Electronic Journal of Statistics
Vol. 17 (2023) 2745–2809
ISSN: 1935-7524
https://doi.org/10.1214/23-EJS2163

On nonparametric estimation for cross-sectional sampled data under stationarity

Kwun Chuen Gary Chan

Department of Biostatistics, University of Washington, Seattle, United States
e-mail: kcgchan@uw.edu

Hok Kan Ling

Department of Mathematics and Statistics, Queen's University, Kingston, Canada
e-mail: brian.ling@queensu.ca

Sheung Chi Phillip Yam

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

Abstract: We study the nonparametric estimation of the underlying survival function of a survival time in a study with cross-sectional sampling without any follow-up. Under a stationarity assumption on disease incidence rate in the population, the survival function S_0 is related to the observed density of the backward recurrence time, f_0, via the relationship $S_0(x) = f_0(x)/f_0(0)$. As $f_0(x)$ is non-decreasing, it is well-known that the nonparametric maximum likelihood estimator of f_0 at $x = 0$ is inconsistent. In this article, we establish the asymptotic distributions of the estimators of $S_0(x)$ when different consistent estimators of $f_0(0)$ are used. Such results are currently missing in the literature. Another contribution is the establishment of a local Kiefer-Wolfowitz-type result of the form $\sup_{x\in[0,y]} |\hat{F}_n(x) - \mathbb{F}_n(x)| = O_p(n^{-2/3}(\log n)^{2/3})$ that makes use of weaker assumptions than existing results, where $\mathbb{F}_n$ and $\hat{F}_n$ are the empirical distribution function and its least concave majorant, respectively.

MSC2020 subject classifications: Primary 62G07, 62G20.
Keywords and phrases: Backward recurrence time, current duration data, cross-sectional sampling, Grenander estimator, shape-constrained inference, Kiefer-Wolfowitze-type result.

Received October 2022.

Contents

1. Introduction

Cross-sectional surveys often collect information on time since a certain incident event, but there is no follow-up of the study participants to collect the end-points of interest. The time from the initial event to the sampling time is known as the backward recurrence time or current duration data. Current duration data are collected in studies of time to pregnancy [11], length of residency [28], duration between episodes of psychiatric disorders [17], among others. However, the goal is often to estimate the survival function of the full duration between the initial event and an endpoint of interest, but the endpoints are never observed or always censored as there is no active follow-up.

An additional complication of current duration data is that the survival times from the population do not have the same chance of being sampled. In particular, subjects with a longer survival time are more likely to be sampled. Let Y be the unobserved survival time, which is the time from a disease incidence event to a failure event of interest. An assumption that the disease incidence rate in the population remains constant over calendar time and independent of individual survival time, often referred to as a stationary disease incidence assumption, are often assumed for cross-sectional sampling [26], because under the assumption cross-sectional sampling is length-biased [2]. Let Y^* be the length-biased version of Y which is potentially observed from the cross-sectional sample when there is complete follow-up. As there is no follow-up in our consideration, all the survival times are immediately censored at recruitment and we only observe the backward recurrence time X, which can be represented as $X = Y^*U$ [25, 24], where U is a standard uniform random variable on $[0,1]$ and it is also independent of Y^*. The form of the distortion Y^*U is also known as multiplicative censoring because the incompleteness of the Y^* results from Y being scaled down by an independent random variable U [25]. Recall S_0 denotes the survival function of Y. The density of X, f_0, is then given by the following key formula [2]:

$$f_0(x) = \frac{S_0(x)}{\int_0^\infty S_0(y)dy},$$

where we assume throughout that $\mathbb{E}(Y) = \int_0^\infty S_0(y)dy < \infty$.

In this article, we study the nonparametric estimation of S_0. Certainly,

$$S_0(x) = \frac{f_0(x)}{f_0(0)} \tag{1.1}$$

for any $x \geq 0$, and we could estimate S_0 whenever an estimate of f_0 is available. Since S_0 is by definition decreasing, f_0 is also decreasing. The nonparametric maximum likelihood estimator (NPMLE) of f_0 is given by the Grenander estimator $\hat{f}_n$ [5] of the backward recurrence times. However, [27] pointed out that $\hat{f}_n(0) := \hat{f}_n(0+)$ is not a consistent estimator of $f_0(0)$. As a result, $\hat{S}_n(x) := \hat{f}_n(x)/\hat{f}_n(0)$ will also not be a consistent estimator of $S_0(x)$. Different alternative consistent estimators of $f_0(0)$ has been proposed in the literature: based on a penalized likelihood [27], using the first bin of a simple histogram [8], a local [14] and a smoothed Grenender estimator [8].

With a consistent estimator of $f_0(0)$, say $\tilde{f}_n(0)$, we can obtain a consistent estimator of $S_0(x)$ of the form $\tilde{S}_n(x) = \tilde{f}_n(x)/\tilde{f}_n(0)$, where $\tilde{f}_n(x)$ is a consistent estimator of $f_0(x)$ using, for example, the Grenander estimator, its penalized or smoothed version. Nevertheless, to the best of our knowledge, theoretical properties of $\tilde{S}_n(x)$ have not yet been studied. Our main result in this paper is to fill in this gap by establishing the asymptotic distributions of $\tilde{S}_n(x)$ at any interior point of the support of Y, where $\tilde{f}_n(0)$ can be the different estimators introduced above.

The main crux of establishing the asymptotic distribution of $\tilde{S}_n(x)$ is to observe that $\tilde{f}_n(x)$ and $\tilde{f}_n(0)$ are asymptotically independent for a variety of choice of $\tilde{f}_n(x)$ and $\tilde{f}_n(0)$. This insight results from the switch relation (3.2) and the Hungarian approximation [13], so that the estimators $\tilde{f}_n(x)$ and $\tilde{f}_n(0)$ will be related to Brownian motions in non-overlapping intervals asymptotically, which is the source of independence.

Another contribution of this paper is the establishment of a local Kiefer-Wolfowitz-type result that requires weaker assumptions than those in [12, 4]. Roughly speaking, [12] proved in their Theorem 1 that, if f_0 is bounded away from 0 with a continuous first derivative f_0' that is bounded and away from 0, then, with probability one, the maximum absolute distance between $\hat{F}_n$ and $\mathbb{F}_n$ is of the order $n^{-2/3}\log n$. [4] considered Grenander-type estimators for monotone functions in a more general setting. This kind of Kiefer-Wolfowitz-type result has been applied to study the asymptotic optimality of shape-constrained estimators and bootstrap theory (see, e.g., [8] and the references therein) which could be of independent interest. However, under our setting, the condition in [12] or [4] corresponds to that S_0 is bounded away from 0, and this must lead to problematic formulation because S_0 is a survival function and should be ultimately equal to 0 at the endpoint. We weakened this assumption but allow the result to be valid over a local region.

The organization of this paper is as follows. In Section 2, we first discuss the non-uniqueness and inconsistency of nonparametric maximum likelihood estimation and then review four different classes of consistent estimators, including estimators based on the penalized estimator, the histogram estimator, the local Grenander estimator and the smoothed Grenander estimator. Next, we establish the asymptotic distributions of the above estimators at a fixed interior point as well as a local Kiefer-Wolfowitz resul under relaxed conditions in Section 3. Section 4 provides some numerical comparisons of different estimators. A real data application is given in Section 5. Some discussions are provided in Section 6. Proofs of the theoretical results are given in the appendix.

2. Nonparametric estimators and their consistencies

In this article, we focus on the nonparametric estimation of the survival function S_0. We first note that the NPMLE for S_0 is not necessarily the plug-in estimator obtained by using (1.1) with f_0 being replaced by $\hat{f}_n$, the NPMLE for f_0, unless there is an additional constraint. In addition, it is well-known that $\hat{f}_n(0) := \hat{f}_n(0+)$ is inconsistent, which also leads to the inconsistency of the plug-in estimator for $S_0(x)$; see [27]. We then review four different consistent estimators of $f_0(0)$ proposed in the literature and formally define and establish the consistency of the corresponding estimators of $S_0(x)$. This then enables us to establish the asymptotic distribution of the estimators of $S_0(x)$.

2.1. Nonparametric maximum likelihood estimation

Given a random sample of current duration data $X_1, \ldots, X_n$ as described in Section 1, the likelihood function in an arbitrary survival function S is

$$L_n(S) := \prod_{i=1}^{n} \frac{S(X_i)}{\int_0^\infty S(y)dy}, \tag{2.1}$$

subject to $S \in \mathcal{S} := \{H : H \text{ is a survival function on } [0, \infty)\}$. On the other hand, the likelihood function in an arbitrary density function f is

$$L_{n,2}(f) := \prod_{i=1}^{n} f(X_i), \tag{2.2}$$

subject to $f \in \mathcal{F} := \{h : (0, \infty) \to [0, \infty) : h \text{ is decreasing and } \int_0^\infty h(y)dy = 1\}$. The NPMLE of (2.2) subject to $f \in \mathcal{F}$ is the well-known Grenander estimator [5], and we denote it by $\hat{f}_n$. The Grenander estimator can be characterized as the left-continuous slope of the least concave majorant of the empirical distribution function of X_i's. Assuming that all the observations are distinct, it can also be computed using the following formula:

$$\hat{f}_n(X_{(k)}) = \min_{0 \le i < k} \max_{k \le j \le n} \frac{(j-i)/n}{X_{(j)} - X_{(i)}}, \qquad \text{for } k = 1, \ldots, n,$$

where $X_{(i)}$'s are the order statistics of X_i's with $X_{(1)} < \cdots < X_{(n)}$; see [20]. In view of (1.1), a natural nonparametric estimator of S_0 would be the plug-in estimator:

$$\hat{S}_n(x) := \begin{cases} \frac{\hat{f}_n(x)}{\hat{f}_n(0)}, & x \ge 0; \\ 1, & x < 0, \end{cases} \tag{2.3}$$

which is well-received in the community; for instance, see [11]; here, $\hat{f}_n(0) = \hat{f}_n(0+)$, which is equal to $\hat{f}_n(X_{(1)})$.

However, it is not immediate that $\hat{S}_n$ is the NPMLE of (2.1) although $\hat{f}_n$ is the NPMLE of (2.2); in fact, if $S \in \mathcal{S}$ maximizes L_n, then for any $c > 1$, define

$$S_c(x) := \begin{cases} 1, & \text{if } x = 0; \\ S(x)/c, & \text{if } x > 0. \end{cases}$$

Then, $S_c \in \mathcal{S}$ and $L_n(S) = L_n(S_c)$. This means that the maximizer of L subject to $S \in \mathcal{S}$ is not unique without some additional appropriate constraints. This is unlike the case of density estimation under monotonicity, where the constraint that the density integrates to 1 alone ensures the uniqueness of the NPMLE. A natural choice is to confine that

$$S(X_{(1)}) = 1, \tag{2.4}$$

as $X_{(1)}$ is physically the first observation. In the following Lemma 2.1, we shall show that $\hat{S}_n$ defined in (2.3) is the unique maximizer of L_n under the additional constraint of (2.4).

Lemma 2.1. *Any possible maximizer of L_n is in the form of:*

$$S_{n,c}(x) := \begin{cases} \frac{\hat{f}_n(x)}{c}, & \text{if } x > 0; \\ 1, & \text{if } x \le 0, \end{cases} \tag{2.5}$$

for some $c \ge \hat{f}_n(0)$. Hence, the unique maximizer of L_n subject to (2.4) is $\hat{S}_n$ defined in (2.3).

2.2. Consistent nonparametric estimators

Let $\tau := \inf\{x \in (0,\infty) : f_0(x) = 0\}$. While $\hat{f}_n(x)$ is a consistent estimator of $f_0(x)$ for any $x \in (0,\tau)$ (see Corollary 3.1 in [8]), [27] noted that $\hat{f}_n(0)$ is not a consistent estimator for $f_0(0)$; from which, in fact, if $f_0(0) < \infty$, $\hat{f}_n(0)/f_0(0) \xrightarrow{d} \sup_{1\le j<\infty} j/\Gamma_j \stackrel{d}{=} 1/U$, where Γ_j's are partial sums of independent and identically standard exponential random variables and U is a uniform random variable on $[0,1]$. Furthermore, $\hat{f}_n(0)$ is simply too big as $\mathbb{P}(\sup_{1\le j<\infty}(j/\Gamma_j) > 1) = 1$ by the strong law of large numbers. As a result, the estimator $\hat{S}_n(x) = \hat{f}_n(x)/\hat{f}_n(0)$ is not a consistent estimator of $S_0(x) = f_0(x)/f_0(0)$ for any $x \in (0,\tau)$. In this subsection, we shall discuss four different classes of consistent nonparametric estimators for $S_0(x)$, $x > 0$, based on various consistent estimators and their variants of $f_0(0)$ proposed in the literature. The corresponding asymptotic distributions of these estimators will be established in Section 3.

2.2.1. Assumptions for asymptotic results

We first state the conditions under which the asymptotic results in Subsection 2.2 and Section 3 will be derived. Let $f_0^{(k)}(0) := \lim_{x\downarrow 0} f_0^{(k)}(x)$ and $S_0^{(k)}(0) := \lim_{x\downarrow 0} S_0^{(k)}(x)$. Fix $x_0 \in (0,\tau)$.

- C_f1 f_0 is decreasing and $0 < f_0(0) < \infty$;
- C_f2 $0 < |f_0'(x_0)| < \infty$;
- C_f3 $0 < |f_0''(x_0)| < \infty$;
- C_f4 for some $k \ge 1$, $0 < |f_0^{(k)}(0)| < \infty$ and $f_0^{(i)}(0) = 0$ for $1 \le i \le k-1$, or
- C_f4' for some $k \ge 1$, $0 < |f_0^{(k)}(0)| \le \sup_{s\ge 0} |f_0^{(k)}(s)| < \infty$ and $f_0^{(i)}(0) = 0$ for $1 \le i \le k-1$;

C_f2 is used for establishing the asymptotic distribution of the Grenander estimator at an interior point of the support; see [6]. C_f3 is used for establishing the asymptotic distribution based on smoothed Grenander estimator, which is a standard assumption for density estimation using kernel smoothing when deriving the asymptotic distribution; see [15]. The asymptotic behaviours of different consistent estimators of $f_0(0)$ depend on the smoothness of f_0 at 0. Clearly, C_f4 is strictly weaker than C_f4'. C_f4' is used in [14], where they studied the behavior of the Grenander estimator $\hat{f}_n$ near the boundaries of the support of a decreasing density. The canonical examples that satisfies $C_f 4'$ (and therefore also for $C_f 4$) for $k = 1$ and $k = 2$ are the exponential distributions and the half-normal distributions, respectively.

The followings are the corresponding equivalent conditions in terms of S_0.

- $C_S 1$ $0 < \int_0^\infty S_0(y)dy < \infty$;
- $C_S 2$ $0 < |S_0'(x_0)| < \infty$;
- $C_S 3$ $0 < |S_0''(x_0)| < \infty$;
- $C_S 4$ for some $k \geq 1$, $0 < |S_0^{(k)}(0)| < \infty$ and $S_0^{(i)}(0) = 0$ for $1 \leq i \leq k-1$;
- $C_S 4'$ for some $k \geq 1$, $0 < |S_0^{(k)}(0)| \leq \sup_{s \geq 0} |S_0^{(k)}(s)| < \infty$, and $S_0^{(i)}(0) = 0$ for $1 \leq i \leq k-1$;

Recall that g_0 is the density of Y, the underlying variable of interest. That is $g_0'(x) = -S_0'(x)$. Hence, for densities such that $g_0(0) \in (0, \infty)$, $C_S 4$ is satisfied with $k = 1$ (e.g., exponential and half-normal distribution). For densities such that $g_0(0) = 0$ but $|g_0'(0)| \in (0, \infty)$, $C_S 4$ is satisfied with $k = 2$ (e.g., gamma distribution with shape and scale parameters being equal to 2).

2.2.2. Penalized NPMLE

As mentioned in Section 2.1, $\hat{f}_n(0)$ is simply too big and inconsistent. [27] proposed a penalized version of the log-likelihood by maximizing

$$l_{n,2,\alpha}(f) := \sum_{i=1}^n \log f(X_i) - n\alpha f(0+), \tag{2.6}$$

where $\alpha > 0$ is a penalty parameter to be determined, but we assume throughout that $\alpha < X_{(n)}$. From now on, we denote $\hat{f}_n^P$ to be the penalized NPMLE for f_0. It is also shown in [27] that the penalized NPMLE has the same form that can be achieved as the unpenalized one but with transformed variables $\tilde{X}_1, \ldots, \tilde{X}_n$ from X_i's in the following way. Given a penalty parameter α, let $\tilde{X}_0 := 0$ and $\tilde{X}_i = \tilde{X}_i(\alpha) := \alpha + \gamma_\alpha X_{(i)}$, for $i = 1, \ldots, n$, where γ_α is the unique, positive number satisfying the following equation:

$$\gamma = \min_{1 \leq i \leq n} \left\{ 1 - \frac{\alpha i/n}{\alpha + \gamma X_{(i)}} \right\}.$$

The penalized NPMLE is a step function with jumps possibly at X_i's with its value at the point $X_{(i)}$ being equal to the value of the Grenander estimator computed from the transformed sample of $\tilde{X}_1, \ldots, \tilde{X}_n$ at $\tilde{X}_i$.

With the penalized estimator $\hat{f}_n^P$, it has been noted that (e.g. [11]) one can use $\hat{f}_n^P(x)/\hat{f}_n^P(0)$ to estimate $S_0(x)$. To formally recognize that this estimator is the NPMLE of a penalized log-likelihood, we define the penalized log-likelihood as

$$l_{n,\alpha}(S) := \sum_{i=1}^n \log \left\{ \frac{S(X_i)}{\int_0^\infty S(y)dy} \right\} - n\alpha \frac{S(0+)}{\int_0^\infty S(y)dy}, \tag{2.7}$$

subject to $S \in \mathcal{S}$, as $f(0+) = \frac{S(0+)}{\int_0^\infty S(y)dy}$. As in the case without penalization, $l_{n,\alpha}$ does not admit a unique maximizer without any further constraint; henceforth, we impose the same constraint of (2.4) as in the unpenalized case. We shall then show that the maximizer is given by

$$\hat{S}_n^P(x) := \begin{cases} \frac{\hat{f}_n^P(x)}{\hat{f}_n^P(0)}, & x \geq 0; \\ 1, & x < 0, \end{cases} \tag{2.8}$$

where $\hat{f}_n^P(0) := \hat{f}_n^P(0+)$.

Lemma 2.2. *Any maximizer of $l_{n,\alpha}$ is in the form of:*

$$S_{n,c}^P(x) := \begin{cases} \frac{\hat{f}_n^P(x)}{c}, & x > 0; \\ 1, & x \leq 0, \end{cases} \tag{2.9}$$

for some $c \geq \hat{f}_{n,\alpha}(0)$. Hence, the unique maximizer of $l_{n,\alpha}$ subject to (2.4) is $\hat{S}_n^P$ as defined in (2.8).

The consistency of $\hat{S}_n^P(x)$ follows from that of $\hat{f}_n^P(x)$ and $\hat{f}_n^P(0)$.

Lemma 2.3. *If $\alpha_n \to 0$ and $n\alpha_n \to \infty$ as $n \to \infty$, and S_0 is strictly decreasing near zero, then $\hat{S}_n^P(x) \xrightarrow{\mathbb{P}} S_0(x)$ for $x \in [0, \tau)$.*

2.2.3. Using a histogram estimator

Let $\mathbb{F}_n$ be the empirical distribution of $X_1, \ldots, X_n$. To estimate $f_0(0)$, a natural and simple histogram estimator is suggested in [8] (see also [10]):

$$\hat{f}_n^H(0) = \hat{f}_{n,b_n}^H(0) := \frac{\mathbb{F}_n(b_n)}{b_n},$$

where the bin width $\{b_n\}$ is a vanishing sequence of positive numbers. Assuming $f_0(0) < \infty$ and $|f_0'(0)| < \infty$, the asymptotic mean square error optimal choice for b_n can be shown to be $\{2f_0(0)/f_0'(0)^2\}^{1/3}n^{-1/3}$; see [10]. The following lemma gives a more general result on the asymptotic mean square error optimal choice of b_n for $\hat{f}_n^H(0)$ under different regularity conditions of f_0 at 0 as stated in C_f4. The corresponding statement in [8] corresponds to the case when $k = 1$. Given two sequences $\{p_n\}$ and $\{q_n\}$, the notation $p_n \sim q_n$ means $p_n/q_n \to 1$ as $n \to \infty$.

Lemma 2.4. *Under C_f4 for some $k \geq 1$, $E(\hat{f}_n^H(0)) - f_0(0) \sim f_0^{(k)}(0)b_n^k/k!$, $Var(\hat{f}_n^H(0)) \sim \frac{1}{nb_n}f_0(0)$, and the asymptotic mean square error optimal choice of b_n for $\hat{f}_n^H(0)$ is*

$$b_n^{opt} := \left\{ \frac{f_0(0)}{2k\{f_0^{(k)}(0)/k!\}^2} \right\}^{\frac{1}{2k+1}} n^{-\frac{1}{2k+1}}.$$

The corresponding estimator of $S_0(x)$ using $\hat{f}_n^H(0)$ can be defined as

$$\hat{S}_n^H(x) := \begin{cases} \frac{\hat{f}_n(x)}{\hat{f}_n^H(0)}, & \text{for } x \geq 0 \text{ such that } \hat{f}_n(x) < \hat{f}_n^H(0); \\ 1, & \text{for } x \geq 0 \text{ such that } \hat{f}_n(x) \geq \hat{f}_n^H(0). \end{cases} \tag{2.10}$$

Lemma 2.5. *If $b_n \to 0$ and $nb_n \to \infty$ as $n \to \infty$, then $\hat{f}_n^H(0) \xrightarrow{\mathbb{P}} f_0(0)$. As a result, $\hat{S}_n^H(x) \xrightarrow{\mathbb{P}} S_0(x)$ for $x \in [0, \tau)$.*

2.2.4. Using a local Grenander estimator

[14] studied the behavior of the Grenander estimator $\hat{f}_n$ near the boundaries of the support of a decreasing density. They established the asymptotic distribution of estimators of the form $\hat{f}_n^N(0) := \hat{f}_n(cn^{-\alpha})$, where $\alpha \in (0, 1)$ and $c > 0$ under $C_f 4'$. In particular, the result implies that $\hat{f}_n(cn^{-\alpha})$ is a consistent estimator of $f_0(0)$ if $1/(2k+1) \leq \alpha < 1$, $c > 0$. The corresponding estimator of $S_0(x)$ using $\hat{f}_n(cn^{-\alpha})$ can be defined as

$$\hat{S}_n^N(x) := \begin{cases} \frac{\hat{f}_n(x)}{\hat{f}_n(cn^{-\alpha})}, & x > cn^{-\alpha}; \\ 1, & x \leq cn^{-\alpha}. \end{cases}$$

Lemma 2.6. *Suppose that $C_S 4'$ holds and $1/(2k+1) \leq \alpha < 1$, then $\hat{S}_n^N(x) \xrightarrow{\mathbb{P}} S_0(x)$ for $x \in [0, \tau)$.*

2.2.5. Using a smoothed Grenander estimator

Let $K : [-1, 1] \to \mathbb{R}$ be a nonnegative kernel that is symmetric around 0, satisfies $\int_{-1}^{1} K(u)du = 1$, and has a bounded derivative. While standard kernel density estimators lead to inconsistency problems at the boundary, it is well-known that the use of boundary kernels can correct the bias of a kernel estimator and the smoothed Grenander estimator using boundary kernel is consistent at 0, see [3] and [8] and the references therein. Here, we consider boundary kernels as in [3] and describe the use of boundary kernels when $\tau < \infty$. For $a, b \in \mathbb{R}$, denote $a \wedge b := \min\{a, b\}$ and $a \vee b := \max\{a, b\}$. To be more precise, we define the smoothed Grenander estimator $\hat{f}_{n,h}^S(x)$ by

$$\hat{f}_{n,h}^S(x) := \int_{(x-h)\vee 0}^{(x+h)\wedge\tau} \frac{1}{h} K_x\left(\frac{x-u}{h}\right) \hat{f}_n(u)du, \quad x \in [0, \tau],$$

where $h > 0$ is the bandwidth and the boundary kernel K_x is defined by

$$K_x(u) := \begin{cases} \phi(x/h)\, K(u) + \psi(x/h)\, uK(u), & x \in [0, h]; \\ K(u), & x \in (h, \tau - h); \\ \phi((\tau - x)/h)\, K(u) - \psi((\tau - x)/h)\, uK(u), & x \in [\tau - h, \tau]. \end{cases} \tag{2.11}$$

The coefficients $\phi(s)$ and $\psi(s)$, for $s \in [0,1]$, are determined by the requirements that

$$\phi(s)\int_{-1}^{s} K(u)du + \psi(s)\int_{-1}^{s} uK(u)du = 1, \tag{2.12}$$

$$\phi(s)\int_{-1}^{s} uK(u)du + \psi(s)\int_{-1}^{s} u^2K(u)du = 0. \tag{2.13}$$

For $h_1, h_2 > 0$, a smoothed estimator of S is then given by

$$\hat{S}^S_{n,h_1,h_2}(x) := \begin{cases} \frac{\hat{f}^S_{n,h_1}(x)}{\hat{f}^S_{n,h_2}(0)}, & \text{for } x \text{ such that } \hat{f}^S_{n,h_1}(x) < \hat{f}^S_{n,h_2}(0); \\ 1, & \text{for } x \text{ such that } \hat{f}^S_{n,h_1}(x) \geq \hat{f}^S_{n,h_2}(0). \end{cases}$$

In the definition above, we allow one to use two different bandwidths for estimating f_0 at x and 0. As stated in the following Lemma 2.7, $\hat{S}^S_{n,h_1,h_2}(x)$ will also be a consistent estimator if the same bandwidth $h_n = h_{1n} = h_{2n}$ is used whenever $h_n \to 0$ and $h_n n^{1/2} \to \infty$.

Remark 1. *We define $\hat{S}^S_{n,h_1,h_2}(x)$ using two different bandwidths because the rates of convergence of the smoothed estimator at x and 0 can be different, depending on the smoothness assumptions at the corresponding locations; see Theorem 3.4. From another point of view, one may want to use a smaller bandwidth in a region with more data and vice versa, following the idea of adaptive kernel estimates; see, e.g., [22]. In the present case, we have a decreasing density so that there will be more data at 0 than at $x > 0$. For simplicity, we only use one bandwidth in the simulation studies.*

Lemma 2.7. *Under Condition C_S1, assume that $\tau < \infty$ and S_0 is continuous on $[0,\tau]$, $h_j = h_{jn} \to 0$ and $h_j n^{1/2} \to \infty$ for $j = 1, 2$, then $\hat{S}^S_{n,h_1,h_2}(x) \xrightarrow{\mathbb{P}} S_0(x)$ for $x \in [0,\tau]$.*

Remark 2. *Smoothing $\hat{S}_n$ directly will not give a consistent estimator of S because $\hat{S}_n$ is everywhere inconsistent.*

2.3. Conditional distribution of Y given $Y > y_0$

For a small enough y_0, an alternative way to avoid the inconsistency problem is to consider the conditional survival function of Y given $Y > y_0$ (see, e.g., [11]), denoted by $S_{|y_0}(x)$. Then, $S_{|y_0}(y) = f_0(y)/f_0(y_0)$ for any $y \geq y_0$. A consistent nonparametric estimator is then given by

$$\hat{S}_{|y_0}(y) := \frac{\hat{f}_n(y)}{\hat{f}_n(y_0)},$$

because $\hat{f}_n(y) \xrightarrow{\mathbb{P}} f_0(y)$ and $\hat{f}_n(y_0) \xrightarrow{\mathbb{P}} f_0(y_0)$. Note that in the case when $y_0 = y_{0n} \downarrow 0$ at a certain rate in the sense that $y_0 = cn^{-\alpha}$ for some $c > 0$ and $\alpha > 0$, $\hat{S}_{|y_0}(y)$ is the same as the estimator $\hat{S}^N_n(y)$ for any $y \geq y_{0n}$.

