$$

where, for $l = 1, 2$,

$$U_{1n}^{(l)} = \sqrt{n}(\mathbb{P}_n - P)\left[W_n^{(k)}(Y)\left\{\hat{\lambda}_n^{(l)}(Y) - \lambda_0(Y)\right\}\right],$$

$$U_{2n}^{(l)} = \sqrt{n}P\left[\left(W_n^{(k)}(Y) - W(Y)\right)\left\{\hat{\lambda}_n^{(l)}(Y) - \lambda_0(Y)\right\}\right],$$

$$U_{3n}^{(l)} = \sqrt{n_l}P\left[W(Y)\left\{\hat{\lambda}_n^{(l)}(Y) - \lambda_0(Y)\right\}\right].$$

Firstly consider $U_{1n}^{(l)} = \sqrt{n}(P_n - P)\left[W_n^{(k)}(Y)\{\hat{\lambda}_n^{(l)}(Y) - \lambda_0(Y)\}\right]$. Set

$$\mathcal{G} = \{\xi : [0, b] \to [0, \tau]\},$$

where τ is the uniform upper bound of weight process $W_n^{(k)}$, $k = 1, 2, 3, 4$. Let

$$\psi_\lambda(\xi, \mathbf{D}) = \xi(Y)\{\lambda(Y) - \lambda_0(Y)\},$$

where $\xi \in \mathcal{G}$, $\lambda \in \mathcal{F}_\eta$ and $\mathcal{F}_\eta = \{\lambda | \lambda \in \psi_{l,\mathcal{T}}, d(\lambda, \lambda_0) \leq \eta\}$. For a fixed $\xi \in \mathcal{G}$, let

$$\Psi_\eta(\xi) = \{\psi_\lambda(\xi, \mathbf{D}) : \lambda \in \mathcal{F}_\eta\},$$

where $\eta > 0$. By the conclusion of Theorem 1, $\hat{\lambda}_n^{(l)} \in \mathcal{F}_\eta$ for any $\eta > 0$ and sufficiently large n. Note that it follows from Corollary 2.7.2 of van der Vaart and Wellner (1996, page 157) that

$$N_{[\,]}(\varepsilon, \mathcal{F}_\eta, L_2(P)) \leq e^{c_1/\varepsilon^{1/2}},$$

for some constant c_1. Then, we have

$$N_{[\,]}(\varepsilon, \Psi_\eta(\xi), L_2(P)) \leq e^{c_1/\varepsilon^{1/2}}.$$

It can be easily shown that $|\psi_\lambda(\xi, \mathbf{D})| \lesssim \eta$, and $P\psi_\lambda^2(\xi, \mathbf{D}) \lesssim \eta^2$. Thus,

$$J_{[\,]}(\eta, \Psi_\eta(\xi), L_2(P)) = \int_0^\eta \sqrt{\log N_{[\,]}(\varepsilon \|\psi\|_{P,2}, \Psi_\eta(\xi), L_2(P)) + 1} d\varepsilon \lesssim \eta.$$

Hence, from Theorem 2.14.2 of van der Vaart and Wellner (1996), we have

$$E^*\left\{\sup_{\psi_\lambda(\xi,X)\in\Psi_\eta(\xi)} \left|\sqrt{n}(\mathbb{P}_n - P)\psi_\lambda(\xi, X)\right|\right\} \lesssim \left[J_{[\,]}(\eta, \Psi_\eta(\xi), L_2(P))\|\psi\|_{P,2} + \sqrt{n}P\psi\{\psi > \sqrt{n}a(\eta)\}\right.$$
$$\left. + \|\psi\|_{P,2}\sqrt{\log N_{[\,]}(\eta\|\psi\|_{P,2}, \Psi_\eta(\xi), L_2(P)) + 1}\right],$$

where

$$a(\eta) = \eta\|\psi\|_{P,2}/\sqrt{\log N_{[\,]}(\eta\|\psi\|_{P,2}, \Psi_\eta(\xi), L_2(P)) + 1}.$$

Then, it is easily shown that

$$\limsup_{n\to\infty} E^*\left\{\sup_{\psi_\lambda(\xi,X)\in\Psi_\eta(\xi)} \left|\sqrt{n}(\mathbb{P}_n - P)\psi_\lambda(\xi, X)\right|\right\} \lesssim \eta^{1/2}.$$

It follows from $d(\hat{\lambda}_n^{(l)}, \lambda_0) \overset{a.s.}{\to} 0$ that

$$\limsup_{n\to\infty} E\left\{|\sqrt{n}(\mathbb{P}_n - P)\psi_{\hat{\lambda}_n^{(l)}}(W_n^{(k)}, X)|\right\} \lesssim \eta^{1/2}.$$

Let $\eta \to 0$ to see

$$\lim_{n\to\infty} E\left\{|\sqrt{n}(\mathbb{P}_n - P)\psi_{\hat{\lambda}_n^{(l)}}(W_n^{(k)}, X)|\right\} = 0,$$

which yields $U_{1n}^{(l)} = o_p(1)$.