3. Asymptotic distributions

In this section, we establish the asymptotic distributions of the four estimators $\hat{S}_n^P(x_0)$, $\hat{S}_n^H(x_0)$, $\hat{S}_n^0(x_0)$, and $\hat{S}_{n,h_1,h_2}^S(x_0)$ for $x_0 \in (0,\tau)$ and that of $\hat{S}_{|y_0}(x_0)$ for $x_0 \in (y_0,\tau)$ and $y_0 > 0$. To this end, we first review some technical tools used in the proofs in the next subsection and establish a local Kiefer-Wolfowitz result under relaxed conditions in Subsection 3.2.

3.1. Preliminaries

We first recall a few technical tools that shall be used in the proofs of the results in this section. As the estimators of $S_0(x)$ depend on the estimators of $f_0(x)$ and $f_0(0)$, the study of the asymptotic distribution of the former ones depend on the asymptotic joint distribution of the latter two. Instead of studying $\hat{f}_n$ directly, the established approach by [7] studies its inverse process, which is more tractable. For $a \geq 0$, the inverse process of $\hat{f}_n$ is defined by

$$U_n(a) := \sup\{t \geq 0 : \mathbb{F}_n(t) - at \text{ is maximal}\}. \tag{3.1}$$

Then, with probability one, we have the following switch relation (see [7] or [8]):

$$\hat{f}_n(x) \leq a \Leftrightarrow U_n(a) \leq x. \tag{3.2}$$

Let $\mathbb{B}_{\text{loc}}$ be the space of all locally bounded real functions on $\mathbb{R}$ endowed with the topology of uniform convergence on compacta. That is, for $h_n, h \in \mathbb{B}_{loc}(\mathbb{R})$, h_n converges to h if for every $M > 0$, $\sup_{t\in[-M,M]} |h_n(t) - h(t)| \to 0$ as $n \to \infty$. Let $\mathbb{C}_{\max}(\mathbb{R})$ denote the (separable) subset of continuous functions x in $\mathbb{B}_{\text{loc}}(\mathbb{R})$ which satisfies $x(t) \to -\infty$ as $|t| \to \infty$, and x achieves its maximum at a unique point in $\mathbb{R}$.

Proposition 3.1 (Theorem 6.1 in [9])**.** *Let (J_{1n}, J_{2n}) be a sequence of a pair of random mappings valued in $\mathbb{B}_{loc}(\mathbb{R}) \times \mathbb{B}_{loc}(\mathbb{R})$ and (T_{1n}, T_{2n}) be another sequence of random mappings into $\mathbb{R} \times \mathbb{R}$ such that:*

(i) $(J_{1n}, J_{2n}) \xrightarrow{d} (J_1, J_2)$, $\mathbb{P}((J_1, J_2) \in \mathbb{C}_{max}(\mathbb{R}) \times \mathbb{C}_{max}(\mathbb{R})) = 1$;
(ii) $T_{1n}, T_{2n} = O_p(1)$;
(iii) $J_{1n}(T_{1n}) \geq \sup_t J_{1n}(t) - \beta_{1n}$, *and* $J_{2n}(T_{2n}) \geq \sup_t J_{2n}(t) - \beta_{2n}$ *where* $\beta_{1n}, \beta_{2n} = o_p(1)$.

Then $(T_{1n}, T_{2n}) \xrightarrow{d} (T_1, T_2) := (\arg\max(J_1), \arg\max(J_2))$.

Another tool that simplifies the establishment of asymptotic theory related to the Grenander estimator is the Hungarian approximation ([13]): There are Brownian motions W_n and Brownian bridges $\mathbb{B}_n$ for which $\mathbb{B}_n(t) = W_n(t) - tW_n(1)$, for all $0 \leq t \leq 1$ and $n \geq 1$, and

$$\mathbb{F}_n(x) - F_0(x) = \frac{1}{\sqrt{n}}\mathbb{B}_n(F_0(x)) + R_n(x), \quad 0 \leq x < \infty, \tag{3.3}$$

where with probability one,

$$\sup_x |R_n(x)| = O\left(\frac{\log n}{n}\right). \tag{3.4}$$

Denote $\{W(t) : t \in [0,\infty)\}$ to be a standard Brownian motion over the positive real line with $W(0) = 0$ and $\{\mathbb{W}(t) : t \in \mathbb{R}\}$ to be a standard two-sided Brownian motion with $\mathbb{W}(0) = 0$. Let $\mathbb{Y}$ be distributed as $\arg\max_{t\in\mathbb{R}}\{\mathbb{W}(t) - t^2\}$, the (almost surely unique) location of the maximum of two-sided Brownian motion minus the parabola $t \mapsto t^2$.

Lastly, we also need the asymptotic distribution of $\hat{f}_n(x_0)$ at $x_0 \in (0,\infty)$ with $f_0'(x_0) < 0$, which is established in [19].

Proposition 3.2 (Theorem 6.3 in [19]; see also Theorem 2.1 in [7])**.** *Let* $x_0 \in (0,\infty)$. *If* $f_0'(x_0) < 0$, *then*

$$n^{1/3}\left|\frac{1}{2}f_0(x_0)f_0'(x_0)\right|^{-1/3} \cdot \{\hat{f}_n(x_0) - f_0(x_0)\} \xrightarrow{d} 2\mathbb{Y}. \tag{3.5}$$

Since the asymptotic distribution of $n^{1/3}f_0^{-1}(0)\{\hat{f}_n(x_0) - f_0(x_0)\}$ will appear frequently in the next few sections, for the sake of notational simplicity, we denote it by

$$\begin{aligned}\mathbb{Y}_0(S_0) &:= f_0^{-1}(0)\left|\frac{1}{2}f_0(x_0)f_0'(x_0)\right|^{1/3} 2\mathbb{Y} \\ &= \left|4S_0(x_0)S_0'(x_0)\int_0^\infty S_0(y)dy\right|^{1/3}\mathbb{Y}\end{aligned}$$

in terms of S_0. We also write $\mathbb{Y}_0 = \mathbb{Y}_0(S_0)$ without cause of ambiguity. In the rest of the article, $\{W(t) : t \in [0,\infty)\}$ and $\mathbb{Y}_0$ are defined on the same probability space and are independent of each other.

3.2. A Local Kiefer-Wolfowitz-type result

To establish the asymptotic distribution of $\hat{S}^S_{n,h_1,h_2}(x_0)$, we shall need an estimate of the difference between the empirical distribution function $\mathbb{F}_n$ and its least concave majorant $\hat{F}_n$ in a neighbourhood of x_0. The Kiefer-Wolfowitz result refers to the following statement in [12] concerning the maximum absolute distance between $\mathbb{F}_n$ and $\hat{F}_n$, where the underlying true distribution function F_0 is concave.

Theorem 3.3 (Theorem 1 in [12])**.** *Suppose that* $\tau < \infty$ *and* $\sup\{x : F_0(x) = 0\} = 0$. *If* F_0 *is concave, twice continuously differentiable on* $(0,\tau)$*,*

$$\frac{\sup_{0<x<\tau}(-f_0'(x))}{\inf_{0<x<\tau} f_0^2(x)} < \infty \text{ and } \inf_{0<x<\tau}\frac{(-f_0'(x))}{f_0^2(x)} > 0. \tag{3.6}$$

Then, with probability one,

$$\sup_{0\le x\le\tau} |\hat{F}_n(x) - \mathbb{F}_n(x)| = O(n^{-2/3}\log n).$$

Roughly speaking, this theorem holds if f_0 has a bounded support, and it is also bounded away from 0 with a continuous first order derivative f_0' whose negative values are both bounded from above and bounded away from 0. Since we only wish to establish the asymptotic distribution of the estimator of $S_0(x)$ at an interior point x, it may not be necessary to assume conditions over the whole support of f_0. More importantly, their proposed condition that f_0 is bounded away from 0 corresponds to the condition, under our setting, that S_0 is bounded away from 0, and this must lead to problematic formulation because S_0 is a survival function and should be ultimately equal to 0 at the endpoint; in the following, we aim to modify the conditions, so that a similar claim remains valid.

In view of this critical matter, we establish a local Kiefer-Wolfowitz result in the following Theorem 3.4 where it suffices to assume the following Condition C_f5 instead of (3.6).

C_f5 f_0 is continuously differentiable on $[0, x_0+2\delta]$ for some $\delta > 0$ and $0 < \inf_{t\in[0,x_0+2\delta]} |f_0'(x)| \le \sup_{t\in[0,x_0+2\delta]} |f_0'(x)| < \infty$.

Theorem 3.4. *Under C_f5, we have the rate of convergence,*

$$\sup_{x\in[0,x_0+\delta]} |\hat{F}_n(x) - \mathbb{F}_n(x)| = O_p(n^{-2/3}(\log n)^{2/3}).$$

This theorem is a direct consequence of a more general local Kiefer-Wolfowitz-type theorem given in Appendix B. Note that although a local Kiefer-Wolfowitz-type theorem is established in Theorem 2.5 of [4], it also requires the assumption that the density has a bounded support and to be bounded away from 0, which is thus not possible in our present setting. However, our proof follows a similar argument as that for Theorem 2.2 in [4] with some modifications where we only consider a region at a distance away from the right endpoint of the support.

3.3. Penalized NPMLE

In this subsection, we shall establish the asymptotic distribution of the penalized NPMLE $\hat{S}_n^P(x_0)$. Because $\hat{S}_n^P(x_0) = \hat{f}_n^P(x_0)/\hat{f}_n^P(0)$, to establish the asymptotic distribution of $\hat{S}_n^P(x_0)$, we shall first establish the joint asymptotic distribution of $\hat{f}_n^P(x_0)$ and $\hat{f}_n^P(0)$ in the following Theorem 3.5. Note that for the marginal distribution of $\hat{f}_n^P(x_0)$, Theorem 4 of [27] implies that under C_f4 with $\alpha_n = cn^{-(k+1)/(2k+1)}$ for some $k \ge 1$ and $c > 0$, we have

$$n^{\frac{k}{2k+1}}\{\hat{f}_n^P(0) - f_0(0)\} \xrightarrow{d} \sup_{t>0} \frac{W(t) - (c + \beta_{k+1}t^{k+1})}{t}, \tag{3.7}$$

where $\beta_{k+1} := -f_0^{(k)}(0) f_0^k(0)/(k+1)!$.

Theorem 3.5. *Under Conditions C_f1, C_f2, and C_f4 for some $k \ge 1$, if $\alpha_n = cn^{-(k+1)/(2k+1)}$ for some constant $c \in (0,\infty)$, then*

$$(n^{\frac{k}{2k+1}}\{\hat{f}_n^P(0) - f_0(0)\}, n^{1/3}\{\hat{f}_n^P(x_0) - f_0(x_0)\}) \tag{3.8}$$

converges in distribution to

$$\left(\sup_{t>0}\frac{W(t)-(c+\beta_{k+1}t^{k+1})}{t}, f_0(0)\mathbb{Y}_0\right).$$

Recall that $\{W(t):t\geq 0\}$ and $\mathbb{Y}_0$ are independent. The same holds for the other propositions that follow.

Corollary 3.1. *Under Conditions C_S1, C_S2, and C_S4 for some $k\geq 1$, and also assume that $\alpha_n = cn^{-(k+1)/(2k+1)}$ for some constant $c\in(0,\infty)$, then the followings hold.*

(a) If $k=1$, then $n^{1/3}\{\hat{S}_n^P(x_0)-S_0(x_0)\}$ converges in distribution to

$$-S_0(x_0)\int_0^\infty S_0(y)dy\sup_{0<t<\infty}\left\{\frac{W(t)-(c+\beta_2^S t^2)}{t}\right\}+\mathbb{Y}_0,$$

where $\beta_2^S := -S_0'(0)/[2\{\int_0^\infty S_0(y)dy\}^2]$.

(b) If $k>1$, then

$$n^{1/3}\{\hat{S}_n^P(x_0)-S_0(x_0)\}\xrightarrow{d}\mathbb{Y}_0.$$

Note that the constant c for the asymptotically optimal penalty parameter in the mean-squared error sense for $\hat{S}_n^P(x_0)$ is the same as that for $\hat{f}_n^P(0)$ for all $k\geq 1$, If $k=1$, this is true because of the independence between $\sup_{0<t<\infty}\{\frac{W(t)-(c+\beta_2^S t^2)}{t}\}$ and $\mathbb{Y}_0$ due to that of $\{W(t):t\geq 0\}$ and $\mathbb{Y}_0$. For $k>1$, the constant c does not appear in the limiting distribution.

Remark 3. *(1) In [27], the following condition is used in deriving the asymptotic distribution of the penalized estimator $\hat{f}_{n,\alpha}$:*

$$F_0(x)=f_0(0)x-f_1x^p+o(x^p)\ \text{as}\ x\downarrow 0, \tag{3.9}$$

where $0\leq f_1<\infty$ and $p>1$ and does not have to be a positive integer. If C_f4 holds, then (3.9) holds with $p=k+1$ and $f_1=-f_0^{(k)}(0)/(k+1)!$ by Taylor's theorem. On the other hand, there are examples where (3.9) holds but C_f4 is not satisfied. For example, if $f_0(x)=\frac{p}{p-1}(1-x^{p-1})I(x\in[0,1])$ for $p\in(1,2)$, then for $x\in[0,1]$,

$$F_0(x)=\frac{p}{p-1}x-\frac{1}{p-1}x^p,$$

but $f_0'(x)=-p/x^{2-p}$ for $x\in(0,1)$ so that $f_0'(0)=\infty$.

(2) The corresponding condition of (3.9) in terms of S_0 is

$$\int_0^x S_0(y)dy = x-S_1x^p+o(x^p)\ \text{as}\ x\downarrow 0, \tag{3.10}$$

where $0\leq S_1<\infty$. Similar to Remark 3 (1) above, if C_S4 holds, then (3.10) holds with $p=k+1$ and $S_1=-S_0^{(k)}(0)/(k+1)!$ by Taylor's theorem.

On the other hand, there are examples where (3.10) holds but C_S4 is not satisfied. For example, suppose that $S_0(x) = 1 - x^{p-1}$ for $x \in [0,1]$. Then $\int_0^x S_0(y)dy = x - x^p/p$ and $g_0(x) = (p-1)x^{p-2}$ for any $x \in [0,1]$. (3.10) is satisfied for any $p > 1$. If $p \in (1,2)$, then $S_0'(0) = -g_0(0) = -\infty$ so that C_S4 is not satisfied.

(3) *To simplify the presentation of the conditions and asymptotic properties of the different estimators considered here, we only consider C_f4 (resp. C_S4) or C_f4' (resp. C_S4') but not (3.9) (resp. (3.10)).*

3.4. Using a histogram estimator

Similar to the previous subsection, to establish the asymptotic distribution of $\hat{S}_n^H(x_0)$, we first establish the joint asymptotic distribution of the histogram estimator $\hat{f}_n^H(0)$ and the Grenander estimator at a fixed interior point $\hat{f}_n(x_0)$. The rate of convergence and the asymptotic distribution depend on the order of the bin width as well as the assumptions near $f_0(0)$.

Let $Z(\mu, \sigma^2)$ denote a normal random variable with mean μ and variance σ^2. In this article, Z is independent of $\mathbb{Y}_0$.

Theorem 3.6. *Under Conditions C_f1, C_f2, and C_f4 for some $k \geq 1$, if $b_n = cn^{-1/(2k+1)}$ for some constant $c \in (0,\infty)$, then,*

$$\Big(n^{\frac{k}{2k+1}}\{\hat{f}_n^H(0) - f_0(0)\}, n^{1/3}\{\hat{f}_n(x_0) - f_0(x_0)\}\Big)$$
$$\xrightarrow{d} \left(Z\left(\frac{f_0^{(k)}(0)}{(k+1)!}c^k, c^{-1}f_0(0)\right), f_0(0)\mathbb{Y}_0\right).$$

Corollary 3.2. *Under Conditions C_S1, C_S2, and C_S4 for some $k \geq 1$ and if $b_n = cn^{-1/(2k+1)}$ for some constant $c \in (0,\infty)$, then the followings hold.*

(a) *If $k = 1$, $n^{\frac{1}{3}}\{\hat{S}_n^H(x_0) - S_0(x_0)\}$ converges in distribution to*

$$Z\bigg(-\frac{c}{2}S_0(x_0)S_0'(0), c^{-1}\{S_0(x_0)\}^2\int_0^\infty S_0(y)dy\bigg) + \mathbb{Y}_0.$$

(b) *If $k > 1$,*

$$n^{\frac{1}{3}}\{\hat{S}_n^H(x_0) - S_0(x_0)\} \xrightarrow{d} \mathbb{Y}_0.$$

Note that when $k = 1$, the constant c for the asymptotically optimal bin width for $\hat{S}_n^H(x_0)$ in the mean-squared error sense is the same as that for $\hat{f}_n^H(0)$. This is true because of the independence between the normal random variable Z in Corollary 3.2 (a) and $\mathbb{Y}_0$. For $k > 1$, the constant c does not appear in the limiting distribution.

Remark 4. *In general, we can also consider the asymptotic results under (3.10) instead of C_S4 and the cases when $b_n = cn^{-\gamma}$, where γ belongs to one of the followings: (i) $1/(2p-1) < \gamma < 1$, (ii) $\gamma = 1/(2p-1)$ and (iii) $0 < \gamma < 1/(2p-1)$, where $p \in (1,\infty)$ as in (3.9); see Appendix D.1 for these considerations.*

3.5. Using a Grenander Estimator near 0

In [14], the asymptotic distribution of $\hat{f}_n(cn^{-\alpha})$ is established for different range of values of $\alpha \in (0,1)$ under $C_f 4'$. In the following Theorem 3.7, we first establish the joint asymptotic distribution of $\hat{f}_n(cn^{-\alpha})$ and $\hat{f}_n(x_0)$. Using the notation in [14], define $D[\mathcal{Z}(t)](a)$ as the right derivative of the least concave majorant (LCM) on $\mathbb{R}$ of the process $\mathcal{Z}$ at the point $t = a$, and define D_R similarly, where the LCM is restricted to the set $\{t \geq 0\}$. For the same value of k in $C_f 4'$ or $C_S 4'$, define

$$\begin{aligned} B_{2k} &:= \{f_0(0)^{1/2}|f_0^{(k)}(0)|^{-1}(k+1)!\}^{2/(2k+1)} \\ &= \left\{(k+1)!|S_0^k(0)|^{-1}\left(\int_0^\infty S_0(y)dy\right)^{1/2}\right\}^{2/(2k+1)}; \\ A_{2k} &:= \{B_{2k}/f_0(0)\}^{1/2} = \left\{B_{2k}\int_0^\infty S_0(y)dy\right\}^{1/2}. \end{aligned}$$

Theorem 3.7. *Under Conditions $C_f 1$, $C_f 2$, and $C_f 4'$ for some $k \geq 1$, if $\alpha = 1/(2k+1)$, then the couple $(n^{k/(2k+1)}\{\hat{f}_n(cn^{-\alpha}) - f_0(0)\}, n^{1/3}\{\hat{f}_n(x_0) - f_0(x_0)\})$ as a random vector converges in distribution to*

$$(A_{2k}^{-1}D_R[W(t) - t^{k+1}](cB_{2k}^{-1}), f_0(0)\mathbb{Y}_0).$$

Again, note that as a functional purely dependent on $W(\cdot)$, $D_R[W(t) - t^{k+1}](cB_{2k}^{-1})$ and $\mathbb{Y}_0$ are independent of each other.

Corollary 3.3. *Under Conditions $C_S 1$, $C_S 2$, and $C_S 4'$ for some $k \geq 1$, if $\alpha = 1/(2k+1)$, then $n^{1/3}\{\hat{S}_n^N(x_0) - S_0(x_0)\} \xrightarrow{d}$ converges in distribution to*

$$\begin{cases} S_0(x_0)\int_0^\infty S_0(y)dy \cdot A_{2k}^{-1} \cdot D_R[W(t) - t^{k+1}](cB_{2k}^{-1}) + \mathbb{Y}_0, & \text{if } k = 1; \\ \mathbb{Y}_0, & \text{if } k > 1, \end{cases}$$

where as before $D_R[W(t) - t^{k+1}](cB_{2k}^{-1})$ and $\mathbb{Y}_0$ are independent.

Remark 5. *Again, for $k = 1$, the two summands are independent of each other. It is also possible to establish the asymptotic results when $1/(2k+1) < \alpha < 1$; see Appendix D.2 for details.*

3.6. Based on smoothing $\hat{f}_n$

To establish the asymptotic distribution of $\hat{S}^S_{n,h_1,h_2}(x_0)$, we shall first obtain the joint asymptotic distribution of $(\hat{f}^S_{n,h_1}(x_0) - f_0(x_0), \hat{f}^S_{n,h_2}(0) - f_0(0))$. Recall that K_0 is the boundary kernel defined in (2.11) when $x = 0$. Fix $c_1 > 0$ and $c_2 > 0$. Denote

$$\mu_{x_0}^{(2)} := \frac{1}{2}c_1^2 f_0''(x_0)\int_{-1}^1 y^2 K(y)dy,$$

$$\mu_0^{(k)} := \frac{(-1)^k}{k!} c_2^k f_0^{(k)}(0) \int_{-1}^{0} y^k K_0(y) dy.$$

Let D be a 2×2 diagonal matrix with diagonal elements

$$d_{11} := c_1^{-1} f_0(x) \int_{-1}^{1} K^2(y) dy,$$
$$d_{22} := c_2^{-1} f_0(0) \int_{-1}^{0} K_0^2(y) dy.$$

Since we assume C_f5 to obtain an estimate of the difference between $\hat{F}_n$ and $\mathbb{F}_n$, we are going to impose a modified C_f4:

C_f4'' for some $k \geq 2$, $0 < |f_0^{(k)}(0)| < \infty$ and $f_0^{(i)}(0) = 0$ for $2 \leq i \leq k-1$.

Theorem 3.8. *Under Conditions C_f1, C_f3, C_f4'' for some $k \geq 2$, and C_f5, if $h_1 = h_{1n} = c_1 n^{-1/5}$ and $h_2 = h_{2n} = c_2 n^{-1/(2k+1)}$, then the sequence of random vectors $\left(n^{2/5}\{\hat{f}_{n,h_1}^S(x_0) - f_0(x_0)\}, n^{k/(2k+1)}\{\hat{f}_{n,h_2}^S(0) - f_0(0)\}\right)$ converges in distribution to $Z((\mu_{x_0}^{(2)}, \mu_0^{(k)})^\top, D)$.*

To state the main result in this subsection, we define additional notations. Denote

$$\mu_1^{(2)} := \frac{1}{2} c_1^2 S_0''(x_0) \int_{-1}^{1} y^2 K(y) dy,$$
$$\sigma_1^2 := c_1^{-1} S_0(x_0) \int_0^{\infty} S_0(y) dy \int_{-1}^{1} K^2(y) dy,$$
$$\mu_2^{(k)} := \frac{(-1)^{k+1}}{k!} c_2^k S_0^k(0) S_0(x_0) \int_{-1}^{0} y^k K_0(y) dy,$$
$$\sigma_2^2 := c_2^{-1} S_0^2(x_0) \int_0^{\infty} S_0(y) dy \int_{-1}^{0} K_0^2(y) dy.$$

The corresponding conditions of C_f4'' and C_f5 in terms of S_0 are

C_S4'' for some $k \geq 2$, $0 < |S_0^{(k)}(0)| < \infty$ and $S_0^{(i)}(0) = 0$ for $2 \leq i \leq k-1$;

C_S5 S_0 is continuously differentiable on $[0, x_0 + 2\delta]$ for some $\delta > 0$ and $0 < \inf_{t \in [0, x_0+2\delta]} |S_0'(x)| \leq \sup_{t \in [0, x_0+2\delta]} |S_0'(x)| < \infty$.

Corollary 3.4. *Under Conditions C_S1, C_S3, C_S4'' for some $k \geq 2$, and C_S5, if $h_1 = h_{1n} = c_1 n^{1/5}$, $h_2 = h_{2n} = c_2 n^{-1/(2k+1)}$, then $n^{2/5}\{\hat{S}_{n,h_1,h_2}^S(x_0) - S_0(x_0)\}$ converges in distribution to*

$$\begin{cases} Z(\mu_1^{(2)} + \mu_2^{(k)}, \sigma_1^2 + \sigma_2^2), & \text{if } k = 2; \\ Z(\mu_1^{(2)}, \sigma_1^2), & \text{if } k > 2. \end{cases}$$

Remark 6. *(i) It is also possible to establish the asymptotic results with the same h_1 but $h_2 = c_2 n^{-\alpha}$ under the two different cases (i): $1/(2k+1) < \alpha < 1/3$ and (ii): $0 < \alpha < 1/(2k+1)$; see Appendix D.3 for details.*

(ii) It is also possible to consider estimator in the form

$$\frac{\hat{f}_n(x_0)}{\hat{f}^S_{n,h}(0)},$$

in which we only use the NPMLE $\hat{f}_n$ with smoothing effect at 0. The corresponding asymptotic distribution can also be established using the methods developed here. In particular, $\hat{f}_n(x_0)$ and $\hat{f}^S_{n,h}(0)$ will be asymptotically independent as n goes to infinity. The argument follows that found in the proof of Theorem 3.8 and the details are omitted.

3.7. Conditional distribution of Y given $Y > y_0$

Similar to previous sections, it can be shown that $\hat{f}_n(x_0)$ and $\hat{f}_n(y_0)$ are asymptotically independent when $x_0 \neq y_0$. Hence, it is straightforward to obtain the following result regarding the asymptotic distribution of the estimator of the conditional distribution of Y given $Y > y_0$.

Theorem 3.9. *For fixed $y_0 \in (0, \tau)$ and $x_0 \in (y_0, \tau)$, assume that $|S'(x_0)| \in (0, \infty)$ and $|S'(y_0)| \in (0, \infty)$. Then*

$$\begin{aligned} n^{1/3}\{\hat{S}_{|y_0}(x_0) - S_{|y_0}(x_0)\} \xrightarrow{d} \frac{1}{S_0(y_0)} &\left|4S_0(x_0)S_0'(x_0)\int S_0(y)dy\right|^{1/3} \mathbb{Y}_1 \\ &+ \frac{S_0(x_0)}{S_0^2(y_0)}\left|4S_0(y_0)S_0'(y_0)\int S_0(y)dy\right|^{1/3} \mathbb{Y}_2, \end{aligned}$$

where $\mathbb{Y}_1, \mathbb{Y}_2 \stackrel{d}{=} \arg\max_{t\mathbb{R}}\{\mathbb{W}(t) - t^2\}$, and $\mathbb{Y}_1$ and $\mathbb{Y}_2$ are independent of each other.

4. Simulation studies

In this section, we perform some simulation studies to investigate some properties and compare the performance of the four estimators considered in this paper. All the estimators require some tuning parameters. We consider two methods of choosing a suitable tuning parameter. The first one is using the asymptotic mean-square optimal results and using some plug-in estimate to estimate the constant involved. Note that the asymptotic mean-square optimal choice also depends on the conditions near $f_0(0)$, in particular, whether $f_0'(0) = 0$ or $f_0'(0) < 0$, which is generally unknown. In this simulation studies, we simply use the rate that is optimal assuming $f_0'(0) < 0$. The second method we consider is to use least-squares cross validation. For this method, we use f_0 as the criterion function although we are interested in S_0 because it is easier to find a nearly unbiased estimator of f_0. For the estimator based on the smoothed MLE, we only use the second method based on [8].

From Section 3, we know that all the estimators at 0 and the estimators at x_0 considered in this paper are asymptotically independent, where x_0 is in the interior support of f_0. In Table 1, we present the sample correlations of the estimators at 0 and x_0 based on different methods at $x_0 = 0.25, 0.5, 1$, where the samples are from the standard exponential distribution. The sample correlations computed are based on 10000 replications of sample sizes $n = 50, 200, 500, 10000$, where the tuning parameters for $\hat{f}_n^P(0)$, $\hat{f}_n^H(0)$, $\hat{f}_n^N(0)$ and $\hat{f}_n^S(0)$ are $0.5n^{-2/3}, 0.5n^{-1/3}$, $0.5n^{-1/3}$, $\min(2n^{-1/5}, \max_{i=1,\ldots,n} X_i/3)$, respectively. From Table 1, we see that the correlations are getting closer to 0 as the sample size increases as expected. It can also be seen that the SMLE $\hat{f}_n^S(x_0)$ at $x_0 = 0.25$ has higher correlation with $\hat{f}_n^S(0)$ compared with other types of estimators. This is probably because for small to moderate sample sizes, the bandwidth is not too small so that $\hat{f}_n^S(x_0)$ at $x_0 = 0.25$ and $\hat{f}_n^S(0)$ both use the values of $\hat{f}_n(x)$ over a large common interval. In addition, the correlation between the SMLE $\hat{f}_n^S(x_0)$ and $\hat{f}_n^S(0)$ is higher, in absolute value, in almost all cases.

TABLE 1
Sample correlations of the estimators at 0 *and* x_0 *based on different methods.*

	$n = 50$			$n = 200$		
x_0	0.25	0.5	1	0.25	0.5	1
$\mathrm{Cor}(\hat{f}_n^P(x_0), \hat{f}_n^P(0))$	0.16	−0.15	−0.20	−0.01	−0.13	−0.12
$\mathrm{Cor}(\hat{f}_n(x_0), \hat{f}_n^H(0))$	−0.02	−0.20	−0.20	−0.04	−0.15	−0.13
$\mathrm{Cor}(\hat{f}_n(x_0), \hat{f}_n^N(0))$	0.36	−0.12	−0.28	0.12	−0.18	−0.16
$\mathrm{Cor}(\hat{f}_n^S(x_0), \hat{f}_n^S(0))$	0.85	0.26	−0.55	0.70	−0.13	−0.38
	$n = 500$			$n = 10000$		
x_0	0.25	0.5	1	0.25	0.5	1
$\mathrm{Cor}(\hat{f}_n^P(x_0), \hat{f}_n^P(0))$	−0.03	−0.11	−0.09	−0.04	−0.05	−0.03
$\mathrm{Cor}(\hat{f}_n(x_0), \hat{f}_n^H(0))$	−0.08	−0.10	−0.07	−0.06	−0.04	−0.02
$\mathrm{Cor}(\hat{f}_n(x_0), \hat{f}_n^N(0))$	−0.04	−0.13	−0.11	−0.06	−0.05	−0.05
$\mathrm{Cor}(\hat{f}_n^S(x_0), \hat{f}_n^S(0))$	0.54	−0.27	−0.26	−0.11	−0.15	−0.10

Since the estimate of $S_0(x_0)$ depends strongly on that of $f_0(0)$, we first compare the estimators of $f_0(0)$ consider in this paper. Similar simulation studies have been performed in [14] and [10]. [14] compared the penalized estimator using the adaptive choice considered in [23] with the estimator using Grenander estimator near 0. In addition to the two estimators considered in [14], [10] also considered the histogram estimator and the Bayesian estimator proposed in their paper. In this paper, we shall only compare the four frequentist estimators considered here. Note that all the estimators require the specification of some tuning parameters. We follow similar specifications in [14] and [10] by using the adaptive method in [23] for the penalized estimator and estimating the asymptotic mean-square optimal choice of the tuning parameter assuming $|f_0'(0)| > 0$ for $\hat{f}_n^H(0)$ and $\hat{f}_n^N(0)$. From the asymptotic independence results in Section 3, we also know that the asymptotic mean-sqaure optimal choice for $\hat{S}_n^H(0)$ and $\hat{S}_n^N(0)$ are the same as those for $\hat{f}_n^H(0)$ and $\hat{f}_n^N(0)$, respectively. We also provide some numerical comparison by choosing the tuning parameters using least-squares cross-validation (LSCV). To be specific, with a slight abuse of notation, we define the following:

1. Denote $\hat{f}^P_{n,0.5} = \hat{f}^P_{n,\alpha_n}(0)$ with $\alpha_n = 0.649\hat{\beta}_n^{-1/3}n^{-2/3}$ with
$$\hat{\beta}_n = \max\left\{\hat{f}^P_{n,c_0 n^{-2/3}}(0)\frac{\hat{f}^P_{n,c_0 n^{-2/3}}(0) - \hat{f}^P_{n,c_0 n^{-2/3}}(x_m)}{2x_m}, n^{-1/3}\right\},$$
where x_m is the second point of jump of $\hat{f}^P_{n,c_0 n^{-2/3}}$ and $c_0 = 0.5$. Similarly, define $\hat{f}^P_{n,1}$ to be the same estimator as $\hat{f}^P_{n,0.5}$ except that $c_0 = 1$. Here $c_0 n^{-2/3}$ is an initial tuning parameter that is used to obtain $\hat{\beta}_n$, an estimate of the unknown quantity $\beta = \frac{1}{2}f_0(0)|f_0'(0)|$ that appears in the asymptotic mean-square optimal choice.
2. Denote $\hat{f}^H_n = \hat{f}^H_{n,b_n}(0)$, where $b_n = 2^{-1/3}\hat{B}_{21}n^{-1/3}$, where
$$\hat{B}_{21} = 4^{1/3}\hat{f}_n(n^{-1/3})^{1/3}|\hat{f}_n'(0)|^{-2/3}, \tag{4.1}$$
and
$$\hat{f}_n'(0) = \min\{n^{1/6}\{\hat{f}_n(n^{-1/6}) - \hat{f}_n(n^{-1/6})\}, -n^{-1/3}\}.$$
3. Denote $\hat{f}^N_n = \hat{f}^N_{n,c_n}(0)$ to be the estimator using Grenander estimator near 0, where $c_n = 0.345\hat{B}_{21}n^{-1/3}$ with $\hat{B}_{21}$ as in (4.1).
4. Denote $\hat{f}^P_{n,CV} = \hat{f}^P_{n,\alpha_{CV}}(0)$, where
$$\alpha_{CV} := \arg\min_{\alpha}\left\{\int_0^\infty \hat{f}^2_{n,\alpha}(x)dx - 2\sum_{i=1}^n \hat{f}^{-i}_{n,\alpha}(X_i)\right\},$$
with $\hat{f}^{-i}_{n,\alpha}$ being the penalized Grenander estimator without using the ith observation. We search α over a grid from $0.2n^{-2/3}$ to $2n^{-2/3}$.
5. Denote $\hat{f}^H_{n,CV} = \hat{f}^H_{n,b_{CV}}(0)$. When using the histogram estimator to estimate $f_0(0)$, there are multiple ways to normalize the modified Grenander estimator. Given $\hat{f}^H_{n,b_n}(0)$ for any b_n, we first obtain the corresponding estimator $\hat{S}^H_n$ of S_0 through (2.10). Then the normalized estimator of $f_0(x)$, denoted by $\hat{f}^H_{n,b_n}(x)$ is given by $\hat{S}^H_{n,b_n}(x)/\int_0^\infty \hat{S}^H_{n,b_n}(y)dy$. Then, b_{CV} is defined as
$$b_{CV} := \arg\min_{b}\left\{\int_0^\infty \hat{f}^H_{n,b}(x)^2dx - 2\sum_{i=1}^n \hat{f}^{H,-i}_{n,b}(X_i)\right\},$$
where $\hat{f}^{H,-i}_{n,b}$ is the estimator without the ith observation. We search b over a grid from $0.2n^{-1/3}$ to $2n^{-1/3}$.
6. $\hat{f}^N_{n,CV}$ is defined similarly as $\hat{f}^H_{n,CV}$ and we search c over a grid from $0.2n^{-1/3}$ to $2n^{-1/3}$.

7. For simplicity, we only use one bandwidth for the smoothed estimator at all values. Denote $\hat{f}^S_{n,CV} = \hat{f}^S_{n,h_{CV}}(0)$. Following equation (9.40) in [8], h_{CV} is chosen to minimize

$$\int_0^\infty \hat{f}^S_{n,h}(x)^2 dx - \frac{2}{n-1}\sum_{i=1}^n \hat{f}^S_{n,h}(X_i) + \frac{2K(0)}{(n-1)h}.$$

In the implementation, we evaluate the above expression with h at grid points from 0.1 to $\max_{i=1,\ldots,n} X_i/2 - 0.1$.

We considered two probability distributions to generate the samples. The first one is the exponential distribution, where $f_0'(0) < 0$. For the rate parameter λ, we consider $\lambda = 1$ and $\lambda = 2$. The second one is the half-normal distribution where $f_0'(0) = 0$ but $|f_0''(0)| > 0$. For the scale parameter σ, we consider $\sigma = 1$ and $\sigma = 0.5$. In Tables 2 and 3, we see that suitable choice of tuning parameters or preliminary estimates is needed to obtain good results. For example, $\hat{f}^H_n$ and $\hat{f}^N_n$ work well in the half-normal distribution with $\sigma = 1$, but have large biases for the other distributions. The estimators with tuning parameters chosen using cross-validation tend to have smaller biases and larger variances than their counterparts. While the smoothed estimator has a faster rate of convergence, it does not always outperform the other estimators in terms of MSE for the half-normal distributions. It tends to give a larger bias. This can be due to the fact that the first derivative vanishes at 0 for the half-normal distribution. In such case, kernel smoothing with boundary correction using the reflection method could perform better. Another possible reason is that the bandwidth is chosen using the L^2-loss of the whole density, not the single point at 0.

TABLE 2
Bias, variance and mean-squared error (MSE) of each estimator at 0 *where the samples are from an exponential distribution.*

Exp(1)		$\hat{f}^P_{n,0.5}$	$\hat{f}^P_{n,1}$	$\hat{f}^H_n$	$\hat{f}^N_n$	$\hat{f}^P_{n,CV}$	$\hat{f}^H_{n,CV}$	$\hat{f}^N_{n,CV}$	$\hat{f}^S_{n,CV}$
$n = 50$	Bias	0.012	−0.054	−0.176	−0.214	−0.061	−0.012	−0.115	0.024
	Var	0.118	0.072	0.032	0.044	0.128	0.163	0.098	0.091
	MSE	0.118	0.075	0.063	0.090	0.132	0.163	0.111	0.092
$n = 200$	Bias	0.015	−0.042	−0.139	−0.133	−0.023	−0.025	−0.076	−0.027
	Var	0.063	0.036	0.018	0.024	0.083	0.055	0.040	0.026
	MSE	0.063	0.038	0.038	0.042	0.084	0.056	0.046	0.027
$n = 500$	Bias	0.010	−0.034	−0.107	−0.096	−0.007	−0.011	−0.052	−0.014
	Var	0.036	0.021	0.012	0.014	0.056	0.036	0.025	0.013
	MSE	0.036	0.022	0.023	0.023	0.056	0.036	0.028	0.013
Exp(2)									
$n = 50$	Bias	−0.131	−0.252	−0.618	−0.692	−0.191	−0.115	−0.291	0.046
	Var	0.299	0.199	0.096	0.170	0.396	0.397	0.299	0.394
	MSE	0.316	0.262	0.478	0.649	0.433	0.410	0.384	0.396
$n = 200$	Bias	−0.091	−0.173	−0.384	−0.380	−0.176	−0.151	−0.242	−0.037
	Var	0.145	0.097	0.058	0.096	0.193	0.167	0.133	0.109
	MSE	0.153	0.127	0.206	0.240	0.224	0.190	0.192	0.11
$n = 500$	Bias	−0.075	−0.135	−0.265	−0.242	−0.138	−0.128	−0.186	−0.045
	Var	0.083	0.057	0.032	0.054	0.115	0.088	0.073	0.045
	MSE	0.089	0.075	0.102	0.113	0.134	0.105	0.108	0.047

TABLE 3

Bias, variance and mean-squared error (MSE) of each estimator at 0 *where the samples are from a half-normal distribution.*

Hnorm(1)		$\hat{f}^P_{n,0.5}$	$\hat{f}^P_{n,1}$	$\hat{f}^H_n$	$\hat{f}^N_n$	$\hat{f}^P_{n,CV}$	$\hat{f}^H_{n,CV}$	$\hat{f}^N_{n,CV}$	$\hat{f}^S_{n,CV}$
$n=50$	Bias	0.112	0.071	0.000	−0.002	0.072	0.086	0.037	0.246
	Var	0.066	0.036	0.020	0.025	0.089	0.086	0.043	0.056
	MSE	0.078	0.041	0.020	0.025	0.094	0.094	0.044	0.117
$n=200$	Bias	0.074	0.041	−0.006	0.013	0.047	0.037	0.030	0.112
	Var	0.031	0.016	0.008	0.010	0.037	0.031	0.017	0.012
	MSE	0.036	0.018	0.008	0.010	0.039	0.032	0.017	0.024
$n=500$	Bias	0.060	0.032	−0.007	0.014	0.049	0.040	0.037	0.076
	Var	0.020	0.009	0.005	0.006	0.025	0.020	0.012	0.005
	MSE	0.023	0.010	0.005	0.006	0.027	0.022	0.013	0.011
Hnorm(0.5)									
$n=50$	Bias	0.113	0.041	−0.174	−0.162	0.007	0.076	−0.036	0.552
	Var	0.158	0.100	0.080	0.109	0.159	0.185	0.137	0.286
	MSE	0.171	0.102	0.111	0.135	0.159	0.191	0.138	0.590
$n=200$	Bias	0.075	0.025	−0.096	−0.050	0.002	0.050	−0.002	0.232
	Var	0.071	0.042	0.042	0.042	0.066	0.074	0.060	0.040
	MSE	0.076	0.043	0.051	0.044	0.066	0.076	0.060	0.094
$n=500$	Bias	0.056	0.017	−0.058	−0.017	0.021	0.039	0.009	0.167
	Var	0.040	0.023	0.025	0.022	0.050	0.044	0.032	0.023
	MSE	0.043	0.023	0.028	0.022	0.051	0.045	0.033	0.051

TABLE 4

Estimates of $E(\int(\tilde{f}_n - f_0)^2)$, *where* $\tilde{f}_n$ *is the density obtained using* $\hat{f}^P_{n,0.5}$, $\hat{f}^P_{n,1}$, $\hat{f}^H_n$, $\hat{f}^N_n$, $\hat{f}^P_{n,CV}$, $\hat{f}^H_{n,CV}$, $\hat{f}^N_{n,CV}$, $\hat{f}^S_{n,CV}$.

Exp(1)	$\hat{f}^P_{n,0.5}$	$\hat{f}^P_{n,1}$	$\hat{f}^H_n$	$\hat{f}^N_n$	$\hat{f}^P_{n,CV}$	$\hat{f}^H_{n,CV}$	$\hat{f}^N_{n,CV}$	$\hat{f}^S_{n,CV}$
$n=50$	0.036	0.034	0.031	0.034	0.044	0.039	0.041	0.014
$n=200$	0.014	0.014	0.014	0.014	0.015	0.014	0.015	0.004
$n=500$	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.002
Exp(2)								
$n=50$	0.071	0.068	0.066	0.084	0.083	0.071	0.080	0.027
$n=200$	0.029	0.028	0.028	0.032	0.033	0.029	0.032	0.009
$n=500$	0.015	0.015	0.015	0.016	0.018	0.016	0.017	0.004
Hnorm(1)								
$n=50$	0.031	0.029	0.028	0.029	0.034	0.031	0.031	0.019
$n=200$	0.012	0.011	0.011	0.012	0.013	0.013	0.012	0.005
$n=500$	0.006	0.006	0.006	0.006	0.007	0.007	0.006	0.003
Hnorm(0.5)								
$n=50$	0.059	0.056	0.052	0.057	0.067	0.058	0.065	0.039
$n=200$	0.023	0.022	0.022	0.023	0.024	0.023	0.024	0.009
$n=500$	0.012	0.012	0.012	0.012	0.013	0.013	0.013	0.005

Tables 4 and 5 give the estimates of the expected L_2-error of estimating f_0 and S_0 using different estimators, respectively. For the former one, the smoothed MLE gives the smallest error in all the scenarios. For the L_2-error of estimating S_0, smoothed MLE gives smaller errors in the exponential distributions but similar errors in the half-normal distributions as the other estimators. Tables 6 and 7 show the biases, variances and mean-squared errors of the estimators $\hat{S}_{|y_0}(x_0)$ of the conditional survival function $S_{|y_0}(x_0)$ at two different values of x_0 and y_0. As expected, the biases are small.

TABLE 5

Estimates of $E(\int(\tilde{S}_n - S_0)^2)$, where $\tilde{S}_n$ is the survival function obtained using $\hat{f}^P_{n,0.5}$, $\hat{f}^P_{n,1}$, $\hat{f}^H_n$, $\hat{f}^N_n$, $\hat{f}^P_{n,CV}$, $\hat{f}^H_{n,CV}$, $\hat{f}^N_{n,CV}$, $\hat{f}^S_{n,CV}$.

Exp(1)	$\hat{S}^P_{n,0.5}$	$\hat{S}^P_{n,1}$	$\hat{S}^H_n$	$\hat{S}^N_n$	$\hat{S}^P_{n,CV}$	$\hat{S}^H_{n,CV}$	$\hat{S}^N_{n,CV}$	$\hat{S}^S_{n,CV}$
$n = 50$	0.069	0.069	0.068	0.097	0.098	0.087	0.100	0.024
$n = 200$	0.032	0.032	0.035	0.040	0.040	0.037	0.040	0.010
$n = 500$	0.019	0.018	0.021	0.022	0.025	0.021	0.023	0.006
Exp(2)								
$n = 50$	0.039	0.043	0.055	0.089	0.062	0.045	0.065	0.011
$n = 200$	0.017	0.017	0.023	0.028	0.025	0.020	0.025	0.006
$n = 500$	0.009	0.010	0.011	0.013	0.014	0.011	0.013	0.003
Hnorm(1)								
$n = 50$	0.065	0.056	0.049	0.053	0.069	0.077	0.064	0.059
$n = 200$	0.034	0.027	0.023	0.025	0.035	0.040	0.030	0.022
$n = 500$	0.021	0.016	0.013	0.014	0.022	0.024	0.018	0.014
Hnorm(0.5)								
$n = 50$	0.030	0.029	0.028	0.033	0.040	0.033	0.039	0.031
$n = 200$	0.014	0.012	0.013	0.013	0.017	0.016	0.015	0.012
$n = 500$	0.008	0.007	0.007	0.007	0.010	0.009	0.008	0.007

TABLE 6

Bias, variance and mean-squared error of $\hat{S}_{|y_0}(x_0)$ where f_0 is the exponential distribution.

Exp(1)		$\hat{S}_{0.1}(0.5)$	$\hat{S}_{0.1}(1)$	$\hat{S}_{0.2}(0.5)$	$\hat{S}_{0.2}(1)$
$n = 50$	Bias	−0.027	−0.01	0.007	0.016
	Var	0.049	0.031	0.047	0.038
	MSE	0.05	0.031	0.047	0.038
$n = 200$	Bias	−0.01	−0.005	0.008	0.008
	Var	0.024	0.011	0.024	0.013
	MSE	0.024	0.011	0.024	0.013
$n = 500$	Bias	−0.007	−0.007	−0.002	−0.003
	Var	0.013	0.006	0.014	0.007
	MSE	0.013	0.006	0.014	0.007
Exp(2)					
$n = 50$	Bias	0.015	−0.009	0.037	−0.002
	Var	0.036	0.007	0.05	0.012
	MSE	0.036	0.007	0.051	0.012
$n = 200$	Bias	0.004	−0.003	0.015	0.001
	Var	0.014	0.003	0.021	0.004
	MSE	0.014	0.003	0.021	0.004
$n = 500$	Bias	0	0.002	0.004	0.004
	Var	0.006	0.001	0.011	0.002
	MSE	0.006	0.001	0.011	0.002

5. Application to a health survey without follow-up

Most surveys are administered cross-sectionally without follow-up data, due to cost and logistic reasons. Partial survival information in the form of backward recurrence times are frequently collected in health surveys. However, without any prospective follow-up, survival times being collected are all censored and conventional statistical methods are not applicable. These partial survival information being collected are therefore seldom analyzed. An exception in given in [17], who examined the associations between childhood adversities and the dura-

Table 7
Bias, variance and mean-squared error of $\hat{S}_{|y_0}(x_0)$ where f_0 is the half-normal distribution.

Hnorm(1)		$\hat{S}_{0.1}(0.5)$	$\hat{S}_{0.1}(1)$	$\hat{S}_{0.2}(0.5)$	$\hat{S}_{0.2}(1)$
$n = 50$	Bias	−0.109	−0.097	−0.056	−0.06
	Var	0.042	0.036	0.034	0.039
	MSE	0.054	0.046	0.037	0.043
$n = 200$	Bias	−0.074	−0.045	−0.035	−0.017
	Var	0.019	0.019	0.016	0.02
	MSE	0.025	0.021	0.018	0.02
$n = 500$	Bias	−0.045	−0.029	−0.022	−0.013
	Var	0.011	0.008	0.01	0.009
	MSE	0.013	0.009	0.01	0.009
Hnorm(0.5)					
$n = 50$	Bias	−0.052	−0.027	−0.01	−0.019
	Var	0.041	0.007	0.045	0.009
	MSE	0.044	0.007	0.045	0.009
$n = 200$	Bias	−0.019	−0.007	0.001	−0.003
	Var	0.019	0.003	0.022	0.003
	MSE	0.019	0.003	0.022	0.003
$n = 500$	Bias	−0.009	−0.005	0.006	−0.002
	Var	0.009	0.002	0.011	0.002
	MSE	0.009	0.002	0.011	0.002

tions of adult mental disorders using backward recurrence times collected from a nationally representative sample from the National Comorbidity Survey Replication. We analyzed a different survival outcome collected in the same survey, the duration between suicidal thoughts. Although suicidal thoughts are recurrent events, cross-sectional surveys like the one we have usually collect the most recent onset of the recurrent events. The survey was administered in 2001–2002, and collected the time of last suicidal thoughts from 1010 respondents with ages between 18 and 91.

The left panel of Figure 1 shows a histogram of the observed backward recurrence data, where the observed data apparently has a decreasing density. The right panel of Figure 1 shows the estimator based on the penalized Grenander estimator $\hat{S}_{p,CV}$, where the tuning parameter is chosen using cross-validation (the step function in black), and the one based on the smoothed Grenander estimator (the smoothed curve in red). Only these two estimators are shown in the figure because the other non-smooth estimates are similar to $\hat{S}_{p,CV}$. Table 8 shows the estimated values of the survival functions using different estimators at different times. The values are in generally similar for various non-smooth estimators.

6. Discussion

In this article, we considered four different classes of estimators of the survival function for current duration data under stationarity. We established the asymptotic distribution of those estimators under various regularity conditions of S_0 at 0 and the assumption that $S_0'(x_0) < 0$. If S_0 contains a flat region so that $S_0'(x) = 0$, then the corresponding f_0 will have a flat region and the Grenander

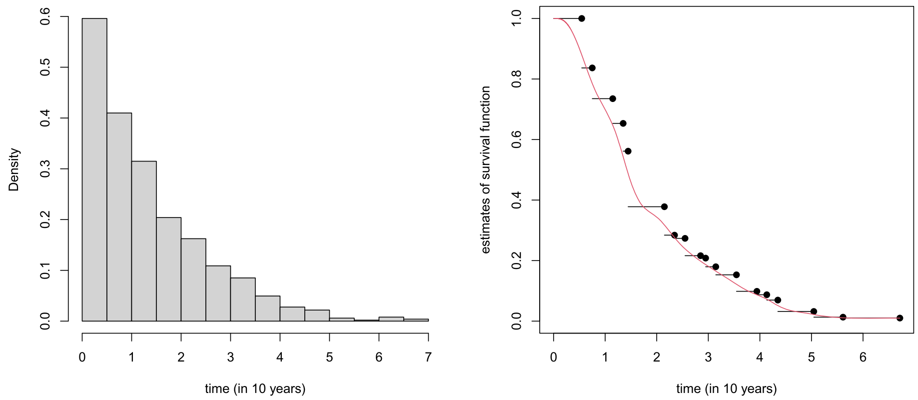

FIG 1. *The left figure shows a histogram of the observed data, where the observed data apparently has a decreasing density. The right figure shows the estimator of the survival function based on the penalized Grenander estimator (the step function in black) and the one based on the smoothed Grenander estimator (the smoothed curve in red).*

TABLE 8
Estimated values of the survival functions at different times (in 10 years) using different estimators.

Time	$\hat{S}^P_{n,0.5}$	$\hat{S}^P_{n,1}$	$\hat{S}^H_n$	$\hat{S}^N_n$	$\hat{S}^P_{n,CV}$	$\hat{S}^H_{n,CV}$	$\hat{S}^N_{n,CV}$	$\hat{S}^S_{n,CV}$
0.25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.984
0.5	1.000	1.000	1.000	0.977	1.000	1.000	1.000	0.901
0.75	0.737	0.737	0.769	0.713	0.735	0.741	0.729	0.785
1	0.737	0.737	0.769	0.713	0.735	0.741	0.729	0.699
2	0.379	0.379	0.395	0.367	0.378	0.381	0.375	0.344
3	0.180	0.180	0.188	0.174	0.180	0.181	0.178	0.181
4	0.087	0.087	0.091	0.085	0.087	0.088	0.087	0.083
5	0.032	0.032	0.033	0.031	0.032	0.032	0.032	0.023

estimator at any point inside the flat region converges at $\sqrt{n}$ instead of $n^{1/3}$ according to Theorem 6.4 in [1]. When such a case is considered, the asymptotic distribution of the estimators of $S_0(x_0)$ discussed in this article can be established easily because $\hat{f}_n(x_0)$ converges at a faster rate than any of the estimator of $f_0(0)$ considered here. On the other hand, the implication of S_0 having a flat region is that the survival time cannot occur in this region. However, the observed variable can nevertheless occur at that point due to multiplicative censoring. In establishing the asymptotic results, we also relaxed a bounded density condition for a local Kiefer-Wolfowitz-type result in a general setting which is of independent interest.