Next consider $U_{2n}^{(l)} = \sqrt{n}P\left[(W_n^{(k)}(Y) - W(Y))\left\{\hat{\lambda}_n^{(l)}(Y) - \lambda_0(Y)\right\}\right]$.

$$\begin{aligned}
U_{2n}^{(l)} &= \sqrt{n}P\left\{\left(W_n^{(k)}(Y) - W(Y)\right)\left[\hat{\lambda}_n^{(l)}(Y) - \lambda_0(Y)\right]\right\} \\
&\le \sqrt{n}\int \left|W_n^{(k)}(t) - W(t)\right|\left|\hat{\lambda}_n^{(l)}(Y) - \lambda_0(Y)\right| dF^*(t) \\
&\lesssim \sqrt{n}\left\{\int_0^b \left(W_n^{(k)}(t) - W(t)\right)^2 dF^*(t)\right\}^{1/2}\left\{\int_0^b \left(\hat{\lambda}_n^{(l)}(Y) - \lambda_0(Y)\right)^2 dF^*(t)\right\}^{1/2}.
\end{aligned}$$

Since

$$\left[\int_a^b \left\{W_n^{(k)}(t) - W(t)\right\}^2 dt\right]^{1/2} = o_p\left(n^{-\frac{1}{2(1+2r)}}\right)$$

and

$$\left\{\int_0^b \left(\hat{\lambda}_n^{(l)}(Y) - \lambda_0(Y)\right)^2 dF^*(t)\right\}^{1/2} = O_p(n^{-\frac{r}{1+2r}}),$$

we have $U_{2n}^{(l)} = o_p(1)$.

From the result of Theorem 3, we have , for $l = 1, 2$,

$$U_{3n}^{(l)} = \sqrt{n_l}(\mathbb{P}_{n_l} - P)\left[\phi(\lambda_0; X)[h]\right] + o_p(1) = Z_n^{(l)} + o_p(1),$$

where $\mathbb{P}_{n_l} f = \frac{1}{n_l}\sum_{i\in S_l} f(Z_i)$ and S_l denotes the set of indices for subjects in group l, $l = 1, 2$. Moreover, $Z_n^{(l)}$'s converge to U_w in distribution as $n \to \infty$, where U_w has a normal distribution with mean zero and variance $\sigma^2 = E\left[\phi^2(\lambda_0; X)[h]\right]$. Evidently, $Z_n^{(l)}$'s are independent and identically distributed, because $\mathbb{P}_{n_l}$ is the empirical measure based on group l respectively. Hence, we have

$$U_n = \sqrt{\frac{n}{n_1}}Z_n^{(1)} - \sqrt{\frac{n}{n_2}}Z_n^{(2)} + o_p(1),$$

where U_n convergences in distribution to $N(0, \frac{1}{p(1-p)}\sigma^2)$. Thus it follows that U_n has an asymptotic normal distribution $N(0, \sigma_w^2)$, where

$$\sigma_w^2 = \frac{1}{p(1-p)}E\{\phi^2(\lambda_0; X)[h_w]\}.$$

To show that $\hat{\sigma}_w^2 - \sigma_w^2 = o_p(1)$. We set ${\sigma_w}^2 = P\phi^2(\lambda_0; X)[h_w]$ and $\hat{\sigma}_w^2 = \mathbb{P}_n\phi^2(\hat{\lambda}_n; X)[\hat{h}_w]$. Note that

$$\begin{aligned}\hat{\sigma}_w^2 - \sigma_w^2 =& \mathbb{P}_n\phi^2(\hat{\lambda}_n; X)[W_n^{(k)}\hat{\lambda}_n^2] - P\phi^2(\lambda_0; X)[W\lambda_0^2] \\ =& \mathbb{P}_n\left\{\phi^2(\hat{\lambda}_n; X)[W_n^{(k)}\hat{\lambda}_n^2] - \phi^2(\lambda_0; X)[W_n^{(k)}\lambda_0^2]\right\} + (\mathbb{P}_n - P)\phi^2(\lambda_0; X)[W\lambda_0^2] \\ &+ \mathbb{P}_n\left\{\phi^2(\lambda_0; X)[W_n^{(k)}\lambda_0^2] - \phi^2(\lambda_0; X)[W\lambda_0^2]\right\}.\end{aligned}$$

It can be easily shown that

$$\mathbb{P}_n\left\{\phi^2(\hat{\lambda}_n; X)[W_n^{(k)}\hat{\lambda}_n^2] - \phi^2(\lambda_0; X)[W_n^{(k)}\lambda_0^2]\right\} = o_p(1)$$

and

$$(\mathbb{P}_n - P)\phi^2(\lambda_0; X)[W\lambda_0^2] = o_p(1).$$

On the other hand, based on the conditions imposed on W_n and W, we have

$$\left|\phi(\lambda_0; X)[W_n^{(k)}\lambda_0^2] - \phi(\lambda_0; X)[W\lambda_0^2]\right| = \left|\phi(\lambda_0; X)[(W_n^{(k)} - W)\lambda_0^2]\right| = o_p(1),$$

and

$$\left|\phi(\lambda_0; X)[W_n^{(k)}\lambda_0^2] + \phi(\lambda_0; X)[W\lambda_0^2]\right| = \left|\phi(\lambda_0; X)[(W_n^{(k)} + W)\lambda_0^2]\right| = O(1).$$

The above two displays imply that

$$\left|\phi^2(\lambda_0; X)[W_n^{(k)}\lambda_0^2] - \phi^2(\lambda_0; X)[W\lambda_0^2]\right| \lesssim \left|\phi(\lambda_0; X)[W_n^{(k)}\lambda_0^2] - \phi(\lambda_0; X)[W\lambda_0^2]\right| = o_p(1).$$

Therefore,

$$\mathbb{P}_n\left\{\left|\phi^2(\lambda_0; X)[W_n^{(k)}\lambda_0^2] - \phi^2(\lambda_0; X)[W\lambda_0^2]\right|\right\} = o_p(1).$$

References

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993), *Statistical Models Based on Counting Processes*, Springer US.

Balakrishnan, N. and Mitra, D. (2011), “Likelihood Inference for Lognormal Data with Left Truncation and Right Censoring with an Illustration,” *Statistical Planning and Inference*, 141(11), 3536–3553.

— (2012), “Left Ttuncated and Right Censored Weibull Data and Likelihood Inference with an Illustration,” *Computational Statistics and Data Analysis*, 56(12), 4011–4025.

— (2013), “Likelihood Inference Based on Left Truncated and Right Censored Data from a Gamma Distribution,” *IEEE Transactions on Reliability*, 62(3), 679–688.

— (2014), “Some Further Issues Concerning Likelihood Inference for Left Truncated and Right Censored Lognormal Data,” *Communications in Statistics–Simulation and Computation*, 43 (2), 400–416.

Balakrishnan, N. and Zhao, X. Q. (2009), "New Multi-Sample Nonparametric Tests for Panel Count Data," *The Annals of Statistics*, 37(3), 1112–1149.

Carrión, A., Solano, H., Gamiz, M. L., and Debón, A. (2010), "Evaluation of the Reliability of a Water Supply Network from Right-Censored and Left-Truncated Break Data," *Water Resources Management*, 24(12), 2917–2935.

Emura, T. and Shiu, S. (2016), "Estimation and Model Selection for Left-Truncated and Right-Censored Lifetime Data with Application to Electric Power Transformers Analysis," *Communications in Statistics - Simulation and Computation*, 45(9), 3171–3189.

Frydman, H. (1994), "A Note on Nonparametric Estimation of the Distribution Function from Interval-Censored and Truncated Observations," *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(1), 71–74.

Hong, Y., Duan, Y., Meeker, W., Stanley, D., and Gu, X. (2015), "Statistical Methods for Degradation Data with Dynamic Covariates Information and An Application to Outdoor Weathering Data," *Technometrics*, 57, 180–193.

Hong, Y., Meeker, W. Q., and Mccalley, J. D. (2009), "Prediction of Remaining Life of Power Transformers Based on Left Truncated and Right Censored Lifetime Data," *The Annals of Applied Statistics*, 3(2), 857–879.

Hua, L. and Zhang, Y. (2012), "Spline-based Semiparametric Projected Generalized Estimating Equation Method for Panel Count Data," *Biostatistics*, 13(3), 440–454.

Kevin, C. C., Siobán, D. H., Roderick, J. L., Nan, B., Matheos, Y., Taffe, J. R., and Michael, R. E. (2011), "Bias Due to Left Truncation and Left Censoring in Longitudinal Studies of Developmental and Disease Processes," *American Journal of Epidemiology*, 173(9), 1078–1084.