As expected from the faster rate of convergence of the smoothed estimator compared with the non-smoothed one, our numerical results confirm that the smoothed estimator can provide a more accurate estimate for both f_0 and S_0. Therefore, if one is willing to impose more smoothness assumptions on the underlying survival function, the smoothed estimator will be preferred. The bandwidth chosen using least-squares cross-validation also seem to work well in that case. Another advantage of the smoothed estimator is that the estimated survival probability will not be equal to 1 at values close to 0 so that it can give more sensible estimate. For the non-smoothed estimators, they performed similarly in simulation studies and share the same rate of convergence, with the estimators based on the histogram estimator and local Grenander estimator perform slightly better than the penalized estimator in terms of the L_2-loss.

Regarding the derivative condition of the order k in Condition C_f4 (or C_S4), similar conditions are also imposed in conventional kernel density estimation problems at the interior point. However, it is in generally difficult to determine this order k in any kernel density estimation problems, so many papers just fix a certain k such as $k = 1$. The presented theoretical results are mainly for understanding the limiting convergence rates of the estimators under different precise conditions. In addition, the unknown constant multiple in the optimal bandwidth affect practical performance as in conventional kernel density estimation. We therefore recommend using data-adaptive approaches like least-squares cross-validation illustrated in the simulation studies to choose the tuning parameters instead of trying to estimate the order of k. In addition, the higher the value of the k is, the faster the convergence will be (see, e.g., Theorems 3.5 and 3.6). Thus, the conservative approach will be to use the minimal value like $k = 1$ in Condition C_f4. If the true k is 2 and one assumed $k = 1$, the estimator will still be consistent but the convergence will no longer be the optimal one. Section D in the appendix gives additional results when the order of the tuning parameter is not the optimal one.

Alternative consisent estimators of $f_0(0)$ may be used. For example, [21] and [10] studied Bayesian estimation of a decreasing density, where the former derived posterior consistency of the Bayesian estimator of $f_0(0)$ and the latter derived a contraction rate equal to $n^{-2/9}$ (up to log factors) under some assumptions on the prior distribution.

We only considered the case without covariates. When a regression model is considered, we expect that the NPMLE will be inconsistent as in the nonparametric estimator and suitable smoothing or penalization will be needed. We shall devote the further study in a future work.

Appendix A: Proofs for Section 2

A.1. Proof for Section 2.1

Proof of Lemma 2.1. For an arbitrary survival function S in $\mathcal{S}$, define $f(x) = S(x)/\int_0^\infty S(y)dy$. Then $f \in \mathcal{F}$ and $L_n(S) = L_{n,2}(f)$. Hence, $\sup_{S\in\mathcal{S}} L_n(S) \leq \sup_{f\in\mathcal{F}} L_{n,2}(f) = L_{n,2}(\hat{f}_n)$. If $\bar{S}$ is of the form (2.5), then $L_n(\bar{S}) = L_{n,2}(\hat{f}_n)$ and so $\bar{S}$ is a maximizer. Now, suppose $\bar{S} \in \mathcal{S}$ is not of the form (2.5). Define $\bar{f}(x) := \frac{\bar{S}(x)}{\int_0^\infty \bar{S}(y)dy}$ for $x > 0$. Then $\bar{f} \in \mathcal{F}$ and $\bar{f} \neq \hat{f}_n$. Since $\hat{f}_n$ is the unique maximizer of $L_{n,2}(f)$ subject to $f \in \mathcal{F}$,

$$L_n(\bar{S}) = L_{n,2}(\bar{f}) < L_{n,2}(\hat{f}_n) = L_n(S_{n,c}),$$

for any $c \geq \hat{f}_n(0)$.

Hence, any possible maximizer of L_n must be of the form (2.5). The uniqueness of $\hat{S}_n$ follows from the fact that the constraint (2.4) implies that it can only be achieved if and only if $c = \hat{f}_n(X_{(1)}) = \hat{f}_n(0)$. □

A.2. Proofs for Section 2.2.2

Proof of Lemma 2.2. The proof is similar to that in Lemma 2.1. For an arbitrary survival function S in $\mathcal{S}$, define $f(x) = S(x)/\int_0^\infty S(y)dy$. Then $f \in \mathcal{F}$ and $l_{n,\alpha}(S) = l_{n,2,\alpha}(f)$. Hence, $\sup_{S\in\mathcal{S}} l_{n,\alpha}(S) \leq \sup_{f\in\mathcal{F}} l_{n,2,\alpha}(f) = l_{n,2,\alpha}(\hat{f}_{n,\alpha})$. If $\bar{S}$ is of the form (2.9), then $l_{n,\alpha}(\bar{S}) = l_{n,2,\alpha}(\hat{f}_{n,\alpha})$ and so $\bar{S}$ is a maximizer. Now, suppose $\bar{S} \in \mathcal{S}$ is not of the form (2.9). Define $\bar{f}(x) := \frac{\bar{S}(x)}{\int_0^\infty \bar{S}(y)dy}$. Note that $\bar{f} \in \mathcal{F}$ and $\bar{f}(0+) = \bar{S}(0+)\{\int_0^\infty \bar{S}(y)dy\}^{-1}$, where the limits exist as $\bar{S}$ is a bounded monotone function. Since $\hat{f}_{n,\alpha}$ is the unique maximizer of $l_{n,\alpha,2}(f)$ subject to $f \in \mathcal{F}$,

$$l_{n,\alpha}(\overline{S}) = l_{n,2,\alpha}(\overline{f}) < l_{n,2,\alpha}(\hat{f}_{n,\alpha}) = l_{n,\alpha}(S_{n,\alpha,c}),$$

for any $c \geq \hat{f}_{n,\alpha}(0)$.

Hence, any possible maximizer of $l_{n,\alpha}$ must be of the form (2.9). The uniqueness of $\hat{S}_{n,\alpha}$ follows from the fact that the constraint (2.4) implies that it can only be achieved if and only if $c = \hat{f}_{n,\alpha}(X_{(1)}) = \hat{f}_{n,\alpha}(0)$. □

Proof of Lemma 2.3. For any $x \in (0, \tau)$, by Corollary 2 and Proposition 1 in [27], and the fact that $\hat{f}_n(x)$ is consistent (Corollary 3.1 in [8]),

$$\hat{S}_{n,\alpha_n}(x) = \frac{\hat{f}_{n,\alpha_n}(x)}{\hat{f}_{n,\alpha_n}(0)} = \frac{\hat{f}_n(x) + O_p(\alpha_n)}{\hat{f}_{n,\alpha_n}(0)} \xrightarrow{\mathbb{P}} \frac{f_0(x)}{f_0(0)} = S_0(x).$$

The case when $x = 0$ is trivial because $\hat{S}_{n,\alpha_n}(0) = 1$ and $S_0(0) = 1$. □

A.3. Proofs for Section 2.2.3

Proof of Lemma 2.4. Similar to the established approach in kernel methods, to find an optimal choice of the bin width, we strike a balance between the square of the bias and the variance. First, the variance is

$$\mathrm{Var}(\hat{f}_n^H(0)) = \frac{1}{nb_n^2} F_0(b_n)(1 - F_0(b_n)).$$

Hence, $\mathrm{Var}(\hat{f}_n^H(0)) \sim \frac{1}{nb_n} f_0(0)$. The bias is

$$\mathbb{E}(\hat{f}_n^H(0)) - f_0(0) = \frac{F_0(b_n)}{b_n} - f_0(0) = \frac{f_0^{(k)}(0)}{k!} b_n^k + o(b_n^k).$$

Hence, $\mathbb{E}(\hat{f}_n^H(0)) - f_0(0) \sim f_0^{(k)}(0) b_n^k / k!$. The mean square error is therefore

$$\mathrm{MSE}(\hat{f}_n^H(0)) \sim \left\{ \frac{f_0^{(k)}(0)}{k!} \right\}^2 b_n^{2k} + \frac{f_0(0)}{nb_n}$$

and the optimal choice of b_n in the sense of mean square error can be found accordingly by minimizing the RHS of the above relation. □

Proof of Lemma 2.5. By Taylor's theorem, the bias is $o(1)$ as $b_n \downarrow 0$. Hence,

$$\mathrm{MSE}(\hat{f}_n^H(0)) = o(1) + \frac{1}{nb_n} \frac{F_0(b_n)}{b_n} (1 - F_0(b_n)) \to 0$$

as $nb_n \to \infty$. As a result, $\hat{f}_n^H(0) \xrightarrow{\mathbb{P}} f_0(0)$. The consistency of $\hat{S}_n^H(x)$ then follows from the consistency of $\hat{f}_n^H(0)$ and $\hat{f}_n(x)$ (Corollary 3.1 in [8]). □

A.4. Proofs for Section 2.2.4

Proof of Lemma 2.6. Under $C_f 4'$, Theorem 3.1 in [14] implies that $\hat{f}_n(cn^{-\alpha}) \xrightarrow{\mathbb{P}} f_0(0)$. The consistency of $\hat{S}_n^N(x)$ then follows from the consistency of $\hat{f}_n(cn^{-\alpha})$ and $\hat{f}_n(x)$. □

A.5. Proofs for Section 2.2.5

The proof of Lemma 2.7 requires the consistency of $\hat{f}_{n,h}^S$ as given in the following Lemma A.1. The result is standard when dealing with the kernel smoothed isotonic estimators, see, for instance, [15] and [18]. For completeness, we provide a proof.

Lemma A.1. *Let $K : [-1,1] \to \mathbb{R}$ be a nonnegative kernel that is symmetric around 0, satisfies $\int_{-1}^1 K(u)du = 1$, and has a bounded derivative. Suppose that $h = h_n \to 0$ and $hn^{1/2} \to \infty$. If f_0 is a decreasing density and continuous on $[0, \tau]$,*

$$\sup_{x \in [0,\tau]} |\hat{f}_{n,h}^S(x) - f_0(x)| \xrightarrow{\mathbb{P}} 0.$$

To prove the above lemma, we first state some properties of the boundary kernel defined in (2.11) and (2.12). For $x \in [0, h]$, by (2.12),

$$\int_{-1}^{x/h} K_x(y)dy = 1 \text{ and } \int_{-1}^{x/h} yK_x(y)dy = 0. \tag{A.1}$$

Similarly, by (2.12) and the fact that K is symmetric, for $x \in [\tau - h, \tau]$,

$$\int_{\frac{x-\tau}{h}}^{1} K_x(y)dy = 1, \tag{A.2}$$

indeed, for $x \in [\tau - h, \tau]$, by symmetry of K,

$$\begin{aligned}\int_{\frac{x-\tau}{h}}^{1} K_x(y)dy &= \int_{\frac{x-\tau}{h}}^{1} \left\{\phi\left(\frac{\tau - x}{h}\right)K(y) - \psi\left(\frac{\tau - x}{h}\right)yK(y)\right\}dy \\ &= \int_{-1}^{\frac{\tau-x}{h}} \left\{\phi\left(\frac{\tau - x}{h}\right)K(y) + \psi\left(\frac{\tau - x}{h}\right)yK(y)\right\}dy = 1,\end{aligned}$$

where the last equality follows from (2.12). Furthermore, it can be shown that ϕ and ψ are bounded from above (see the supplementary of [3]).

Proof of Lemma A.1. For simplicity, we write $\hat{f}_n^S = \hat{f}_{n,h}^S$. For $x \in [0, \tau]$, write

$$\hat{f}_n^S(x) - f_0(x) = \{\hat{f}_n^S(x) - f_n^S(x)\} + \{f_n^S(x) - f_0(x)\},$$

where

$$f_n^S(x) := \int_{(x-h)\vee 0}^{(x+h)\wedge\tau} \frac{1}{h} K_x\left(\frac{x-u}{h}\right) f_0(u)du.$$

We first consider the three ranges of x according to the definition of K_x.

(i) Fix $x \in [0, h]$. By the change of variable $y = (x - u)/h$ and (A.1),

$$f_n^S(x) - f_0(x) = \int_{-1}^{x/h} K_x(y)\{f_0(x - hy) - f_0(x)\}dy.$$

Therefore, by the uniform continuity of f_0,

$$\begin{aligned}&\sup_{x\in[0,h]} |f_n^S(x) - f_0(x)| \\ &\leq \sup_{x\in[0,h]} \int_{-1}^{x/h} K_x(y)|f_0(x - hy) - f_0(x)|dy \to 0\end{aligned}$$

as $n \to \infty$. By using integration by parts and a change-of-variable,

$$\begin{aligned}\hat{f}_n^S(x) - f_n^S(x) &= \int_0^{x+h} \frac{1}{h} K_x\left(\frac{x-u}{h}\right) d(\hat{F}_n - F_0)(u) \\ &= -\int_0^{x+h} \frac{1}{h}\frac{\partial}{\partial u} K_x\left(\frac{x-u}{h}\right) \{\hat{F}_n(u) - F_0(u)\}du \\ &= \frac{1}{h}\int_{-1}^{x/h} \frac{\partial}{\partial y} K_x(y)\{\hat{F}_n(x - hy) - F_0(x - hy)\}dy.\end{aligned}$$

Since K is assumed to be continuously differentiable, K and K' are bounded on the compact set of $[-1, 1]$. By the boundedness of ϕ, ψ, K and K', we have

$$\sup_{x\in[0,h]} |\hat{f}_n^S(x) - f_n^S(x)| \precsim \frac{1}{h}\|\hat{F}_n - F_0\|_\infty,$$

where $\precsim$ indicates an inequality up to a constant multiple that does not depend on n. By Marshall's lemma [16], $\|\hat{F}_n - F_0\|_\infty \le \|\mathbb{F}_n - F_0\|_\infty = O_p(n^{-1/2})$. Since $hn^{1/2} \to \infty$, $\sup_{x\in[0,h]} |\hat{f}_n^S(x) - f_0(x)| = o_p(1)$.

(ii) Fix $x \in (\tau, \tau - h)$. By the change-of-variable $y = (x - u)/h$ and the fact that $\int_{-1}^1 K(y)dy = 1$,

$$f_n^S(x) - f_0(x) = \int_{-1}^1 K(y)\{f_0(x - hy) - f_0(x)\}dy.$$

By the uniform continuity of f_0,

$$\begin{aligned}&\sup_{x\in(h,\tau-h)} |f_n^S(x) - f_0(x)|\\ &\le \sup_{x\in(h,\tau-h)} \int_{-1}^1 K(y)|f_0(x - hy) - f_0(x)|dy \to 0.\end{aligned}$$

By using integration by parts and a change-of-variable,

$$\begin{aligned}\hat{f}_n^S(x) - f_n^S(x) &= \int_{x-h}^{x+h} \frac{1}{h}K\left(\frac{x-u}{h}\right) d(\hat{F}_n - F_0)(u)\\ &= -\int_{x-h}^{x+h} \frac{1}{h}\frac{d}{du}K\left(\frac{x-u}{h}\right)\{\hat{F}_n(u) - F_0(u)\}du\\ &= \frac{1}{h}\int_{-1}^1 \frac{d}{dy}K(y)\{\hat{F}_n(x - hy) - F_0(x - hy)\}dy.\end{aligned}$$

Similar to (i), by the boundedness of K',

$$\sup_{x\in(h,\tau-h)} |\hat{f}_n^S(x) - f_n^S(x)| \precsim \frac{1}{h}\|\hat{F}_n - F_0\|_\infty = o_p(1).$$

(iii) Fix $x \in [\tau - h, \tau]$. By the change-of-variable $y = (x - y)/h$ and (A.2),

$$f_n^S(x) - f_0(x) = \int_{\frac{x-\tau}{h}}^1 K_x(y)\{f_0(x - hy) - f_0(x)\}dy.$$

Therefore, by the uniform continuity of f_0,

$$\sup_{x\in[\tau-h,\tau]} |f_n^S(x) - f_0(x)|$$

$$\leq \sup_{x\in[\tau-h,\tau]} \int_{\frac{x-\tau}{h}}^{1} K_x(y)|f_0(x-hy) - f_0(x)|dy \to 0,$$

as $n \to \infty$. By using integration by parts and a change-of-variable,

$$\begin{aligned}
\hat{f}_n^S(x) - f_n^S(x) &= \int_{x-h}^{\tau} \frac{1}{h} K_x\left(\frac{x-u}{h}\right) d(\hat{F}_n - F_0)(u) \\
&= -\int_{x-h}^{\tau} \frac{1}{h}\frac{\partial}{\partial u} K_x\left(\frac{x-u}{h}\right)\{\hat{F}_n(u) - F_0(u)\}du \\
&= \frac{1}{h}\int_{\frac{x-\tau}{h}}^{1} \frac{\partial}{\partial y} K_x(y)\{\hat{F}_n(x-hy) - F_0(x-hy)\}dy.
\end{aligned}$$

By the boundedness of ϕ, ψ, K and K', we have

$$\sup_{x\in[\tau-h,\tau]} |\hat{f}_n^S(x) - f_n^S(x)| \precsim \frac{1}{h}\|\hat{F}_n - F_0\|_\infty = o_p(1).$$

Combining (i)–(iii), the result is proven. □

Proof of Lemma 2.7. Since S_0 is continuous, so does f_0 by (1.1). By Lemma A.1, $\sup_{x\in[0,\tau]} |\hat{f}_{n,h}^S(x) - f_0(x)| \xrightarrow{\mathbb{P}} 0$ for any $h \to 0$ and $hn^{1/2} \to \infty$. Thus,

$$\hat{S}_{n,h_1,h_2}^S(x) = \min\left(1, \frac{\hat{f}_{n,h_1}^S(x)}{\hat{f}_{n,h_2}^S(0)}\right) \xrightarrow{\mathbb{P}} S_0(x)$$

for any $x \in [0,\tau]$. □

Appendix B: A local Kiefer-Wolfowitz-type result

In this section, we follow the general setting and notation in [4] and establish a local version of Kiefer-Wolfowitz-type result under weaker conditions on the underlying function of interest; namely by removing the condition of the lower bound being away from zero; see Theorem B.3. Suppose F_n is a cadlag step estimator for a concave function $F : [a,b] \to \mathbb{R}$, where a and b are known and finite. Note that the argument and results in this section are also valid when we take b to infinity, by then $F : [a,\infty) \to \mathbb{R}$. For simplicity, we only state the results when F is defined on $[a,b]$. The corresponding results when F is defined on $[a,\infty)$ can be obtained by replacing "$b]$" by "$\infty)$" without any major change of arguments. Assume that F is continuously differentiable with $F(a) = 0$ and denote by f its derivative. Fix an interior point $y \in (a,b)$.

We impose the following conditions on f:

(A1) The function $f : [a,b] \to \mathbb{R}$ is decreasing and continuously differentiable on $[0, y+2\delta_0] \subset [a,b]$ for some $\delta_0 > 0$ such that $0 < \inf_{t\in[0,y+2\delta_0]} |f'(t)| \leq \sup_{t\in[0,y+2\delta_0]} |f'(t)| < \infty$.

Note that [4] assumes instead $0 < \inf_{t\in[a,b]} |f'(t)| \le \sup_{t\in[a,b]} |f'(t)| < \infty$ on the whole support. Furthermore, assume that the cadlag estimator F_n can be approximated in the sense that

$$\sup_{t\in[a,b]} |F_n(t) - F(t) - n^{-1/2} B_n \circ L_n(t)| = O_p(\gamma_n), \tag{B.1}$$

where $\gamma_n \to 0$ as $n \to \infty$, $L : [a, b] \to \mathbb{R}$ is non-decreasing, and B_n is a process on $[L(a), L(b)]$ that satisfies the following two conditions for a given $\tau \in [0, 4)$:

(A2) There are positive constants K_1, K_2 such that for all $x \in [L(a), L(b)]$, $u \in (0, 1]$, and $\nu > 0$,

$$\mathbb{P}\left(\sup_{|x-y|\le u} |B_n(x) - B_n(y)| > \nu\right) \le K_1 \exp(-K_2\nu^2 u^{-\tau});$$

(A3) There are positive constants K_1, K_2 such that for all $x \in [L(a), L(b)], u \in (0, 1]$, and $\nu > 0$,

$$\mathbb{P}\left(\sup_{z\ge u}\{B_n(x - z) - B_n(x) - \nu z^2\} > 0\right) \le K_1 \exp(-K_2\nu^2 u^{4-\tau}).$$

Finally, we impose the following smoothness condition on L that does not require L' to be bounded away from 0 on $[a, b]$.

(A4) The function $L : [a, b] \to \mathbb{R}$ is continuously differentiable on $[0, y + 2\delta_0] \subset [a, b]$ such that $0 < \inf_{t\in[0,y+2\delta_0]} L'(t) \le \sup_{t\in[0,y+2\delta_0]} L'(t) < \infty$.

As explained in [4], a typical example that satisfies Conditions (A1)–(A4) is during the estimation procedure for a monotone density f. By the Hungarian approximation, F_n can be approximated by a sequence of Brownian bridges B_n with L being the cumculative distribution function F corresponding to f, and $\gamma_n = (\log n)/n$. In our application, the function $S(x) = f(x)/f(0)$ cannot be uniformly bounded away from 0 when x approaches the ceiling of the support. Since $L' = f$ in this monotone density estimation, we also need to relax the condition that L' is bounded away from 0, which was necessarily required in [4].

Let $F_n^B := F + n^{-1/2} B_n \circ L$. Denote

$$c_n := \left(\frac{c_0 \log n}{n}\right)^{1/(4-\tau)} \tag{B.2}$$

for some $c_0 > 0$. For $x \in [a, b]$, let $\hat{F}^{(B,x)}_{n,c_n}(\cdot)$ be the least concave majorant of the process $\{F_n^B(\eta) : \eta \in [x - 2c_n, x + 2c_n] \cap [a, b]\}$. The following Lemma B.1 is similar to Lemma 2.1 in [4]. However, instead of considering the whole support $[a, b]$ of f, we only consider a local region around a fixed point $y \in (a, b)$ that allows us to relax the assumption of L' being bounded away from 0.

Lemma B.1. *Under Conditions (A1)–(A4), there exist positive numbers K_1, K_2, C_0 independent of n, such that for $c_0 \ge C_0$,*

$$\mathbb{P}\left(\sup_{x\in[y-\delta_0,y+\delta_0]} |\hat{F}^B_n(x) - \hat{F}^{(B,x)}_{n,c_n}(x)| \ne 0\right) \le K_1 n^{-c_0 K_2}.$$

Proof of Lemma B.1. The proof is similar to that of Lemma 2.1 in [4]. We still include here to illustrate the major differences. Without loss of generality, let $a = 0$. Denote $I_y := [y - \delta_0, y + \delta_0]$ and $I_{y,2} := [y - 2\delta_0, y + 2\delta_0]$. For all $x \in I_y$, let

$$\tilde{x}_i := \inf\{u \geq (x - 2c_n) \vee (y - \delta_0) : \hat{F}_n^B(u) = \hat{F}_{n,c_n}^{(B,x)}(u)\},$$

with the convention that the infimum of an empty set is $(x + 2c_n) \wedge (y + \delta_0)$; also let

$$\tilde{x}_s := \sup\{u \leq (x + 2c_n) \wedge (y + \delta_0) : \hat{F}_n^B(u) = \hat{F}_{n,c_n}^{(B,x)}(u)\},$$

with the convention that the supremum of an empty set is $(x - 2c_n) \vee (y - \delta_0)$. If $\hat{F}_n^B(u) = \hat{F}_{n,c_n}^{(B,x)}(u)$ for some $u \leq x$, and $\hat{F}_n^B(v) = \hat{F}_{n,c_n}^{(B,x)}(v)$ for some $v \geq x$, then we must have the common values of $\hat{F}_n^B = \hat{F}_{n,c_n}^{(B,x)}$ on the interval $[u, v]$. Therefore, if for some $x \in I_y$, we have $\hat{F}_n^B(x) \neq \hat{F}_{n,c_n}^{(B,x)}(x)$, then we must have either $\tilde{x}_i > x$ or $\tilde{x}_s < x$. Thus,

$$\begin{aligned}\mathbb{P}\left(\sup_{x \in I_y} |\hat{F}_n^B(x) - \hat{F}_{n,c_n}^{(B,x)}(x)| \neq 0\right) &\leq \mathbb{P}(\tilde{x}_i > x \text{ for some } x \in I_y) \\ &\quad + \mathbb{P}(\tilde{x}_s < x \text{ for some } x \in I_y). \qquad \text{(B.3)}\end{aligned}$$

Our goal is to show both probabilities on the right hand side are tending to be arbitrarily small. Comparing with Lemma 2.1 of [4], they aim to show that the probability $\mathbb{P}(\sup_{x \in [a,b]} |\hat{F}_n^B(x) - \hat{F}_{n,c_n}^{(B,x)}(x)| \neq 0)$ is converging to 0. As a result, instead of considering two probabilities on the right hand side of (B.3), they need to consider $\mathbb{P}(\tilde{x}_i > x \text{ for some } x \in [2c_n, b])$ and $\mathbb{P}(\tilde{x}_s < x \text{ for some } x \in [0, b - 2c_n])$, where they assumed f is bounded away from 0 to provide an upper bound of these probabilities. In our case, we only need to consider these probabilities when $x \in I_y$ and so we do not need to require f to be bounded away from 0. We shall first show that

$$\mathbb{P}(\tilde{x}_i > x \text{ for some } x \in I_y) \leq K_1 n^{-K_2 c_0}, \qquad \text{(B.4)}$$

for all $c_0 \geq C_0$ for some sufficiently large C_0. If $\tilde{x}_i > x$ for some $x \in I_y$, then by definition of $\tilde{x}_i$,

$$\hat{F}_n^B(u) \neq \hat{F}_{n,c_n}^{(B,x)}(u),$$

for all $x - 2c_n \leq u \leq x$. In that case, there exist $0 \leq y \leq x - 2c_n$ and $x \leq z \leq (x+2c_n) \wedge b$, such that the line segment joining $(y, F_n^B(y))$ and $(z, F_n^B(z))$ is above $(t, F_n^B(t))$ for all $t \in (y, z)$. In particular, this line segment is above the point $(x - c_n, F_n^B(x - c_n))$, which implies that the slope of the straight line segment joining the points $(y, F_n^B(y))$ and $(x - c_n, F_n^B(x - c_n))$ is smaller than the slope of another line segment joining the points $(z, F_n^B(z))$ and $(x - c_n, F_n^B(x - c_n))$; this implies that

$$\frac{F_n^B(y) - F_n^B(x - c_n)}{y - (x - c_n)} < \frac{F_n^B(z) - F_n^B(x - c_n)}{z - (x - c_n)}.$$

At a time, for every $\alpha \in \mathbb{R}$, this further implies that either

$$\frac{F_n^B(y) - F_n^B(x-c_n)}{y-(x-c_n)} < \alpha \text{ or } \alpha < \frac{F_n^B(z) - F_n^B(x-c_n)}{z-(x-c_n)}.$$

In particular, with the choice of $\alpha_x := f(x) + c_n|f'(x)|$, we have

$$\mathbb{P}(\tilde{x}_i > x \text{ for some } x \in [2c_n, 1]) \leq \mathbb{P}_1 + \mathbb{P}_2,$$

where

$$\mathbb{P}_1 := \mathbb{P}(\exists x \in I_y, \exists y \in [0, x-2c_n] : F_n^B(y) - F_n^B(x-c_n) > (y-x+c_n)\alpha_x),$$

and

$$\mathbb{P}_2 := \mathbb{P}(\exists x \in I_y, \exists z \in [x, (x+2c_n) \wedge b] : F_n^B(z) - F_n^B(x-c_n) > (z-x+c_n)\alpha_x).$$

Furthermore, with $t_x := c_n^2 f'(x)/4$, we can easily argue that $\mathbb{P}_1 \leq \mathbb{P}_{1,1} + \mathbb{P}_{1,2}$, where

$$\mathbb{P}_{1,1} := \mathbb{P}(\exists x \in I_y : F_n^B(x) - F_n^B(x-c_n) > c_n\alpha_x + t_x),$$

and

$$\mathbb{P}_{1,2} := \mathbb{P}(\exists x \in I_y, \exists y \in [0, x-2c_n] : F_n^B(x) - F_n^B(y) < (x-y)\alpha_x + t_x).$$

We first estimate $\mathbb{P}_{1,1}$. From (A1), f' is uniform continuous on $[0, y+\delta_0]$. Thus, by using Taylor's theorem, as $n \to \infty$,

$$F(x) - F(x-c_n) = c_n f(x) + \frac{c_n^2}{2}(|f'(x)| + o(1)),$$

where the $o(1)$ term is uniform, in $x \in I_y$ in the order of small-o of c_n^2. Therefore, with $M_n^B := F_n^B - F$, we obtain

$$\begin{aligned} \mathbb{P}_{1,1} &\leq \mathbb{P}\left(\exists x \in I_y : (M_n^B(x) - M_n^B(x-c_n)) > \frac{c_n^2}{2}(|f'(x)| + o(1))\right) \\ &\leq \mathbb{P}\left(\sup_{x \in I_y}(M_n^B(x) - M_n^B(x-c_n)) > \frac{c_n^2}{8}\inf_{t \in I_y}|f'(t)|\right), \end{aligned}$$

provided n is sufficiently large. By definition of F_n^B, $M_n^B = n^{-1/2}B_n \circ L$. Moreover, $|L(x) - L(x-c_n)| \leq c_n \sup_{t \in I_{y,2}} L'(t)$ for all $x \in I_y$, where by (A4), $\|L'\|_{I_{y,2}} := \sup_{t \in I_{y,2}} L'(t) < \infty$. Using Lemma 5.1 in [4], we conclude that with $\varepsilon := \inf_{t \in I_y}|f'(t)| > 0$ and $J := [L(0), L(y+\delta_2)]$,

$$\begin{aligned} \mathbb{P}_{1,1} &\leq \mathbb{P}\left(\sup_{x \in J}\sup_{|x-y| \leq c_n\|L'\|_{I_{y,2}}}(B_n(x) - B_n(y)) > \frac{c_n^2\sqrt{n}}{8}\varepsilon\right) \\ &\leq K_1\|L'\|_{I_{y,2}}^{-1} c_n^{-1} \exp\left(-\frac{K_2\varepsilon^2}{64\|L'\|_{I_{y,2}}^{\tau}} n c_n^{4-\tau}\right) \end{aligned}$$

$$= K_1 \|L'\|_{I_y,2}^{-1} \left(\frac{n}{c_0 \log n} \right)^{1/(4-\tau)} n^{-K_2 \varepsilon^2 c_0/(64\|L'\|_{I_y,2}^{\tau})}.$$

For a sufficiently large c_0, we see that

$$\left(\frac{n}{c_0 \log n} \right)^{1/(4-\tau)} n^{-K_2 \varepsilon^2 c_0/(64\|L'\|_{I_y,2}^{\tau})} \leq n^{-K_2 \varepsilon^2 c_0/(65\|L'\|_{I_y,2}^{\tau})}.$$

Therefore we can find a K_1 such that for c_0 sufficiently large and all large n,

$$\mathbb{P}_{1,1} \leq K_1 n^{-K_2 \varepsilon^2 c_0/(65\|L'\|_{I_y,2}^{\tau})}.$$

Replacing $K_2 \varepsilon^2 c_0/(65\|L'\|_{I_y,2}^{\tau})$ by K_2, we conclude that there are positive numbers K_1 and K_2 which depend only on f, L and C_0 such that

$$\mathbb{P}_{1,1} \leq K_1 n^{-K_2 c_0},$$

for all large n, provide that $c_0 \geq C_0$ for some sufficiently large C_0.

Next, consider $\mathbb{P}_{1,2}$. For all $x \in I_y$ and $z \in [1, x/(2c_n)]$, let $Y_n(x,z)$ be defined by

$$Y_n(x,z) := F_n^B(x - 2c_n z) - F_n^B(x) + 2c_n \alpha_x z + t_x,$$

so that

$$\mathbb{P}_{1,2} = \mathbb{P}(\exists x \in I_y, \exists z \in [1, x/(2c_n)] : Y_n(x,z) > 0). \tag{B.5}$$

Denote $\varepsilon_2 := \inf_{t \in [0, y+\delta_2]} |f'(t)|$. Let $\delta > 0$ such that $\delta \varepsilon_2 > 2 \sup_{t \in I_y} |f'(t)|$ (that gives $\delta \geq 2$). Now, distinguish between two possible cases $z \in [1, \delta]$ or $z \in [\delta, x/(2c_n)]$.

For $z \in [\delta, x/(2c_n)]$, by Taylor's theorem and the definition of α_x,

$$F(x - 2c_n z) - F(x) = -2c_n z(\alpha_x - c_n |f'(x)|) + 2c_n^2 z^2 f'(x_n),$$

for some x_n lying between $x - 2c_n z$ and x. Thus, as $x \geq x_n \geq x - 2c_n z \geq 0$ for $z \in [\delta, x/(2c_n)]$ and $f'(x_n) < 0$,

$$\begin{aligned} F(x - 2c_n z) - F(x) + 2c_n z \alpha_x &\leq 2c_n^2 z \sup_{t \in I_y} |f'(t)| - 2c_n^2 z^2 \varepsilon_2 \\ &\leq 2c_n^2 z \sup_{t \in I_y} |f'(t)| - c_n^2 z \delta \varepsilon_2 - c_n^2 z^2 \varepsilon_2 \\ &\leq -\varepsilon_2 c_n^2 z^2, \end{aligned} \tag{B.6}$$

where the second inequality follows as $z \geq \delta$ and the last inequality follows from the choice of δ.

Define $A_n := \{(x,z) : x \in I_y, z \in [\delta, x/(2c_n)]\}$. From (B.6), we have

$$\begin{aligned} &\mathbb{P}\left(\sup_{(x,z) \in A_n} Y_n(x,z) > 0 \right) \\ &\leq \mathbb{P}\left(\sup_{(x,z) \in A_n} \{M_n^B(x - 2c_n z) - M_n^B(x) - \varepsilon_2 c_n^2 z^2\} > \frac{c_n^2 \varepsilon_2}{4} \right) \end{aligned}$$

$$= \mathbb{P}\left(\sup_{(x,z)\in A_n}\{B_n\circ L(x-2c_nz)-B_n\circ L(x)-\varepsilon_2c_n^2\sqrt{n}z^2\} > \frac{c_n^2\varepsilon_2\sqrt{n}}{4}\right),$$

where the last equality follows as $M_n^B = n^{-1/2}B_n\circ L$.

Define $A_n' := \{(t,u) : t = L(x), u = (L(x)-L(x-2c_nz))/(2c_n), (x,z)\in A_n\}$. Then, for $(t,u)\in A_n'$, by mean-value theorem, $u = L'(x_n)z$ for some x_n lying between x and $x-2c_nz$. Thus,

$$z^2 = \frac{u^2}{(L'(x_n))^2} \geq \frac{u^2}{\|L'\|_{[0,y+\delta_0]}^2},$$

where $\|L'\|_{[0,y+\delta_0]} := \sup_{t\in[0,y+\delta_0]} L'(t)$. Thus, following the arguments as above, it yields

$$\mathbb{P}\left(\sup_{(x,z)\in A_n} Y_n(x,z) > 0\right)$$
$$\leq \mathbb{P}\left(\sup_{(t,u)\in A_n'}\left\{B_n(t-2c_nu)-B_n(t)-\frac{c_n^2\varepsilon_2\sqrt{n}u^2}{\|L'\|_{[0,y+\delta_0]}^2}\right\} > \frac{c_n^2\varepsilon_2\sqrt{n}}{4}\right).$$

Now, denote by $k_n := \lfloor c_n^{-1}\rfloor$, the integer part of c_n^{-1} and for all $j = 0,1,\ldots,k_n$, let $t_j := L(y-\delta_1)+j(L(y+\delta_2)-L(y-\delta_1))/k_n$. If for some $(t,u)\in A_n'$, one has

$$B_n(t-2c_nu)-B_n(t)-\frac{c_n^2\varepsilon_2\sqrt{n}u^2}{\|L'\|_{[0,y+\delta_0]}^2} > \frac{c_n^2\varepsilon_2\sqrt{n}}{4}, \tag{B.7}$$

then, for $j = 1,2,\ldots,k_n$, such that $t\in[t_{j-1},t_j]$, one either has

$$B_n(t_j-2c_nu)-B_n(t_j)-\frac{c_n^2\varepsilon_2\sqrt{n}u^2}{\|L'\|_{[0,y+\delta_0]}^2} > 0,$$

or

$$B_n(t-2c_nu)-B_n(t_j-2c_nu)-B_n(t)+B_n(t_j) > \frac{c_n^2\varepsilon_2\sqrt{n}}{4};$$

indeed, if not, then both

$$B_n(t_j-2c_nu)-B_n(t_j)-\frac{c_n^2\varepsilon_2\sqrt{n}u^2}{\|L'\|_{[0,y+\delta_0]}^2} \leq 0,$$

and

$$B_n(t-2c_nu)-B_n(t_j-2c_nu)-B_n(t)+B_n(t_j) \leq \frac{c_n^2\varepsilon_2\sqrt{n}}{4},$$

which contradicts (B.7) after adding these last two inequalities.

Note that for any $j = 0,1,\ldots,k_n$,

$$|B_n(t)-B_n(t_j)| \leq \sup_{u\in[L(y-\delta_0),L(y+\delta_0)]}\ \sup_{|u-v|\leq k_n^{-1}} |B_n(u)-B_n(v)|. \tag{B.8}$$

Furthermore, for $(t,u) \in A_n'$, we have $t - 2c_n u = L(x - 2c_n z) \in J = [L(0), L(y+\delta_0)]$, so that

$$|B_n(t - 2c_n u) - B_n(t_j - 2c_n u)| \le \sup_{t \in J} \sup_{|t-y| \le k_n^{-1}} |B_n(t) - B_n(y)|. \tag{B.9}$$

Hence, from (B.8), (B.9) and a simple use of triangle inequality, it follows that

$$\begin{aligned}&\sup_{t \in [t_{j-1}, t_j]} \sup_{u \ge \delta \inf_{t \in [0, y+\delta_0]} L'(t)} \{B_n(t - 2c_n u) - B_n(t_j - 2c_n u) - B_n(t) + B_n(t_j)\} \\ &\le 2 \sup_{t \in J} \sup_{|t-y| \le k_n^{-1}} |B_n(t) - B_n(y)|.\end{aligned}$$

We conclude that

$$\begin{aligned}&\mathbb{P}\left(\sup_{(x,z) \in A_n} Y_n(x,z) > 0 \right) \\ &\le \mathbb{P}\left(2 \sup_{t \in J} \sup_{|t-y| \le k_n^{-1}} |B_n(t) - B_n(y)| > \frac{c_n^2 \varepsilon_2 \sqrt{n}}{4} \right) \\ &\quad + \sum_{j=1}^{k_n} \mathbb{P}\left(\sup_{u \ge \delta \inf_{t \in [0, y+\delta_0]} L'(t)} \left\{ B_n(t_j - 2c_n u) - B_n(t_j) - \frac{c_n^2 \varepsilon_2 \sqrt{n} u^2}{\|L'\|_{[0,y+\delta_0]}^2} \right\} > 0 \right).\end{aligned}$$

Using Lemma 5.1 in [4] with $I = J, u = k_n^{-1}, \nu = \frac{c_n^2 \varepsilon_2 \sqrt{n}}{4}$, we have

$$\begin{aligned}&\mathbb{P}\left(2 \sup_{t \in J} \sup_{|t-y| \le k_n^{-1}} |B_n(t) - B_n(y)| > \frac{c_n^2 \varepsilon_2 \sqrt{n}}{4} \right) \\ &\le K_1 k_n \exp\left(-\frac{K_2 \varepsilon_2^2 n}{16} c_n^4 k_n^\tau \right) \le K_1 c_n^{-1} \exp\left(-\frac{K_2 \varepsilon_2^2 n}{16} \cdot c_n^4 \cdot \frac{c^{-\tau}}{2^\tau} \right) \\ &= K_1 \left(\frac{n}{c_0 \log n} \right)^{1/(4-\tau)} n^{-K_2 c_0 \varepsilon_2^2 / 2^{4+\tau}},\end{aligned}$$

by definitions of c_n and k_n, since $k_n \le c_n^{-1}$ and $k_n \ge c_n^{-1}/2$ for those sufficiently large n. Hence, there exist positive numbers K_1 and K_2 that depend only on f, L and C_0 such that

$$\mathbb{P}\left(2 \sup_{t \in J} \sup_{|t-y| \le k_n^{-1}} |B_n(t) - B_n(y)| > \frac{c_n^2 \varepsilon_2 \sqrt{n}}{4} \right) \le K_1 n^{-K_2 c_0},$$

for all n, provided $c_0 \ge C_0$ for some sufficiently large C_0. Furthermore, with (A3), we have

$$\sum_{j=1}^{k_n} \mathbb{P}\left(\sup_{u \ge \delta \inf_{t \in [0, y+\delta_0]} L'(t)} \left\{ B_n(t_j - 2c_n u) - B_n(t_j) - \frac{c_n^2 \varepsilon_2 \sqrt{n} u^2}{\|L'\|_{[0,y+\delta_0]}^2} \right\} > 0 \right)$$

$$
\begin{aligned}
&= \sum_{j=1}^{k_n} \mathbb{P}\left(\sup_{z \geq 2c_n\delta \inf_{t\in[0,y+\delta_0]} L'(t)} \left\{ B_n(t_j - z) - B_n(t_j) - \frac{c_n^2 \varepsilon_2 \sqrt{n} z^2}{4\|L'\|^2_{[0,y+\delta_0]}} \right\} > 0 \right) \\
&\leq \sum_{j=1}^{k_n} K_1 \exp\left(-\frac{K_2 \varepsilon_2^2 n}{16\|L'\|^4_{[0,y+\delta_0]}} \left(2c_n\delta \inf_{t\in[0,y+\delta_0]} L'(t) \right)^{4-\tau} \right) \\
&\leq K_1 \left(\frac{n}{c_0 \log n} \right)^{1/(4-\tau)} n^{-K_2\varepsilon_2^2 c_0 (2\delta \inf_{t\in[0,y+\delta_0]} L'(t))^{4-\tau}/(16\|L'\|^4_{[0,y+\delta_0]})},
\end{aligned}
$$

by definitions of c_n and k_n. Renaming K_1 and K_2, the expressions on the right hand side in the last inequality is bounded from above by $K_1 n^{-K_2 c_0}$ for all n, provided $c_0 \geq C_0$ for some sufficiently large C_0, where K_1 and K_2 depend only on f, L and C_0. We conclude that there exist constants K_1, K_2 such that

$$
\mathbb{P}\left(\sup_{(x,z)\in A_n} Y_n(x,z) > 0 \right) \leq K_1 n^{-K_2 c_0},
$$

for all n, provided c_0 is sufficiently large.

To show that $\mathbb{P}_{1,2} \leq K_1 n^{-K_2 c_0}$, we also need to show that

$$
\mathbb{P}\left(\sup_{x\in I_y} \sup_{z\in[1,\delta]} Y_n(x,z) > 0 \right) \leq K_1 n^{-K_2 c_0}.
$$

For any $x \in I_y$ and $z \in [1,\delta]$, by Taylor's theorem, for some x_n lying between $x - 2c_n z$ and x,

$$
F(x - 2c_n z) - F(x) = -2c_n z f(x) + 2c_n^2 z^2 f'(x_n).
$$

By the definition of α_x and as f' is uniform continuous on I_y,

$$
F(x - 2c_n z) - F(x) + 2c_n \alpha_x z = 2c_n^2 |f'(x)| z(1-z) + o(c_n^2),
$$

where $2c_n^2|f'(x)|z(1-z) \leq 0$ because $z \geq 1$, and $o(c_n^2)$ is uniform in $z \in [1,\delta]$ and $x \in I_y$. Recall that $t_x = c_n^2 f'(x)/4 \leq 0$. Thus,

$$
F(x - 2c_n z) - F(x) + 2c_n \alpha_x z + t_x \leq -c_n^2 |f'(x)|/8,
$$

for all $z \in [1,\delta]$ and $x \in I_y$ for all sufficiently large n. Hence,

$$
\begin{aligned}
&\mathbb{P}\left(\sup_{x\in I_y} \sup_{z\in[1,\delta]} Y_n(x,z) > 0 \right) \\
&\leq \mathbb{P}\left(\sup_{x\in I_y} \sup_{z\in[1,\delta]} (M_n^B(x - 2c_n z) - M_n^B(x)) > \frac{c_n^2}{8} \inf_{x\in I_y} |f'(t)| \right) \\
&= \mathbb{P}\left(\sup_{x\in I_y} \sup_{z\in[1,\delta]} (B_n(L(x - 2c_n z)) - B_n L((x))) > \frac{c_n^2 \sqrt{n}\varepsilon}{8} \right)
\end{aligned}
$$

$$\leq \mathbb{P}\left(\sup_{t\in[L(y-\delta_0),L(y+\delta_0)]}\ \sup_{|t-y|\leq 2c_n\|L'\|_{I_y,2}\delta}|B_n(t)-B_n(y)|>\frac{c_n^2\sqrt{n}\varepsilon}{8}\right).$$

Using Lemma 5.1 in [4], we have

$$\begin{aligned}&\mathbb{P}\left(\sup_{x\in I_y}\ \sup_{z\in[1,\delta]}Y_n(x,z)>0\right)\\&\leq K_1 2^{-1}c_n^{-1}\|L'\|_{I_y,2}^{-1}\delta^{-1}e^{-K_2c_n^{4-\tau}n\varepsilon^2 2^{-\tau}\|L'\|_{I_y,2}^{-\tau}\delta^{-\tau}/64}.\end{aligned}$$

By renaming K_1 and K_2, the last displayed term is bounded from above by $K_1n^{-K_2c_0}$ for all n, provided $c_0 \geq C_0$ for some sufficiently large C_0. Thus, we have shown that

$$\mathbb{P}_{1,2}\leq K_1n^{-K_2c_0}.$$

Using similar argument as above, $\mathbb{P}_2$ can also be shown to be bounded from above by $K_1n^{-K_2c_0}$. It remains to show that $\mathbb{P}(\tilde{x}_s < x \text{ for some } x\in I_y)\leq K_1n^{-K_2c_0}$. If for some $x\in I_y$, $\tilde{x}_s<x$, then by the definition of $\tilde{x}_s$,

$$\hat{F}_n^B(u)\neq \hat{F}_{n,c_n}^{(B,x)}(u) \tag{B.10}$$

for all $x\leq u\leq x+2c_n$. Let $x_n := x+c_n$. Note that $\hat{F}_{n,c_n/4}^{(B,x_n)}$ is the least concave majorant of the process $\{F_n^B(\eta):\eta\in[x_n-\frac{c_n}{2},x_n+\frac{c_n}{2}]\}$. Define

$$\tilde{x}_{i,n} := \inf\{u\geq (x_n-c_n/2)\vee(y-\delta_0):\hat{F}_n^B(u)=\hat{F}_{n,c_n/4}^{(B,x_n)}(u)\},$$

with the convention that the infimum of an empty set is $(x_n+c_n/2)\wedge(y+\delta_0)$. From (B.10), we also have

$$\hat{F}_n^B(u)\neq \hat{F}_{n,c_n/4}^{(B,x_n)}(u)$$

for all $x\in[x_n-c_n/2,x_n+c_n/2]$. Hence,

$$\mathbb{P}(\tilde{x}_s<x \text{ for some } x\in I_y)\leq \mathbb{P}(\tilde{x}_{i,n}>x_n \text{ for some } x_n\in I_y).$$

The last probability can be shown to be bounded from above by $K_1n^{-K_2c_0}$ as (B.4). □

Lemma B.2. *Under Conditions (A1)–(A4), we have*

$$\sup_{x\in[y-\delta_0,y+\delta_0]}|\hat{F}_n^B(x)-F_n^B(x)|=O_p\left(\frac{\log n}{n}\right)^{2/(4-\tau)}.$$

Proof of Lemma B.2. The proof is similar to that for Theorem 2.1 in [4]. For any intervals $I\subset\mathbb{R}$, denote by CM_I the operator that maps a bounded function $h: I\to\mathbb{R}$ into the least concave majorant of h on I.

Denote $I_y := [y-\delta_0, y+\delta_0]$ and $I_{y,2} := [y-2\delta_0, y+2\delta_0]$. Fix $x \in I_y \subset (a,b)$. It suffices to consider all sufficiently large n such that $a < x-2c_n$ and $x+2c_n < b$. Define

$$T_n^{(B,x)}(\eta) := F_n^B(x+c_n\eta) - F_n^B(x),$$

for all $\eta \in [-2,2]$. Then, by the definition of $\hat{F}_{n,c_n}^{(B,x)}(x)$,

$$\hat{F}_{n,c_n}^{(B,x)}(x) - F_n^B(x) = (CM_{[-2,2]}T_n^{(B,x)})(0).$$

With $M_n^B = F_n^B - F$, we have

$$T_n^{(B,x)}(\eta) = M_n^B(x+c_n\eta) - M_n^B(x) + F(x+c_n\eta) - F(x).$$

Since $\sup_{x\in I_{y,2}} |f'(x)| < \infty$ and $|\eta| \le 2$, it follows from Taylor's theorem that

$$T_n^{(B,x)}(\eta) = M_n^B(x+c_n\eta) - M_n^B(x) + Y_n^{(B,x)}(\eta) + O(c_n^2), \tag{B.11}$$

where $Y_n^{(B,x)}(\eta) := c_n\eta f(x)$, and the big O-term is uniform in $\eta \in [-2,2]$ and $x \in I_y$ for all large enough n so that $2c_n \le \delta_0$. Because the process $Y_n^{(B,x)}$ is linear, its least concave majorant on $[-2,2]$ is $Y_n^{(B,x)}$ itself. Using the fact that the maximal distance between the least concave majorants of processes is less than or equal to the maximum possible distance between the processes themselves, we have

$$\begin{aligned}
|\hat{F}_{n,c_n}^{(B,x)}(x) - F_n^B(x)| &= |(CM_{[-2,2]}T_n^{(B,x)})(0)| \\
&\le |Y_n^{(B,x)}(0)| + |(CM_{[-2,2]}T_n^{(B,x)})(0) - Y_n^{(B,x)}(0)| \\
&\le \sup_{\eta\in[-2,2]} |(CM_{[-2,2]}T_n^{(B,x)})(\eta) - Y_n^{(B,x)}(\eta)| \\
&\le \sup_{\eta\in[-2,2]} |T_n^{(B,x)}(\eta) - Y_n^{(B,x)}(\eta)|.
\end{aligned}$$

Using (B.11), we obtain for any $x \in I_y$ and all large enough n,

$$|\hat{F}_{n,c_n}^{(B,x)}(x) - F_n^B(x)| \le \sup_{\eta\in[-2,2]} |M_n^B(x+c_n\eta) - M_n^B(x)| + O(c_n^2).$$

Hence, for $A > 0$ sufficiently large and all large enough n,

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{x\in I_y} |\hat{F}_{n,c_n}^{(B,x)}(x) - F_n^B(x)| > Ac_n^2\Big) \\
&\le \mathbb{P}\Big(\sup_{x\in I_y}\sup_{\eta\in[-2,2]} |M_n^B(x+c_n\eta) - M_n^B(x)| > Ac_n^2/2\Big) \\
&= \mathbb{P}\Big(\sup_{x\in I_y}\sup_{\eta\in[-2,2]} |B_n\circ L(x+c_n\eta) - B_n\circ L(x)| > Ac_n^2\sqrt{n}/2\Big) \\
&\le \mathbb{P}\Big(\sup_{x\in[L(y-\delta_0),L(y+\delta_0)]}\ \sup_{|x-y|\le 2c_n\|L'\|_{I_{y,2}}} |B_n(x) - B_n(y)| > Ac_n^2\sqrt{n}/2\Big),
\end{aligned}$$

where $\|L'\|_{I_{y,2}} := \sup_{t\in I_{y,2}} |L'(t)|$. Using Lemma 5.1 in [4], for sufficiently large A and n,

$$\begin{aligned}
&\mathbb{P}\bigg(\sup_{x\in I_y} |\hat{F}^{(B,x)}_{n,c_n}(x) - F^B_n(x)| > Ac_n^2\bigg)\\
&\le \frac{K_1}{2c_n\|L'\|_{I_{y,2}}} \exp\bigg\{-K_2A^2 2^{-2-\tau}\|L'\|^{-\tau}_{I_{y,2}} nc_n^{4-\tau}\bigg\}\\
&= \frac{K_1}{2\|L'\|_{I_{y,2}}}\bigg(\frac{n}{c_0\log n}\bigg)^{1/(4-\tau)} n^{-K_2A^2 2^{-2-\tau}\|L'\|^{-\tau}_{I_{y,2}} c_0}.
\end{aligned}$$

Since the above upper bound tends to 0 as $n\to\infty$ provided that A is sufficiently large,

$$\sup_{x\in I_y} |\hat{F}^{(B,x)}_{n,c_n}(x) - F^B_n(x)| = O_p(c_n^2). \tag{B.12}$$

Finally, the proof is completed in view of Lemma B.1, (B.12) and the triangle inequality that

$$\sup_{x\in I_y} |\hat{F}^B_n(x) - F^B_n(x)| \le \sup_{x\in I_y} |\hat{F}^B_n(x) - \hat{F}^{(B,x)}_{n,c_n}(x)| + \sup_{x\in I_y} |\hat{F}^{(B,x)}_{n,c_n}(x) - F^B_n(x)|. \qquad \square$$

Theorem B.3. *Under Conditions (A1)–(A4), we have*

$$\sup_{x\in[y-\delta_0,y+\delta_0]} |\hat{F}_n(x) - F_n(x)| = O_p(\gamma_n) + O_p\bigg(\frac{\log n}{n}\bigg)^{2/(4-\tau)}.$$

Proof of Theorem B.3. The proof is similar to the proof for Theorem 2.2 in [4]. Write

$$\hat{F}_n - F_n = (\hat{F}_n - \hat{F}^B_n) + (F^B_n - F_n) + (\hat{F}^B_n - F^B_n).$$

Since the maximal distance between least concave majorant processes is less than or equal to the maximum possible distance between the processes themselves,

$$\sup_{x\in[a,b]} |\hat{F}_n(x) - \hat{F}^B_n(x)| \le \sup_{x\in[a,b]} |F_n(x) - F^B_n(x)|.$$

Denote $I_y := [y-\delta_0, y+\delta_0]$. The triangle inequality gives

$$\begin{aligned}
&\sup_{x\in I_y} |\hat{F}_n(x) - F_n(x)|\\
&\le \sup_{x\in I_y} |\hat{F}_n(x) - \hat{F}^B_n(x)| + \sup_{x\in I_y} |F^B_n(x) - F_n(x)| + \sup_{x\in I_y} |\hat{F}^B_n(x) - F^B_n(x)|\\
&\le \sup_{x\in[a,b]} |\hat{F}_n(x) - \hat{F}^B_n(x)| + \sup_{x\in[a,b]} |F^B_n(x) - F_n(x)| + \sup_{x\in I_y} |\hat{F}^B_n(x) - F^B_n(x)|\\
&\le 2\sup_{x\in[a,b]} |F_n(x) - F^B_n(x)| + \sup_{x\in I_y} |\hat{F}^B_n(x) - F^B_n(x)|.
\end{aligned}$$

The required result then follows from (B.1) and Lemma B.2. $\square$

Appendix C: Proofs for Section 3

C.1. Proofs for Section 3.2

Proof of Theorem 3.4. By Theorem B.3, it suffices to verify Conditions (A1)–(A4). For the problem of estimating a decreasing density, F_n is the empirical distribution function. By the Hungarian approximation, F_n can be approximated by a sequence of Brownian bridges B_n with L being the distribution function F corresponding to f, and $\gamma_n = (\log n)/n$. Under C_f5, Conditions (A1) and (A4) are satisfied. For Conditions (A2) and (A3), Corollary 3.1 in [4] has verified that Brownian bridges will satisfy these two conditions with the choice of $\tau = 1$. Hence, the required results follow from Theorem B.3 with y taking values $(2j+1)\delta_0$ for $j = 0, 1, \dots, \lfloor \frac{x_0}{2\delta_0} \rfloor$, the integer part of $\frac{x_0}{2\delta_0}$, and $y = x_0$. $\square$

C.2. Proofs for Section 3.3

Proof of Theorem 3.5. We only provide the proof when $k = 1$ as the other cases can be proven similar. To derive the asymptotic joint distribution in (3.8), we shall make use of some intermediate result of [23] as well as the method of proof (see, e.g., [7] and [8]) of deriving the asymptotic distribution of the Grenander estimator at a fixed interior point; so, some of the details will be omitted. With the Hungarian approximation in (3.3), denote

$$Z_n(c) := \sup_{x>0} \left[n^{1/3} \frac{\frac{1}{\sqrt{n}}\mathbb{B}_n(F_0(x)) - f_0(0)\alpha_n - \{f_0(0)x - F_0(x)\} + R_n(x)}{\alpha_n + x} \right],$$
$$Z_n^{\#}(c) := \sup_{t>0} \frac{n^{1/6} f_0^{-1}(0) W_n(f_0^2(0)tn^{-1/3}) - c - \beta_2 t^2}{t}.$$

By Lemma 2 in [23], we have

$$\sup_{0 \le c \le C_n} |n^{1/3}\{\hat{f}_{n,\alpha_n}(0) - f_0(0)\} - Z_n(c)| \xrightarrow{\mathbb{P}} 0, \tag{C.1}$$

for whatever $C_n = o_p(n^{1/3})$. Furthermore, from the proof of Proposition 2 in Woodroofe and Sun (1996), for any $0 < c_1 < c_2 < \infty$,

$$\sup_{c_1 \le c \le c_2} |Z_n(c) - Z_n^{\#}(c)| \xrightarrow{\mathbb{P}} 0. \tag{C.2}$$

In view of (C.1) and (C.2), to study the asymptotic distribution of (3.8), it suffices to consider that of $(Z_n^{\#}(c), n^{1/3}\{\hat{f}_n(x_0) - f_0(x_0)\})$.