Lu, M., Zhang, Y., and Huang, J. (2007), "Estimation of the Mean Function with Panel Count Data using Monotone Polynomial Splines," *Biometrika*, 94(3), 705–718.

— (2009), "Semiparametric Estimation Methods for Panel Count Data using Monotone B-splines," *Journal of the American Statistical Association*, 104(484), 1060–1070.

Pepe, M. S. and Fleming, T. R. (1989), "Weighted Kaplan-Meier Statistics: A Class of Distance Tests for Censored Survival Data," *Biometrics*, 45(2), 497–507.

Ramsay, J. O. (1988), "Monotone Regression Splines in Action," *Statistical Science*, 3(4), 425–441.

Rosenberg, P. S. (1995), "Hazard Function Estimation Using B-Splines," *Biometrics*, 51(3), 874–887.

Schumaker, L. (2007), *Spline Functions: Basic Theory*, Cambridge Mathematical Library.

Shen, P. S. (2007), "A General Class of Test Procedures for Left - Truncated and Right - Censored Data," *Communications in Statistics - Theory and Methods*, 36(16), 2913–2925.

— (2014), "Nonparametric Tests for Left-Truncated and Interval-Censored Data," *Journal of Statistical Computation and Simulation*, 85(8), 1544–1553.

Shen, P. S. and Yan, Y. F. (2008), "Nonparametric Estimation of the Bivariate Survival Function with Left-Truncated and Right-Censored Data," *Journal of Statistical Planning and Inference*, 138(12), 4041 – 4054.

Shen, X. T. and Wong, W. H. (1994), "Convergence Rate of Sieve Estimates," *The Annals of Statiatics*, 22(2), 580–615.

Stone, C. J. (1994), "The Use of Polynomial Splines and Their Tensor Products in Multivariate Function Estimation," *The Annals of Statistics*, 22(1), 118–184.

Su, Y. and Wang, J. (2012), "Modeling Left-Truncation and Right-Censored Survival Data with Longitudinal Covariates," *The Annals of Statistics*, 40(3), 1465–1488.

Tsai, W.-Y., Jewell, N. P., and Wang, M.-C. (1987), "A Note on the Product-Limit Estimator Under Right Censoring and Left Truncation," *Biometrika*, 74(4), 883–886.

Turnbull, B. W. (1976), "The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data," *the Royal Statistical Society: Series B (Statistical Methodology)*, 38(3), 290–295.

van der Vaart, A. W. and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, Springer.

Wang, J. L. (2005), "Smoothing Hazard Rate," *Encyclopedia of Biostatistics, 2nd Edition*, 7, 4986–4997.

Wang, M.-C. (1987), "Product Limit Estimates: A Generalized Maximum Likelihood Study," *Communications in Statistics - Theory and Methods*, 16(11), 3117–3132.

Wang, M.-C., Jewell, N. P., and Tsai, W.-Y. (1986), "Asymptotic Properties of the Product Limit Estimate Under Random Truncation," *The Annals of Statistics*, 14(4), 1597–1605.

Wellner, J. A. and Zhang, Y. (2007), "Two Likelihood-Based Semiparametric Estimation Methods for Panel Count Data with Covariates," *The Annals of Statiatics*, 35(5), 2106–2142.

Wu, Y. and Ying, Z. (2012), "Partially Monotone Tensor Spline Estimation of the Joint Distribution Function with Bivariate Current Status Data," *The Annals of Statistics*, 40(3), 1609–1636.

Xie, Y., King, C., Hong, Y., and Yang, Q. (2018), "Semi-Parametric Models for Accelerated Destructive Degradation Test Data Analysis," *Technometrics*, in press.

Ye, Z.-S., Hong, Y., and Xie, Y. (2013), "How Do Heterogeneities in Operating Environments Affect Field Failure Predictions and Test Planning?" *The Annals of Applied Statistics*, 7(4), 2249–2271.

Ye, Z.-S. and Tang, L.-C. (2016), "Augmenting the Unreturned for Field Data with Information on Returned Failures Only," *Technometrics*, 58(4), 513–523.

Zhao, X., Deng, S., Liu, L., and Liu, L. (2013), "Sieve Estimation in Semiparametric Modeling of Longitudinal Data with Informative Observation Times," *Biostatistics*, 15(1), 140–153.

Zhao, X. and Zhang, Y. (2017), "Asymptotic Normality of Nonparametric M-Estimations with Applications to Hypothesis Testing for Panel Count Data," *Statistica Sinica*, 931–950.