Let $a, b \in \mathbb{R}$. Using the switch relation in (3.2) and the fact that adding or multiplying a constant will not affect the location of the maximum of a process, we obtain, with the probability one, that

$$n^{1/3}\{\hat{f}_n(x_0) - f_0(x_0)\} \le b \tag{C.3}$$

$$\Leftrightarrow n^{1/3}\{U_n(n^{-1/3}b + f_0(x_0)) - x_0\} \leq 0$$
$$\Leftrightarrow \sup\{t \geq -x_0 n^{1/3} : \mathbb{F}_n(x_0+tn^{-1/3}) - n^{-2/3}bt - f_0(t_0)tn^{-1/3} \text{ is maximal}\} \leq 0$$
$$\Leftrightarrow \sup\{t \geq -x_0 n^{1/3} : V_n(t,b) \text{ is maximal}\} \leq 0,$$

where for $t \geq -x_0 n^{1/3}$,

$$V_n(t,b) := n^{2/3}\{\mathbb{F}_n(x_0 + tn^{-1/3}) - \mathbb{F}_n(x_0) - f_0(x_0)tn^{-1/3}\} - bt. \tag{C.4}$$

By the same Hungarian approximation in (3.3), define

$$\begin{aligned}\widetilde{V}_n(t,b) := {} & n^{1/6}\{W_n(F_0(x_0 + tn^{-1/3})) - W_n(F_0(x_0))\} \\ & + n^{2/3}\{F_0(x_0 + tn^{-1/3}) - F_0(x_0) - f_0(x_0)n^{-1/3}t\} - bt\end{aligned} \tag{C.5}$$

and note that

$$\begin{aligned} V_n(t,b) &= n^{1/6}\{\mathbb{B}_n(F_0(x_0 + tn^{-1/3})) - \mathbb{B}_n(F_0(t_0))\} \\ &\quad + n^{2/3}\{F_0(x_0 + tn^{-1/3}) - F_0(x_0) - f_0(x_0)n^{-1/3}t\} \\ &\quad + n^{2/3}\{R_n(x_0 + tn^{-1/3}) - R_n(x_0)\} - bt \\ &= \widetilde{V}_n(t,b) - n^{1/6}\{F_0(x_0 + tn^{-1/3}) - F_0(x_0)\}W_n(1) \\ &\quad + n^{2/3}\{R_n(x_0 + tn^{-1/3}) - R_n(x_0)\} \\ &= \widetilde{V}_n(t,b) + o_p(1), \end{aligned}$$

where the $o_p(1)$ term is uniform in t over any compacta as $W_n(1) = O_p(1)$, $n^{1/6}\{F_0(x_0 + tn^{-1/3}) - F_0(x_0)\} \to 0$ uniform in t over any compacta, and $\sup_{t\in\mathbb{R}} |n^{2/3}\{R_n(x_0+tn^{-1/3}) - R_n(x_0)\}| = o_p(1)$ by (3.4). Using Proposition 3.1, it can be shown that $\sup\{t : V_n(t,b) \text{ is maximal}\}$ and $\sup\{t : \widetilde{V}_n(t,b) \text{ is maximal}\}$ are asymptotically equivalent in the sense that the difference between them converges to 0 in probability. As a result, it suffices to consider $\widetilde{V}_n$ in establishing the asymptotic distribution of (3.8).

Note that we can write

$$\begin{aligned} Z_n^{\#}(c) &= \sup_{t>0}\left\{\frac{\widetilde{W}_n(t) - (c + \beta_2 t^2)}{t}\right\} \\ \widetilde{V}_n(t,a) &= f_0(0)\left\{\widetilde{W}_n\left(\frac{n^{1/3}F_0(x_0 + tn^{-1/3})}{f_0^2(0)}\right) - \widetilde{W}_n\left(\frac{n^{1/3}F_0(x_0)}{f_0^2(0)}\right)\right\} \\ &\quad + n^{2/3}\{F_0(x_0 + tn^{-1/3}) - F_0(x_0) - f_0(x_0)n^{-1/3}t\}, \end{aligned} \tag{C.6}$$

where

$$\widetilde{W}_n(t) := \frac{n^{1/6}}{f_0(0)} W_n\left(\frac{f_0^2(0)t}{n^{1/3}}\right)$$

is again a Brownian motion. Denote

$$H_{n1} := \sup\left\{t > 0 : \frac{\widetilde{W}_n(t) - (c + \beta_2 t^2)}{t} \text{ is maximal}\right\},$$

$$H_{n2} := \sup\{t : \widetilde{V}_n(t,b) \text{ is maximal}\}.$$

Let

$$A_n := \{H_{n1} \le K_n, |H_{n2}| \le K_n\},$$

where $\{K_n\}$ is any positive sequence such that $K_n \to \infty$ but $K_n = o(n^{1/3})$. Note that $H_{n1} = O_p(1)$ and $H_{n2} = O_p(1)$. The former one can be seen by, for example, using the law of iterated logarithm of Brownian motions. Thus, $\mathbb{P}(A_n) \to 1$ as $n \to \infty$. Note that

$$\mathbb{P}(Z_n^{\#}(C) \le a, \sup\{t : \widetilde{V}_n(t,b) \text{ is maximal}\} \le 0) =: D_{n1} + D_{n2}, \tag{C.7}$$

where

$$D_{n1} := \mathbb{P}(Z_n^{\#}(C) \le a, \sup\{t : \widetilde{V}_n(t,b) \text{ is maximal}\} \le 0, A_n),$$
$$D_{n2} := \mathbb{P}(Z_n^{\#}(C) \le a, \sup\{t : \widetilde{V}_n(t,b) \text{ is maximal}\} \le 0, A_n^c).$$

Clearly, since $H_{n1}, H_{n2} = O_p(1)$ and $K_n \to \infty$,

$$\limsup_{n\to\infty} D_{n2} \le \limsup_{n\to\infty} \mathbb{P}(A_n^c) = 0. \tag{C.8}$$

Also, define a sequence of events

$$B_n := \Bigg\{\omega : \sup_{0<t\le K_n} \left\{\frac{\widetilde{W}_n(t) - (c+\beta_2 t^2)}{t}\right\} \le a \text{ and}$$
$$\sup\{t \in [-K_n, K_n] : \widetilde{V}_n(t,b) \text{ is maximal}\} \le 0\Bigg\}.$$

Observe that on A_n,

$$\sup_{0<t\le K_n} \left\{\frac{\widetilde{W}_n(t) - (c+\beta_2 t^2)}{t}\right\} = \sup_{t>0} \left\{\frac{\widetilde{W}_n(t) - (c+\beta_2 t^2)}{t}\right\}.$$

Therefore, we have

$$D_{n1} = \mathbb{P}(B_n \cap A_n) = \mathbb{P}(B_n) + o(1), \tag{C.9}$$

as $\mathbb{P}(A_n) \to 1$. Recall the definition of $\tilde{V}_n$ in (C.6). Note that for all sufficiently large enough n,

$$K_n \le \frac{n^{1/3} F_0(x_0 - K_n n^{-1/3})}{f_0^2(0)}.$$

Therefore, using the independent increment property of Brownian motions, for all sufficiently large enough n,

$$\sup_{0<t\le K_n} \left\{\frac{\widetilde{W}_n(t) - (c+\beta_2 t^2)}{t}\right\} \text{ and } \sup\{t \in [-K_n, K_n] : \widetilde{V}_n(t,b) \text{ is maximal}\}$$

are independent. Hence, for all sufficiently large enough n,

$$\mathbb{P}(B_n) = \mathbb{P}\left(\sup_{0<t\le K_n}\left\{\frac{\widetilde{W}_n(t)-(c+\beta_2 t^2)}{t}\right\}\le a\right) \\ \cdot\,\mathbb{P}(\sup\{t\in[-K_n,K_n]:\widetilde{V}_n(t,b)\text{ is maximal}\}\le 0). \tag{C.10}$$

Therefore,

$$\begin{aligned}
&\mathbb{P}\biggl(\sup_{0<t\le K_n}\left\{\frac{\widetilde{W}_n(t)-(c+\beta_2 t^2)}{t}\right\}\le a\biggr)\\
&=\mathbb{P}\biggl(\sup_{0<t\le K_n}\left\{\frac{\widetilde{W}_n(t)-(c+\beta_2 t^2)}{t}\right\}\le a, A_n\biggr)\\
&\quad+\mathbb{P}\biggl(\sup_{0<t\le K_n}\left\{\frac{\widetilde{W}_n(t)-(c+\beta_2 t^2)}{t}\right\}\le a, A_n^c\biggr)\\
&=\mathbb{P}\biggl(\sup_{0<t<\infty}\left\{\frac{\widetilde{W}_n(t)-(c+\beta_2 t^2)}{t}\right\}\le a\biggr)+o(1),
\end{aligned}\tag{C.11}$$

where the last equality follows as $\mathbb{P}(A_n)\to 1$. Similarly, we have

$$\begin{aligned}
&\mathbb{P}(\sup\{t\in[-K_n,K_n]:\widetilde{V}_n(t,b)\text{ is maximal}\}\le 0)\\
&=\mathbb{P}(\sup\{t:\widetilde{V}_n(t,b)\text{ is maximal}\}\le 0)+o(1).
\end{aligned}\tag{C.12}$$

In view of (C.7)–(C.12), for all sufficiently large n,

$$\begin{aligned}
&\mathbb{P}(Z_n^{\#}(C)\le a,\sup\{t:\widetilde{V}_n(t,b)\text{ is maximal}\}\le 0)\\
&=\mathbb{P}\biggl(\sup_{0<t<\infty}\left\{\frac{\widetilde{W}_n(t)-(c+\beta_2 t^2)}{t}\right\}\le a\biggr)\\
&\quad\cdot\,\mathbb{P}(\sup\{t:\widetilde{V}_n(t,b)\text{ is maximal}\}\le 0)+o(1).
\end{aligned}$$

Note that

$$Z_n^{\#}(c)\overset{d}{=}\sup_{t>0}\left\{\frac{W(t)-(c+\beta_2 t^2)}{t}\right\}$$

and by using Proposition 3.1, it can be shown that

$$\sup\{t:\widetilde{V}_n(t,b)\text{ is maximal}\}\xrightarrow{d}\mathcal{S}(b),$$

where $\mathcal{S}(b):=\sup\{t:\sqrt{f_0(x_0)}W(t)+\frac{1}{2}f_0'(x_0)t^2-bt\text{ is maximal}\}$. Therefore,

$$\begin{aligned}
&\lim_{n\to\infty}\mathbb{P}(n^{1/3}\{\hat{f}_{n,\alpha_n}(0)-f_0(0)\}\le a, n^{1/3}\{\hat{f}_n(x_0)-f_0(x_0)\}\le b) \qquad \text{(C.13)}\\
&=\lim_{n\to\infty}\mathbb{P}(Z_n^{\#}(c)\le a,\sup\{t:\widetilde{V}_n(t,b)\text{ is maximal}\}\le 0)\\
&=\mathbb{P}\biggl(\sup_{t>0}\left\{\frac{W(t)-(c+\beta_2 t^2)}{t}\right\}\le a\biggr)\cdot\mathbb{P}(\mathcal{S}(b)\le 0)
\end{aligned}$$

$$= \mathbb{P}\left(\sup_{t>0} \left\{ \frac{W(t) - (c + \beta_2 t^2)}{t} \right\} \le a \right)$$
$$\cdot \mathbb{P}\left(|4 f_0(x_0) f_0'(x_0)|^{1/3} \, \mathbb{Y} \le b \right). \qquad \square$$

Proof of Corollary 3.1. Firstly, we write

$$\hat{S}_{n,\alpha_n}(x_0) - S_0(x_0) = I_{n1} + I_{n2} + I_{n3}, \tag{C.14}$$

where

$$I_{n1} := \hat{f}_n(x_0) \left\{ \frac{1}{\hat{f}_{n,\alpha_n}(0)} - \frac{1}{f_0(0)} \right\};$$
$$I_{n2} := \frac{1}{f_0(0)} \{ \hat{f}_n(x_0) - f_0(x_0) \};$$
$$I_{n3} := \frac{\hat{f}_{n,\alpha_n}(x_0) - \hat{f}_n(x_0)}{\hat{f}_{n,\alpha_n}(0)}.$$

(a) Note that under C_S4, f_0 is strictly decreasing near 0. Hence, by Corollary 2 and Proposition 1 in [27], and the choice that $\alpha_n = cn^{-2/3}$,

$$n^{1/3} I_{n3} = n^{1/3} O_p(\alpha_n) = O_p(n^{-1/3}) = o_p(1).$$

Now, note that the rates of convergence of I_{n1} and I_{n2} are both $n^{1/3}$. With Lemma 3.5, continuous mapping theorem, and Slutsky's theorem, we obtain

$$n^{1/3}(I_{n1} + I_{n2}) \xrightarrow{d} -\frac{f_0(x_0)}{\{f_0(0)\}^2} \cdot \sup_{t>0} \left\{ \frac{W(t) - (c + \beta_2 t^2)}{t} \right\} + \frac{1}{f_0(0)} |4 f_0(x_0) f_0'(x_0)|^{1/3} \, \mathbb{Y},$$

where $W(\cdot)$ and $\mathbb{Y}$ are independent. The result then follows as $f_0(x_0)/f_0(0) = S_0(x_0)$, $1/f_0(0) = \int_0^\infty S_0(y)dy$, and $f_0'(x_0)/f_0(0) = S_0'(x_0)$.

(b) For I_{n1}, by (3.7),

$$n^{1/3} I_{n1} = n^{1/3} O_p(n^{-\frac{k+1}{2k+1}}) = o_p(1),$$

as $k > 1$. For I_{n3}, note that $\alpha_n = cn^{-(k+1)/(2k+1)} = o(n^{-1/3})$. Thus, by Corollary 2 and Proposition 1 in [27],

$$n^{1/3} \left(\frac{\hat{f}_{n,\alpha_n}(x_0) - \hat{f}_n(x_0)}{\hat{f}_{n,\alpha_n}(0)} \right) = n^{1/3} \left(\frac{O_p(\alpha_n)}{\hat{f}_{n,\alpha_n}(0)} \right) = o_p(1).$$

For I_{n2}, by Proposition 3.2, we have

$$n^{1/3} I_{n2} \xrightarrow{d} 2\{f_0(0)\}^{-1} \left| \frac{1}{2} f_0(x_0) f_0'(x_0) \right|^{1/3} \mathbb{Y}$$

$$
\begin{aligned}
&= \left|4\frac{f_0(x_0)}{f_0(0)} \cdot \frac{f_0'(x_0)}{f_0(0)} \cdot \frac{1}{f_0(0)}\right|^{1/3} \mathbb{Y} \\
&= \left|4S_0(x_0)S_0'(x_0)\int_0^\infty S_0(y)dy\right|^{1/3} \mathbb{Y}. \qquad \square
\end{aligned}
$$

C.3. Proofs of Section 3.4

Proof of Theorem 3.6. Using C_f4 and the Hungarian approximation (3.3)–(3.4), we have

$$
\begin{aligned}
&\frac{\mathbb{F}_n(b_n)}{b_n} - f_0(0) \\
=&\left\{\frac{\mathbb{F}_n(b_n)}{b_n} - \frac{F_0(b_n)}{b_n}\right\} + \left\{\frac{F_0(b_n)}{b_n} - f_0(0)\right\} \\
=&\frac{1}{\sqrt{nb_n}} \cdot \frac{1}{\sqrt{b_n}} W_n(F_0(b_n)) + \frac{F_0(b_n)}{b_n}\frac{W_n(1)}{\sqrt{n}} + O_p\left(\frac{\log n}{nb_n}\right) \\
&\quad + \frac{f_0^{(k)}(0)}{(k+1)!} b_n^k + o(b_n^k) \\
=&c^{-1/2} n^{-\frac{k}{2k+1}} \frac{W_n(F_0(b_n))}{\sqrt{b_n}} + O_p(n^{-1/2}) + O_p\left(n^{-\frac{2k}{2k+1}} \log n\right) \\
&\quad + \frac{f_0^{(k)}(0)}{(k+1)!} b_n^k + o(b_n^k).
\end{aligned}
$$

Hence,

$$
n^{\frac{k}{2k+1}}\left\{\frac{\mathbb{F}_n(b_n)}{b_n} - f_0(0)\right\} = \frac{f_0^{(k)}(0)}{(k+1)!}c^k + \frac{c^{-1/2}}{\sqrt{b_n}} W_n(F_0(b_n)) + o_p(1). \qquad \text{(C.15)}
$$

Thus, it suffices to consider the joint asymptotic distribution of the term $W_n(F_0(b_n))/\sqrt{b_n}$ and $n^{1/3}\{\hat{f}_n(x_0) - f_0(x_0)\}$. For the latter one, it suffices to consider $\sup\{t : \tilde{V}_n(t,b) \text{ is maximal}\}$, where $\tilde{V}_n$ is defined in (C.6). With H_{n2} and K_n as defined in the proof of Lemma 3.5, let

$$
E_n := \{|H_{n2}| \leq K_n\}.
$$

From there, we have $\lim_{n\to\infty} \mathbb{P}(E_n) = 1$. For $a, b \in \mathbb{R}$, we have

$$
\begin{aligned}
J_n :=& \mathbb{P}\left(\frac{1}{\sqrt{b_n}} W_n(F_0(b_n)) \leq a, \sup\{t : \tilde{V}_n(t,b) \text{ is maximal}\} \leq 0\right) \\
=& \mathbb{P}\left(\frac{1}{\sqrt{b_n}} W_n(F_0(b_n)) \leq a, \sup\{t : \tilde{V}_n(t,b) \text{ is maximal}\} \leq 0, E_n\right) + o(1) \\
=& \mathbb{P}\left(\frac{1}{\sqrt{b_n}} W_n(F_0(b_n)) \leq a, \sup\{t \in [-K_n, K_n] : \tilde{V}_n(t,b) \text{ is maximal}\} \leq 0\right)
\end{aligned}
$$

$$+ o(1). \tag{C.16}$$

Note that for all large enough n,

$$F_0(b_n) \le F_0(x_0 - K_n n^{-1/3}).$$

Hence, by the independent increment property of Browinan motions, we have

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{\sqrt{b_n}} W_n(F_0(b_n)) \le a, \sup\{t \in [-K_n, K_n] : \tilde{V}_n(t,b) \text{ is maximal}\} \le 0\right) \\
&= \mathbb{P}\left(\frac{1}{\sqrt{b_n}} W_n(F_0(b_n)) \le a\right) \cdot \mathbb{P}\left(\sup\{t \in [-K_n, K_n] : \tilde{V}_n(t,b) \text{ is maximal}\} \le 0\right) \\
&= \mathbb{P}\left(\sqrt{\frac{F_0(b_n)}{b_n}} Z \le a\right) \cdot \left\{\mathbb{P}(\sup\{t : \tilde{V}_n(t,b) \text{ is maximal}\} \le 0) + o(1)\right\}.
\end{aligned} \tag{C.17}$$

Thus, from (C.16) and (C.17) (see also (C.13)),

$$\lim_{n\to\infty} J_n = \mathbb{P}\big(\sqrt{f_0(0)} Z \le a\big) \cdot \mathbb{P}\Big(|4 f_0(x_0) f_0'(x_0)|^{1/3}\, \mathbb{Y} \le b\Big). \tag{C.18}$$

In view of (C.15) and (C.18),

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}\left(n^{\frac{k}{2k+1}} \left\{b_n^{-1}\mathbb{F}_n(b_n) - f_0(0)\right\} \le a, n^{1/3}\{\hat{f}_n(x_0) - f_0(x_0)\} \le b\right) \\
&= \lim_{n\to\infty} \mathbb{P}\Bigg(\frac{f_0^{(k)}(0)}{(k+1)!} c^k + \frac{c^{-1/2}}{\sqrt{b_n}} W_n(F_0(b_n)) \le a, \\
&\qquad\qquad \sup\{t : \tilde{V}_n(t,b) \text{ is maximal}\} \le 0\Bigg) \\
&= \mathbb{P}\left(\frac{f_0^{(k)}(0)}{(k+1)!} c^k + \sqrt{f_0(0)/c} Z \le a\right) \cdot \mathbb{P}(|4 f_0(x_0) f_0'(x_0)|^{1/3}\, \mathbb{Y} \le b). \qquad \square
\end{aligned}$$

Proof of Corollary 3.2. Write

$$\hat{S}_n^H(x_0) - S_0(x_0) = \mathcal{I}_{1n} + \mathcal{I}_{2n},$$

where

$$\begin{aligned}
\mathcal{I}_{1n} &:= -\frac{\hat{f}_n(x_0)}{\hat{f}_0^H(0) f_0(0)} \left\{\hat{f}_n^H(0) - f_0(0)\right\}, \\
\mathcal{I}_{2n} &:= \frac{1}{f_0(0)} \{\hat{f}_n(x_0) - f_0(x_0)\}.
\end{aligned}$$

(a) If $k = 1$, both the rates of convergence to 0 of $\mathcal{I}_{1n}$ and $\mathcal{I}_{2n}$ are $n^{1/3}$. By Theorem 3.6,

$$n^{\frac{1}{3}}\{\hat{S}_n^H(x_0) - S_0(x_0)\} \xrightarrow{d} -\frac{f_0(x_0)}{f_0(0)^2} Z\left(\frac{c f_0'(0)}{2}, c^{-1} f_0(0)\right) + \mathbb{Y}_0.$$

The result follows from the relationship between f_0 and S_0.

(b) If $k > 1$, $\mathcal{I}_{1n} = O_p\left(n^{-k/(2k+1)}\right) = o_p(n^{-1/3})$. Hence, by Theorem 3.6, the result follows. $\square$

C.4. Proofs of Section 3.5

The following elementary fact is used in the Proof of Theorem 3.7.

Lemma C.1. *Let $\{E_{1n}\}$ and $\{E_{2n}\}$ be two sequences of events. Suppose that $\{A_n\}$ is a sequence of events such that $\mathbb{P}(A_n) \to 1$ and for all large enough n, E_{1n} and E_{2n} are independent given A_n, then $\mathbb{P}(E_{1n} \cap E_{2n}) = \mathbb{P}(E_{1n})\mathbb{P}(E_{2n}) + o(1)$.*

Proof. For all large enough n,

$$\begin{aligned}
\mathbb{P}(E_{1n} \cap E_{2n}) &= \mathbb{P}(E_{1n} \cap E_{2n}|A_n)\mathbb{P}(A_n) + o(1) \\
&= \mathbb{P}(E_{1n}|A_n)\mathbb{P}(E_{2n}|A_n)\mathbb{P}(A_n) + o(1) \\
&= \mathbb{P}(E_{1n} \cap A_n)\mathbb{P}(E_{2n} \cap A_n)/\mathbb{P}(A_n) + o(1) \\
&= \{\mathbb{P}(E_{1n})(1 + o(1))\}\{\mathbb{P}(E_{2n})(1 + o(1))\}/(1 + o(1)) + o(1) \\
&= \mathbb{P}(E_{1n})\mathbb{P}(E_{2n}) + o(1).
\end{aligned}$$

□

Proof of Theorem 3.7. For $1/(2k + 1) \le \alpha < 1$, $t \ge 0$, and $x \in \mathbb{R}$, define

$$V_{n1}(x,t) := n^{(1+\alpha)/2}\{\mathbb{F}_n(tn^{-\alpha}) - f_0(0)tn^{-\alpha}\} - xt$$

and recall the definition of V_n in (C.4). For $x > 0, y \in \mathbb{R}$, using the switch relation (3.2), and the fact that adding or multiplying a constant does not affect the location of the maximum of a process,

$$\begin{aligned}
&\mathbb{P}(n^{(1-\alpha)/2}\{\hat{f}_n(cn^{-\alpha}) - f_0(0)\} \le x, n^{1/3}\{\hat{f}_n(x_0) - f_0(x_0)\} \le y) \\
&= \mathbb{P}(n^{\alpha}U_n(f_0(0) + xn^{-(1-\alpha)/2}) \le c, n^{1/3}\{U_n(f_0(0) + yn^{-1/3}) - x_0\} \le 0) \\
&= \mathbb{P}(\sup\{t \ge 0 : V_{n1}(x,t) \text{ is maximal}\} \le c, \\
&\qquad \sup\{t \ge -x_0 n^{1/3} : V_n(x,t) \text{ is maximal}\} \le 0). \qquad \text{(C.19)}
\end{aligned}$$

For $t \ge 0$, by the Hungarian approximation (3.3) and Condition C_f3,

$$\begin{aligned}
V_{n1}(x,t) &= n^{\alpha/2}W_n(F_0(tn^{-\alpha})) - n^{\alpha/2}F_0(tn^{-\alpha})W_n(1) \\
&\quad + n^{(1+\alpha)/2}\frac{f_0^{(k)}(0)}{(k+1)!}(tn^{-\alpha})^{k+1} - xt \\
&= \tilde{V}_{n1}(x,t) + o_p(1),
\end{aligned}$$

where

$$\tilde{V}_{n1}(x,t) := n^{\alpha/2}W_n(F_0(tn^{-\alpha})) - xt, \quad t \ge 0$$

and the $o_p(1)$ term is uniform on compacta. To apply Proposition 3.1, we first extend the process $V_{n1}(x,\cdot)$ and $\tilde{V}_{n1}(x,\cdot)$ to the real line by defining $V_{n1}(x,t) = t$ and $\tilde{V}_{n1}(x,t) = t$ for $t < 0$. Then, by Proposition 3.1, we can show that $\sup\{t : V_{n1}(x,t) \text{ is maximal}\}$ and $\sup\{t : \tilde{V}_{n1}(x,t) \text{ is maximal}\}$ are asymptotically equivalent in the sense that the difference between them converges to 0 in probability. The establishment of the joint convergence and checking of the tightness conditions required in Proposition 3.1 are essentially the same as in the proof of Lemmas 3.1 and 3.2 in [14]. Note also that $\sup\{t : V_n(t, y) \text{ is maximal}\}$ and $\sup\{t : \tilde{V}_n(t, b) \text{ is maximal}\}$ are asymptotically equivalent; see the proof of Theorem 3.5. Thus, from (C.19), we obtain that

$$\begin{aligned} &\lim_{n\to\infty} \mathbb{P}(n^{(1-\alpha)/2}\{\hat{f}_n(cn^{-\alpha}) - f_0(0)\} \le x, n^{1/3}\{\hat{f}_n(x_0) - f_0(x_0)\} \le y) \\ &= \lim_{n\to\infty} \mathbb{P}(\sup\{t : \tilde{V}_{n1}(x, t) \text{ is maximal}\} \le c, \sup\{t : \tilde{V}_n(y, t) \text{ is maximal}\} \le 0). \end{aligned} \tag{C.20}$$

Define

$$\begin{aligned} A_n := \big\{&|\sup\{t : \tilde{V}_{n1}(x, t) \text{ is maximal}\}| \le K_n, \\ &|\sup\{t : \tilde{V}_n(x, t) \text{ is maximal}\}| \le K_n\big\}, \end{aligned}$$

where $K_n \to \infty$ but $K_n = o(n^{\min(\alpha,1/3)})$. Since $\sup\{t : \tilde{V}_{n1}(x, t) \text{ is maximal}\}$ and $\sup\{t : \tilde{V}_n(x, t) \text{ is maximal}\}$ are both $O_p(1)$, $\mathbb{P}(A_n) \to 1$. On A_n, for all large enough n,

$$\sup\{t : \tilde{V}_{n1}(x, t) \text{ is maximal}\} = \sup\{0 \le t \le K_n : \tilde{V}_{n1}(x, t) \text{ is maximal}\};$$
$$\sup\{t : \tilde{V}_n(x, t) \text{ is maximal}\} = \sup\{-x_0 n^{-1/3} \le t \le K_n : \tilde{V}_n(x, t) \text{ is maximal}\};$$

and

$$F_0(K_n n^{-\alpha}) < F_0(x_0 - K_n n^{-1/3}).$$

By the independent increment property of Brownian motions, $\sup\{t : \tilde{V}_{n1}(x, t) \text{ is maximal}\}$ and $\sup\{t : \tilde{V}_n(x, t) \text{ is maximal}\}$ are independent conditional on A_n for all large enough n. Therefore, by Lemma C.1,

$$\begin{aligned} &\lim_{n\to\infty} \mathbb{P}(\sup\{t : \tilde{V}_{n1}(x, t) \text{ is maximal}\} \le c, \sup\{t : \tilde{V}_n(y, t) \text{ is maximal}\} \le 0) \\ &= \lim_{n\to\infty} \mathbb{P}(\sup\{t : \tilde{V}_{n1}(x, t) \text{ is maximal}\} \le c) \\ &\quad \cdot \lim_{n\to\infty} \mathbb{P}(\sup\{t : \tilde{V}_n(y, t) \text{ is maximal}\} \le 0) \\ &= \lim_{n\to\infty} \mathbb{P}(\sup\{t : V_{n1}(x, t) \text{ is maximal}\} \le c) \\ &\quad \cdot \lim_{n\to\infty} \mathbb{P}(n^{1/3}\{\hat{f}_n(x_0) - f_0(x_0)\} \le y)) \\ &= \mathbb{P}(\{f_0(0)/c\}^{1/2} D_R[W(t)](1) \le x)\mathbb{P}(f_0(0)\mathbb{Y}_0 \le y), \end{aligned} \tag{C.21}$$

where the convergence of the first term in the last line is proven in the proof of Theorem 3.1 in [14] while the second one follows from (3.5). For $x \le 0$,

$$\begin{aligned} &\mathbb{P}(n^{(1-\alpha)/2}\{\hat{f}_n(cn^{-\alpha}) - f_0(0)\} \le x, n^{1/3}\{\hat{f}_n(x_0) - f_0(x_0)\} \le y) \\ &\le \mathbb{P}(n^{(1-\alpha)/2}\{\hat{f}_n(cn^{-\alpha}) - f_0(0)\} \le 0) \to 0, \end{aligned} \tag{C.22}$$

where the convergence on the right hand side is obtained in the proof of Theorem 3.1 in [14]. Combining (C.20), (C.21), and (C.22), the required result is proven. □

Proof of Corollary 3.3. Write

$$\hat{S}_n^N(x_0) - S_0(x_0) = -\frac{\hat{f}_n(x_0)}{\hat{f}_n(cn^{-\alpha})f_0(0)}\{\hat{f}_n(cn^{-\alpha}) - f_0(0)\} + \frac{1}{f_0(0)}\{\hat{f}_n(x_0) - f_0(x_0)\}.$$

The rest of the proof is similar to that for Corollary 3.2 by using Theorem 3.6, we just omit here. □

C.5. Proofs of Section 3.6

To prove Theorem 3.8, we first expand $\hat{f}^S_{n,h}(x_0)$ and $\hat{f}^S_{n,h}(0)$ in the following (C.23) and (C.24), and then study the asymptotic behavior of the terms in the two expansions in the following Lemmas C.2, C.3 and C.4.

For all sufficiently large n, $[x_0 - h, x_0 + h] \subset (0, \tau)$ and so the boundary kernel reduces to the usual kernel K for $\hat{f}^S_{n,h}(x_0)$. Write

$$\hat{f}^S_{n,h}(x_0) = M_{x_0,1,n}(h) + M_{x_0,2,n}(h) + M_{x_0,3,n}(h), \tag{C.23}$$

where

$$\begin{aligned}
M_{x_0,1,n}(h) &:= \int_{x_0-h}^{x_0+h} \frac{1}{h}K\left(\frac{x_0-u}{h}\right) dF_0(u),\\
M_{x_0,2,n}(h) &:= \int_{x_0-h}^{x_0+h} \frac{1}{h}K\left(\frac{x_0-u}{h}\right) d(\mathbb{F}_n - F_0)(u),\\
M_{x_0,3,n}(h) &:= \int_{x_0-h}^{x_0+h} \frac{1}{h}K\left(\frac{x_0-u}{h}\right) d(\hat{F}_n - \mathbb{F}_n)(u).
\end{aligned}$$

Similarly, write

$$\hat{f}^S_{n,h}(0) = M_{0,1,n}(h) + M_{0,2,n}(h) + M_{0,3,n}(h), \tag{C.24}$$

where

$$\begin{aligned}
M_{0,1,n}(h) &:= \int_0^h \frac{1}{h}K_0\left(-\frac{u}{h}\right) dF_0(u),\\
M_{0,2,n}(h) &:= \int_0^h \frac{1}{h}K_0\left(-\frac{u}{h}\right) d(\mathbb{F}_n - F_0)(u),\\
M_{0,3,n}(h) &:= \int_0^h \frac{1}{h}K_0\left(-\frac{u}{h}\right) d(\hat{F}_n - \mathbb{F}_n)(u).
\end{aligned}$$

The following lemma establishes the asymptotic behaviour of the first term in the expansion of $\hat{f}^S_{n,h}(x_0)$ and $\hat{f}^S_{n,h}(0)$. In particular, the deterministic terms $M_{x_0,1,n}$ and $M_{0,1,n}$ will contribute to the asymptotic biases of $\hat{f}^S_{n,h}(x_0)$ and $\hat{f}^S_{n,h}(0)$, respectively.

Lemma C.2. *Under Conditions C_f1, C_f3 and C_f4 for some integer $k \geq 2$, if $h = h_n \to 0$, then we have the followings.*

(i) As $n \to \infty$,

$$\frac{1}{h^2}\left\{M_{x_0,1,n}(h) - f_0(x_0)\right\} \to \frac{1}{2}f_0''(x_0)\int_{-1}^{1} y^2K(y)dy;$$

(ii) As $n \to \infty$,

$$\frac{1}{h^k}\left\{M_{0,1,n}(h) - f_0(0)\right\} \to \frac{(-1)^k}{k!}f_0^{(k)}(0)\int_{-1}^{0} y^kK_0(y)dy.$$

Proof. (i) By the change-of-variables formula, the facts that $\int_{-1}^{1} K(u)du = 1$ and K is symmetric,

$$\begin{aligned}
&M_{x_0,1,n}(h) - f_0(x_0)\\
&= \int_{-1}^{1} K(y)\{f_0(x_0 - hy) - f_0(x_0)\}dy\\
&= \int_{-1}^{1} K(y)\{-f_0'(x_0)hy + \frac{1}{2}f_0''(\xi_{n,y})h^2y^2\}dy\\
&= \frac{h^2}{2}\int_{-1}^{1} f_0''(\xi_{h,y})y^2K(y)dy,
\end{aligned}$$

where $\xi_{h,y}$ lies between x_0 and $x_0 - hy$. The result then follows from the dominated convergence theorem.

(ii) By the change-of-variables formula again,

$$\begin{aligned}
&M_{0,1,n}(h) - f_0(0)\\
&= \int_{-1}^{0} K_0(y)\{f_0(-hy) - f_0(0)\}dy\\
&= \int_{-1}^{0} K_0(y)\left\{-f_0'(0)hy + f_0^{(k)}(\xi_{h,y})\frac{(-hy)^k}{k!}\right\}dy\\
&= \frac{(-1)^kh^k}{k!}\int_{-1}^{0} K_0(y)f_0^{(k)}(\xi_{h,y})y^kdy,
\end{aligned}$$

where Taylor's theorem has been applied to an extension of f_0 to an interval containing 0 as an interior point and the last equality follows from (A.1). The result then follows from the dominated convergence theorem. □

The following lemma establishes the joint asymptotic distribution of the second term in the expansion (C.23) and (C.24) of $\hat{f}^S_{n,h_1}(x_0)$ and $\hat{f}^S_{n,h_2}(0)$.

Lemma C.3. *Under Condition $C_f 1$, if $h_j = h_{jn} \to 0$ and $\sqrt{nh_{jn}}/\log n \to \infty$ for $j = 1, 2$, then we have*

$$\left(\sqrt{nh_1} M_{x_0,2,n}, \sqrt{nh_2} M_{0,2,n}\right) \xrightarrow{d} Z(0, \tilde{D}), \tag{C.25}$$

where $\tilde{D}$ is a diagonal matrix with elements $\tilde{d}_{11} = f_0(x_0) \int_{-1}^{1} K^2(y)dy$ and $\tilde{d}_{22} = f_0(0) \int_{-1}^{0} \{K_0(y)\}^2 dy$.

Proof of Lemma C.3. By the change-of-variables formula, $(\sqrt{nh_1} M_{x_0,2,n}$, $\sqrt{nh_2} M_{0,2,n})$ can be written as

$$\left(\int_{-1}^{1} K(y)dW_{1n}(y), \int_{-1}^{0} K_0(y)dW_{2n}(y)\right),$$

where

$$W_{1n}(y) := \sqrt{\frac{n}{h_1}} \{\mathbb{F}_n(x_0 - h_1 y) - \mathbb{F}_n(x_0) - F_0(x_0 - h_1 y) + F_0(x_0)\}, \quad y \in [-1, 1],$$
$$W_{2n}(y) := \sqrt{\frac{n}{h_2}} \{\mathbb{F}_n(-h_2 y) - F_0(-h_2 y)\}, \quad y \in [-1, 0].$$

Using the Hungarian approximation (3.3) and (3.4) for the following $R_n(\cdot)$, as $h_1 \to 0$ and $\sqrt{nh_1}/\log n \to \infty$,

$$\begin{aligned} W_{1n}(y) &= \sqrt{\frac{n}{h}} [n^{-1/2} \{\mathbb{B}_n(F_0(x_0 - h_1 y)) - \mathbb{B}_n(F_0(x_0))\} + R_n(x_0 - h_1 y) - R_n(x_0)] \\ &= \frac{1}{\sqrt{h_1}} \{W_n(F_0(x_0 - h_1 y)) - W_n(F_0(x_0))\} \\ &\quad + \sqrt{h_1} \frac{F_0(x_0 - h_1 y) - F_0(x_0)}{h_1} W_n(1) + O_p\left(\frac{\log n}{\sqrt{nh_1}}\right) \\ &= \frac{1}{\sqrt{h_1}} \{W_n(F_0(x_0 - h_1 y)) - W_n(F_0(x_0))\} + o_p(1), \end{aligned} \tag{C.26}$$

where the $o_p(1)$ term is uniformly small in y. Similarly, as $h_2 \to 0$ and $\sqrt{nh_2}/\log n \to \infty$, we have

$$\begin{aligned} W_{2n}(y) &= \sqrt{\frac{n}{h_2}} [n^{-1/2} \mathbb{B}_n(F_0(-h_2 y)) + R_n(-h_2 y)] \\ &\quad \times \frac{1}{\sqrt{h_2}} W_n(F_0(-h_2 y)) - \sqrt{h_2} \frac{F_0(-h_2 y)}{h_2} W_n(1) + O_p\left(\frac{\log n}{\sqrt{nh_2}}\right) \\ &= \frac{1}{\sqrt{h}} W_n(F_0(-hy)) + o_p(1), \end{aligned} \tag{C.27}$$

where the $o_p(1)$ term is again uniformly small in y. Since Brownian motions have independent increments, for all large n, $\frac{1}{\sqrt{h_1}}\{W_n(F_0(x_0 - h_1 y)) - W_n(F_0(x_0))\}$ and $\frac{1}{\sqrt{h_2}} W_n(F_0(-h_2 y))$ are independent as $F_0(-hy)$ is smaller than $\min(F_0(x_0 -$

$h_1y), F_0(x_0))$ so that the increments do not overlap. In addition, $y \mapsto \frac{1}{\sqrt{h}}\{W_n(F_0(x_0 - h_1y)) - W_n(F_0(x_0))\}$ has the same distribution as the process

$$\mathbb{W}_1\left(\frac{F_0(x_0 - h_1y) - F_0(x_0)}{h_1}\right), \quad y \in [-1, 1]$$

and $y \mapsto \frac{1}{\sqrt{h_2}}W_n(F_0(-h_2y))$ has the same distribution as the process

$$\mathbb{W}_2\left(\frac{F_0(-h_2y)}{h_2}\right), \quad y \in [-1, 0],$$

where $\mathbb{W}_1$ is a standard two-sided Brownian motion and $\mathbb{W}_2$ is a standard Brownian motion, and $\mathbb{W}_1$ are independent of $\mathbb{W}_2$. Therefore, for all large n, $y_1 \in [-1, 1]$ and $y_2 \in [-1, 0]$,

$$\begin{aligned}&\left(\frac{1}{\sqrt{h_1}}\{W_n(F_0(x_0 - h_1y_1)) - W_n(F_0(x_0))\}, \frac{1}{\sqrt{h_2}}W_n(F_0(-h_2y_2))\right)\\ &\stackrel{d}{=} \left(\mathbb{W}_1\left(\frac{F_0(x_0 - h_1y_1) - F_0(x_0)}{h_1}\right), \mathbb{W}_2\left(\frac{F_0(-h_2y_2)}{h_2}\right)\right). \end{aligned} \tag{C.28}$$

By the uniform continuity of $\mathbb{W}_1$ and $\mathbb{W}_2$ on compact intervals, we have

$$\begin{aligned}\sup_{y_1\in[-1,1]}\left|\mathbb{W}_1\left(\frac{F_0(x_0 - h_1y_1) - F_0(x_0)}{h_1}\right) - \mathbb{W}_1(f_0(x_0)y_1)\right| &\xrightarrow{\mathbb{P}} 0,\\ \sup_{y_2\in[-1,0]}\left|\mathbb{W}_2\left(\frac{F_0(-h_2y_2)}{h_2}\right) - \mathbb{W}_2(f_0(0)y_2)\right| &\xrightarrow{\mathbb{P}} 0.\end{aligned} \tag{C.29}$$

By (C.26), (C.27), (C.28), and (C.29), we obtain that

$$(W_{n1}(y_1), W_{n2}(y_2)) \xrightarrow{d} (\sqrt{f_0(x_0)}\mathbb{W}_1(y_1), \sqrt{f_0(0)}\mathbb{W}_2(y_2))$$

as a process on $[-1, 1] \times [-1, 0]$. Accordingly, as $n \to \infty$,

$$\begin{aligned}&\left(\int_{-1}^{1} K(y_1)dW_{n1}(y_1), \int_{-1}^{0} K_0(y_2)dW_{n2}(y_2)\right)\\ &\xrightarrow{d} \left(\sqrt{f_0(x_0)}\int_{-1}^{1} K(y_1)d\mathbb{W}_1(y_1), \sqrt{f_0(0)}\int_{-1}^{0} K_0(y_2)d\mathbb{W}_2(y_2)\right) \sim Z(0, \tilde{D}),\end{aligned}$$

where $\tilde{D}$ is defined as in the statement in the theorem. This is because the expected values of $\int_{-1}^{1} K(y_1)d\mathbb{W}_1(y_1)$ and $\int_{-1}^{0} K_0(y_2)d\mathbb{W}_2(y_2)$ are both 0, the variance of $\int_{-1}^{1} K(y_1)d\mathbb{W}_1(y_1)$ is $\int_{-1}^{1} K^2(y_1)dy_1$, the variance of $\int_{-1}^{0} K_0(y_2)d\mathbb{W}_2(y_2)$ is $\int_{-1}^{0} K_0(y_2)^2dy_2$, and $\mathbb{W}_1$ and $\mathbb{W}_2$ are independent so that the off-diagonal elements in $\tilde{D}$ are 0. $\square$

The following lemma establishes the order of the third term in the expansion of $\hat{f}^S_{n,h}(x_0)$ and $\hat{f}^S_{n,h}(0)$.

Lemma C.4. *Under Conditions C_f1 and C_f5, if $h = h_n \to 0$, then we have the followings.*

(i) $M_{x_0,3,n} = O_p(h^{-1}n^{-2/3}(\log n)^{2/3})$;
(ii) $M_{0,3,n} = O_p(h^{-1}n^{-2/3}(\log n)^{2/3})$.

Proof. (i) By using integration by parts, the change-of-variables formula, and Theorem 3.4, for all sufficiently large n,

$$\begin{aligned}
M_{x_0,3,n} &= \int_{x_0-h}^{x_0+h} \frac{1}{h^2} K'\left(\frac{x_0-u}{h}\right)\{\hat{F}_n(u) - \mathbb{F}_n(u)\}du \\
&= \frac{1}{h}\int_{-1}^{1}\{\hat{F}_n(x_0-hy) - \mathbb{F}_n(x_0-hy)\}K'(y)dy \\
&\le \frac{1}{h}\sup_{y\in[0,x_0+\delta]}|\hat{F}_n(y) - \mathbb{F}_n(y)|\int_{-1}^{1}K'(y)dy \\
&= O_p(h^{-1}n^{-2/3}(\log n)^{2/3}).
\end{aligned}$$

(ii) Similar to (i), by using integration by parts, the change-of-variables, and Theorem 3.4, for all sufficiently large n,

$$\begin{aligned}
&M_{0,3,n} \\
&= \frac{1}{h^2}\int_0^h \left\{\phi(0)K'\left(-\frac{u}{h}\right) + \psi(0)K'\left(-\frac{u}{h}\right)\left(-\frac{u}{h}\right) + \psi(0)K\left(-\frac{u}{h}\right)\right\} \\
&\quad \{\hat{F}_n(u) - \mathbb{F}_n(u)\}du \\
&= \frac{1}{h}\int_{-1}^{0}\{\phi(0)K'(y) + \psi(0)K'(y)y + \psi(0)K(y)\}\{\hat{F}_n(-hy) - \mathbb{F}_n(-hy)\}dy \\
&\le \frac{1}{h}\sup_{y\in[0,x_0+\delta]}|\hat{F}_n(y) - \mathbb{F}_n(y)|\int_{-1}^{0}\{\phi(0)K'(y) + \psi(0)K'(y)y + \psi(0)K(y)\}dy \\
&= O_p(h^{-1}n^{-2/3}(\log n)^{2/3}).
\end{aligned}$$

□

Proof of Theorem 3.8. First, for $\hat{f}^S_{n,h_1}(x_0)$, by Lemma C.2 (i), Lemma C.3, and Lemma C.4 (i), we have

$$\begin{aligned}
n^{2/5}\{\hat{f}^S_{n,h_1}(x_0) - f_0(x_0)\} &= n^{2/5}M_{x_0,1,n} + n^{2/5}M_{x_0,2,n} + n^{2/5}M_{x_0,3,n} \\
&\xrightarrow{d} \mu^{(2)}_{x_0} + Z(0, d_{11}),
\end{aligned}$$

as $n^{2/5}M_{x_0,3,n} = n^{2/5}O_p(n^{1/5}n^{-2/3}(\log n)^{2/3}) = o_p(1)$. For $\hat{f}^S_{n,h_2}(0)$, by Lemma C.2 (ii), Lemma C.3, and Lemma C.4 (ii),

$$\begin{aligned}
&n^{k/(2k+1)}\{\hat{f}^S_{n,h_2}(0) - f_0(0)\} \\
&= n^{k/(2k+1)}M_{0,1,n} + n^{k/(2k+1)}M_{0,2,n} + n^{k/(2k+1)}M_{0,3,n} \\
&\xrightarrow{d} \mu^{(k)}_0 + Z(0, d_{22}),
\end{aligned}$$

as $n^{k/(2k+1)}M_{0,3,n} = n^{k/(2k+1)}O_p(n^{1/(2k+1)}n^{-2/3}(\log n)^{2/3}) = o_p(1)$.

The required joint convergence follows as (i) $M_{x_0,3,n}$ and $M_{0,3,n}$ are of smaller order than the other two terms in the expansion (C.23) and (C.24) of $\hat{f}^S_{n,h_1}(x_0)$ and $\hat{f}^S_{n,h_2}(0)$, respectively; and (ii) $\sqrt{nh_1}M_{x_0,2,n}$ and $\sqrt{nh_2}M_{0,2,n}$ are asymptotically independent by Theorem C.3. □

Proof of Corollary 3.4. The proof is similar to the proof of Corollary 3.2 using Theorem 3.8. Write

$$\hat{S}^S_{n,h_1,h_2}(x_0) - S_0(x_0) = H_{1n} + H_{2n},$$

where

$$H_{1n} := -\frac{f_0(x)}{\hat{f}^S_{n,h_2}(0)f_0(0)}\{\hat{f}^S_{n,h_1}(0) - f_0(0)\}, \tag{C.30}$$

$$H_{2n} := \frac{1}{\hat{f}^S_{n,h_2}(0)}\{\hat{f}^S_{n,h_1}(x_0) - f_0(x_0)\}. \tag{C.31}$$

Since we assume $|S_0''(x_0)| > 0$ and $h_1 = c_1 n^{-1/5}$, the rate of convergence of H_{2n} is $n^{2/5}$. On the other hand, the rate of convergence of H_{1n} is $n^{k/(2k+1)}$.

(i) If $k = 2$, both H_{1n} and H_{2n} have the same rate of convergence. By Theorem 3.8,

$$n^{2/5}\{\hat{S}^S_{n,h_1,h_2}(x_0) - S_0(x_0)\} \xrightarrow{d} \frac{1}{f_0(0)}Z(\mu^{(2)}_{x_0}, d_{11}) - \frac{f_0(x_0)}{f_0(0)^2}Z(\mu^{(k)}_0, d_{22}).$$

(ii) If $k > 2$, H_{2n} dominates and

$$n^{2/5}\{\hat{S}^S_{n,h_1,h_2}(x_0) - S_0(x_0)\} \xrightarrow{d} \frac{1}{f_0(0)}Z(\mu^{(2)}_{x_0}, d_{11}).$$

The results then follow using the relationship between f_0 and S_0. □

C.6. Proofs of Section 3.7

Proof of Theorem 3.9. Similar to previous sections, it can be shown that $\hat{f}_n(x_0)$ and $\hat{f}_n(y_0)$ are asymptotically independent when $x_0 \neq y_0$. Write

$$\begin{aligned}
&n^{1/3}\{\hat{S}_{|y_0}(x_0) - S_{|y_0}(x_0)\}\\
&= \frac{1}{\hat{f}(y_0)}n^{1/3}\{\hat{f}_n(x_0) - f_0(x_0)\} - \frac{f_0(x_0)}{f_0(y_0)\hat{f}_n(y_0)}n^{1/3}\{\hat{f}_n(y_0) - f_0(y_0)\}\\
&\xrightarrow{d} \frac{1}{f_0(y_0)}|4f_0(x_0)f_0'(x_0)|^{1/3}\mathbb{Y}_1 + \frac{f_0(x_0)}{f_0^2(y_0)}|4f_0(y_0)f_0'(y_0)|^{1/3}\mathbb{Y}_2\\
&= \frac{f_0(0)}{f_0(y_0)}\left|4\cdot\frac{f_0(x_0)}{f_0(0)}\cdot\frac{f_0'(x_0)}{f_0(0)}\cdot\frac{1}{f_0(0)}\right|^{1/3}\mathbb{Y}_1\\
&\quad + \frac{f_0(x_0)}{f_0(0)}\cdot\frac{f_0(0)^2}{f_0(y_0)^2}\left|4\cdot\frac{f_0(y_0)}{f_0(0)}\cdot\frac{f_0'(y_0)}{f_0(0)}\cdot\frac{1}{f_0(0)}\right|^{1/3}\mathbb{Y}_2.
\end{aligned}$$

The result then follows using the relationship between f_0 and S_0. □

Appendix D: Additional results

In this section, we provide additional asymptotic results under different regimes defined by underlying smoothness (p or k) and the tuning parameter (γ or α).

D.1. Additional Results for Section 3.4

Theorem D.1. *Under Conditions C_f1, C_f2, and (3.9) for some $p > 1$, if $b_n = cn^{-\gamma}$, where $c > 0$ and $\gamma \in (0, 1)$, then the followings hold.*

(a) If $\frac{1}{2p-1} < \gamma < 1$, then

$$\Big(n^{\frac{1}{2}(1-\gamma)}\{\hat{f}_n^H(0) - f_0(0)\}, n^{1/3}\{\hat{f}_n(x_0) - f_0(x_0)\}\Big) \xrightarrow{d} (Z(0, c^{-1}f_0(0)), f_0(0)\mathbb{Y}_0).$$

(b) if $\gamma = \frac{1}{2p-1}$, then

$$\Big(n^{\frac{p-1}{2p-1}}\{\hat{f}_n^H(0) - f_0(0)\}, n^{1/3}\{\hat{f}_n(x_0) - f_0(x_0)\}\Big) \xrightarrow{d} (Z(-f_1c^{p-1}, c^{-1}f_0(0)), f_0(0)\mathbb{Y}_0).$$

(c) if $0 < \gamma < \frac{1}{2p-1}$,

$$\Big(n^{\gamma(p-1)}\{\hat{f}_n^H(0) - f_0(0)\}, n^{1/3}\{\hat{f}_n(x_0) - f_0(x_0)\}\Big) \xrightarrow{d} (-f_1c^{p-1}, f_0(0)\mathbb{Y}_0).$$

Proof of Theorem D.1. Using the Hungarian approximation (3.3), (3.4) and (3.9),

$$\begin{aligned}
&\frac{\mathbb{F}_n(b_n)}{b_n} - f_0(0)\\
=&\left\{\frac{\mathbb{F}_n(b_n)}{b_n} - \frac{F_0(b_n)}{b_n}\right\} + \left\{\frac{F_0(b_n)}{b_n} - f_0(0)\right\}\\
=&\frac{1}{\sqrt{nb_n}}\cdot\frac{W_n(F_0(b_n))}{\sqrt{b_n}} + \frac{F_0(b_n)}{b_n}\cdot\frac{W_n(1)}{\sqrt{n}} + O_p\left(\frac{\log n}{nb_n}\right) - f_1b_n^{p-1} + o(b_n^{p-1})\\
=&c^{-1/2}n^{\frac{1}{2}(\gamma-1)}\frac{W_n(F_0(b_n))}{\sqrt{b_n}} + O_p(n^{-1/2}) + O_p(n^{\gamma-1}\log n)\\
&\quad - f_1c^{p-1}n^{-\gamma(p-1)} + o(n^{-\gamma(p-1)}).
\end{aligned}$$

Note that

(i) If $\frac{1}{2p-1} < \gamma < 1$, then $-\gamma(p-1) < -\frac{p-1}{2p-1} < \frac{1}{2}(\gamma-1) < 0$. Therefore,

$$n^{\frac{1}{2}(1-\gamma)}\left\{\frac{\mathbb{F}_n(b_n)}{b_n} - f_0(0)\right\}$$

$$
\begin{aligned}
=&\frac{c^{-1/2}}{\sqrt{b_n}} W_n(F_0(b_n)) + O_p(n^{\frac{1}{2}(1-\gamma)-\frac{1}{2}}) + O_p(n^{\frac{1}{2}(1-\gamma)+\gamma-1}\log n) \\
&+ O(n^{\frac{1}{2}(1-\gamma)-\gamma(p-1)}) \xrightarrow{d} Z(0, c^{-1} f_0(0)).
\end{aligned}
$$

(ii) If $\gamma = \frac{1}{2p-1}$, $-\gamma(p-1) = -\frac{p-1}{2p-1} = \frac{1}{2}(\gamma - 1)$. Therefore,

$$
\begin{aligned}
n^{\frac{p-1}{2p-1}} \left\{ \frac{\mathbb{F}_n(b_n)}{b_n} - f_0(0) \right\} &= \frac{c^{-1/2}}{\sqrt{b_n}} W_n(F_0(b_n)) - f_1 c^{p-1} + o_p(1) \\
&\xrightarrow{d} Z(-f_1 c^{p-1}, c^{-1} f_0(0)).
\end{aligned}
$$

(iii) If $0 < \gamma < \frac{1}{2p-1}$, then $\frac{1}{2}(\gamma - 1) < -\frac{p-1}{2p-1} < -\gamma(p-1) < 0$. Therefore,

$$
n^{\gamma(p-1)} \left\{ \frac{\mathbb{F}_n(b_n)}{b_n} - f_0(0) \right\} = -f_1 c^{p-1} + o_p(1) \xrightarrow{\mathbb{P}} -f_1 c^{p-1}.
$$

The joint convergence can be established similar to the proof of Theorem 3.6 and we omit the details. □

To state the following Corollary D.1, denote

$$
\begin{aligned}
D_1 &= D_1(S_0, x_0, c, p) := c^{p-1} S_1 S_0(x_0), \\
D_2 &= D_2(S_0, c) := Z\left(0, c^{-1} S_0^2(x_0) \int_0^\infty S_0(y)dy\right).
\end{aligned}
$$

Corollary D.1. *Under Conditions $C_S 1$, $C_S 2$, and (3.10) for some $p > 1$, if $b_n = cn^{-\gamma}$, where $c > 0$ and $\gamma \in (0,1)$, then the followings hold.*

(a) If $\frac{1}{2p-1} < \gamma < 1$, then

$$
\begin{aligned}
n^{\frac{1}{2}(1-\gamma)} \{\hat{S}_n^H(x_0) - S_0(x_0)\} &\xrightarrow{d} D_2, \text{ if } \gamma > \frac{1}{3}; \\
n^{\frac{1}{3}} \{\hat{S}_n^H(x_0) - S_0(x_0)\} &\xrightarrow{d} D_2 + \mathbb{Y}_0, \text{ if } \gamma = \frac{1}{3}; \\
n^{\frac{1}{3}} \{\hat{S}_n^H(x_0) - S_0(x_0)\} &\xrightarrow{d} \mathbb{Y}_0, \text{ if } \gamma < \frac{1}{3}.
\end{aligned}
$$

(b) If $\gamma = \frac{1}{2p-1}$, then

$$
\begin{aligned}
n^{\frac{p-1}{2p-1}} \{\hat{S}_n^H(x_0) - S_0(x_0)\} &\xrightarrow{d} D_1 + D_2, \text{ if } \gamma > \frac{1}{3}; \\
n^{\frac{1}{3}} \{\hat{S}_n^H(x_0) - S_0(x_0)\} &\xrightarrow{d} D_1 + D_2 + \mathbb{Y}_0, \text{ if } \gamma = \frac{1}{3}; \\
n^{\frac{1}{3}} \{\hat{S}_n^H(x_0) - S_0(x_0)\} &\xrightarrow{d} \mathbb{Y}_0, \text{ if } \gamma < \frac{1}{3}.
\end{aligned}
$$

(c) If $0 < \gamma < \frac{1}{2p-1}$*, then*

$$n^{\gamma(p-1)}\{\hat{S}_n^H(x_0) - S_0(x_0)\} \xrightarrow{d} D_1,\ \textit{if}\ \gamma(p-1) < \frac{1}{3};$$
$$n^{\frac{1}{3}}\{\hat{S}_n^H(x_0) - S_0(x_0)\} \xrightarrow{d} D_1 + \mathbb{Y}_0,\ \textit{if}\ \gamma(p-1) = \frac{1}{3};$$
$$n^{\frac{1}{3}}\{\hat{S}_n^H(x_0) - S_0(x_0)\} \xrightarrow{d} \mathbb{Y}_0,\ \textit{if}\ \gamma(p-1) > \frac{1}{3}.$$

In (a) and (b), D_2 *and* $\mathbb{Y}_0$ *are independent.*

Proof of Corollary D.1. Write

$$\hat{S}_n^H(x_0) - S_0(x_0) = H_{1n} + H_{2n},$$

where

$$H_{1n} := -\frac{\hat{f}_n(x_0)}{\hat{f}_0^H(0) f_0(0)} \left\{ \hat{f}_n^H(0) - f_0(0) \right\},$$
$$H_{2n} := \frac{1}{f_0(0)} \{\hat{f}_n(x_0) - f_0(x_0)\}.$$

With Theorem D.1, the proof remains to compare the rates of convergence to 0 of H_{1n} and H_{2n}. Since $f_0'(x_0) < 0$, the rate of convergence of H_{2n} is $n^{1/3}$.

(a) If $1/(2p-1) < \gamma < 1$, then the rate of convergence of H_{n1} is $n^{(1-\gamma)/2}$.

(i) If $\gamma > 1/3$, then $(1-\gamma)/2 < 1/3$. Hence, H_{1n} dominates and

$$n^{(1-\gamma)/2}\{\hat{S}_n^H(x_0) - S_0(x_0)\} \xrightarrow{d} -\frac{f_0(x_0)}{f_0(0)^2} Z(0, c^{-1} f_0(0)) = D_2.$$

(ii) If $\gamma = \frac{1}{3}$, then $(1-\gamma)/2 = 1/3$. Both H_{1n} and H_{2n} have the same rate of convergence. Using the asymptotic joint distribution of the two random variables in Theorem D.1,

$$n^{1/3}\{\hat{S}_n^H(x_0) - S_0(x_0)\} \xrightarrow{d} D_2 + \mathbb{Y}_0,$$

where D_2 and $\mathbb{Y}_0$ are independent.

(iii) If $\gamma < 1/3$, then $(1-\gamma)/2 > 1/3$. Then H_{2n} dominates and

$$n^{1/3}\{\hat{S}_n^H(x_0) - S_0(x_0)\} \xrightarrow{d} \mathbb{Y}_0.$$

(b) Suppose that $\gamma = 1/(2p-1)$. The rate of convergence of H_{1n} is $n^{\frac{p-1}{2p-1}}$.

(a) If $\gamma > 1/3$, then $p \in (1,2)$ and $(p-1)/(2p-1) < 1/3$. Hence, H_{1n} dominates and

$$n^{\frac{p-1}{2p-1}}\{\hat{S}_n^H(x_0) - S_0(x_0)\} \xrightarrow{d} -\frac{f_0(x_0)}{f_0(0)^2} Z(-f_1 c^{p-1}, c^{-1} f_0(0)) = D_1 + D_2.$$

(b) If $\gamma = 1/3$, then $p = 2$ and $(p-1)/(2p-1) = 1/3$. Both H_{1n} and H_{2n} have the same rate of convergence and

$$n^{1/3}\{\hat{S}_n^H(x_0) - S_0(x_0)\} \xrightarrow{d} D_1 + D_2 + \mathbb{Y}_0,$$

where D_2 and $\mathbb{Y}_0$ are independent.

(c) If $\gamma < 1/3$, then $p > 2$ and $(p-1)/(2p-1) > 1/3$. Hence, H_{2n} dominates and

$$n^{1/3}\{\hat{S}_n^H(x_0) - S_0(x_0)\} \xrightarrow{d} \mathbb{Y}_0.$$

(c) The proof is similar to (a) and (b), and it is omitted. □

D.2. Additional Results for Section 3.5

For $c > 0$, define

$$A_1 := \{c/f_0(0)\}^{1/2} = \left\{c \int_0^\infty S_0(y)dy\right\}^{1/2}.$$

Theorem D.2. *Under Conditions C_f1, C_f2, and C_f4' for some $k \geq 1$, if $1/(2k+1) < \alpha < 1$, then $(n^{(1-\alpha)/2}\{\hat{f}_n(cn^{-\alpha}) - f_0(0)\}, n^{1/3}\{\hat{f}_n(x_0) - f_0(x_0)\})$ converges in distribution to $(A_1^{-1}D_R[W(t)](1), f_0(0)\mathbb{Y}_0)$.*

Again, note that as a functional purely dependent on $W(\cdot)$, $D_R[W(t)](1)$ and $\mathbb{Y}_0$ are independent of each other.

Proof of Theorem D.2. First, note that Theorem 3.1 in [14] implies that

$$A_1 n^{(1-\alpha)/2}\{\hat{f}_n(cn^{-\alpha}) - f_0(0)\} \xrightarrow{d} D_R[W(t)](1).$$

The require joint convergence can be proven similar to that in the proof of Theorem 3.7 □

Corollary D.2. *Under Conditions C_S1, C_S2, and C_S4' for some $k \geq 1$, if $1/(2k+1) < \alpha < 1$, then*

$$n^{(1-\alpha)/2}\{\hat{S}_n^N(x_0) - S_0(x_0)\} \xrightarrow{d} S_0(x_0)\int_0^\infty S_0(y)dyA_1^{-1}D_R[W(t)](1), \textit{ for } \alpha > \frac{1}{3};$$

$$n^{1/3}\{\hat{S}_n^N(x_0) - S_0(x_0)\} \xrightarrow{d} S_0(x_0)\int_0^\infty S_0(y)dyA_1^{-1}D_R[W(t)](1) + \mathbb{Y}_0, \textit{ for } \alpha = \frac{1}{3};$$

$$n^{1/3}\{\hat{S}_n^N(x_0) - S_0(x_0)\} \xrightarrow{d} \mathbb{Y}_0, \textit{ for } \alpha < \frac{1}{3}.$$

Proof of Corollary D.2. Write

$$\hat{S}_n^N(x_0) - S_0(x_0) = -\frac{\hat{f}_n(x_0)}{\hat{f}_n(cn^{-\alpha})f_0(0)}\{\hat{f}_n(cn^{-\alpha}) - f_0(0)\} + \frac{1}{f_0(0)}\{\hat{f}_n(x_0) - f_0(x_0)\}.$$

The rest of the proof is similar to that for Corollary D.1 by using Theorem D.2, we just omit here. □

D.3. Additional Result for Section 3.6

Theorem D.3. *Under Conditions C_f1, C_f3, C_f4'' for some $k \geq 2$, and C_f5, if $h_1 = h_{1n} = c_1 n^{-1/5}$ and $h_2 = h_{2n} = c_2 n^{-\alpha}$, where $\alpha \in (0,1)$, then the followings hold.*

(i) If $1/(2k+1) < \alpha < 1/3$, then

$$\left(n^{2/5}\{\hat{f}^S_{n,h_1}(x_0) - f_0(x_0)\}, n^{(1-\alpha)/2}\{\hat{f}^S_{n,h_2}(0) - f_0(0)\}\right)$$

converges in distribution to $(Z(\mu^{(2)}_{x_0}, 0)^\top, D)$.

(ii) If $0 < \alpha < 1/(2k+1)$, then

$$\left(n^{2/5}\{\hat{f}^S_{n,h_1}(x_0) - f_0(x_0)\}, n^{k\alpha}\{\hat{f}^S_{n,h_2}(0) - f_0(0)\}\right)$$

converges in distribution to $(Z(\mu^{(2)}_{x_0}, d_{11}), \mu_0)$

Proof of Theorem D.3. First, for $\hat{f}^S_{n,h_1}(x_0)$, by Lemma C.2 (i), Lemma C.3, and Lemma C.4 (i), we have

$$\begin{aligned} n^{2/5}\{\hat{f}^S_{n,h_1}(x_0) - f_0(x_0)\} &= n^{2/5}M_{x_0,1,n} + n^{2/5}M_{x_0,2,n} + n^{2/5}M_{x_0,3,n} \\ &\xrightarrow{d} \mu^{(2)}_{x_0} + Z(0, d_{11}), \end{aligned}$$

as $n^{2/5}M_{x_0,3,n} = n^{2/5}O_p(n^{1/5}n^{-2/3}(\log n)^{2/3}) = o_p(1)$. For $\hat{f}^S_{n,h_2}(0)$, by Lemma C.2 (ii), Lemma C.3, and Lemma C.4 (ii), we have

(i) if $1/(2k+1) < \alpha < 1/3$,

$$\begin{aligned} &n^{(1-\alpha)/2}M_{0,1,n} = O(n^{(1-\alpha)/2-\alpha k}) = o(1), \\ &n^{(1-\alpha)/2}M_{0,2,n} \xrightarrow{d} Z(0, d_{22}), \\ &n^{(1-\alpha)/2}M_{0,3,n} = n^{(1-\alpha)/2}O_p(n^{\alpha}n^{-2/3}(\log n)^{2/3}) = o_p(1). \end{aligned}$$

Thus,

$$n^{(1-\alpha)/2}\{\hat{f}^S_{n,h_2}(0) - f_0(0)\} \xrightarrow{d} Z(0, d_{22}).$$

(ii) if $0 < \alpha < 1/(2k+1)$,

$$\begin{aligned} &n^{k\alpha}M_{0,1,n} \to \mu^{(k)}_0, \\ &n^{k\alpha}M_{0,2,n} = n^{k\alpha}O_p(n^{-1/2+\alpha/2}) = o_p(1), \\ &n^{k\alpha}M_{0,3,n} = n^{k\alpha}O_p(n^{\alpha}n^{-2/3}(\log n)^{2/3}) = o_p(1). \end{aligned}$$

Thus,

$$n^{k\alpha}\{\hat{f}^S_{n,h_2}(0) - f_0(0)\} \xrightarrow{d} \mu^{(k)}_0.$$

The required joint convergence follows as (i) $M_{x_0,3,n}$ and $M_{0,3,n}$ are of smaller order than the other two terms in the expansion (C.23) and (C.24) of $\hat{f}^S_{n,h_1}(x_0)$ and $\hat{f}^S_{n,h_2}(0)$, respectively; and (ii) $\sqrt{nh_1}M_{x_0,2,n}$ and $\sqrt{nh_2}M_{0,2,n}$ are asymptotically independent by Theorem C.3. □

Corollary D.3. *Under Conditions C_S1, C_S3, C_S4'' for some $k \geq 2$, and C_S5, if $h_1 = h_{1n} = c_1 n^{-1/5}$, $h_2 = h_{2n} = c_2 n^{-\alpha}$, where $\alpha \in (0,1)$, then the followings hold.*

(a) If $1/(2k+1) < \alpha < 1/3$, then

$$n^{(1-\alpha)/2}\{\hat{S}^S_{n,h_1,h_2}(x_0) - S_0(x_0)\} \xrightarrow{d} Z(0, \sigma_2^2), \text{ if } \alpha > \frac{1}{5};$$
$$n^{2/5}\{\hat{S}^S_{n,h_1,h_2}(x_0) - S_0(x_0)\} \xrightarrow{d} Z(\mu_1^{(2)}, \sigma_1^2 + \sigma_2^2), \text{ if } \alpha = \frac{1}{5};$$
$$n^{2/5}\{\hat{S}^S_{n,h_1,h_2}(x_0) - S_0(x_0)\} \xrightarrow{d} Z(\mu_1^{(2)}, \sigma_1^2), \text{ if } \alpha < \frac{1}{5}.$$

(b) If $0 < \alpha < 1/(2k+1)$, then

$$n^{\alpha k}\{\hat{S}^S_{n,h_1,h_2}(x_0) - S_0(x_0)\} \xrightarrow{d} \mu_2, \text{ if } \alpha k < \frac{2}{5};$$
$$n^{2/5}\{\hat{S}^S_{n,h_1,h_2}(x_0) - S_0(x_0)\} \xrightarrow{d} Z(\mu_1^{(2)} + \mu_2^{(k)}, \sigma_1^2), \text{ if } \alpha k = \frac{2}{5};$$
$$n^{2/5}\{\hat{S}^S_{n,h_1,h_2}(x_0) - S_0(x_0)\} \xrightarrow{d} Z(\mu_1^{(2)}, \sigma_1^2), \text{ if } \alpha k > \frac{2}{5}.$$

Proof of Corollary D.3. Recall the definition of H_{1n} and H_{2n} in the proof of Corollary 3.4.

(a) If $1/(2k+1) < \alpha < 1/3$, then the rate of convergence of H_{1n} is $n^{(1-\alpha)/2}$.

 (i) If $\alpha > 1/5$, then $(1-\alpha)/2 < 2/5$. Hence, H_{1n} dominates and

$$n^{(1-\alpha)/2}\{\hat{S}^S_{n,h_1,h_2}(x_0) - S_0(x_0)\} \xrightarrow{d} -\frac{f_0(x_0)}{f_0(0)^2} Z(0, d_{22}).$$

 (ii) If $\alpha = 1/5$, then $(1-\alpha)/2 = 2/5$. Both H_{1n} and H_{2n} have the same rate of convergence. By Theorem 3.8,

$$n^{2/5}\{\hat{S}^S_{n,h_1,h_2}(x_0) - S_0(x_0)\} \xrightarrow{d} \frac{1}{f_0(0)} Z(\mu^{(2)}_{x_0}, d_{11}) - \frac{f_0(x_0)}{f_0(0)^2} Z(0, d_{22}).$$

 (iii) If $\alpha < 1/5$, then $(1-\alpha)/2 > 2/5$. Hence, H_{2n} dominates and

$$n^{2/5}\{\hat{S}^S_{n,h_1,h_2}(x_0) - S_0(x_0)\} \xrightarrow{d} \frac{1}{f_0(0)} Z(\mu^{(2)}_{x_0}, d_{11}).$$

 The results then follow using the relationship between f_0 and S_0.

(b) If $0 < \alpha < 1/(2k+1)$, the rate of convergence of H_{1n} is $n^{\alpha k}$.

(i) If $\alpha k < 2/5$, then H_{1n} dominates and

$$n^{\alpha k}\{\hat{S}^S_{n,h_1,h_2}(x_0) - S_0(x_0)\} \xrightarrow{d} -\frac{f_0(x_0)}{f_0(0)^2}\mu_0.$$

(ii) If $\alpha k = 2/5$, both H_{1n} and H_{2n} have the same rate of convergence. By Theorem 3.8,

$$n^{2/5}\{\hat{S}^S_{n,h_1,h_2}(x_0) - S_0(x_0)\} \xrightarrow{d} \frac{1}{f_0(0)}Z(\mu^{(2)}_{x_0}, d_{11}) - \frac{f_0(x_0)}{f_0(0)^2}\mu_0.$$

(iii) If $\alpha k > 2/5$, then H_{2n} dominates and

$$n^{2/5}\{\hat{S}^S_{n,h_1,h_2}(x_0) - S_0(x_0)\} \xrightarrow{d} \frac{1}{f_0(0)}Z(\mu^{(2)}_{x_0}, d_{11}).$$

The results follow using the relationship between f_0 and S_0. □

Acknowledgments

Gary Chan acknowledges the support by US National Institutes of Health Grant R01HL122212 and US National Science Foundation Grant DMS1711952. Hok Kan Ling thanks for the hospitality at the Chinese University of Hong Kong during the final draft of this work and acknowledges the support by NSERC Grant RGPIN/03124-2021. Phillip Yam acknowledges the financial support from HKGRF – Project Number 14300319 with the project title: Shape-constrained Inference: Testing for Monotonicity. He also thanks Columbia University for the kind invitation to be a visiting faculty member in the Department of Statistics during his sabbatical leave.

References

[1] Carolan, C. and Dykstra, D. (1999). Asymptotic behavior of the Grenander estimator at density flat regions. *Canadian Jorunal of Statistics* **27** 557–566. MR1745821

[2] Cox, D. R. (1962). *Renewal Theory*. MR0153061

[3] Durot, C., Groeneboom, P. and Lopuhaä, H. P. (2013). Testing equality of functions under monotonicity constraints. *Journal of Nonparametric Statistics* **25** 939–970. MR3174305

[4] Durot, C. and Lopuhaä, H. P. (2014). A Kiefer-Wolfowitz type of result in a general setting, with an application to smooth monotone estimation. *Electronic Journal of Statistics* **8** 2479–2513. MR3285873

[5] Grenander, U. (1956). On the theory of mortality measurement: Part II. *Scandinavian Actuarial Journal* **1956** 125–153. MR0093415

[6] Groeneboom, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Vol. II (Berkeley, Calif., 1983)*, 539–555. Wadsworth, Belmont, CA. MR822052

[7] Groeneboom, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Vol. II*, 539–555. MR0822052

[8] Groeneboom, P. and Jongbloed, G. (2014). *Nonparametric Estimation Under Shape Constraints* **38**. Cambridge University Press. MR3445293

[9] Huang, J. and Wellner, J. A. (1995). Estimation of a monotone density or monotone hazard under random censoring. *Scandinavian Journal of Statistics* 3–33. MR1334065

[10] Jongbloed, G., van der Meulen, F. and Pang, L. (2021). Bayesian estimation of a decreasing density. *Brazilian Journal of Probability and Statistics* **35** 392–420. MR4255164

[11] Keiding, N., Kvist, K., Hartvig, H., Tvede, M. and Juul, S. (2002). Estimating time to pregnancy from current durations in a cross-sectional sample. *Biostatistics* **3** 565–578.

[12] Kiefer, J. and Wolfowitz, J. (1977). Asymptotically minimax estimation of concave and convex distribution functions. II. In *Statistical Decision Theory and Related Topics* 193–211. Elsevier. MR0443202

[13] Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent RV'-s, and the sample DF. I. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **32** 111–131. MR0375412

[14] Kulikov, V. N. and Lopuhaä, H. P. (2006). The behavior of the NPMLE of a decreasing density near the boundaries of the support. *The Annals of Statistics* **34** 742–768. MR2283391

[15] Lopuhaä, H. P. and Musta, E. (2017). Smooth estimation of a monotone hazard and a monotone density under random censoring. *Statistica Neerlandica* **71** 58–82. MR3605364

[16] Marshall, A. W. (1970). Discussion of Barlow and van Zwet's paper. In *Nonparametric Techniques in Statistical Inference. Proceedings of the First International Symposium on Nonparametric Techniques held at Indiana University, June, 1969*, 174–176.

[17] McLaughlin, K. A., Green, J. G., Gruber, M. J., Sampson, N. A., Zaslavsky, A. M. and Kessler, R. C. (2010). Childhood adversities and adult psychiatric disorders in the national comorbidity survey replication II: Associations with persistence of DSM-IV disorders. *Archives of General Psychiatry* **67** 124–132.

[18] Nane, G. F. (2013). Shape constrained nonparametric estimation in the Cox model.

[19] Prakasa Rao, B. L. S. (1969). Estimation of a unimodal density. *Sankhya Ser. A* **31** 23—36. MR0267677

[20] Robertson, T. (1988). *Order Restricted Statistical Inference. Wiley Series in Probability and Statistics*. Wiley.

[21] Salomond, J.-B. (2014). Concentration rate and consistency of the posterior distribution for selected priors under monotonicity constraints. *Electron. J. Stat.* **8**, no. 1, 1380–1404. MR3263126

[22] Silverman, B. W. (1986). *Density Estimation for Statistics and Data*

Analysis **26**. CRC Press. MR0848134

[23] Sun, J. and Woodroofe, M. (1996). Adaptive smoothing for a penalized NPMLE of a non-increasing density. *Journal of Statistical Planning and Inference* **52** 143–159. MR1392133

[24] Van Es, B., Klaassen, C. A. and Oudshoorn, K. (2000). Survival analysis under cross-sectional sampling: length bias and multiplicative censoring. *Journal of Statistical Planning and Inference* **91** 295–312. MR1814785

[25] Vardi, Y. (1989). Multiplicative censoring, renewal processes, deconvolution and decreasing density: nonparametric estimation. *Biometrika* **76** 751–761. MR1041420

[26] Wang, M.-C. (1991). Nonparametric estimation from cross-sectional survival data. *Journal of the American Statistical Association* **86** 130–143. MR1137104

[27] Woodroofe, M. and Sun, J. (1993). A penalized maximum likelihood estimate of $f(0+)$ when f is non-increasing. *Statistica Sinica* 501–515. MR1243398

[28] Yamaguchi, K. (2003). Accelerated failure–time mover–stayer regression models for the analysis of last–episode data. *Sociological Methodology* **33** 81–110.